Professional development in formative assessment: Effects on teacher classroom practice and student achievement

Catarina Andersson
# 3.6 Summary

## 4. Methodology

### 4.1 The overall research design

#### 4.1.1 The framework and analytical tool

#### 4.1.2 Mixed methods

#### 4.1.3 Triangulation

### 4.2 Overview of the research project

### 4.3 Research methods

#### 4.3.1 Participants

#### 4.3.2 Procedures in data collection and data analysis

##### 4.3.2.1 Study I and II - Classroom observations and Teacher interviews

##### 4.3.2.2 Study III – Teacher interviews and Teacher questionnaires

##### 4.3.2.3 Study IV – Student mathematics pre- and post-tests

### 4.4 Philosophical considerations

#### 4.4.1 A pragmatic approach

#### 4.4.2 Values, beliefs and attitudes

### 4.5 The role of the researcher

### 4.6 Ethical considerations

#### 4.5.1 Information

#### 4.5.2 Consent

#### 4.5.3 Confidentiality

#### 4.5.4 Use

#### 4.5.5 Additional ethical considerations

### 4.6 Summary

## 5. Results

### 5.1 Article 1

### 5.2 Article 2

### 5.3 Article 3

### 5.4 Article 4

## 6. Conclusions

## 7. Discussion

### 7.1 Contribution of the research presented in this thesis

### 7.2 A unity of integrated formative assessment strategies

### 7.3 The value of characterising high quality formative assessment classroom practice

### 7.4 Important characteristics of professional development programs in formative assessment

### 7.5 Implications for educational practice

### 7.6 Implications for the educational research field of formative assessment

### 7.7 Future research

### 7.8 Summary and final words

## Acknowledgements

## References
Abstract

The potential of formative assessment, evident in several research reviews, has raised the interest in many countries to invest in reform initiatives to develop the use of formative assessment. However, implementation of formative assessment is not straightforward and there is a lack of knowledge about how to design appropriate professional development. The intervention study presented in this thesis aimed to see if a random selection of teachers, participating in a professional development program with many contact hours and substantial support of an expert, implemented formative assessment in a way that increased their students’ learning in mathematics. It also aimed to examine the reasons for the teachers’ changes in their classroom practice.

The twenty-two year 4 teachers attended a professional development program in formative assessment in mathematics. A mixed methods approach used classroom observations, teacher interviews, questionnaire surveys and student mathematics tests to investigate the effects on teacher classroom practice and student achievement.

It was found that the teachers trained in formative assessment built on their previous formative classroom practice and added new formative assessment activities into their mathematics classroom practice to a level that had significant impact on student achievement in mathematics ($p = .036, d = .66$). The teachers developed their formative assessment practice in three dimensions: key processes in teaching and learning, agents in the classroom, and the length of the formative assessment cycle.

The reasons for teachers’ implementation of new formative assessment activities were well explained by the expectancy-value theory of achievement motivation. Important aspects of the professional development program were: (1) A formative and process-oriented character; (2) Activities directly useable in classrooms; (3) Experience of using formative assessment activities; (4) Connection between theory and practice; (5) Time; and (6) Knowledgeable support.

The thesis shows that it was possible to provide sufficient support to a random selection of teachers for them to develop their formative assessment practice in a way that improved student achievement. However, this thesis also indicates that it can be expected that teachers would need substantial time and support to achieve such developments in their classroom practice.
List of original articles

The thesis consists of the following articles (also addressed as studies I–IV), that at the time of printing of this thesis have not yet been submitted for publication.

**Article 1**
*Formative assessment in Swedish mathematics classroom practices.*
Authors: Catarina Andersson and Erika Boström.

**Article 2**
The characteristics of formative assessment that enhance student achievement in mathematics.
Authors: Catarina Andersson and Torulf Palm.

**Article 3**
*Reasons for teachers' successful development of a formative assessment practice through professional development - a motivation perspective.*
Authors: Catarina Andersson and Torulf Palm.

**Article 4**
The impact of formative assessment on student achievement: A study of the effects of changes in classroom practice after a comprehensive professional development program.
Authors: Catarina Andersson and Torulf Palm.
Svensk sammanfattning

Inledning
Flertalet forskningsöversikter har visat på de effekter på elevers lärande som är möjliga genom att använda formativ bedömning. I flera länder har denna forskning ökat intresset för att satsa på reformer för att utveckla användningen av formativ bedömning. Att utveckla sitt sätt att använda formativ bedömning är dock inte en okomplicerad process och det behövs ytterligare kunskap om hur man utformar lärarförtbildning i formativ bedömning. Interventionsstudien som presenteras i denna avhandling syftar till att se om vanliga (slumpvis utvalda) lärare, som deltog i en förtbildning med många kontakttimmar och ett betydande stöd från en expert, kunde utveckla sin användning av formativ bedömning på ett sätt som ökade deras elevers lärande i matematik. Ett annat syfte var att undersöka varför lärarna valde att göra förändringar i sin undervisning och vilka egenskaper i förtbildningen de ansåg vara viktiga för att de skulle göra förändringar i sin undervisning.

Avhandlingens fyra delstudier (Artikel 1-4) har undersökt: lärarnas matematikundervisning, med avseende på formativ bedömning, innan förtbildningen; förtbildningens effekter på lärarnas undervisning; orsakerna till den förändrade undervisningen, och vilka aspekter i förtbildningen som var viktiga för att förändringarna i undervisningen skulle ske; samt förtbildningens effekter på elevernas matematikprestation.

Vi har använt ett ramverk som definierar effektiv och ändamålsenlig formativ bedömning som en klassrumspraktik grundad på en tillämpning av en grundläggande idé (the big idea), som handlar om att använda information om elevernas lärande för att justera undervisningen för att bättre möta elevernas behov, samt en kompetent användning av följande fem nyckelstrategier (William, 2010; Wiliam & Thompson, 2008.):

1. Klargöra, dela och skapa förståelse för lärandemål och kriterier för framsteg
2. Åstadkoma effektiva klassrumsdiskussioner, aktiviteter och inlärningsuppgifter som visar att lärande har skett
3. Ge feedback som för lärandet framåt
4. Aktivera eleverna till att bli undervisningsresurser för varandra
5. Aktivera eleverna till att äga sitt eget lärande
Metod

Resultat
Resultaten i den första studien visar på vilka sätt mellanstadielärarna (samt en ytterligare grupp slumpvis utvalda högstadielärare) använde aktiviteter för formativ bedömning i sin matematikundervisning innan fortbildningen, men också på möjligheter för lärarna att vidareutveckla sin användning av formativ bedömning.

Resultaten i den andra studien visar att alla de 22 lärare som deltagit i fortbildningen i formativ bedömning hade utvecklat sin användning av formativ bedömning i sin matematikundervisning. Lärarna hade, i olika utsträckning, infört nya formativa aktiviteter som de integrerat med sina tidigare aktiviteter för formativ bedömning. I denna artikel karakteriserades lärarnas utveckling av formativ bedömning utifrån tre dimensioner: nyckelprocesser i undervisning och lärande, agenter i klassrummet (läraren, eleven, kamrater), och längden av bedömningscykeln för användningen av den grundläggande idén. Artikeln föreslår också en koppling mellan dessa dimensioner och möjligheternas för elevers lärande.

Den tredje studien visar att lärarna var motiverade att utveckla sin användning av formativ bedömning i sin matematikundervisning. Utifrån motivationsteorin expectancy-value theory of achievement motivation förklarades både varför lärarna valde att utveckla sin formativa bedömningspraktik och varför lärarna valde vissa aktiviteter för formativ bedömning framför andra. Lärarna såg ett värde i att använda formativ bedömning, för dem själva och för sina elever. De såg också sig själva som kompetenta att klara av att använda nya aktiviteter för formativ bedömning. Lärarna valde aktiviteter med högt värde i förhållande till kostnader. Sex egenskaper i fortbildningen som lärarna ansåg vara särskilt viktiga för förändringarna med hänseende till formativ bedömning var: (1) en formativ och processorienterad fortbildning (2); aktiviteter direkt användbara i klassrum; (3) erfarenheterna av att använda aktiviteter för formativ
Resultaten från den fjärde studien visade att fortbildningen hade haft effekt på elevernas resultat i matematik. Eleverna till lärarna i interventionsgruppen presterade signifikant \( p = 0.036 \) högre resultat på ett test i slutet av årskurs 4 än eleverna till lärarna i kontrollgruppen jämfört med resultaten på ett test i början av skolåret. Effektstorleken på klassnivå var 0,66 (Cohen’s d).

**Slutsatser**

Den första studien visar att lärarna använde formativ bedömning redan innan fortbildningen, men att många aktiviteter inte användes på ett effektivt sätt. Följaktligen fanns möjligheter för lärarna att utveckla sin användning av formativ bedömning. Eftersom lärarna som deltog i fortbildningen var slumpmässigt utvalda ur samma population som lärarna i kontrollgruppen kan det antas att kontrollgruppens lärare använder formativ bedömning på liknande sätt som interventionsgruppslärarna gjorde innan de deltog i fortbildningen. Lärarnas tidigare användning av formativ bedömning är viktig för att visa på behovet av utveckling av formativ bedömning och vilka utvecklingsmöjligheter som finns. Denna information kan användas vid planering av fortbildning i formativ bedömning för lärare.

Den andra studien visar att fortbildningen hade kunde stötta lärarna att förändra sin undervisning på ett sådant sätt som i den fjärde studien visat sig ha påverkat elevernas prestationer i matematik. Lärarna hade inte bara infört nya aktiviteterna i formativ bedömning i sin undervisning, de använde också dessa aktiviteter utifrån den grundläggande idén och de allmänna principerna för formativ bedömning.

Den tredje studien visar att fortbildningen hade gett lärarna möjligheter att uppleva värdet av formativ bedömning och att skapa förväntningar på att de skulle klara av att använda formativ bedömning på ett nytt sätt. Studien identifierade sex viktiga egenskaper i fortbildningen som lärarna upplevde varit viktiga för att kunna utveckla sin formativa bedömningspraktik. Dessa egenskaper kan ha bidragit till deras uppfattningar om värde av, och förväntningar att lyckas med, att utveckla deras formativa bedömning och därmed motivationen att engagera sig i sådan undervisningsutveckling.

Den fjärde studien visar att fortbildningen hade effekt även på elevernas lärande. Eftersom lärarna i interventionsgruppen var ett slumpmässigt urval ur samma population som kontrollgruppen lärare kan de antas ha bedrivit liknande undervisning innan fortbildningen. Baserat på detta antagande och den stora förändring av undervisningen lärarna i interventionsgruppen gjorde med avseende på formativ bedömning dras slutsatsen att elevernas förbättrade resultat beror på denna utvecklade undervisning som bygger på en användning av principerna för formativ bedömning.
Diskussion
Interventionsstudien som presenteras i denna avhandling visar att dessa slumpmässigt utvalda mellanstadielärare, som deltagit i en fortbildning som var utformad utifrån en grundidé om att användning av formativ bedömning med hög kvalitet är svårt, utvecklade sin formativa bedömningspraktik som i sin tur resulterade i inlärningsvinster för deras elever. Studien visar också på att det är viktigt att beakta lärarnas motivation för förståelse av varför lärare väljer att förändra sin undervisning. Resultaten från studierna kan bidra till att öka förståelsen för hur man kan stödja lärarnas införande och användning av formativ bedömning i matematikundervisningen. Den mest framträdande viktiga egenskap hos fortbildningen i formativ bedömning var fortbildningens processorienterade och formativa karaktär. De andra fem identifierade viktiga egenskaperna skapar förutsättningar och bidrar till att fortbildningen kan få denna övergripande karakter.

1. Introduction

There is a large amount of research, including several comprehensive research reviews, showing that the use of formative assessment in classroom practice is one of the most educationally effective ways of increasing students’ achievement (e.g. Black & Wiliam, 1998a; Hattie, 2009; National Mathematics Advisory Panel, 2008). However it is also clear from these reviews, and from other studies (Black & Atkin, 1996; Black & Wiliam, 1998b) that achieving this is by no means straightforward. Much is unknown about the institutional conditions that help teachers to learn new classroom practices (James, Black, McCormick, Pedder, & Wiliam, 2006). The task of applying research into practice is much more than a simple process of ‘translating’ the findings of researchers into the classroom (Black & Wiliam, 2003). In particular, a strong research base supporting how to effectively help teachers to implement a formative assessment practice is lacking (Schneider & Randel, 2010; Wiliam, 2010). Thus, there is a need for intervention studies in formative assessment:

Research that provides precise estimates of the impact of professional development in formative classroom assessment on student outcomes or teacher outcomes is just beginning. This type of research is essential to determine if professional development in formative classroom assessment is effective in raising student achievement and changing teacher practice. This type of research is also needed to better understand for whom professional development in formative classroom assessment is effective and under what conditions (Schneider & Randel, 2010, p. 268).

Few empirical studies have investigated the impact from professional development programs in formative assessment concerning both teacher practice and student achievement (Schneider & Randel, 2010; Tierney, 2006; Vescio, Ross, & Adams, 2008). Consequently, the relation between professional development programs’ impact on teachers’ classroom practice and impact on student achievement has not been well investigated. Literature on professional development in formative assessment shows that despite many national reform initiatives in formative assessment (Tierney, 2006), many countries still ask for more extensive use of formative assessment in schools (e.g. Black & Wiliam, 1998b; Carless, 2005; Cizek, 2010; Smith, 2011; Stiggins, 2002). In addition, empirical studies report distorted understandings and superficial implementations of formative assessment (Hume & Coll, 2009; James & McCormick, 2009; Marshall & Jane Drummond, 2006; Torrance, 2007), also reflected in theoretical reflections on the research field (Klenowski, 2009; Organisation for economic co-operation and development [OECD], 2005; Swaffield, 2011;
Torrance, 2012). This research can be seen as an indication of the difficulties and complexity involved in the implementation of formative assessment.

The education system calls for both general successful ways of teaching (e.g. using formative assessment in classroom practice) and subject-specific ways of teaching (e.g. in mathematics). One aim of educational science is to provide research to be used in schools for the purpose of enhancing student learning. Educational research needs to conduct research that is “more useful to practitioners and to policymakers, allowing the latter to make better-informed, less-speculative decisions that will improve practice more reliably” (Burkhardt & Schoenfeld, 2003, p.3). Because of the research evidence showing that the use of formative assessments can be remarkably successful for raising students’ achievement, politicians as well as principals and teachers, in Sweden and elsewhere, show a great interest in formative assessment. Therefore, many reform initiatives in formative assessment are taken. Thus, strong research evidence about characteristics of successful professional development programs in formative assessment is of extreme value for education systems trying to improve teaching and students’ learning. The need of developing, implementing and scrutinizing professional development programs in formative assessment is of essence. We need to know more about how to design effective professional development in formative assessment and what kind of formative assessment to include in such programs.

My thesis reports on an intervention study of a professional development program in formative assessment that was implemented in a middle-sized municipality in Sweden. I have examined the effects and the teachers’ experiences of the program given to 22 primary school teachers in spring 2011. This group of teachers, with their classes, served as informants in all four sub-studies. The program was designed to offer the participants a large amount of time and substantial support to offer them a chance to overcome obstacles in the process of implementing formative assessment activities and to experience the value of using formative assessment.

The aims and research questions guiding this study:

1.1 Aims

The aim of the intervention study presented in this thesis was to see if a random selection of teachers, who got the chance to participate in a professional development program with many contact hours and substantial support of an expert, would implement formative assessment in a way that
could increase their students' learning in mathematics. Another aim was to examine what motivated the teachers to implement new formative assessment activities into their mathematics classroom practices. In this way the thesis contributes to the understanding about how to design professional development that helps teachers to implement formative assessment into their mathematics classroom practice, which has been argued to be the main priority for research on formative assessment (Wiliam, 2010), valid for research purposes as well as for educational practice. The aims are specified in four research questions, which correspond to the four articles (studies) that the thesis is based upon:

1) Do year 4 mathematics teachers in the municipality use formative assessment in their classroom practices [before the professional development program], and if so how? (Study I);

2) What are the characteristics of the changes the teachers made in their formative classroom practice due to participating in the professional development program [after participating in the professional development program]? (Study II);

3) How can the effect, in form of changed teacher practices, be explained and what are the characteristics of the professional development program that the teachers experienced as most important for their development of a formative classroom practice? (Study III); and

4) What effect on student mathematics achievement did the professional development program in formative assessment have? (Study IV).

The next section describes the content of and the coherence between the four articles.

1.2 The coherence between the four articles

The four articles in this thesis describe the four sub-studies that were completed within the same research project. These articles present results about: how formative assessment was used in a random sample of traditional mathematics classroom practices in one municipality (Study I); what impact a comprehensive professional development program in formative assessment had on teachers' mathematics classroom practice (Study II); the reasons for the impact in form of changed teacher classroom practices and the characteristics of the professional development program that the teachers experienced as most important for their development of a formative
classroom practice (Study III); and what impact the professional development program in formative assessment had on student achievement in mathematics (Study IV). The first article is independent from the professional development program in contrast to the other three articles. Also in contrast to the other articles, the sample in the first article included both the same primary school teachers as in the other three studies, but also a group of secondary school teachers.

Next, the intervention, the professional development program, is described.

1.3 The intervention, the professional development program

In this section I describe the professional development program in formative assessment set up as a part of the total research project. The professional development program given to the teachers was implemented in collaboration between the municipality and the university. The idea was to design an intervention, a professional development program that gave a group of teachers a fair chance to learn about and implement formative classroom assessment into the mathematics classroom practice. The group of teachers participating in the professional development program forms the intervention group in the research studies described in this thesis.

1.3.1 Collaboration between municipality and the university

The municipality was responsible for the funding of substitute teachers and the initial communication with the schools. The university was responsible for the design and organization of the research studies and for carrying out the professional development program. Through the cooperation the municipality got the opportunity to receive a well-documented and well-evaluated in-service training for a group of teachers in the municipality. The university got the opportunity to accomplish a large-scale research study about professional development in formative assessment in mathematics.

The research leader from the university employed two doctoral students into the research group and functioned as supervisor for both students during the doctoral studies. The author of this thesis was employed to be responsible for the studies on primary school level and the other doctoral student for the studies on secondary school level. Together these three persons formed the research group. The research group has expanded since then, but in this thesis I refer to the original group. When I refer to the total research project I
mean the work conducted within the original group, even though the focus in this thesis is on the research conducted on primary school level.

As responsible for conducting the professional development program, the research group designed the program. The research leader was the tutor of the weekly meetings in the program. All school leaders responsible for the teachers in the intervention group had agreed on the arrangement of reduction in the teachers’ workload by 20%. One day each week for one term was allocated to meetings at the university. The program is presented in the following.

1.3.2 Design of the professional development program

The core ideas for the design of the professional development program was based on the expectancies that the implementation of high quality formative assessment into classroom practice is a difficult process, and therefore the teachers would need substantial support and a large amount of time to be able to accomplish a successful implementation of formative assessment. Thus, the professional development program was designed to offer support from an expert, many contact hours and a long duration of time. A long duration of time makes possible the practice of formative assessment in the classroom and iterative processes of reflections on those practical experiments. The expert would add the theoretical perspective into those reflections, and the experiences from the practical experiments would be shared and reflected in the group of teachers. In this way, the expert can support the teachers, but at the same time that support is led by the needs implicated by the teachers’ experiences of using formative assessment in their classroom practice. The design was inspired by previous literature about professional development and formative assessment.

The program started in January 2011 and was supposed to end in June the same year. For organizational reasons two days were moved to the beginning of September. The teachers appreciated this move because of the heavy workload in the end of a school year (June) and the possibility to get extra inspiration in the beginning of the school year (starts in the end of August in Sweden). During spring 2011 the 22 primary school teachers, once a week, came to six hours long meetings led by the research leader (144 hours in total). In addition, the teachers had available another 72 hours for reading literature, and planning and reflecting over new formative assessment activities. Together this time would replace the 20% in reduction of school workload offered this spring.
During the meetings, the tutor (the research leader) presented theory about formative assessment conceptualized as one fundamental idea and five key strategies (see Section 2.3 for a description of this conceptualization). The theoretical presentation included the goal of each key strategy complemented with examples of practical formative assessment activities. Another main component in the professional development program was the teachers’ use of these formative assessment activities in their mathematics classroom practices between the weekly meetings. Continuously during the meetings, the teachers discussed the theory presented in literature and lectures, and the practical use of formative assessment in the mathematics classroom.

Much meeting time was used to collaborate and plan for the use of the formative activities between the meetings and for the follow-up discussions after the teachers had practiced the activities in their classrooms. Those discussions included exchange of experiences and emphasized a collective problem solving approach. In this sharing of experiences and ideas the teachers collectively helped each other to overcome perceived setbacks. The teachers discussed and worked in smaller groups and in whole group settings. The collaboration in smaller groups was sometimes arranged from the teachers’ own choices of different mathematical goals and areas to work on. One example is the creation of rubrics, which can be used in the classroom to clarify, share and reach a shared understanding of learning intentions and criteria for success (Key strategy 1). The work from the small groups were later shared and discussed in the whole group of teachers, both before and after the experimentation in the classroom.

Most formative assessment activities presented in the professional development program were elaborated during the meetings, but dependent on the teachers’ needs the time spent on each activity varied. After some months the teachers wanted to make a common document of the presented formative assessment activities. The teachers then used this document in the end of the professional development program to plan for the implementation of formative assessment in their new (year 4) class they would meet in autumn 2011. The research group did not influence these plans or control the implementation of formative assessment during the following school year.

The next section describes the Swedish context in which the professional development program and the research studies were conducted.
1.4 Mathematics education and formative assessment in the Swedish context

It is important to provide the Swedish context for the studies in this thesis, because the context in which an intervention study is situated is always an influencing factor on the results. Knowledge about the context supports the understanding of results and the estimation of the possibilities to transfer the findings to other contexts.

Between 2009 and 2011 funds were allocated to development programs specifically focused on implementing formative assessment in mathematics (Skolverket, 2012a). More recently, major investments on a national level were launched to raise the quality of mathematics education. The amount of scheduled time allocated to mathematics is raised and guidelines provided how to use the extended time to strengthen students' mathematical knowledge (Skolverket, 2012c). A comprehensive national development program in mathematics (Matematiklyftet) is going on from 2012 until 2016 (Skolverket, 2012c) in which formative assessment is embedded in several modules in the program. These reform initiatives show that policy makers in Sweden perceive both mathematics and formative assessment as important areas.

Formative assessment is well promoted in the Swedish school system. The national policy documents clearly announce that assessment should be used for different purposes including a formative use of information, both by the teachers and the students. For example, the goals for assessment and grades in the national curriculum for compulsory school demand schools to strive for all pupils to develop increasingly greater responsibility for their studies and to develop the ability to assess their results themselves and to place their own and others’ assessment in relation to their own achievements and circumstances (Skolverket, 2011c, p. 19).

Sweden follows the trend in several other countries concerning the debate about education and assessment, provoked by the results from international evaluations such as TIMSS and PISA. A decrease in Sweden in student achievement in mathematics is recognised (e.g. Skolverket, 2007, 2009, 2012b). In addition, Sweden is influenced by the international trend of an expanded concept of assessment to include the use of assessment in teaching (i.e. formative assessment) and documents are produced and used to motivate government investments in formative assessment as a mean of improving the quality of primarily mathematics and science education (Levinsson, 2013; Levinsson, Hallström, & Claesson, 2013).
The debate about assessment in Sweden includes decision-making and governance of the school, as well as the assessment in the classroom as a teaching tool and as self-assessment (Skolverket, 2011a). A new approach to students, teachers and learning, in Swedish and international policy, describes the teachers as professional leaders and the students as self-regulating and responsible for their own lifelong learning (Sjöberg, 2009). In Sweden, assessment is seen as needed to bind together the entire teaching process in the goal-oriented school system, and to help students to understand and approach the learning goals (Skolverket, 2011a). In the same text, the use of various assessment strategies, models and tools (e.g. authentic assessment tasks, rubrics, peer-assessment and self-assessment) is promoted to make teaching more effective; to offer all students a chance to review and go on with their progress in school; and for students to use each other as resources for their learning.

A recent publication by the Swedish Schools Inspectorate (Skolinspektionen, 2012) synthesize educational research and formulated keys to success in teaching. The identified success factors were later described in a book for teachers (Häkansson & Sundberg, 2012). Based on the research synthesis, formative assessment is put forward as one of seven dimensions of quality in teaching. Another example of promotion of formative assessment in Sweden is a publication by the national organization for “Swedish Association of Local Authorities and Regions”. They produced their own edition of John Hattie’s (2009) famous overview “Visible learning” that pay special attention to the impact of formative assessment (Sveriges kommuner och landsting, 2011).

The Swedish National Agency for Education has produced a great deal of general recommendations and supportive materials for teachers, included support of teachers’ use of formative assessment (e.g. Skolverket, 2000; 2003; 2011b; 2014). One supportive material for assessment includes one chapter named “Assessment for learning”, but the general focus in the whole material is assessment as observations and information, gathered and interpreted, to lead to some kind of decision and consequence (Skolverket, 2011b, p. 6). The most prominent objectives of such assessments are expressed as to: map skills; assess knowledge; give feedback for learning; highlight the practical knowledge; and to evaluate teaching. Another material uses a model to illustrate the way instruction and assessment are linked (Skolverket, 2014, p. 5). The model emphasizes the importance of what students already know from the beginning, what the teaching aims at, and how to assess in relation to those two factors. Further, the text directs the teacher to clarify the assessment criteria several times during the period of instruction and relate those criteria to the work of the students. The
teacher’s ongoing assessment is outlined to give the teacher feedback about
the effect of the instruction so that adjustments can be done.

The promotion of formative assessment can be connected to the general
trend of evidence-based education (Levinsson, 2013). Reviews of the effects
of different formative assessment strategies (e.g. Black & Wiliam, 1998a;
Hattie, 2009; Hattie & Timperley, 2007; OECD, 2005) have been highlighted
as examples of the value of evidence-based education believed to lead to
improved school outcomes (in Sweden see Prop., 2012/13:30, p. 101) and
leading to extensive investment of implementation of formative assessment
in many countries. Such promotions of evidence-based practice have in
particular been focused on teaching and learning in mathematics (Levinsson,
2013).

Recently the Swedish Government (Regeringsuppdrag, U2013/6845/S)
commissioned the Swedish Research Council to conduct surveys of Swedish
and international research findings relevant to the school system (SFS
2014:1578). The background of the commission was the new national School
Research Institute that the government intended to set up. Each survey
would form the basis for the work in the institute and facilitate the work to
be relevant to schools and preschools. The survey meant to compile results
from studies in different area, but also to describe what kind of research that
was conducted and not conducted in those areas. In mathematics education
three subareas were reviewed. One area was mathematics teaching and
subareas within this area, chosen because of their high relevance for
understanding and improving Swedish mathematics education and students’
knowing of mathematics, were: formative assessment, classroom teaching,
and curriculum programs in mathematics education. Based on the Swedish
context (e.g. Hemmi & Ryve, 2014; Boesen et al., 2014), international
research (e.g. Hattie, 2009; Stein & Smith, 2011) and the findings of the
survey, the authors of the survey recommended the new national School
Research Institute to focus on two aspects of great importance for developing
students’ knowing of mathematics: to synthesize research that support actors
(e.g. teachers and principals) to act in school practices; and to support the
actors to institutionalize developments that has been initiated and
implemented. Formative assessment was mentioned as a specific area
important in which to support teachers (Ryve, in press).

In summary, formative assessment is well highlighted and promoted in
Sweden: in national steering documents for school education; national
support materials; literature reviews and summaries; and reform initiatives.
Special attention is paid to reforms in mathematics education, because of the
decreasing Swedish results in international comparative studies such as TIMSS and PISA.

1.5 Summary

In this chapter I have introduced and contextualized the research studies in this thesis by describing the educational motives as well as research motives, the aims of the thesis and research questions for the four research studies presented in the thesis, and the practical background of the total research project. Moreover, the chapter has described the coherence between the four studies and the professional development that was set up as a part of the total research project.

Sweden follows the trend in several other countries concerning the debate about education and assessment provoked by the results presented in international evaluations. The research-based potential of enhancing student achievement by using formative assessment is reflected in Swedish policy documents, but still a more extensive use of formative assessment in schools seems to be lacking in Sweden similar to many other countries.

The research field asks for research about what characteristics of formative assessment that produces significant learning gains, and research that links characteristics of professional development in formative assessment to successful development of such teacher practice (Schneider & Randel, 2010). The research studies presented in this thesis were designed to meet that request.

The following two chapters review the literature in formative assessment and professional development. I present an extensive description of the conceptualization of formative assessment used in the research presented in the thesis and provide an historical background and context for that conceptualization. In addition, I reflect on the challenges that have emerged during the process of development in the research field. In the chapter about professional development I outline literature on professional development in general and in formative assessment in particular. The chapter about professional development also presents the specific motivation theory used to investigate the reasons for the teachers’ development of their classroom practice made due to the professional development program. Both chapters start with a presentation of the procedures used for searching the literature.


2. Formative assessment

Assessment is used in schools to examine the students’ progress in learning. The need for assessment derives from the unpredictable relation between teaching and learning (Wiliam, 2010). Teacher instruction does not always lead to intended student learning, and consequently: “It is only through assessment that we can find out whether a particular sequence of instructional activities has resulted in the intended learning outcomes” (Wiliam, 2011b, p. 3).

Commonly assessment is seen as a process. In American assessment standards for school mathematics the definition of assessment is:

*The process of gathering evidence about a student’s knowledge of, ability to use, and disposition towards mathematics and of making inferences from that evidence for a variety of purposes. Assessment is a term that has often been used interchangeably with the terms testing, measurement, and evaluation, or to distinguish between student assessment and program evaluation. In this document assessment is used as defined above to emphasize understanding and description of both qualitative and quantitative evidence in making judgements and decisions (NCTM, 1995, p. 87).*

Assessment seen as a process can have different purposes. Some authors have defined formative assessment from the purpose of the assessment and make a distinction between summative assessment and formative assessment (e.g. Cowie & Bell, 1999; Harlen, 2005). Summative assessment has the purpose to summarize a student’s accumulated knowledge at a specific occasion, commonly at the end of working with a specific subject area, while formative assessment has the purpose to form and strengthen students’ learning process during the work (Harlen, 2005). Both forms of assessment are justified and have different roles in education (Wiliam & Thompson, 2008). Other definitions of formative assessment emanate from the function of the assessment (Black & Wiliam, 2009). In such a definition, the assessment is defined as formative when evidence of student learning actually is used in the teaching and learning in the classroom (Black, Harrison, Lee, Marshall, & Wiliam, 2004). Using assessment in this way is sometimes presented as the bridge between teaching and learning (Wiliam, 2013). In other words, the function of the bridge has to be used in formative assessment, it is not sufficient to have a purpose of using it.

This chapter presents the conceptualization of formative assessment used in the total research project. The chapter also briefly outline the historical development and the wider field of research in formative assessment to contextualize the conceptualization we used. The subchapters are: procedures used for searching literature in formative assessment (2.1);
development of the field and conceptualization of formative assessment (2.2); the definition and framework of formative assessment used in the research project (2.3); formative assessment and student learning (2.4); challenges and difficulties in implementation of formative assessment (2.5); and challenges within the research field of formative assessment (2.6). The chapter ends with a summary (2.7).

2.1 Procedures for searching literature in formative assessment

The literature relevant for the research in this thesis lies in the overlap of three fields of interest: formative assessment, mathematics education and professional development. The most relevant literature lies in the area where all three fields overlap and least relevant are literature not connected to formative assessment at all. In this subchapter, the general procedures are described for searching literature related to all three research fields, although the specific procedures for the literature review in the field of professional development is described in the beginning of the next chapter about professional development (Chapter 3).

Three methods are used in this thesis to search for literature: (A) snowball method, (B) selected database and journal search, and (C) an extended systematic search. The author of this thesis used the first two methods. Four people (included the author of this thesis) in the research group of formative assessment used the last method. Before I present those three methods, I give some general comments on the procedures.

The literature review was guided by the following questions, which corresponds to the aims of the thesis and cover the relevant areas of the thesis:

1: To what extent and in what ways are formative assessment used in Swedish mathematics classroom practices?

2: What are the characteristics of formative assessment practices shown important for successful student learning?

3: What ingredients in professional development programs are important to support teachers to use effective formative assessment in their classroom practice?

4: What is the impact of formative assessment on student learning in mathematics?
Scanning the literature in the field of formative assessment is challenging because of the various terms that are used for the same concept (e.g. assessment for learning) and the different meanings authors put into the term formative assessment (this will be described below). In addition, formative assessment includes important strategies for formative assessment. The term formative assessment is sometimes used to describe one strategy, and sometimes used to describe more than one of the strategies. Both formative assessment (and alternative terms) and strategies (e.g. feedback, self-regulation, self-assessment, and peer-assessment) are used in these literature search procedures. Most relevant for the research project is the literature that conceptualized formative assessment as including all strategies.

The snowball method (A) was used from the very beginning in 2010. The search in databases and journals (B), as well as the extended systematic search (C), were conducted during 2014. Methods B and C used search terms such as: formative assessment; assessment for learning; classroom assessment; formative feedback; feedback; self-regulation; self-regulated learning; self-assessment; and peer-assessment; peer-assisted learning. All search terms were not used in all the searches. The search term “mathematics” was included in searches in databases that are not limited to mathematics.

Next I describe the procedures for each of the three search methods.

### 2.1.1 Snowball method

This method is specifically useful when finding literature that is very important for your research. The method enables both to look for old and new literature on the same subject. The first type of the snowball method leads to earlier publications by looking at the literature list of a publication. The second type, finding publications that have cited a given publication, will lead you to later publications on the same subject.

From the start of the research project in 2010 I used the first form of snowball method leading to earlier publications. First, the handbook in formative assessment (Cizek, 2010) and other anthologies were used. Second, articles found in scientific databases (e.g. ERIC) were used (where I used search terms but not as systematic as in method B and C). In both cases the reference list in the book chapter or the article were used to identify publications of potential interest for the project. The second form of snowball method searched for more recent publications on the same subject.
by using the citation list in a database. This method was used in a later stage but not to a great extent.

2.1.2 Searches in selected databases and journals

In this second method I searched for literature in selected databases and journals. The two databases Mathematics Education Database (Math Educ) and ERIC were chosen because of their relevance to the field of interest. Four journals were chosen because of their high rank and/or as possible prospective journals to publish our own articles in: Journal for Research in Mathematics Education (JRME), Journal of Mathematics Teacher Education (JMTE), Nordisk matematikk didaktikk/Nordic Studies in Mathematics Education (NOMAD), and Educational Studies in Mathematics (ESM). In ERIC the search term mathematics was included and in searching NOMAD also Swedish search terms were used. I first present the result in form of number of publications. After that I present the procedures used in identifying the relevant literature.

Table 1. Number of publications found in the literature search in databases (Math Educ; ERIC) and journals (JRME; JMTE; NOMAD; ESM). The number of publications found for Swedish search words is presented in brackets.

<table>
<thead>
<tr>
<th>Search term</th>
<th>Math Educ</th>
<th>ERIC</th>
<th>JRME</th>
<th>JMTE</th>
<th>NOMAD</th>
<th>ESM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formative assessment</td>
<td>146</td>
<td>603</td>
<td>790</td>
<td>17</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Assessment for learning</td>
<td>1246</td>
<td>2024</td>
<td>265</td>
<td>103 (2)</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Classroom assessment</td>
<td>663</td>
<td>1281</td>
<td>252</td>
<td>86 (0)</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Feedback/Formative feedback</td>
<td>43/687</td>
<td>54/6</td>
<td>8/0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-regulation/Self-regulated learning</td>
<td>14/12</td>
<td>11/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-assessment</td>
<td>50</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The search terms were used without restriction to title or key words. This kind of search method generates lots of irrelevant posts. (For example in NOMAD all individual word in assessment generate publications). I did not scan more than 200 publications in any of the search results presented in Table 1. At this point I found that the publications were not relevant because of date of publication or not within the desired field. The most relevant publications were read through completely. Preferences for what publications were most relevant are described above, with a remark that all preferences were not demanded for each individual publication. Publications (article or book chapter) that were not possible to find in any easy way (through the library or Google scholar) were not included in the review.

2.1.3 An extensive systematic search in Web of Science and Swepub

From the Swedish Government, the Swedish Research Council was commissioned to conduct surveys of Swedish and international research findings relevant to the school system. In mathematics education, formative assessment was identified as one of three subareas of high relevance for understanding and improving the Swedish mathematics education and students’ performance in mathematics (see Chapter 1.4). Our research group in formative assessment in Umeå was asked to make the literature survey in formative assessment. In addition to the research group presented earlier in this thesis, one more doctoral student joined us in this assignment.

The research questions in this survey corresponded to the guiding questions above with one additional question asking: “What type of research is used to answer the other four questions?” In the survey we used the databases Web of Science to search for international publications and Swepub to search for Swedish publications. In the Swedish review we included journal articles, books, conference papers and dissertations, but in the international review we only included journal articles. We limited the articles to be published between 2005-2014. The search terms were chosen to find articles that considered formative assessment together with the school subject mathematics and classroom practice (i.e. teaching and/or student learning).
The procedure for using the search terms had two strands. One strand used the different search terms for formative assessment together with student (student*). The other strand used the same search terms together with professional development (professional development*). The methodological design was inspired by the ten processes for systematic literature reviews introduced by Gough, Olives, and Thomas (2013). All details in the methods used in this literature review are presented in the final report (Palm, Andersson, Boström & Vingsle, in press).

The subchapters below present some of the literature that was found through these three literature search methods.

2.2 Development of the research field and concept of formative assessment

Early work in formative assessment concerns evaluation of curricula. Scriven (1967) was the first to propose the distinction between the summative and formative roles of evaluation. This distinction was made in the context of school programs and curricula evaluation. Bloom and colleagues (1971) extended the use of the term to also consider teaching and learning. In doing so, he made a distinction between evaluation and assessment. Evaluation refers to the act of ascribing worth or merits, for example score giving. Assessment is broader and refers to the whole process of planning, gathering and using information for a certain purpose. Formative assessment is sometimes defined by the absence of the evaluation act (Cizek, 2010).

Bloom’s distinction between summative and formative assessment and his recognition of feedback in the instruction process were later elaborated. Ramaprasad (1983) emphasized feedback as a system and noted that: “Feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way” (Ramaprasad, 1983, p. 4). Sadler (1989) pointed out that the information becomes feedback only when the information actually is used to alter the gap (Sadler, 1989, p. 121). All researchers do not agree on such a restriction, but acknowledge that the use of assessment information to improve learning cannot be separated from the instructional system within which it is provided (Wiliam, 2011b, p. 4). Black and Wiliam (1998a) noted that for assessment to function formatively, the feedback information must be used and consequently formative assessment cannot be detached from the learning milieu in which it is undertaken. In summary, it was not until the late 1980s that classroom assessment practices were seen as possible to both afford and constrain student learning and it was during the 1990s that many
researchers started to study the idea of using assessment in instruction (Wiliam, 2011b).

Several studies have investigated ways of integrating instruction and assessment without using the term formative assessment or alike. Best known might be the Cognitively Guided Instruction (CGI, see e.g. Franke, Carpenter, Levi, & Fennema, 2001). In addition, a number of reviews highlight the importance of using assessment to inform instruction (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Black & Wiliam, 1998a; Crooks, 1988; Fuchs & Fuchs, 1986; Natriello, 1987). Black and Wiliam are known by their research synthesis that drew together a wide range of research findings relevant to the notion of formative assessment defined as “encompassing all those activities undertaken by the teachers, and/or by their students, which provide information to be used as feedback to modify teaching and learning activities in which they are engaged” (Black & Wiliam, 1998a, p. 7-8). This article that highlighted the potential of formative assessment also contributed to the raise of interest for the concept and stimulated further research.

However, as Bennett (2011) points out, the effects of such changes in practice were unclear due to the lack of consensus about the term formative assessment. Researchers using the term formative assessment have focused on different ways of doing formative assessment. Consequently the term formative assessment has different meanings for different researchers. In addition, alternative terms are used. One example is “assessment for learning” first used by Harry Black in 1986 to emphasize the function of the assessment and later brought to a wider audience by Mary James at the annual meeting of the Association for Supervision and Curriculum Development (ASCD) in 1992 (Wiliam, 2011b). The term assessment for learning was later popularized in North America by Rick Stiggins (2005).

The Assessment Reform Group (ARG) in the United Kingdom considered that the openness in the interpretation of the term formative assessment often reduced the meaning to no more than frequent assessment and assessment planned at same time as the teaching (Broadfoot et al., 1999). They suggested using the alternate term assessment for learning to overcome the ambiguous use of the term formative assessment. They defined assessment for learning as “the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there” (Assessment Reform Group, 2002, p. 2; Swaffield, 2011, p. 436). In this definition the intention behind the assessment forms the distinction between different purposes of assessment. Assessing in order to teach better is assessment for
learning. Assessing in order to grade students, to rank them or to give them a score on a test, is assessment of learning. Thus, even if assessment of learning can and do affect learning it is more appropriate to talk about effectiveness in assessment when the intention of the assessment is for learning (Swaffield, 2011).

The emerging interest and use of the concept formative assessment has changed the view of assessment as exclusively a teacher activity (Brookhart 2011). The student-centred aspects have attracted more attention in recent years. Such a shift of focus implies significant consequences. For example, viewing the student as the key agent and main user of the information, clearly demand the teacher feedback to be understandable to the student (Stiggins, 2005). The interaction in the assessment process was put forward in a definition adopted at an international conference on assessment for learning in Dunedin in 2009: “Assessment for Learning is part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning” (Klenowski, 2009, p. 264).

Much work has been done to define the characteristics of such classroom practice. Cizek (2010, p. 7) has synthesized the recent literature to establish key characteristics of formative assessment, all identified by their potential to maximize the achievement, development and instructional benefits. Those characteristics harmonize with the key strategies used in the research presented in this thesis. However, Black, Wiliam and colleagues (Black et al., 2004) have promoted an important distinction between the terms assessment for learning and formative assessment. They mean that you can find classrooms with formative intention but with very little formative action. These authors argue that assessment for learning becomes formative assessment only when the evidence of student learning is actually used to adapt the teaching to meet students' learning needs. If you are not using the information to do something that you could not have done without the information, you are not doing formative assessment (ibid.). Thus, Black and Wiliam (1998b), but also other authors (Cowie and Bell, 1999; OECD, 2005; Shephard, 2005) have emphasized the function in formative assessment before the characteristics of formative assessment.

Black and Wiliam had not developed their extensive framework of five key strategies before the review (Black & Wiliam, 1998a) and the following summary (Black & Wiliam, 1998b) that became a landmark for research on the potential use of formative assessment to increase student learning. The studies in the review by Black and Wiliam were diverse, but the authors ascertained principles about how to create effective and successful classroom
practices. The review resulted in the conclusion that attention to the use of assessment to inform instruction, particularly at the classroom level, in many cases doubles the speed of student learning (Black & Wiliam, 1998a). Their recent work has attempted to integrate the different ways of doing formative assessment into a unified whole and to specify classroom strategies and practical techniques that teachers can use to improve the quality of evidence on which the instructional decisions they, and their students make (Black & Wiliam, 2009; Wiliam, 2011b).

The research presented in this thesis used a theoretical framework that focuses on the function of formative assessment and encapsulates all the identified strategies in formative assessment. In the four empirical studies presented in this thesis formative assessment is conceptualized as one fundamental idea (or “big idea”) and five key strategies (Wiliam & Thompson, 2008). In the next section that framework is described in detail.

2.3 Definition and framework of formative assessment used in the research project

In this section I outline the definition of formative assessment and the framework of five key strategies and a fundamental idea (or “big idea”) used in the research project. I describe how these five key strategies can interact and strengthen each other to support the fundamental idea in formative assessment. Further, I suggest that from this conceptualization it is also possible to view three dimensions in formative assessment. Two dimensions are connected to the five key strategies and one to the length of the assessment cycle. These three dimensions indicate three possible directions to develop a formative assessment classroom practice, which are suggested to create new opportunities for student learning.

2.3.1 Definition of formative assessment

Commonly, formative assessment is seen as a strategy of instruction, where assessment is used with a main purpose of supporting learning and a function of using the information from the assessment to adjust teaching to better meet the needs of the students. Wiliam and colleagues’ theorizing work was influenced by their extensive research on teachers’ classroom practice in the King’s-Medway-Oxfordshire Formative Assessment Project (the KMOFAP project, Black, Harrison, Lee, & Marshall, 2003; Black & Wiliam, 2003). In recent work, Wiliam and colleagues have elaborated on their previous definition of formative assessment (Black & Wiliam, 1998a) into a more detailed definition:
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 next steps in instruction that are likely to be better, or be better founded, than the decisions they would have taken in the absence of evidence that was elicited (Black & Wiliam, 2009, p. 9).

This definition comprises the diverse ways of using formative assessment in the classroom. Moreover, the definition clearly demands every formative strategy to comply with the fundamental idea of using the evidence of student learning in taking decisions about instruction. The authors make clear that the term “instruction” is used in the sense in which it is used in the United States—the design of learning environments—and the “next steps in instruction” can be taken by teachers, learners, or their peers, or any combination of these three (Wiliam, 2011, p. 11).

In our research project we have used this definition by Black and Wiliam and the framework that conceptualizes formative assessment as a practice based on an adherence to a fundamental idea (“the big idea”) and a use of the five key strategies (Wiliam, 2010; William & Thompson, 2008), which are:

1. clarifying, sharing and understanding learning intentions and criteria for success
2. engineering effective classroom discussions, questions, and tasks that elicit evidence of learning
3. providing feedback that moves learners forward
4. activating students as instructional resources for one another
5. activating students as the owners of their own learning

The matrix in Figure 1 visualizes how three processes (horizontally) and three agents in the classroom (vertically) construct the five key strategies in formative assessment. When constructing the framework, Wiliam and Thompson (2008) drew on Ramaprasad (1983) to formulate the three key processes in teaching and learning: establishing where the learners are going in their learning; establishing where the learners are in their learning; and establishing what needs to be done to get where they are going. These three processes (on the horizontal) are central for the fundamental idea of using evidence of student learning in decisions about how to proceed in the instruction. The three agents (on the vertical) who are responsible for the learning in the classroom are defined as the teacher, the learner and the peers.

20
<table>
<thead>
<tr>
<th>Teacher</th>
<th>Where the learner is going</th>
<th>Where the learner is right now</th>
<th>How to get there</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Clarifying learning intentions and criteria for success</td>
<td>2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding</td>
<td>3. Providing feedback that moves learners forward</td>
</tr>
<tr>
<td>Peer</td>
<td>Understanding and sharing learning intentions and criteria for success</td>
<td>4. Activating students as instructional resources for one another</td>
<td></td>
</tr>
<tr>
<td>Learner</td>
<td>Understanding learning intentions and criteria for success</td>
<td>5. Activating students as the owners of their own learning</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1. Matrix showing the relation between the five key strategies, key processes in teaching and learning, and agents in the classroom (After a figure in Black & Wiliam, 2009, p. 8)*

The definition and framework by Wiliam and colleagues emphasize the teacher’s decisions (instead of intentions) as being at the heart of formative assessment (Wiliam, 2011a). This conceptualization assists in the design of the assessment process. If the assessments are designed without any clear decision in mind, there is a good chance that the information from the assessment will be useless (e.g. when the assessment is planned or designed by others than the teacher). As the author suggests, a feasible way is to design assessments backward from the decisions. The idea is that the teacher searches for relevant evidence to make decisions in a smarter way (ibid.). Then the teacher in advance knows what to do with the information as they have already thought of alternative instructional decisions before information was collected.

The framework can be used by the teachers, but as the authors illustrate the big idea and five key strategies do not constitute all components of instruction: “The five key strategies are, of course, not the only important processes in instruction, but they do appear to be powerful lenses for thinking about practice, and thus for supporting teachers to engage with wider issues of psychology, pedagogy, and curriculum” (Wiliam, 2010, p. 37).
From the broad definition of formative assessment it can be relevant to talk about a formative assessment classroom practice. Thereby, the relevance of using the word “assessment” has sometimes been questioned, but the value of using the term assessment in the terminology refers to “that it is illuminating to draw attention to the fact that the processes under consideration can be thought of assessment processes” (Wiliam, 2011b, p. 11).

The next section shows how the framework that comprises the different ways of doing formative assessment reveals how different strategies work together in the formative assessment process.

2.3.2 Interaction between the key strategies

The different key strategies in formative assessment are connected and sometimes dependent on each other’s existence and performance. In other words, the key strategies interact and can strengthen each other (or weaken if they are not used properly). As an example I describe how Key strategy 1 can be connected to all the other key strategies.

The first key strategy has a central position, which means that the potential of several other key strategies more or less depend on how learning intentions are clarified and shared amongst the teacher and students. For example, the learning intention should be clear to the teacher when the teacher asks questions or provides tasks to the students. Otherwise, the teacher might not get the relevant and needed information about students’ learning. The same works for feedback. The teacher feedback can more effectively help the student to take the next step in learning if the feedback is directed towards a specific learning intention.

The learning intentions also need to be clear to the students if they are to obtain the opportunity to be engaged and involved in the learning process (their own or their peers’). In the learning process the students can receive feedback from the teacher and from their peers. The feedback can also be internal self-feedback. If the students have a clear perception of learning intentions and success criteria, it will be easier for the students to understand and make use of the feedback they receive. It will also be easier for them to provide appropriate feedback to the teacher, their peers and to themselves.
Thus, there are interaction and synergy effects between different key strategies. The key strategies also correspond to the fundamental idea, which I will illustrate in the next section.

### 2.3.3 Three dimensions for development

The conceptualization of formative assessment used in this thesis makes it possible to view how teachers can develop their formative assessment classroom practice in three dimensions. Two of the dimensions in formative assessment are illustrated in Figure 1 above: the horizontal dimension of the processes related to the fundamental idea, and the vertical dimension of agents in the classroom. The third (time) dimension describes the length of the formative assessment cycles. Below I further illuminate those three dimensions.

From the fundamental idea point of view (the horizontal in Figure 1) a formative assessment activity that strengthens any of the three processes in this dimension will raise the potential inherent in formative assessment. Here I call those three processes the directional process (where the learner is going), the baseline process (where the learner is right now) and the process of movement (how to get there). The first key strategy corresponds to the directional process in the fundamental idea and apparent in Figure 1 all the key strategies connect to one or two of the processes of the fundamental idea. For the teacher, the elicitation of students' understanding corresponds to the baseline process and feedback corresponds to the process of movement in the right direction.

Drawing on Perrenoud’s (1998) characterization of formative assessment as regulation of learning, William (2007b) uses the phrase “keeping learning on track” to convey the grounds for formative assessment. For Perrenoud the nature of the tasks planned for a lesson would impact on the scope and potential of subsequent interaction and feedback exchange between teacher and students in that lesson. Perrenoud distinguishes between traditional sequences of activities within lessons, which he describes only allow for ending up with remediation of narrowly prescribed concepts and those lessons where the activities, in order to take the learning forward, are adjusted once they have been initiated (Perrenoud, 1998, p. 88). Citing Kilpatrick, Swafford, & Findell (2001), William refers to how exemplary mathematics teachers design “teachable moments into their lessons” (William, 2007b, p. 1089). To occur, such teachable moments need to be prepared. In summary, to keep learning on track and make use of such teachable moments, the teacher must be clear about what information is
needed, what tasks to use to elicit that information and be prepared how to respond to various information about student learning. These are the three processes in the horizontal dimension in the framework of formative assessment presented above.

The second dimension (the vertical in Figure 1) in formative assessment is represented by the three agents who are responsible for the learning in the classroom. Those agents are defined as the teacher, the learner and the peers. The engagement of students in the assessment process is not directly required for the teachers’ use of the fundamental idea, but lack of student engagement excludes one dimension of formative assessment. For example, if all students do not engage in the learning situation, evidence from all students’ learning cannot be collected. Furthermore, how the teacher acts to offer opportunity for student engagement is crucial for the students’ active agency in the classroom (i.e. people’s capacity to make choices and to impose those choices on the world) (Björklund Boistrup, 2010). Thus, development of this second dimension put demands on the quality of the interaction between the agents in the classroom.

In an optimal formative assessment classroom practice the interaction works and all agents are seen as resources in the joint learning process implicated in the cyclical process of formative assessment. To establish such high quality interactions, the climate of the classroom becomes critical, particularly for errors and disconfirmation to be welcomed and used by the students (and teacher) as leading to future learning (Hattie & Timperley, 2007, p. 100). In the best scenario, it is seen as favorable to identify mistakes or misunderstandings, because it means a potential to elaborate on that mistake to take the learning a step further. To be noted, the responsibility of the teacher is not reduced but shared and complemented with the students’ responsibility. In this way the roles alter between the teacher and the students in the learning situation (Black & Wiliam, 2009). Black and Wiliam (1998a) have pointed out the importance of changing the interaction, but also that such changes are demanding:

> It is hard to see how any innovation in formative assessment can be treated as a marginal change in classroom work. All such work involves some degree of feedback between those taught and the teacher, and this is entailed in the quality of their interactions which is at the heart of pedagogy. The nature of these interactions between teachers and students, and of students with one another, will be key determinants for the outcomes of any changes (p. 7).

The third dimension refers to the length of any cycle of formative assessment, that is, the process of eliciting information of students’ learning and using this information in decision about how to better meet the needs of
the students and take the learning a step further. The length of the formative assessment cycle can vary from shorter than a minute to more than a year. Some authors have argued for excluding long assessment cycles from being formative (see e.g. Swaffield, 2011), but Wiliam and Thompson (2008) prefer to categorize the feedback loop in classrooms as formative assessment used either on a short, medium or long cycle. Short-cycle formative assessment might not always be preferable, but the consensus in the assessment community is that learning benefits are more evident when “results are available quickly enough to enable teachers to adjust how they’re teaching and students to alter how they're trying to learn” (Popham, 2006, p. 86).

It is here suggested that teachers’ development of formative practices in three dimensions create new opportunities for student learning. The next section explores some of the research literature connecting formative assessment to student learning.

### 2.4 Formative assessment and student learning

Even though the research presenting the impact from formative assessment on student achievement has been criticized (see Chapter 2.6), few researchers disagree on the potential in the use of formative assessment. The discussion about possible explanations behind the effect from formative assessment on student learning will vary depending on what learning perspective is used. It is possible to build on cognitive as well as social grounds to explain the mediating role of formative assessment in student learning. Below I show that different learning perspectives are needed to explain the effect from formative assessment on student learning and that commonly used components (or strategies in formative assessment) are feedback and self-regulated learning. Research showing the effect on student achievement often includes those two components.

Below I first examine student learning perspectives in formative assessment and after that I present literature about the effect on student achievement from formative assessment (or assessment for learning) and from individual strategies as feedback, and self-regulation. The focus is on the relationship between formative assessment and student achievement in mathematics.

#### 2.4.1 Student learning perspectives in formative assessment

Formative assessment does not in itself carry any particular view of what should be learnt or how learning takes place (Wiliam, 2007b). Rather, the
complexity inherent in formative assessment implicates the need to use different perspectives: “Thus it is clear that the complexity of the situations in which formative feedback is exchanged is such that they can only be understood in terms of the several theoretical perspectives required to explore the different types of issue involved” (Black and Wiliam, 2009, p. 28).

Many articles about formative assessment consider theories about feedback interactions and/or theories about self-regulation as mediating processes between formative assessment and student learning. For example Clark (2012), from a post-structuralist foundation, explains the mediating process in formative assessment as lying in the empowering feedback circulated among the students. In these feedback processes the learning objectives and the criteria of success become transparent, which strengthen the learning process for the students. According to Clark, the feedback also makes experiential tacit knowledge that is ‘hidden’ within the learner transparent, explicit and available. This view to a great extent corresponds to Hattie’s (2009) promoted classroom practice called “visible learning”.

Clark also describes how self-regulated learning has been moved away from reductionist perspectives toward the holistic study of social context (Clark, 2012, p. 217). The framework of formative assessment by Wiliam and colleagues’ (see Chapter 2.3) exemplifies such a holistic view. In this framework, understandings from the research community are synthesized and used in new ways and put into a functional framework for how to support self-regulated learning strategies (ibid.). Perrenoud (1998), in a comment to Black and Wiliam’s landmark article in 1998, called for the theoretical perspective of learning to be clear to enable understanding of the mediating function of formative assessment. Wiliam and colleagues met this requirement by considering theories that come under the general description of self-regulated learning. They proposed that theories about self-regulated learning and socio-cultural theorization of classrooms complement each other, and that both are needed, because theories about self-regulated learning recognize the learner as an individual thinker and socio-cultural theorization pays attention to learning in the context of discourse.

*Our approach indicates that any evidence of formative interaction must be analysed as reflecting a teacher’s chosen plan to develop learning, the formative interactions which that teacher carries out contingently within the framework of that plan—as realised in the social world of the classroom and school—and the internal cognitive and affective models of each student of which the responses and broader participation of students provide only indirect evidence (Black & Wiliam, 2009, p. 26).*
Wiliam and colleagues have used a model of self-regulated learning by Boekaerts to bring together motivational and cognitive perspectives on self-regulation (Wiliam, 2007b), but there are also other models grouped under different perspectives of learning, such as cognitive behavior modification, direct teaching of skills and strategies, and those that take a socio-cultural approach (see e.g. Boekartes & Corno, 2005).

To advance the proposition that the theory and practice of formative assessment combines cognition, social, and cultural theories, Clark (2012) points out the importance of feedback, self-efficacy and collective efficacy and the issue of context. He means that feedback is not transmitted in a vacuum, free from affect from environment. Related to self-efficacy, the sub-processes of meta-cognition (planning, monitoring, and evaluation) in self-regulation generate internal feedback (ibid.). Such internal feedback is shaped by discourse and the social context. Related to collective efficacy, the involvement of students, in peer groups and whole class discussions, creates more effective social learning of the students (ibid.).

Such a contextual perspective puts the learning environment in centre. Björklund Boistrup places her thesis in a critical and sociocultural paradigm. She has investigated discourses of classroom assessment in mathematics, how these can be construed and what affordances can be connected to students’ active agency (i.e. capacity for people/students to make choices and to impose those choices on the world) and learning. In her study she identified four discourses with different affordance for students’ active agency and learning (Björklund Boistrup, 2010). The notion of agency is seen as something that people have affordances to take. “It is a matter of a person being active or passive” (ibid., p 4), which implicates that the four different discourses supported students’ self-regulation in different degrees.

The activity of the students is salient regardless of whether you focus the individual learner or the learner situated in a learning community. Therefore, it is suggested that the mediating function of formative assessment should be explained by different theorizations.

The next section summarizes and exemplifies literature that empirically link formative assessment classroom practice to impact on student achievement.
2.4.2 The relationship between formative assessment and student achievement

This chapter provides literature showing the relationship between formative assessment and student achievement with a focus on student achievement in mathematics. Both effect-studies and correlation-studies are included. The intention has been to overview, rather than providing details.

Both Black and William’s landmark review (1998a) and Hattie’s meta-analysis (2009) demonstrated that different formative assessment strategies improve student achievement. This improvement has also been noted in research reviews focusing on each specific strategy such as feedback (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008), self-regulated learning (Dignath & Büttner, 2008), self-assessment using rubrics (Panadero & Jönsson, 2013), and peer-assisted learning (Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003).

Also, research reviews with a focus on mathematics have shown strong relationships between student achievement and teachers’ adjustment of instructional activities based on evidence of student learning gathered from short tests (National Mathematics Advisory Panel, 2008; Yeh, 2009). The review by Palm et al. (in press) reviewed studies on the relation between several formative assessment strategies and student achievement in mathematics. The review found several studies showing significant impact of teachers’ adjustment of instructional activities, based on information from short tests, on student achievement in mathematics.

Other studies report empirically found relationships between other individual formative assessment strategies and student achievement in mathematics. Such studies include investigations about the positive impact of feedback (e.g. Brosvic, Dihoff, Epstein, & Cook, 2006; Roschelle et al., 2010). Other studies found positive relationships between self-regulated learning strategies and student achievement in mathematics, both from intervention studies and correlation studies (Brookhart, Andolina, Zuza, & Furman, 2004; Dresel & Haugwitz, 2008; Kistner, Rakoczy, Otto, Dignath-van Ewijk, Büttner, & Klieme, 2010; Kramarski & Mizrachi, 2006; Kramarski & Revach, 2009; Lazakidou & Retalis, 2010; Metallidou & Vlachou, 2010; Throndsen, 2011). In addition, the positive impact from peer-assisted learning concluded in the review by Rohrbeck et al. (2003) included many studies that focused on students’ mathematics achievement.

The articles analyzed in the research review by Palm et al. (in press) showed that all of the strategies for formative assessment were positively correlated...
with student achievement in mathematics. In studies reporting effect sizes, those were usually between medium and large. However, this literature review, which focused on formative assessment and student achievement in mathematics, also confirms a large variance in the effects reported in different studies and the difficulties in interpreting the results. For example, the meta-analysis by Dignath & Büttner (2008) of 74 intervention studies in self-regulated learning (including 28 in mathematics) reported a high mean effect size (0.96) for grades 1-6, and lower mean effect size (0.23) for grades 7-10.

Summarizing the literature it can be concluded that there is a strong relationship between formative assessment and student achievement. However, few studies used an extended conceptualization of formative assessment that comprises all formative assessment strategies. Most studies investigated the impact on student achievement from individual or a few formative assessment strategies. There are at least two possible reasons for the absence of studies on formative assessment conceptualized as a unity of integrated strategies. One is that the framework is relatively new. The five key strategies was first presented in Leahy, Lyon, Thompson and Wiliam (2005) and presented in a figure by Wiliam and Thompson in 2008. Wiliam and colleagues used a prior version of the framework in their empirical study of putting formative assessment into practice (Wiliam et al., 2004). Another explanation could be the complexity and demands such a classroom practice implicates.

The significance of the design of professional development was also indicated in the meta-analysis by Dignath and Büttner (2008), which showed that the more teaching and training sessions that were part of the intervention, the better students performed in mathematics. The next subchapter more deeply examines the challenges connected to the implementation of formative assessment into classroom practice.

2.5 Challenges in implementation of formative assessment

From research it is clear that implementation of formative assessment is not straightforward (Black & Wiliam, 1998b). This can be explained from the complexity of formative assessment practice in itself. In addition, external factors, such as trends of accountability, makes the implementation of formative assessment more difficult. In this chapter I examine some of the challenges in putting formative assessment into practice in a way that has the potential to enhance student learning and some factors that can impede that process of implementation.
Intervention studies with negative results can provide useful information if it is clear what type of formative assessment is used and what support is provided to the teachers. The intervention study by Schneider and Meyer (2012) showed no impact on students’ achievement in English language arts and mathematics from an intervention in formative assessment. The study examined the teachers’ practice and found that teachers’ assessment knowledge increased. However, this improvement concerned predominantly the collection of evidence rather than the use of information in instruction and feedback. Using assessment information to plan subsequent instruction tends to be the most difficult task for teachers (Heritage, Kim, Vendlinski, & Herman, 2009). In another study (Smith and Gorard, 2005) the missing impact was explained both by teachers’ and students’ responses. The students did not find the feedback from their teacher useful in supporting them to take the next step in learning and the teachers were affected by the dissatisfaction that the intervention had on the students. The authors suggested that the teachers were not given enough support and attention to deal with such setbacks.

The difficulties and complexities in implementation of formative assessment are described in different terms, for example, by teachers’ use of a “convergent assessment” instead of “divergent assessment” (Torrance & Pryor, 2001). A convergent assessment focuses on procedural compliance and misses to find out how learners are learning and what students can do (i.e. divergent assessment) resulting in a distorted understanding of “assessment for learning” to become “assessment as learning” (Torrance, 2007).

Empirical research studies have found that such an “instrumental approach” of formative assessment results in a restricted undesirable classroom practice (e.g. Hume & Coll, 2009). The researchers in a large intervention study (James & McCormick, 2009; Marshall & Drummond, 2006) called an instrumental use of formative assessment “by the letter” in contrast to the use of formative assessment “by the spirit”. From their intervention study of putting assessment for learning (AfL) into practice, they concluded that:

> Although teachers appreciate practical advice, classroom practices can become ritualised and mechanistic if teachers are not stimulated to think about the principles of learning that underpin them. We made a distinction between those teachers who implemented the ‘letter’ of AfL by injecting AfL practices into what they usually do without changing anything more fundamentally, and those who captured the ‘spirit’ of AfL by integrating practices into the flow of lessons to regulate the learning process itself. This latter required some understanding of underlying principles and we concluded that beliefs and practices are inter-related and need to be developed together. It is not sufficient just to tell teachers what to do (James & McCormick, 2009, p. 982).
From a definition of assessment for learning emphasizing the promotion of student autonomy, only 20% of the teachers in their study used formative assessment by the spirit. The group of teachers who used formative assessment by the letter used strategies that superficially looked like formative assessment (e.g. sharing learning intentions and success criteria, and providing comment-only feedback). These teachers rhetorically used formative assessment, which led to instrumental learning and away from the fundamental idea in formative assessment. In addition, the students of these teachers became more reliant on the teacher instead of the opposite goal of student autonomy. The teachers who captured the spirit of formative assessment integrated formative practices into the flow of lessons to regulate the learning process itself. These teachers employed strategies that promoted the deeper formative assessment principles of student motivation and autonomy in learning (e.g. questioning, feedback requiring students to respond to the comments by further work, and peer- and self-assessment) (Marshall & Drummond, 2006). The authors pointed out that using formative assessment by the spirit requires an understanding of underlying principles of formative assessment.

The studies above exemplify the issue of translating theory into practice, and scaling up from positive individual studies to wide implementation and generally improved practice and achievement, an issue that is underestimated (Black & Wiliam, 2005). Thus, previous research indicates an untapped potential in the use of formative assessment and the difficulties in implementation of formative assessment. In addition, researchers highlight the negative consequences on student learning from a mechanical use of activities associated with formative assessment. For example, that static rubrics for knowledge progression can lead away from learning as a process open to students’ perspective, understanding and learning needs (Crossouard, 2012). This reduction in curricula includes an issue about who has the mandate to decide what high quality learning is (ibid.).

As discussed above, the complexity within a formative assessment classroom practice is one explanation for the difficulties in implementation of formative assessment. Another aspect is the complexity of assessment processes and the several sources that influence them (Tierney, 2006). Reforms that emphasize formative assessment can be impeded by trends of accountability (Darling-Hammond & McCloskey, 2008; Klenowski, 2011; OECD, 2005) and the focus on examination and summative assessment (Brown, Kennedy, Fok, Chan, & Yu, 2009; Wiliam, 2006). Even if there is a possibility for summative and formative assessment to coexist (Black, Harrison, Hodgen, Marshall, & Serret, 2010; Brookhart, 2010) it becomes problematic when summative assessment is what counts in assessment and education systems.
(Bennett, 2011; Black & Wiliam, 2005), and formative assessment is spoken of and evaluated in terms from summative assessment (Gallagher, 2010). Even if there, as in Sweden, is a culture with low accountability and with teachers very much in control of the assessment process (Darling-Hammond & McCloskey, 2008), there can be cultural barriers of putting formative assessment into practice (Levinsson et al., 2013). Formative assessment exists within a larger educational context (Bennett, 2011). It is concluded that the effective integration of formative and summative functions of assessment will need to take different forms in different countries, and is likely to be extremely difficult (Black & Wiliam, 2005). Structural issues, practical barriers as well as misconceptions can hinder integration of formative assessment into classroom practice (DeLuca, Luu, Sun, & Klinger, 2012).

Designers of professional development in formative assessment need to consider research about difficulties and risks in formative assessment interventions, as well as the context for the intervention. For example, James & McCormick (2009) in the citation above concluded that telling teachers what to do is not sufficient since a use of formative assessment by the spirit requires an understanding of underlying principles of formative assessment. It is also proposed that such understanding needs to be developed together with teacher beliefs (e. g Delandshere & Jones, 1999; Dixon, 2011; Dixon, Hawe, & Parr, 2011; James & Pedder, 2006) and practices (Schneider & Randel, 2010). Thus, it can be expected that many teachers need substantial support in the process of developing their formative assessment classroom practice and that giving that support requires several considerations.

There are work completed in the research field of formative assessment connected to the challenges in implementation of formative assessment, but as described in the introduction, more work is needed. There are specific challenges pertaining to research on formative assessment, which is described in the next chapter.

2.6 Challenges in the field of formative assessment research

Demands and critique of the research field of formative assessment come from various theoretical and methodological perspectives. Some authors have contested conclusions about the impact of formative assessment on student achievement (Dunn and Mulvenon, 2009; Kingston & Nash, 2011; Bennet, 2011). The critique is based on issues of conceptualization and methodology, and that problems with these limit the scientific evidence of the effect of formative assessment. The different effect sizes become
complicated to interpret due to the lack of an agreed definition (Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012; Filsecker & Kerres, 2012; Kingston & Nash, 2011; McMillan, Venable, & Varier, 2013). Consequently, best practices related to formative assessment become hard to identify, which calls for a sound research-validated framework for such a classroom practice (Dunn & Mulvenon, 2009). In addition, researchers criticize the methods used in research studies intending to demonstrate the positive effect of formative assessment (Bennett, 2011; Dunn & Mulvenon, 2009).

However, this diversity in conceptualization of formative assessment and methodologies used in the investigations can be viewed as an effect of the different research paradigms in educational research (Filsecker & Kerres, 2012). Furthermore, after a long time of development, the consensus of key theoretical conceptualizations of formative assessment overweight the disagreements (Klenowski, 2009). Thus, the work for a consensus in the definition of formative assessment is neither disregarded nor neglected. The requested work from researchers about consensus in the terminology of formative assessment, a sound research-validated framework for best practices in formative assessment, and appropriate methodologies completed in traditional classrooms, is an on-going process.

### 2.7 Summary

This chapter has provided a description of the conceptualization of formative assessment used in the research presented in this thesis and a historical background and context for that conceptualization. Furthermore, it is shown how the framework that comprises different ways of doing formative assessment reveals how different strategies work together in the formative assessment process and makes it possible to view how teachers’ development of formative practices can be characterized. The chapter also connected formative assessment to student learning. It is suggested that the mediating function of formative assessment should be explained by different theorizations. Empirical studies investigating the potential of formative assessment on student achievement, with a focus on studies in mathematics education, were also presented.

There are challenges pertaining to research on formative assessment as well as challenges and difficulties in implementation of formative assessment. Problems related to research are the diversity in conceptualization of formative assessment and reported methodology issues. The complexity within a formative assessment classroom practice is one explanation for the difficulties in implementation of formative assessment. Another aspect is the
complexity of assessment processes and the several sources that influence them. Designers of professional development in formative assessment need to consider research about difficulties and possible risks in formative assessment interventions, as well as the context for the intervention.

Next chapter presents the review of literature on professional development in general and in formative assessment in particular. The chapter also presents the motivation theory used in the study about why the teachers made changes in their classroom practice due to the professional development program (Article 3).
3. Professional development in formative assessment

This chapter briefly contextualizes the intervention study presented in this thesis within the literature about teacher professional development in general and in formative assessment. Especially the coverage of literature on professional development in general is limited. Other research reviews not presented in this thesis (e.g. Darling-Hammond & Richardson, 2009) also provide insights about knowledge, trends and the progress made in the understandings about professional development. In addition, several different models of teacher development are left outside the literature review (e.g. Clarke & Hollingsworth, 2002; Gregoire, 2003; Guskey, 2002). Thus, I do not attempt to give an overview of the field of professional development. Instead, I focus on two reviews, one of them receiving considerable attention in Sweden (Timperley, Wilson, Barrar, & Fung, 2007). That review was used as inspiration in the design of the intervention, the professional development program, in the intervention study presented in this thesis. The other review concerns professional development through organization of teacher learning communities (Vescio, Ross, & Adams, 2008), which is promoted as an advantageous form of professional development in formative assessment (Wiliam, 2007a). Thus, the literature review on professional development in general is focused, rather than comprehensive. The first review was known before the research project started, the second review was identified during searching the literature. The review of literature on professional development in formative assessment is more extensive. As described above, the most relevant literature concerns mathematics education, formative assessment and professional development for teachers, but literature not relevant to all three areas were also included.

The first subchapter (3.1) describes the procedures for searching literature on professional development with a focus on professional development in formative assessment. The second subchapter (3.2) outlines aspects of teacher professional development programs that have been identified as important in two research reviews. The third subchapter (3.3) presents literature on professional development in formative assessment. In the fourth subchapter (3.4) I link the general literature about professional development to professional development literature specifically about formative assessment. In particular, I use a previous literature review (Schneider & Randel, 2010) to support that linkage. The Schneider and Randel review was also a source of inspiration in the design of the professional development program set up as a part of the research presented in this thesis. Motivation is the driving force of human behaviour instigating,
directing and sustaining efforts such as changing teacher instruction. Thus, motivation has the potential to be a useful perspective when trying to understand teachers’ actions in professional development programmes. The last subchapter (3.5) presents the theory of motivation used in Study III (Article 3) as an interpretative lens in the analysis of the reasons for teachers’ implementation of the content (formative assessment) of the professional development program.

3.1 Procedures for searching literature in professional development in formative assessment

In sub chapter 2.1 the general procedures for searching literature were described. This sub chapter specifies the procedures used for searching literature in the research field of professional development in formative assessment.

The same three methods described above (Chapter 2.1) were used to find literature about professional development, with a focus on professional development in formative assessment: (A) snowball method; (B) selected database and journal search; and (C) an extended systematic search.

3.1.1 Snowball method

The use of the snowball method is consistent with the description in Chapter 2.1. When reading through the most relevant publications (see Chapter 2.1), I summarized any information given about the design of the professional development program or conclusion about important aspects to include in professional development programs in formative assessment.

3.1.2 Searches in selected databases

Literature search procedures targeting professional development in formative assessment were used in the two databases (Math Educ and ERIC), but not in the four journals (JRME, JMTE, NOMAD, and ESM). The same search terms previously used were complemented by the search term professional development.
3.1.3 An extended systematic search in Web of Science and Swepub

As described above (Chapter 2.1.3) one strand in the extended systematic literature search, using the databases Web of Science to search for international publications and Swepub for Swedish publications, included a use of the search terms related to professional development. These were: in-service training, professional development, program, implement (or the Swedish equivalents). One search was conducted for each of these terms. The details in the procedures are presented in the final report (Palm, et al., in press).

3.2 Professional development in general

The field of research about professional development is wide, and it is out of scope in this text to cover it all. I have already mentioned the lack of studies investigating the effect from professional development programs on both teacher classroom practice and student achievement (Schneider & Randel, 2010; Tierney, 2006). In addition, I have stated the importance of describing and analyzing (in our case formative) classroom practice that has led to improved student achievement in order to be able to investigate the relations between characteristics of classroom practice and student achievement.

The first of the two reviews presented in this subchapter (Timperley et al., 2007) was based on 97 empirical studies. The review identified a number of conditions and principles associated with professional learning that impacted substantively on student outcomes. Seven contexts were identified necessary, but not necessarily sufficient on their own: extended time for opportunities to learn; external expertise; being engaged in learning rather than volunteering to change; challenging prevailing discourses (of learning and teaching); participating in a professional community of practice; alignment with trends in wider policy and research; and active school leadership. The seven contexts were later reformulated to 10 principles (Timperley, 2008) that integrate a formative assessment classroom practice into the process of teachers' professional learning and development. These principles were:

1. Student outcomes are the focus and the start in a teacher professional development program. (What are the students’ needs in relation to identified learning intentions?)
2. The teachers need to find the content of the professional development program meaningful and worthwhile. (What knowledge and skills do teachers need to help their students take the next step in learning?)

3. Theory and practice need to be integrated in terms of curriculum, teachers’ practice and knowledge of assessment in the areas that are the focus of the development effort.

4. Assessment is used as a base for teachers’ professional inquiry. Assessment needs to be integrated into the teaching and learning in the classroom so that teachers can determine whether the new practice is successful in their particular context.

5. The teachers need multiple opportunities to learn and apply new knowledge and skills, in an environment characterized both of trust and challenge.

6. The approach should be responsive to whether or not new ideas are consistent with the assumptions that currently underpin practice. If the teachers recognize that student achievement enhances, the teachers’ expectations on their students will enhance.

7. Participating teachers must have opportunities to process new learning together with colleagues.

8. External expertise is necessary to challenge existing assumptions and develop the kinds of new knowledge and skills associated with positive outcomes for students.

9. School leaders have a key role in developing expectations for improved student outcomes and they are responsible for organizing regular opportunities to promote teacher development.

10. Maintaining momentum in the professional development program requires an awareness of the effort and time it takes for teachers to change their teaching in a successful and sustainable manner.

From this perspective, the professional learning supports teachers to evaluate the adequacy of tacit knowledge and routines and to negotiate the meaning of new information in relation to existing knowledge and strengths. The clear goal for the teachers in the proposed model of inquiry based on these principles is the improvement of student achievement. The author suggests that sustained improvement in student outcomes requires teachers with good self-regulatory skills to assess the impact of their teaching (Timperley, 2008):
Sustained improvement also depends on teachers developing professional, self-regulatory inquiry skills so that they can collect relevant evidence, use it to inquire into the effectiveness of their teaching, and make continuing adjustments to their practice. Teachers with these crucial self-regulatory skills are able to answer three vital questions: “Where am I going?”, “How am I doing?”, and “Where to next?” (p. 24)

The authors found that teachers with both inquiry skills and content knowledge, who received support from their leaders, were consistently able to do this in terms of impact of their teaching on student learning.

The model the authors propose includes a sequence of inquiries that combine the elements in the model into a co- and self-regulatory learning cycle: “By ‘co- and self-regulatory’, we mean that teachers collectively and individually identify important issues, become the drivers for acquiring the knowledge they need to solve them, monitor the impact of their actions, and adjust their practice accordingly” (Timperley et al., 2007, p. xlii).

In this model, assessment is seen as a component in teachers’ commitment to change and embed formative assessment within the teachers’ learning and development. It is expected that new teaching practices will be reinforced when teachers observe that they are having a positive impact on student outcomes. Maintaining an improvement in student results requires that teacher practice is based on a strong theoretical foundation and a supportive organizational infrastructure is needed, which requires an awareness of the effort and time it takes for teachers to change their teaching in a successful and sustainable manner (Timperley et al., 2007, p. 195).

The second review (Vescio, Ross, & Adams, 2008) of 11 empirical studies concluded that participation in professional learning communities (PLC) impacts teaching practice as teachers become more student centered. In addition, teaching culture was improved because the learning communities increase collaboration, a focus on student learning, teacher authority or empowerment, and continuous learning (ibid., p. 88). It was prominent that the successful PLCs focused on students’ learning and making use of formative assessment in their practice despite the content in the professional learning communities might not be formative assessment: “In the long run, the data across these studies indicated that a key element of successful PLCs is their pervasive attention to meeting the learning needs of their students” (ibid., p. 88).

Wiliam and colleagues (Wiliam, 2007a) have promoted teacher learning communities as the most promising approach for professional development focusing on teacher actions. From the assumption that teachers need
substantial support and guidance to integrate formative assessment into their practice, the authors argue that teacher learning communities have the potential to provide such support and at the same time put the teachers in charge of their own professional development (ibid).

### 3.3 Professional development in formative assessment

I have already argued for the need of studies about what kind of formative assessment to include in professional development programs and how to design programs that can support teachers to implement such formative assessment. In addition, I have presented studies that show the potential of raising student achievement by using formative assessment, but also research studies that indicate the complexity and difficulties in implementing formative assessment. The intention with this subchapter is to bridge the literature review of formative assessment above to the next subchapter (3.4) that link professional development in general to literature specifically about formative assessment. Therefore, there will be some overlap between the subchapters. As pointed out before, the challenges in implementation of formative assessment have to be considered in the design of professional development in formative assessment.

Implementation of formative assessment as a quick fix is dismissed (Black & Wiliam, 1998b) and to be effective, formative assessment has to be integrated into classroom practice, which would require a fundamental reorganization of classroom operations (Black & Wiliam, 1998a). Cautions are issued for underestimated necessary resources and efforts (Cizek, 2010) and simplistic ideas in implementation of formative assessment (Dixon, 2008; Hayward & Spencer, 2010). A problem with lack of time, sometimes expressed as teachers’ onerous workloads, is recognized in several studies (e.g. Carless 2005; Hayward, Priestley, and Young 2004; James and McCormick 2009). Lee and Wiliam (2005) compare asking a teacher to change their assessment practice with asking a golfer to change his or her swing during a tournament. Teachers are required to maintain the success they obtained with their old routines while developing new routines at the same time.

The use of assessment information to plan subsequent instruction is put forward to be the most difficult task by some authors (Heritage et al., 2009), while the change in roles in the classroom and reconstruction of the teaching contract is pointed out as the most difficult by others (Black et al., 2004; Lee & Wiliam, 2005; Perrenoud, 1998). In conclusion, these difficulties demonstrate that the teachers need to be motivated and supported, as Black
and Wiliam (1998b) pointed out: “Teachers will not take up attractive sounding ideas, albeit based on extensive research, if these are presented as general principles which leave entirely to them the task of translating them into everyday practice” (p. 146).

For many teachers, making learning explicit to students, and promoting their learning autonomy requires the teachers to learn new knowledge (about learning), develop new skills, and reassess their roles (James & McCormick, 2009). Teachers need to learn, as well as their pupils, and schools need to support them in this. This requires organizational learning and need for schools to create a learning culture for the teachers as well as for the students. In other words, organizational conditions important for teachers’ uses of assessment for instructional improvement include leadership, administrative support and collaborative norms and structures (ibid.).

Implications for design of professional development programs in formative assessment were found in individual studies but also in research reviews. A review of 24 diverse studies from 2000-2005 (Tierney, 2006) put forward both time and use of time (in contact hours and in duration) as important factors in professional development programs as well as affording opportunity to collaborate. Tierney promote collaboration between research, policy, and practice communities, but also teacher agency rather than autonomy, as needed to support the move toward more constructive classroom assessment practices. The author also found that teachers identify their colleagues as the most important source of information and the teachers’ preference for information that is immediately and contextually relevant. Young and Kim (2010) agree on the collaborative norms and structures as facilitating factors and the need of access to expertise. These authors put forward time as necessary for change of teachers’ orientation toward assessment as integrated in instruction, change of teachers’ conscious practice, and for instructional experimentation. Their research review of empirical research articles from 1980 - 2008 (Young & Kim, 2010) also identified leadership and its multifaceted roles as a critical organizational condition. The authors put forward the challenges inherent in formative assessment practice and the need to support teachers: “Using assessments formatively in the classroom is not a beginner’s skill. It takes a range of foundational content knowledge, pedagogical understanding, instructional skill, and classroom management to effectively use or implement formative assessment practices” (ibid., p. 9).

Reading through the identified most relevant literature I found that in many cases, the professional development was not described and evaluated.
However, implications for design of professional development in formative assessment from a summary of six articles (Black & Wiliam, 1998b; Cooper & Cowie, 2010; James & McCormick, 2009; Lee & Wiliam, 2005; Marshall & Drummond, 2006; Webb & Jones, 2009) identified eight common important characteristics:

1. Flexibility in the professional development, for example not saying to the teachers how to do and letting them find their own ways of incorporating the lessons and ideas.

2. Support from an expert.

3. Support from other teachers, for example in learning communities.

3. Theoretical understanding about formative assessment.

4. Examples of what the theories means in practice, and practical ideas and techniques.

5. Experiences of using such techniques, for example classroom-based collaborative inquiry practices.

6. Discussions, by some authors expressed as reflection on, in and about action.

7. Time set aside and duration of time;

8. Research evidence for effectiveness or evaluation of effectiveness within the program.

All characteristics were not defined in all six articles, but those eight characteristics clearly appeared as outstanding in the summary of the articles. An additional common comment says that professional development in formative assessment needs to deal with teachers’ feelings and beliefs (e.g. James & McCormick, 2009). To be noted, is that the professional development in these studies is defined in a broad sense, which include for example collaborative work between university researchers and teachers.

Publications found in the extended systematic search for literature (Method C, see Subchapter 2.1 and 3.1) confirmed the above-described weaknesses in information provided in the studies. Three studies about formative assessment included professional development in the broad sense (Phelan et al, 2011; Clarke et al, 2014 and Lee et al 2012) but only the study by Phelan both investigated whether the training has an effect on students’ math skills
and describes the characteristics of the intervention. None of the studies about feedback contained any form of teacher training where the effects of the professional development on teacher feedback and student achievement were investigated.

One review (Montague, 2008) and one meta-analysis (Dignath & Büttner, 2008) about self-regulated learning dealt with interventions, but did not include detailed information about the characteristics of the professional development. The same goes for the five individual articles containing intervention studies on self-regulated learning that were found in Web of Science (Dresel & Haugwitz, 2008; Kramarski & Mizrachi, 2006; Lazakidou & Retalis; 2010; Ness & Middleton, 2012). Dignath & Büttner (2008) noted the lack of studies describing the professional development and pointed out the usefulness of such information to get an idea of the difficulties of implementing an instruction about how to support students' self-regulated learning. This is especially important because their meta-analysis shows that interventions to support students' self-regulated learning run by the researchers had a stronger impact on student achievement than those conducted by the teacher. They therefore call for more studies on the effects of teacher professional development on teachers' implementation of an instruction that supports students' self-regulated learning, and the effects of this implementation on student achievement.

The lack of studies investigating the effect of professional development in formative assessment on student achievement is noticed by Schneider & Ranndel (2010). Studies examining the impact of the integrated view of formative assessment are even more rare. However, there do exist a few published studies. One example is the study by Wiliam, Lee, Harrison, and Black (2004) who reported some positive outcomes on student achievement from a practice based on parts of the framework presented by Wiliam and Thompson in 2008. Two other examples below reported no effect on student achievement, with indications of difficulties and challenges both for the teachers and for the designers of professional development programs.

The first example of an integrated view of formative assessment is the Classroom assessment for student learning (CASL) framework by Stiggins, Arter, Chappuis and Chappuis (Arter, 2009). This framework also builds on the idea of using evidence of student skills and understanding to modify teaching and learning, and it also emphasizes the inclusion of the students in this process. CASL includes seven strategies that are similar, but not identical, to the five key strategies of Wiliam and colleagues. Little research exists about the impact of professional development based on this conceptualization. However, Randel et al. (2011) investigated the impact of
teacher professional development of CASL on student achievement in mathematics. The program included an initial videoconference led by a CASL author. The teachers formed learning teams that would meet for 1-2 hours, two to three weeks apart. The teacher leading those meetings attended an extra workshop to be able to provide sufficient assistance for their groups. The gain in achievement of year 4 and 5 mathematics students from a randomized selection of 32 schools was compared to a control group from another 32 schools. The study found no significant impact of the professional development on student achievement. The analysis of teachers’ fidelity to the program showed that many teachers did not complete the program as the developers recommended.

The second example is a study by Bell et al. (2008) investigated the impact of a professional development program in 14 schools based on the five key strategies of Leahy et al. (2005). The design of this Keeping learning on track (KLT) professional development program was similar, but not the same, as the CASL program. It was based on teacher learning communities in which teachers worked in groups to develop their formative assessment practice. The program included a 2-day introductory workshop for the teachers, an additional 2-day workshop for teacher group leaders, and the teachers were supposed to have monthly school-based meetings structured and supported by modules with sufficient content for two years of meetings (provided by the KLT-authors). As in the study of the CASL program, there was no impact on student achievement. There were limited evidence collected about the extent and the fidelity of the implementation of the program, but in conformity with the study by Randel et al. (2008), the available evidence indicated that both the extent and fidelity of program implementation was limited (Bell et al., 2008).

No evidence about the impact of formative assessment on student achievement was presented in those two studies. Rather, the studies provide evidence of the difficulties of generating impact from professional development in formative assessment, and particularly from programs based on the conceptualization of formative assessment as a unity of integrated formative assessment strategies. A specific challenge seems to be how to support the teachers’ motivation to implement the professional development program with sufficient fidelity.

In summary, important ingredients in a professional development program in formative assessment very much coincides with general implications for design of professional development programs for teachers. Nevertheless, the complexity in introducing and using formative assessment makes design of professional development thereof challenging. The next subchapter will link
literature about professional development in general to literature specifically about formative assessment.

### 3.4 Linking literature about professional development in general to literature about professional development in formative assessment

As concluded above, important characteristics of professional development for teachers in general to a great extent coincide with aspects identified as important for professional development in formative assessment. In this chapter I complement my literature review with a more comprehensive review conducted by Schneider and Randel (2010) to link literature on professional development in general to literature on professional development in specifically formative assessment. The Schneider and Randel (2010) review, examine and synthesize seven quasi-experimental studies, six experimental studies and three studies of other types about professional development in formative assessment. The authors highlight the lack of studies investigating the links between professional development, teacher classroom practice and student achievement. The same gap in literature is indicated in the review by Vescio et al., (2008, see above Chapter 3.2), and the review by Tierney (2006), in which the majority of the articles followed a qualitative tradition. Seven of twenty-four studies used mixed methods, and only one was based solely on quantitative data.

Schneider and Randel (2010) link literature across a variety of content areas to studies in formative assessment, and identify seven important factors in professional development in general that are likely to be particularly important in professional development in formative assessment. Those factors were: administrative support; individualization of teachers’ learning goals; content knowledge; time; collaboration; coherence; and active learning. The authors’ comments on those factors specify that higher fidelity of implementation is more likely if the implementation is supported at school and district level, goals should be personalized by the teachers themselves so that teachers develop formative assessment techniques of their own choosing, and that content knowledge is necessary for conducting sound assessment. Moreover, both sufficient time and duration of time is needed, but most important is how the time is used to afford teachers possibility to understand and practice sound formative assessment, and to fully integrate new skills. Collaboration in professional learning communities needs to be designed purposefully and well structured, and the authors point out that coherence and alignment in effort of reform and teacher training is
needed. The last factor, active learning, was most strongly related to changes in practice after controlling for the effects of focusing on content knowledge.

In the summary of the review it was concluded that teachers cannot be passive recipients of information, they need hands-on experience rather than just increased knowledge of formative classroom assessment principles, they need ownership of their own learning and professional development not sustained is unlikely to influence student achievement. Opportunity to review student work, and to have student review their own work and the work of their peers, was supported in the review. The authors suggest that the importance of longer duration may be particularly true for professional development programs in formative assessment, given the changes in knowledge, skills and practice required to develop a more formative classroom practice (Schneider & Randel, 2010, p. 271). The authors argue that in formative assessment practice, content knowledge and pedagogical content knowledge can be both a prerequisite and an outcome.

Furthermore, the review highlights the challenge of conducting randomized trials, which means it is difficult to implement a study from which a strong causal conclusion can be drawn. “Research that provides precise estimates of the impact of professional development in formative classroom assessment on student outcomes or teacher outcomes is just in beginning” (Schneider & Randel, 2010, p. 268). Consequently, no rigorous causal evidence linking student achievement to professional development programs in formative classroom assessment can be made from the review. An additional challenge identified in the review is the lack of instruments available to measure outcomes related to formative classroom assessment, for example teachers’ practice of formative classroom assessment. The authors argue that estimates of teacher outcomes are important to understand the processes or mechanism that are responsible for producing any potential effects (ibid., p. 269).

The authors point out that any conclusions are tentative since researchers investigating effective professional development processes to improve teacher assessment skills are just beginning to use experimental designs, and few experimental or quasi-experimental studies exist (Schneider & Randel, 2010, p. 252). The authors call for: research providing more empirical data on the degree to which teachers are able to implement professional development programs with fidelity, and which elements or aspects of the program are feasible or not; research providing clearly documented contact hours and duration of the professional development program; and research collecting multiple data sources to more fully understand the changes made due to the professional development program. The authors stress the
importance of providing other researchers with sufficient information to inform the design of future studies so that findings across studies can be synthesized.

Next chapter presents the expectancy-value theory of achievement motivation used in Article 3.

3.5 A motivation theory perspective

In the research project we focused on investigating the changes the teachers made in their classroom practice due to the professional development program (Article 2) and we used a motivation theory perspective to shed light on the reasons behind the changes the teachers made (see Article 3). Motivation is the driving force of human behaviour instigating, directing and sustaining efforts such as changing teacher instruction, and can therefore be a useful perspective for understanding teachers’ choice to reform their teaching practice. More specifically, we used expectancy-value theory of achievement motivation (Wigfield & Eccles, 2000) as an interpretive lens in analysis of data about the reasons why the teachers participating in the professional development changed their classroom practice in general and why they chose to make specific changes. In this study we focus on the teachers’ expectancy and value beliefs, but also reflect on the characteristics of the PDP, and how these factors may have influenced the teachers’ expectancy and value beliefs.

According to the expectancy-value theory, expectancies of success and achievement values are the two major determinants of individuals’ motivation to carry out an activity. Expectancy of success can be defined as individuals’ “beliefs about how well they will do on upcoming tasks, either in the immediate or longer term future” (Wigfield & Eccles, 2000, p. 70). Achievement values consist of several different components: attainment value (importance), intrinsic value (interest), utility value (usefulness) of the task, and cost (Eccles et al., 1983; Wigfield & Eccles, 2000). Attainment value refers to the perceived importance of doing well on an activity. It is also linked to the relevance of engaging in an activity to confirm or disconfirm salient aspects of one’s actual or ideal self-schema such as competence in a domain (Wigfield & Eccles, 1992). Such self-schema may involve ideas about what a good teacher is. Intrinsic value refers to the enjoyment of conducting an activity in itself, and can for example arise from the satisfaction of learning and being able to use new teaching methods. Utility value refers to a useful outcome of the activity. The outcome can be useful both from a teacher and a student perspective, and could for example
be enhanced student learning. Cost pertains to a negative value. The activity might demand effort or time that limit the possibility to engage in other valued activities. For example, teachers using extended time for meetings will have less time to spend on other activities such as time for lesson planning.

The two main variables, expectancies and values, are assumed to be influenced by task-specific beliefs such as ability beliefs, the perceived difficulty of different tasks, and individuals’ goals, self-schema, and affective memories. These social cognitive variables, in turn, are influenced by individuals’ perceptions of their own previous experiences and a variety of socialization influences” (Wigfield & Eccles, 2000, p. 69). Expectancy-value theory posits that a number of social, cultural and cognitive factors interact in the development of expectancy and value beliefs. The focus in our research study (Article 3) is how the characteristics of the professional development program, as part of the social factors, may have influenced the teachers’ experiences expectancy and value beliefs.

3.6 Summary

The literature review on professional development in general was focused on presenting two reviews, the review by Timperley et al. (2007) that had inspired the design of the professional development program set up in the research project and the review by Vescio et al. (2008) about professional development through teacher learning communities, promoted as an advantageous form of professional development in formative assessment.

The literature review on professional development in formative assessment referred back to the difficulties in implementation of formative assessment and the challenges this brings to the design of professional development. The chapter outlined some previously identified important aspects of professional development, which largely coincide with professional development in general. However, such studies fail to identify what formative assessment to include in professional development programs and highlight the need of intervention studies investigating the effect from professional development programs on both teacher classroom practice and student achievement.

The review conducted by Schneider and Randel (2010) recognises the lack of intervention studies investigating effects on both teacher classroom practice and student achievement, and stressed the importance of publications including sufficient information to inform the design of future studies so that
findings across studies can be synthesised. This review, which also was a source of inspiration for the design of our professional development program, linked the literature on professional development in general to literature on professional development in specifically formative assessment and identified seven important factors in professional development in formative assessment. In particular, the authors suggest that the importance of longer duration may be particularly true for professional development programs in formative assessment.

From the literature review it is clear that several factors influence assessment processes in schools and that the complexity of assessment processes in education makes the design of a professional development programs in formative assessment extra demanding. Thus there are challenges in the implementation process both outside and within the formative assessment classroom practice. To address such challenges, teachers need to be motivated. Therefore the Expectancy-value theory has the potential of being a useful perspective for understanding teachers’ choices to reform their teaching practice.

From the previous chapters we have some indications on how to design professional development programs in formative assessment. Still there are several gaps in the understandings of the most important ingredients in such in-service training and the most valuable characteristics of formative assessment. In particular, there are needs for: a research-validated framework for best practices in formative assessment, studies with high ecological validity examining classroom practice of common teachers under normal conditions employing appropriate research methods, intervention studies investigating impact on both teachers’ practice and student achievement, and clear descriptions of the support provided in the professional development program.

The research presented in this thesis is targeting at those gaps in the research field of formative assessment. The next chapter outlines the methodologies used in the four research studies.
4. Methodology

Wellington (2000) defines a method of inquiry as a system of strategies and operations, designed at any time – for getting answers to certain questions about events of specific interest. Methodology is the activity of choosing, reflecting upon, evaluating and justifying the methods one uses (ibid.). The aim of the intervention study presented in this thesis was to see if ordinary teachers, who got the chance to participate in a professional development program with many contact hours and substantial support of an expert, would implement formative assessment in a way that could increase their students’ learning in mathematics. Another aim was to examine what motivated the teachers to implement new formative assessment activities into their mathematics classroom practices. In order to achieve those aims we chose to use a mix of methods. As I will describe below, the research was driven by a problem of interest, rather than by a theory.

This methodology chapter focuses more on the choosing, reflection, evaluation and justification of the methods we have used, rather than presenting all details in the design of the methods used. Specific descriptions of the methods are provided in the articles. Instead, I try to repeat as little as possible, but at the same time give an overview of the research project and complement with details not provided in the articles. Throughout this chapter, I describe the division of labour in the different parts of the project and the role of the researcher. In the latter part of the chapter I specifically discuss the underlying approach of the research, problematize the role of the researcher and outline ethical considerations. The subchapters are: the overall research design (4.1), an overview of the research project (4.2), research methods (4.3), philosophical considerations (4.4), the role of the researcher (4.5) and ethical considerations (4.6).

4.1 The overall research design

The overall research design of the research presented in this thesis corresponds to an intervention study. The effect of the intervention is measured both on classroom level and student level (see Figure 2). The main components in the overall design are the intervention, and the use of the framework, mixed methods, and triangulation.
In designing research you need to think backwards (Gorard, 2003). We wanted to evaluate the effect of a professional development program that had a research-based design. We wanted to evaluate the effect on teacher classroom practice (study II) and student achievement (study IV), and investigate what motivated the teachers to implement new formative assessment activities into their mathematics classroom practices (study III). In this way we attempted to contribute to the understandings about how to support teachers in the implementation of high quality formative assessment. Study I investigated the ways the teachers currently conducted formative assessment, before they participated in the professional development program. We could not assume that the teachers did not use any formative assessment before they entered the program. We needed to know if and how this practice could be further developed. Our use of the framework to characterise teachers’ use of formative assessment was a main characteristic in the research design and a way of keeping endpoint in sight (Gorard, 2003):

You cannot possibly design a sensible research instrument without considering in some detail how you will analyse the data you set out to collect. Otherwise you will not know if you asked the right questions, or collected data in the right format. The apparently separate phases of reading, formulating research questions, design, collection of data, analysis and reporting are really concurrent and iterative (p. 8).

This citation very much describes the way we worked in the research group. From previous research literature, logic and common sense, we developed an analytical tool (see below and Article 1, p.9) based on the framework characterising formative assessment as a fundamental idea and five key strategies (see Chapter 2.3.1). This tool helped us to keep the evaluating elements focused.

4.1.1 The framework and analytical tool

In general, a framework provides a structure for conceptualizing and design, recognizes data not possible to catch without the theoretical framework, and
makes it possible to transcend common sense and reach deep understanding (Lester, 2005). Based on the framework comprising the big idea and five key strategies (Wiliam & Thompson, 2008) we developed an analytical tool that comprised categories of formative assessment activities. The use of this analytical tool made it possible to analyse, describe, compare and summarize the changes the teachers made in their classroom practice as a result of the intervention. Using this tool it was possible to characterise each teacher’s classroom practice both from a quantitative and qualitative perspective. In this way we could visualise patterns in the group of teachers and get a deeper understanding of how different strategies in formative assessment were used in the classroom. In other words, we could characterise the formative assessment that came out of the professional development program, the formative assessment that in our hypothesis would have an impact on student achievement in mathematics.

4.1.2 Mixed methods

In three of the studies (I-III), qualitative analyses are mixed with quantitative analysis, while the fourth study is mainly quantitative in character (although the classification of test items was more qualitative). Researchers have criticized a sharp contrast between quantitative and qualitative methods (Bryman, 2008; Gorard, 2003; Pring, 2004). Gorard (2003) claims that all methods of educational research to some extent deal with qualities and all methods of analysis use some form of numbers. Without saying that the distinction between quantitative and qualitative methods is useless for a broad categorizing of research studies, a movement is noticed in mathematics education research from this distinction to a use of mixed methods (Gellert, Hernández, & Chapman, 2013). The difference in the characteristics of the research questions in the studies within this thesis demanded a mix of methods to answer those questions. Therefore, the data analysis used a mix of quantitative and qualitative approaches.

4.1.3 Triangulation

The use of mixed methods affords new possibilities for the researcher. For example, triangulation techniques can be used to map out, or explain more fully, the complexity of peoples’ behaviour by studying it from several view points and in so doing making use of both quantitative and qualitative data (Wellington, 2000). Strategic combinations of qualitative and quantitative methods, approaches, and concepts produce complementary strengths and non-overlapping weaknesses (Johnson, Onwuegbuzie, & Turner, 2007). The
argumentations for using mixed methods mean that quantitative and qualitative approaches give different contributions and complement each other: “Qualitative data give richness and colour, quantitative data provides structure” (Wellington, 2000, p. 19).

Multiple sources of evidence, that is, to employ multiple methods to look at the same phenomena (triangulation), can be used to strengthen the trustworthiness of results and conclusions, whereby observations, questionnaires, and interviews can all be used to challenge, confirm, or expand the information gathered (Schoenfeld, 2007, p. 87). Our data collection using observations, interviews, questionnaires, and mathematics tests is used for such a purpose.

I argue that the results and conclusions in our research studies became richer and more colourful from the use of mixed methods and that the possibilities to triangulate data also raised the trustworthiness of the studies. I also argue that the use of both quantitative and qualitative methods vouch for a balance in the findings, both to be judged as generalizable and applicable (Wellington, 2000).

Below I further describe and reflect on the implementation of the design of the project. First by an overview of the research project and second by an outline of the different methods used for data collection and data analysis.

4.2 Overview of the research project

In this section I present an overview of the total research project including the four research studies. I also outline how different people in the research group conducted the work and what external support was used. The research group consisted of my doctoral student colleague, our supervisor (also research leader), and me. My doctoral student colleague conducted similar studies in a group of secondary school teachers. (The twin study of secondary school teacher is mentioned, but the focus is here on the studies connected to the primary school teachers.) For the overview it can be helpful to know that in Sweden the school year starts with an autumn term lasting from August to December. The spring term starts in January and ends in June.

My doctoral student colleague and the research leader conducted a pilot study in spring 2010. The research leader was responsible for conducting the professional development program (PDP) for the teachers in the pilot study as well as in the main studies (in spring 2011). All preparation work (e.g. information meeting with teachers and school leaders and sampling process)
was completed before the first data collection (Study I) started in November 2010. Figure 3 shows the time, the sample, the methods for data collection, and the methods for data analysis for the four studies.

<table>
<thead>
<tr>
<th></th>
<th>Study I</th>
<th>Study II</th>
<th>Study III</th>
<th>Study IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
<td>Before the PDP</td>
<td>During first year of teaching after the PDP</td>
<td>Directly after the PDP and after one year of teaching.</td>
<td>In the beginning and end of the first year of teaching</td>
</tr>
<tr>
<td><strong>Sample</strong></td>
<td>21 primary teachers (17 secondary teachers)</td>
<td>22* primary teachers</td>
<td>22* primary teachers</td>
<td>362 students in classes of 21 primary teachers in the intervention group 376 students in classes of 24 teachers in a control group</td>
</tr>
<tr>
<td><strong>Method for data collection</strong></td>
<td>Observations Interviews</td>
<td>Observations Interviews</td>
<td>Interviews Questionnaires</td>
<td>Student achievement pre-test Student achievement post-test</td>
</tr>
<tr>
<td><strong>Method for data analysis</strong></td>
<td>Categorization of classroom practices and interview utterances</td>
<td>Categorization of classroom practices and interview utterances</td>
<td>Categorization of interview and utterances. Categorization and statistical analysis (mean values) of questionnaire responses.</td>
<td>Statistical analysis (ANCOVA)</td>
</tr>
</tbody>
</table>

*Figure 3. Overview of the research project*

* one of the teachers from the group of secondary school teachers shifted to teaching at primary school level.

54
The whole research group planned for the data collection and data analysis in Study I. In particular we worked together with the development of the analytical tool to be used for classification of formative assessment activities used by the teachers. The two doctoral students further developed the observation scheme and interview guide that were tested in the pilot study. It was also the doctoral students who in practice carried out the data collection and analysis, in my case for the primary school teachers. Our supervisor supported us during the process. The same division of labour was used in the second study in which we investigated the changes the teachers made after the professional development program. The only difference was that my doctoral student colleague was not in duty during further development of the analytical tool and creation of a new interview guide.

During the analysis of data we calibrated the analytical tool to serve the purpose of identifying recurrently used formative assessment activities in Study I and the (formative assessment) changes implemented after the professional development program in Study II. In practice this means that we together made decisions about the design of the analytical tool during the process of analysis, for example the categorisation of the formative activities for each key strategy and the fundamental idea. The analytical tool had pre-defined categories for expected formative assessment activities within each key strategy and the fundamental idea, but we were open to add new categories emerging from the data. Such decisions were elaborated in the research group. We made decisions about the borders between categories (formative assessment activities) that were similar and about the limitations for what formative assessment activities to include in the results (for example, concerning requirements of frequency for inclusion of the activity). All difficult classification cases were discussed in the research group.

The research design made it particularly important for us to be consistent in our classification. In the first study the practice of both primary and secondary school teachers were presented and compared, so the classification of the practices of both groups of teachers had to be made in the same way. To make it possible in a later stage to compare the impact of the professional development program on the classroom practice of the primary and secondary school teachers it was important that also these classifications of practice were consistent over teacher groups.

Here I have described the division of labour in general and the common work with the analytical tool used in Study I and II. A more detailed description of the data collection and data analysis for all four studies is provided below.
4.3 Research methods

In this section I focus on describing the assumptions and arguments behind choices of methods and procedures used in the different studies. I attempt to repeat as little as possible from the articles, but also to add details not found in the articles. First I describe the sampling procedure. After that I describe the procedures in data collection and data analysis. The procedures in Study I and Study II are described together because of the general similarities between the two studies. Study III and Study IV is described in separate subchapters.

4.3.1 Participants

Participants in the study were teachers from one middle-sized municipality in Sweden who were going to teach a year 4 class in mathematics the school year 2011-2012. That choice of primary teachers that were going to teach a year 4 class was made because we assumed it would be easier for the teachers to make changes in their classroom practice in a new class (in Sweden the students usually change teacher between year three and four). The teachers in the intervention group constituted about half of the teachers in the municipality who would teach year four students that school year. The other half of the year four teachers constituted the control group used in Study IV.

The municipality had a budget set aside for financing schools for the need of substitute teachers to enable the participation in the professional development program, but the municipality also got founding from the Swedish National Agency for Education. To fill the places available (number of places were determined by the size of the budget), more than half of the total number of year 4 teachers were randomly selected to be invited to the program. The schools in the municipality had one, two or three year 4 classes. For each school with two classes, one of the two teachers was randomly selected to be invited to participate in the professional development and the other teacher was included in the control group. Similarly, in the schools with one or three classes (there were only three schools with three year 4 classes), the individual teachers were randomly selected to the possibility to participate in the professional development and to the control group, given the requirement that the number of classes assigned to the study from each school would be as equal as possible to the number of classes not included.
Twenty-eight teachers were randomly selected from all teachers who were going to teach mathematics in year 4 (10-year old students) in a middle-sized Swedish city in the school year 2011-2012. These teachers were invited to participate in the professional development program and the research study. Six of them declined. The reasons for not wanting to participate differed but included soon retirement, already feeling proficient in formative assessment, and other school priorities. Thus, the final sample consisted 22 teachers in the intervention group (14 female and 8 male), and 24 teachers in the control group (18 female and 6 male). One teacher was not included in the year four group in Study I, because he was then a year 7 teacher and participated in the professional development program for year 7 teachers. Another teacher was not included in Study IV because he was not teaching the whole school year after the PDP). The classes belonging to the teachers in the intervention group consisted of 362 students, and the classes belonging to the teachers in the control group consisted of 376 students.

From the beginning, the idea was to randomly select two teachers from each school. This would have been beneficial for the teachers’ possibility to support each other. However, this idea had to be forsaken because all smaller schools (which is the predominant form for primary schools in this municipality) would have been excluded from the study since they don’t have two classes for each school year. To avoid any effect of the professional development program on the teachers in the control group the teachers in the intervention group were asked not to talk about what they had learned during the professional development program with other teachers at their school teaching a year 4 class. If such overspill occurred anyway, the consequence would be a more favourable position for the control group. Since formative assessment is complex and has been proven hard to implement, we consider notable changes of the teachers in the control group unlikely.

### 4.3.2 Procedures in data collection and data analysis

The methods used for data collection were classroom observations, teacher interviews, teacher questionnaires and student mathematics tests. Study I and II used observations and interviews to study teachers’ use of formative assessment in their mathematics classroom practices before and after the professional development program. Study III used interviews and questionnaires to investigate the reasons for the changes the teachers had made in their mathematics classroom practices after the professional development program and to identify the characteristics of the professional development program that the teachers experienced as most important for
their development of a formative classroom practice. In Study IV, a pre-test in the beginning of the year of implementation and a post-test in the end of that school year were used to investigate differences in student achievement between classes belonging to the teachers in the intervention group and classes belonging to the teachers in the control group.

Below we present the methods used in Study I and II together because of their similarities. Study III and Study IV are presented individually.

4.3.2.1 Study I and II - Classroom observations and Teacher interviews

In both Study I and II, for each teacher we used two classroom observations and one interview to investigate the teacher's use of formative assessment in the mathematics classroom practice. Both the classroom observations and the teacher interviews were semi-structured. The observation scheme and interview guide used at all occasions were structured from the conceptualisation of formative assessment described above (Section 2.3). We used the same conceptualization of formative assessment in both studies to analyse what formative assessment activities the teachers used in their classrooms, but in Study II we in particular investigated the changes (established new formative assessment activities) the teachers had made after they participated in a professional development program. Therefore, the interview guide used in Study I was adjusted before it was used in Study II.

The purpose of the data collection was to get as detailed information as possible about teachers' use of formative assessment activities. We wanted to get information about the extent and in which ways any formative assessment activity was used. We also wanted information about the teachers' purpose of using each activity. That information was sometimes necessary to decide which key strategy (or the fundamental idea) the activity belonged.

The classroom observations fulfilled several purposes. One purpose was to get a frame of reference for the interview and facilitate the communication in the interview. The observations facilitated the interview by functioning as a concrete situation to relate to. The observations reduced the risk for misunderstandings and raised the clearness as the teacher was telling about moments from the observed lesson (Goodchild, English & Burton, 2005).

A way to reduce the risk of researcher effect on respondents was to make unannounced observations. This does not mean that we did not trust the
teachers to not follow ordinary ways of teaching. Instead, we wanted to give the teachers a better chance to do so. In addition, at the start of the observed lesson we informed the class about our role, and that our focus was to observe the teacher, not the students. We told the students to ignore us and that we had no intention of disturbing the classroom practice. This was another way of dealing with the observer effect, referring to “unnatural behaviour by the subjects of the research due to the presence of an outsider” (Wellington, 2000, p. 66). We experienced that the students were used to having visitors in their classrooms and did not care much about us.

In the observations for Study II we looked for signs of establishment of new elements in the teaching. For example, we noticed if the mini-whiteboards (which was introduced during the PDP) seemed well used, if the students were familiar with any activity and we sometimes asked the students questions. Those questions were asked to get implications about the students’ understandings about what they were doing, what assistance or resources they might have and if they were used to the situation. These observations were not a structured collection of data, and they were only used to complement and verify other data. In the final categorisation of formative assessment activities, we required that a new established formative assessment activity reported by the teacher was classified as such only if the teacher had given examples or details in the interview, or mentioned briefly in the interview but we also saw signs during the observations of the activity being an established new formative assessment activity. Field notes were taken during the observations and these were checked and completed by the observer the same day as they were taken. Any picture taken or document received from the teacher was added to the field notes.

The semi-structured teacher interviews lasted about one hour for the interviews in Study I, and one and a half hour for the interviews in Study II. Two different interview guides were used since the two situations and the research questions were different. For example, in the second interview the teacher had developed considerable knowledge about formative assessment and we explicitly asked for what changes the teachers had made in their classroom practice during the school year. The interview guide in Study I used questions for eliciting information about the extent, ways and purpose of formative assessment activities used by the teacher. The interview guide in Study II was divided into three parts. In the first part the teachers were prompted to talk more freely about the changes they had made and their purpose behind using any formative assessment activity. The second part in a more structured way asked about (quantitative and qualitative) details about the teachers’ use of any such activity. The third part dealt with
questions for study III, the reasons for the changes the teachers had made in their mathematics classroom practice and their experiences during the process of implementation.

The interview guide was not supposed to be followed from start to end, but to secure that all areas were covered. This gave us, as interviewers, considerable flexibility over the range and order of the questions (Wellington, 2000), but still made the data collection systematic. In such an interview, where the interviewer uses opportunities from one area of the interview guide to another in a logical way, the interview resembles a normal conversation. Following the principles of cognitive interviewing (Kvale & Brinkmann, 2009) we lead the respondent into areas of questions, trying to be sensitive and follow the tracks the conversations took, but without letting it lead away from the area of interest. We used probes and follow-up question to capture details and examples, or for clarification, as well as to keep the focus in the conversation (Patton, 2002). At the end of the interview we asked the teachers if any important area in their mathematics instruction was not brought up during the interview. At that time the teacher got another chance to give additional information, but this question also mediated respect for the teacher’s expertise.

The observations and interviews worked as planned, both in terms of practical implementation and information obtained. Fortunately, no observation or interview was cancelled (although observations sometimes were postponed because of, for example, other events, such as a music performance). By using a conceptual framework to identify characteristics of teachers’ use of formative assessment, the analysis began already in the observations and interviews. As described above, the plan for the analysis was somewhat open in details and the whole research group made decisions during the careful process of dealing with the data. We did not just want to make sure that the process was scholastic, we also wanted the twin studies (the studies on primary school level and on secondary school level) to be performed in the same way so the results could be comparable.

Rigorous procedures were used for the analysis and summary of data, for the individual teacher and the group of teachers. The interviews were recorded and transcribed almost word by word. Retrospectively, from the guiding question “What is useful transcription for my research purpose” (Kvale & Brinkmann, 2009), larger parts of the recordings could have been excluded from being transcribed.

Parts of the relevant text from the transcripts were coded as belonging to one or more headings for the fundamental idea and the five key strategies. Out of
the field notes and the structured transcription we made a narrative for each teacher’s formative classroom practice. From that narrative we made a summary for what formative assessment activities each teacher used for Study I and what changes the teacher had made in Study II. The benefits of making comprehensive narratives are twofold: for understanding of the logic in the teacher’s actions and for detecting incoherence, which could be a sign of misunderstanding between the teacher and researcher. The narratives can be seen as a strategy to secure authenticity and trustworthiness of the findings.

In the narratives it was noted if the data came from the observation or the interview. Any discrepancy between the data sources was noticed, but this was very unusual. In the final categorisation, all data were demanded to be unambiguous. In Study I we used a respondent validation (Wellington, 2000), which means that each teacher checked the accuracy of the data in the narrative about their classroom practice. The same kind of respondent validation was not completed in Study II because of organisational reasons, outside control of the research group. The teachers gave very few comments or corrections on the first narratives, which suggests that the rigor in the procedures worked sufficiently and that few corrections would be expected in a similar respondent validation in Study II. However, another kind of respondent validation was made during the interview (for questions concerning Study I and II as well as Study III). Through the whole conversation we were keen on asking for the respondent to confirm our interpretation of the meaning of their utterances to secure the accuracy of the data (Bassey, 1999).

Any formative assessment activity reported by the teacher was classified as a currently used activity (Study I) or change (Study II) only if the teacher had given examples or details in the interview, or if we had seen signs of that the activity was an established new formative assessment activity during the observations. Thus, it was not sufficient that a teacher said that he or she used an activity in the classroom. The final categorisation of each identified formative assessment activity was based on the main purpose for using the activity. The purpose decided for what key strategy (or the fundamental idea) the activity was used. Thus, each formative assessment activity was only categorised once, as connected to one of the key strategies or to the fundamental idea. This is a somewhat artificial categorisation because teachers can use an activity for several purposes. We chose this form of categorisation to be able to measure the number of formative assessment activities currently used (Study I) or changes made after the professional development program (Study II).
4.3.2.2 Study III – Teacher interviews and Teacher questionnaires

Teacher interviews (the same as in Study II, but another part of the interview) and two questionnaires were used to study the reasons behind the changes the teachers made in their mathematics classroom practice and to identify important characteristics of the professional development program. Questionnaire 1 was administered directly after the PDP, and Questionnaire 2 in the end of the school year after the PDP. The interviews were conducted just before the second questionnaire.

The third part of the interview guide comprised questions for Study III, but also one of two themes in the second part of the guide asked about the reasons for the changes the teachers had made. The questions aimed at understanding the opportunities and barriers the teachers recognised during the process of implementing formative assessment into classroom practice; what drives him or her to desire and apply any formative assessment activity; what prevented him or her; what the teacher expected in terms of value; their own and students’ opportunities to succeed; and expected costs to succeed. Using Kvale’s (1996) metaphors, the work of the interviewer in Study I and II resembles the Miner’s work where knowledge is seen as collected. In Study III, the work resembles the Traveler’s work where the interviewer wanders along with the local inhabitants, asks questions that lead the subjects to tell their own stories of the lived worlds. For this purpose, we were sensible to what the teacher said to capture and explore the teachers’ unique experiences of the implementation, using the strength of a qualitative research approach and the interview as a well-suited method (Cohen, Manion & Morrison, 2011; Kvale, 1996). In such an interview situation the interviewer strives for a rich nuanced description of the phenomena in everyday language, but still with a responsibility for arriving at well-controlled information to be used for meaning condensation to be coded and categorized afterwards (Kvale & Brinkmann, 2009).

The first questionnaire provided to the teachers at the end of the professional development program asked for the reasons behind the changes that the teachers planned for in their classroom practice after the professional development program and their experiences of the professional development program. The second questionnaire provided to the teachers in the end of the first year of implementation asked for the reasons behind the changes the teachers had made during the first year of implementation; what support and conditions were most important for continue the use of the new formative assessment activities in the classroom; and how the teacher had used the individual plan for implementation.
The research group designed the questionnaires. The first questionnaire had 37 items and the second had 13 items. Both questionnaires used items to be answered on a five-point scale, as well as open-ended items. Similar to the questions in the interview guide, the items in the questionnaire aimed at understanding the opportunities, barriers, and driving forces to use formative assessment, as well as expectancy of and opportunities for success and beliefs of value and cost of implementing formative assessment. Thus, some items had direct connection to the expectancy-value theory of achievement motivation, while others were more open. In contrast to the interview one item in the first questionnaire explicitly asked about the teachers' experiences from the professional development program. Details of the design of the questionnaires and examples of items are provided in Article 3.

Almost all teachers, 21 out of 22, answered each of the two questionnaires used. The teacher missing each time was not the same. The doctoral students summarized and analysed the questionnaires. The items to be answered on a scale provided quantitative values of the answers from the group of teachers and the open-ended items were explorative similar to the interview questions. The mean value of the teachers' responses was calculated for the scale response items. The answers from the open-ended items were analysed and structured into categories from common themes using an approach similar to the interviews and could be described by three main components for qualitative data analysis: data reduction, data display, and drawing and verifying conclusions (Miles & Huberman, 1994). The number of teachers for each identified category was calculated and compiled for the group of teachers.

### 4.3.2.3 Study IV – Student mathematics pre- and post-tests

In study IV we used the results on a pre-test and a post-test to examine the impact of the professional development program, and the intervention teachers' use of formative assessment, on student achievement (see below). The achievement of all year 4 teachers' students in the municipality was assessed by a mathematics test in the beginning of the school year (pre-test August 2011) and a second test in the end of the school year (post-test in May 2012). The pre- and post-test were not the same. Both tests were created specifically for the purpose of this study and were developed by the research group in cooperation with people from the department responsible for national tests in mathematics. Since two of the members of the research group had previously worked at that department, the tests were developed by people with a good understanding of the curriculum and test theory. Careful
steps were taken to secure measuring precisely the content of the syllabus that was intended to be measured. For example, the test developers checked that the items tested content that was included in all the textbooks used by the teachers.

The pre-test assessed the mathematics described in the written national syllabi documents and covered in the textbooks until school year 3. Similarly, the post-test assessed the mathematics described in the written national syllabi documents for school years 4-6 and covered in the textbooks for school year 4 (the national syllabi documents does not prescribe what should be learned each school year, but the teachers usually follow their textbooks which are an interpretation of the national syllabi). Thus, the pre-test was designed to provide information about students’ learning up to year 3 and the post-test was designed to provide information about students’ learning in year 4.

A group of teachers with experiences of teaching year 4 supported the design of the items in the tests to ensure that the items would be understandable and have appropriate level for the students. A group of teachers in another municipality supported the piloting procedures of those tests. Those procedures included testing that students understood the items, and that the teachers considered that the test content was relevant in relation to the syllabus.

When administering the tests (in the beginning and end of the school year with year 4 students) the research group provided the teachers with instructions how to introduce and handle the test situation. All teachers (in both the control and intervention group) received letters with information that the results of the tests would not affect the teachers, and that no results at class level would be presented. Only the teacher had access to the results of their own class. This meant that teachers did not have much incentive to influence student performance. The teachers also received information that helped them to use the results in their teaching. This information facilitated a greater focus on learning, and less focus on using the results for comparison and competition. The teachers received the tests only the evening before the day the tests were to be administered to the students. This meant that there was no opportunity for teachers to prepare students by teaching specific mathematics content included in the tests. In addition, the teachers did not grade the test themselves, and consequently had no opportunity to influence the results in that way.

The research group produced scoring instructions for the tests, and a group of experienced mathematics teachers assisted in scoring the tests. An
employed assistant transferred the test data to a statistics software program. During the whole process, all data was treated as confidential.

To compare and test for significant differences in achievement gains between the classes in the intervention group and the classes in the control group, a one-way between groups analysis of covariance (ANCOVA) was conducted. The group variable was professional development (no program or formative assessment program). Since the teachers were considered the unit of analysis, the students’ mean scores on the posttest for each class, constituted the dependent variable in the analysis. The students’ mean scores on the pretest for each class, were used as the covariate. In order to examine the ANCOVA assumptions of normality and homogeneity of variances, the one-sample Kolmogorov test and Levene’s test were used.

Effect size is a way of quantifying the difference between two groups into a standard measure (Cohen et al., 2011). Using effect size makes it possible to estimate how big the effect is, something the p-value does not do. Cohen’s d was used as a measurement of effect size and was computed by the difference in the adjusted mean on the post-test, between the classes of the intervention group and the control group, divided by the root mean squared error, RMSE.

In Study II we did not measure the change of the teacher practices in the control group because we assumed these classroom practices were not affected by the intervention. If these classroom practices would have changed (e.g. by influence from teachers in the intervention group), the conclusion of the effect on students’ achievement is nothing but underestimated. Hence, we assume that the intervention did not have negative effect on those who did not participate.

In this chapter about methodologies I have described the methods we have used with a focus on assumptions and arguments leading us to choose those methods. Next, I describe the more theoretical ideas and assumptions underlying those considerations and decisions.

### 4.4 Philosophical considerations

In the beginning of this chapter I defined methodology as the activity of choosing, reflecting upon, evaluating and justifying the methods one uses (Wellington, 2000). Methodology can also be defined as “a theory of methods - the underlying theoretical framework and the set of epistemological (and ontological) assumptions that determine a way of viewing the world and, hence, that underpin the choice of research methods”
(Ernest, 1997, p. 35). In this definition the underlying theories and assumptions are specified as requirements in the methodological considerations made. In this way methodology refers to the rationale and the theoretical assumptions guiding the research (Gellert et al., 2013).

Even if this thesis is driven by a problem of interest, not a theory, the methods we have used is grounded in some rationale and underlying theoretical assumptions. In this subchapter I propose that the grounds for the research design and methodologies we have used is best described as a pragmatic approach. First I define what is meant by a pragmatic approach and then I describe the values, beliefs and attitudes driving the research design of this study.

4.4.1 A pragmatic approach

During my doctoral studies and writing this thesis I have been asking myself what beliefs, principles, values and assumptions are embedded in the research design and how are these rationales and assumptions best described. In this section of the thesis I will argue for the studies in the research project to fall into a pragmatic approach (Morgan, 2007).

The term paradigm is not definite and there are problems with clear cuts between different paradigms (Morgan, 2007). Morgan chose the term approach instead of paradigm to sort out the confusions around the concept of paradigm as it is used in what he calls the existing “metaphysical paradigm” (see below). From his stance, paradigms are systems of beliefs and practices that influence how researchers select both the questions they study and methods that they use to study them. According to Morgan the metaphysical paradigm, coming up in the 1980s, was a replacement of the “positivist paradigm”. This change included a widespread version in social science methodology, which emphasizes metaphysical issues related to the nature of reality and truth (ontology) (ibid. p. 57). However, it is not clear to what extent this metaphysical paradigm has influenced mathematics education research (Lester, 2005).

Morgan (2007) advocates the “pragmatic approach” as a new guiding paradigm in social science research methods, both as a basis for supporting work that combines qualitative and quantitative methods and as a way to redirect our attention to methodological rather than metaphysical concerns. This, at the same time, provides a new range of opportunities for scholars in the field of social science research methodology (ibid.). In the pragmatic approach, methodology is placed at the centre, with emphasis on the

66
connection between epistemological concerns about the nature of the knowledge that we produce and technical concerns about the methods that we use to generate that knowledge. The research questions then comes into focus, rather than the metaphysical assumptions.

Other authors have argued that mixed methods research is one of the three major “research paradigms” (quantitative research, qualitative research, and mixed methods research) (Johnson et al., 2007) and define mixed methods research as:

Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration. (ibid. p. 123)

I find the pragmatic approach and the research paradigm of mixed methods as very consistent. In both cases issues in epistemology is connected with issues in research design, rather than separating thoughts about the nature of knowledge from the process of producing the knowledge. Seeing epistemology as interlocked with methodology means that values, beliefs and attitudes driving the study, affect choices made in the research design (Burton, 2002).

In the next chapter I clarify the values, beliefs and attitudes driving the research design in this research project.

4.4.2 Values, beliefs and attitudes

The values, beliefs and attitudes driving the research design of this study about formative assessment practices might not be clear and can be misinterpreted in the communication process of the researchers. I will here discuss some potential misinterpretations of the research design and misunderstandings related to words used in that communication process.

As a start, the main purpose of the research project is problematic because it can be interpreted as normative and positions the teachers as an object in the intervention. Although, from the pragmatic approach, we follow Black and Wiliam’s (2003) suggestion that mathematics education research should not seek any eternal truth or universals, but develop principles of certain generality that teachers may integrate into their contextual decision-making in their practice. Formative assessment is concluded to be a way of teaching with strong evidence of being beneficial for student learning. Thus, the
decision to choose formative assessment as the content of the professional development program was a decision of choosing the best alternative available. In addition, we did not expect the teachers to implement formative assessment in any predefined way. We expected the teachers to find their own ways.

Other potential misunderstandings are connected to value-laden words. Since research described in this thesis is driven by a problem, rather than theory, the use of metaphors and language is not framed by a theoretical discourse. In my writings about the research project I use words such as “effective” and “implementation” even if I know they are problematic. Also here the approach is pragmatic. I use the word effective to state that the classroom practice was effective in supporting students learning. This is an obvious goal for research in mathematics education research, although there are other issues within this goal, for example what is to be learnt. The concept of implementation signals an individualistic perspective that considers the teacher as an implementer, based on a linear movement from a starting point to an end point (Skott, 2009). In this perspective the teachers would become the carrier of the problem if the reform was not successful. Skott’s point is that we should take it seriously that contexts matter, not only as starting points for research, but also when we consider the potential impact of research results. My point is that we evaluated the professional development program, not the teachers. This means that any shortcoming in the effects of the intervention includes questioning what support the professional development program afforded the teachers.

My attitude is that teachers (seen as a group) do their very best to create an optimal classroom situation for their students. Therefore, to clarify the attitudes, beliefs and values behind this thesis: The research was driven by an honest desire to establish understandings that can help different groups of professionals to develop a classroom practice that is more advantageous for student learning. I believe that all teachers can improve their classroom practice even if they already are doing a great job in their class. No teacher is so good (or so bad) that he or she cannot improve.

4.5 The role of the researcher

The role of the researcher has to be taken into consideration because of the risk to influence results in a desired direction or in any unconscious way cause bias.
Being the designer of a professional development program and also the evaluator of the same program is problematic. The divided roles between the research leader tutoring the professional development program and the doctoral students collecting the data reduced the risk of an inflated picture. However, my background as a primary school teacher for 23 years could cause another kind of bias in form of over-interpretation of the teachers’ utterances, me hearing what I expected to hear. During the interviews and the data analysis, I was consciously trying to be open for each individual teacher’s intended meaning in the teacher’s responses. To reduce the risk of influence of the researcher on the analysis and summary of each teacher’s classroom practice, I chose to make anonymous the results of student achievement on class level. Meaning, I do not know the performance of individual classes and cannot unconsciously interpret and describe the classroom practice more favourable from knowledge that the class performed well on the achievement test. The interview guides and observation schemes specified what to look at and in that way were helpful in eliminating the influence of myself as a researcher. The consistency in results from the different data sources advocate for that the methods used functioned.

The next section discusses the ethical considerations made during different stages in the research project.

4.6 Ethical considerations

In the initial stage of design, the research project was guided by the Ethical Review Act (Etikprövningslagen) with later addition (SFS 2008:192), which states the rules for what research to be examined, who the research subjects are, and what applies to those persons. In our case, it was not necessary to submit an application for ethical evaluation. In later stages, in the contact with the informants we followed ethical guidelines stated by the Swedish Research Council for research in social science (Vetenskapsrådet, 2002, 2011), which includes requirements in four areas: information, consent, confidentiality and use.

4.5.1 Information

At the meetings before the professional development program had started the teachers received information about the studies planned in the research project, the purpose of the research studies and more specific how they and their class would get involved in those studies. When the teachers accepted
the offer to participate in the professional development program they also agreed to take part in the research studies. This was explicit in the information to the teachers and principals. However, they were informed that they could redraw from the research studies if they changed their minds, and one teacher decided later on not to be observed in the classroom for Study 2. At these meetings the teachers were encouraged to contact us if they had any questions and contact information was provided. In addition, the teachers got the opportunity to ask for information during the professional development program and at the observations and interviews. Some teachers did ask questions to better understand the purpose or how the information would be used. The fact that we met the teachers every week during more than 6 months, made the ethical situation different from many other studies. This gave the teachers several opportunities to ask questions about how the research would be carried out, documented and reported. This reduced the risk of teachers’ insecurity about how any information would be handled.

We also informed about the narratives that we would write and they would get the opportunity to read. The purposes were numerous, but there was an ethical aspect. If the teacher did not recognise himself or herself, this must be taken under consideration, both in validation and from an ethical perspective. In the end, there were very few comments on the narratives in Study I. Unfortunately the same procedure was not accomplished in Study II (out of the scope of our influence), which is unfortunate both from a validation and an ethical perspective.

The students were not research subjects in our study. However, for ethical reasons the class received information before each observation about the purpose and what we as researcher were doing in the classroom. Often the teacher also had given the students this information in advance.

4.5.2 Consent

The teachers had much time to think before they made the decision to participate in the professional development program and research studies. They got the information in May and had to make up their mind in August, which was half a year before the professional development program started. The research team was not involved in this communication about their decision. In addition, the teachers could decline taking part in the research project at any time, without losing the opportunity to participate in the professional development program (one teacher decided to do so).
No consents were collected from students, because data collected in classroom considered the teacher’s actions. The data from the mathematics tests was administrated by the municipality, and was considered as an ordinary occasion of evaluation possible to carry out at any time within a municipality. The tests were designed for a research study, but the tests were also used as an evaluation of the municipal investment in in-service training of teachers. Thus, the situation for the teachers and the students did not differ from other municipal evaluations.

4.5.3 Confidentiality

We kept the participants’ identity confidential during the process of analysing, transcribing, and making the narratives about teachers’ classroom practices. The large number of teachers and not naming the city, the teachers and the schools are considered enough to prevent identification of individual teachers.

The students’ identity in the tests was removed before the data was transferred from the municipality database. Data was handled with caution to not be accessible to anyone outside the research group.

4.5.4 Use

We use the material for research purposes, for this thesis, articles and conference proceedings. The participants were informed that this would be the case.

4.5.5 Additional ethical considerations

Reflection on ethical aspects throughout all phases in the process of research does not mean that the researcher just carry through universal ethical procedures. Especially in interview studies it is much about practical wisdom, and recognizing and responding to what is important for the informant in any situation (Kvale & Brinkmann, 2009). For example, since classroom observations can challenge the teachers’ integrity, it was important in the following interviews to give the teachers a chance to explain the reasons behind choices, behaviours or episodes viewed during the classroom observations. This can be especially important if the teacher did not feel comfortable in the observed situation. It was in our interest that the teachers felt comfortable and motivated to take part in the in-service training. We wanted the teachers to feel that they already did a good job, but
that the new knowledge about formative assessment could be helpful for them.

Another ethical issue concern normativity and who has the mandate to decide on teacher classroom practice. Our approach in the professional development program was to give the mandate to decide ways of teaching to the teachers, but to provide the teachers with guidance from research and support from an expert. However, investigating effects from a professional development program can put oppressive expectancies of fidelity on the teachers.

The publication of our research findings also raises moral questions about what kinds of effect the report leads to. The writing itself creates values, a particular view of reality, and the author in someway speaks on behalf of others (Kvale & Brinkmann, 2009). This issue has been a subject in the research group conversation. Our attitude has been to view the teachers with respect from the view of teaching and learning in classrooms as complex processes not easy to handle.

In the data collection process, we saw the teachers as important co-operators in the process. They were carrier of unique information of importance for us. Thus, we were keen on handling the data collection procedures in a decent way in consideration to both our research interest and teachers’ wellbeing. For example, we asked the respondents about additional information in the end of the interview and asked for feelings about the interview situation. In summary, much of ethical issues were facilitated by the fact that we met the teachers so many times during the professional development program and occasions for data collection.

4.6 Summary

This chapter about methodological considerations has focused on the activities of choosing, reflecting upon, evaluating and justifying the methods used in the research project. The overall research design of the project corresponds to an intervention study, with a use of mix of methods.

The choice to randomly select the teachers for in the professional development program makes generalization outside the sample possible (to mathematics teachers in year 4 in the municipality). Multiple sources of evidence (triangulation) strengthen the trustworthiness of results and conclusions. The use of the framework of formative assessment provided a
structure for conceptualization and design, and made it possible to recognize
data that would be more difficult to catch without the framework.

The research in this thesis is driven by a problem of interest, rather than a
theory. The research design and methodologies used is best described as a
pragmatic approach. This chapter has provided a description of the values,
beliefs and attitudes driving the research design of this study, as well as the
ethical considerations.
5. Results

This chapter summarises the results of the four studies conducted within the research project. More specifically, the chapter presents how formative assessment is used in the classrooms of a random selection of 38 primary and secondary school teachers in a middle-sized Swedish municipality (Study I), the characteristics of the changes in the teachers’ formative classroom practice made after participating in the professional development program (Study II); the reasons for the effect on the teachers’ practices, and the characteristics of the professional development program the teachers experienced as most important for their development of a formative classroom practice (Study III); and the effect from the professional development program on student achievement in mathematics (Study IV).

5.1 Article 1

The aim of the study in this article was to characterize the formative assessment used in mathematics classroom practices in a middle-size Swedish municipality. Two groups of randomly selected teachers participated in the study, 21 primary and 17 secondary school teachers. The research question for the study was: Do mathematics teachers in the municipality use formative assessment in their classroom practices, and if so, in which ways?

The results show that the teachers’ instruction did include aspects of formative assessment. On a group level, the teachers performed activities within all five key strategies and used different ways to adjust instruction based on the information they collected about students’ learning. In general, the two groups of teachers used similar formative activities, although some differences were identified at a more detailed level. For example, primary teachers correct students’ notebooks to elicit evidence of their learning while tests are used to a greater degree by secondary teachers.

This study shows that the random sample of mathematics teachers included in this study do indeed use some aspects of formative assessment. However, it was equally clear that there was room for development in order to reach high-quality formative assessment practice. For example, teachers do collect evidence of student learning and modify instruction, but these modifications are rarely adapted to individual student needs.
5.2 Article 2

The aim of the study in this article was to characterize the kind of formative assessment that came out as a result of a professional development program in formative assessment. This study analyses the changes in the mathematics classroom practice made by a random selection of 22 primary school teachers. The teachers’ use of formative assessment before the professional development program was characterized in another article (Article 1). Still another article showed significant effect (p = 0.036) on student achievement in comparison with a control group (d = 0.66) after one year of teaching with new formative assessment activities (Article 4). The research question for this study was: \textit{What are the characteristics of the changes in teachers’ formative classroom practice that significantly enhanced student achievement?}

The results show that all the teachers had made changes in their mathematics classroom practice after participating in the professional development program. The teachers’ changes span from complementing previous teaching with new activities that enhance the fundamental idea of formative assessment to a classroom practice that is radically developed in its very foundation.

The changes the teachers made, and thus the characteristics of their new formative classroom practice, can be described in relation to three dimensions of formative assessment that regards (a) three key processes in teaching and learning, (b) the agents in the classroom involved in the teaching and learning, and (c) the length of the adjustment cycle. The teachers developed their practice along all of these dimensions, albeit in different ways and to different degrees. In the article it was suggested that this three-dimensional development afforded new opportunities for student learning. First, the integration of the three key processes of teaching and learning may enhance student learning. The second dimension indicates that further learning opportunities may occur by involving all agents (teacher, student, and peers) in these processes. Lastly, shortened adjustment cycles make the formative assessment more time efficient.
The aim of the study in this article was to explain why the teachers changed their formative assessment classroom practice after participating in the professional development program. Except from investigating the reasons for the teachers’ changes, another aim was to identify the characteristics of the professional development program (PDP) that the teachers experienced as most important for their development of formative assessment. A motivational perspective was used to explain why the teachers’ developed their practice in relation to formative assessment. The research questions for the study were: (1) Why did the teachers use the PDP to significantly change their overall teaching towards a more formative classroom practice?; (2) Why did they make the specific formative assessment changes they made, and not others?; (3) Which characteristics of the PDP did the teachers experience as most important for them to be able to implement their new formative classroom practice?; and (4) Which barriers for change did the teachers experience?

The results show that the teachers were very motivated after the professional development program, and all through the teaching the year after, to make changes towards a more formative assessment practice. The teachers had high value beliefs for the outcome of formative classroom practice as well as high expectancies of success to be able to carry out this kind of teaching. The value beliefs included for example high experienced utility value for both themselves and for the students, and only moderate costs in terms of time and effort. According to expectancy-value theory these variables are decisive for the motivation of action.

A similar pattern emerged when analyzing why some formative activities were used to a high extent, and others were not. Those activities that the teachers experienced as most valuable in relation to the cost were mostly used. Thus, the choice of characteristics of formative assessment can also be explained by expectancy-value theory.

The study identified characteristics of the professional development program that the teachers perceived were most important for bringing about this change in practice. It is argued that these characteristics were important for the teachers to experience both an expectancy of succeeding in implementing a more formative classroom practice and its value, which affects their motivation to carry out the change in practice. The following characteristics were identified as most important: (1) formative and process-oriented professional development program, (2) easy and time efficient teaching
activities with immediate positive effects on teachers and students, (3) possibilities of experiencing the benefits of this way of teaching, (4) theory and understanding of formative assessment, and (5) time and (6) knowledgeable support. The teachers’ responses in the interview accentuated the professional development program as a whole, with the mix of input from the lectures and literature together with practice and reflections alone and in groups. The teachers recognized that time was necessary for these features to be possible, both time set aside for the activities and a long duration of time.

Generally, the teachers were satisfied with the conditions they had when participating in the professional development program albeit two conditions at their schools were perceived by some teachers as a threat to the expectancy of succeeding in implementing formative assessment and therefore their motivation of doing so. The first condition was about specific demands concerning grouping of students or concerning characteristics of students (e.g. age or special needs). The second condition referred to time in terms of too high workload or insufficient time to plan and evaluate lessons.
5.4 Article 4

The aim of the study in this article was to investigate whether the teaching conducted by the teachers that participated in the professional development program in formative assessment would lead to higher student achievement in mathematics than the teaching conducted by the control group. Based on the hypothesis that formative assessment practice is advanced and complex, and learning and implementing it require both substantial support and time, we carried out a professional development program that included substantial support for the participating teachers. In another article (Article 2) it was shown that, during the year after the professional development, all of the 22 randomly chosen teachers changed their teaching towards a more formative assessment practice. In the study presented in this article we employed a randomized controlled design to investigate if the changes in teachers' classroom practice had an impact on the students' achievement in mathematics. A one-way between groups analysis of covariance (ANCOVA) was conducted to evaluate the effect on student achievement from the formative classroom practice implemented by the teachers that participated in the professional development program.

The result of the ANCOVA shows that, after adjusting for the pretest scores, there was a significant difference in the scores on the posttest between the intervention group and the control group, $F(1, 42) = 4.71$, $MSE = 7.74$, $p = 0.036$, Cohen's $d = 0.66$. Thus, compared to the control group, the classes that were taught by the teachers who had participated in the professional development program significantly enhanced their mathematics achievement over one year of formative classroom practice.

Additional ANCOVAs were conducted to investigate whether the professional development program had a particular effect on either the classes' proficiency in solving tasks requiring only the application of procedures, or on tasks also requiring other mathematical processes. The analyses show that, after controlling for proficiency on procedural tasks in the pretest, the classes belonging to the intervention group achieved higher results on the posttest on this type of tasks. Similar results occurred on the posttest for the tasks requiring also other solution processes, after controlling for the proficiency on this more conceptual kind of tasks in pretest. However, in both cases the difference failed to reach statistical significance at the 0.05 level, $F(1, 42) = 4.72$, $MSE = 1.42$, $p = .075$, $d = 0.55$, for the procedural tasks and $F(1, 42) = 3.33$, $MSE = 4.28$, $p = 0.075$, $d = 0.56$ for the tasks measuring other processes.
6. Conclusions

The first article concluded that the teachers used formative assessment, but there was also much room for improving this formative assessment classroom practice. For example, the teachers could use more well thought–out questions together with using an all-response system to engage all students and raise the potential to contribute to adjustments of instruction that are tailored to meet student learning needs better. Currently, many of the adjustments were based on information about a few students and the modification was made for the entire class or only for one student at a time, which is not so efficient and time-saving.

The second article concluded that all 22 teachers who had participated in the professional development program in formative assessment had implemented new formative assessment activities into their mathematics classroom practice. It was also concluded that all the teachers used those activities with fidelity to the fundamental principles in formative assessment.

The third article concluded that the expectancy-value theory of achievement motivation functioned well to explain why the teachers’ made substantial changes in in their classroom practice regarding formative assessment. The teachers were motivated, believed they could change their practice, and chose activities that they experienced as most valuable in relation to the cost of using that activity. The study identified six characteristics of the professional development program that were important for teachers to experience an expectancy of success for teaching in this way and experiencing its value, which affects their motivation to carry out the change in practice. The teachers referred to the wholeness of the program as advantageous. In general, the teachers were satisfied with the conditions they had when participating in the professional development program, but identified a few barriers to change in their school environment.

The fourth article concluded that the professional development program afforded the support needed for the teachers to develop their formative assessment practice to such an extent that the promising learning gains the literature indicates as possible with formative assessment was realized.
7. Discussion

The aim of the intervention study presented in this thesis was to see if ordinary teachers, who got the chance to participate in a professional development program with many contact hours and substantial support of an expert, would implement formative assessment in a way that could increase their students’ learning in mathematics. Another aim was to examine what motivated the teachers to implement new formative assessment activities into their mathematics classroom practices.

In this chapter I first discuss the four articles and reflect on how the studies, by themselves and together, contribute to extending the knowledge about how to support teachers to implement high quality formative assessment into the mathematics classroom practice. Second I discuss the findings from three themes: the contribution of using a framework conceptualising formative assessment as a unity of integrated formative assessment strategies; the value of characterising high quality formative assessment; and the identified important aspects of professional development programs in formative assessment. After that I discuss the implications for educational practice as well as for the research field of formative assessment. The chapter ends with my suggestions for future research in the field and some final words.

7.1 Contribution of the research presented in this thesis

The value of the results from the thesis comes from the cumulative contribution from the four studies summarised above. Together the four studies contribute to the understandings about how to support teachers’ implementation of high quality formative assessment. By themselves the studies in different ways have added knowledge of importance for those understandings.

The first article (Study I) provided information and examples of ways the mathematics teachers used formative assessment activities and principles, but also identified potential areas for development in relation to the five key strategies and the fundamental idea. These findings indicate how formative assessment activities and general principles in formative assessment are used in a traditional mathematics classroom (i.e. in classrooms where the teacher has not participated in a comprehensive professional development program in formative assessment in mathematics). Those findings could be
useful in the planning of teacher education and professional development programs for teachers. Moreover, the findings about mathematics teacher’s use of formative assessment in a Swedish municipality could be used for comparison with similar insights that already exist in other countries.

The second article (Study II) characterised the changes made in the classroom practice and the kind of formative assessment that had shown an effect on student achievement (Study IV). Thus, it was concluded that the teachers had changed their use of formative assessment to a level that enhanced their students’ achievement in mathematics.

The teachers in the research study had not just added new formative assessment activities, they used those activities in line with the intended function. None of the teachers seem to have only implemented an instrumental use of new formative assessment activities, which have been reported in several other studies (e.g. Hume & Coll, 2009; James & McCormick, 2009; Marshall & Jane Drummond, 2006; Torrance & Pryor, 2001). The new activities were conducted to strengthen their new focus on a formative assessment classroom practice based on the fundamental idea of using evidence of student learning to modify teaching and learning to better meet student learning needs. This is notable since much professional development has had little impact on classroom practices, implementation of formative assessment is difficult, and that these teachers were randomly selected and thus not a selection of especially interested teachers.

To contribute to the understandings about how to support teachers’ implementation of formative assessment it is important to know what kind of formative assessment to include in a professional development program. In Study II it was suggested that the teachers’ fidelity to the fundamental idea and the principles underlying formative assessment was particularly important for high quality formative assessment. In addition, it was suggested that three dimensions in which the teachers had developed their formative assessment classroom practice could be linked to opportunities for student learning. These two characteristics of changes in the teacher classroom practice, the fidelity to the underlying fundamental idea and principles in formative assessment and the development in three dimensions, are valuable knowledge about what kind of formative assessment to include in a professional development program.

The third article (Study III) showed that the teachers were motivated to make changes in their classroom practice. It was also evident that the teachers actively chose formative assessment activities that they found valuable for them and their students, and did not cost them too much in
form of time and effort. Such insights is most valuable to consider in planning for professional development, so that the program affords the teachers both to experience value of formative activities and expectancies of efficiency to be able to use those activities in the classroom. Thus, the article demonstrated the usefulness of the expectancy-value theory in understanding teachers’ professional development. However, this should not be interpreted as all demanding (high cost) formative assessment activities should not be used, but the motivation perspective implies the need for teachers to have time to prepare for the activities, to experience the value of using the activities and possibilities to overcome setbacks in the implementation of activities.

The third article also identified the six most important characteristics in the professional development program that had supported the teachers’ implementation of formative assessment into their mathematics classroom practice: (1) a formative and process-oriented professional development program; (2) activities directly useable in classrooms; (3) experience of using formative assessment activities; (4) connection between theory and practice; (5) time; and (6) knowledgeable support. Those characteristics very much confirm the previous literature on professional development in general and in formative assessment in particular (see Chapter 3). Such literature was used in the design of the professional development program in the intervention study described in this thesis. The effects of this design were tested in Study II and IV. Study III provided insights about the teachers’ experiences of important characteristics of the professional development program.

The fourth article (Study IV) made evident that the professional development program had been successful to support the teachers to use new formative assessment activities in their mathematics classroom practice to a level that had a significant effect on student achievement. Thus, the teachers’ new formative assessment classroom practices were empirically linked to enhanced student achievement. It is important to investigate the impact of professional development programs on teachers’ classroom practice and student achievement. The lack of such empirical studies have been noticed by several researchers (e.g. Schneider & Randel, 2010; Tierney, 2006; Vescio, Ross, & Adams, 2008).

Research evidence about important characteristics of successful professional development programs in formative assessment is of extreme value for education systems trying to improve teaching and students’ learning. The research presented in this thesis provides such evidence by: the identified important characteristics of the professional development program in
formative assessment; the characterisation of the formative assessment that would be valuable to include in such programs; the dimensions in which a formative assessment classroom practice can be developed to afford new learning opportunities for students; the motivation-based explanation for the implementation made by the teachers; and the impact on student achievement in mathematics. In addition the research in this thesis provided insights about how formative assessment activities and principles are used in a traditional mathematics classroom in the municipality, which is useful in the planning of teacher education and professional development programs for teachers.

In the next section I show the contribution of using the conceptualisation of formative assessment as an integrated unity of strategies in formative assessment (conceptualised as one fundamental idea and five key strategies).

7.2 A unity of integrated formative assessment strategies

Few empirical research studies have used a conceptualisation of formative assessment as a unity of integrated formative assessment strategies. Most research shows the potential of enhancing student learning from individual or few formative assessment strategies (see Subchapter 2.4.2). Since the key strategies are not independent of each other (see Subchapter 2.3.2), this implicates the gains of combining those different strategies.

The framework that integrate formative assessment strategies into a unity was essential for the possibility to analyse and describe the teachers' use of formative assessment activities in the research presented in this thesis. This framework made it possible to show how the different formative assessment strategies worked together and in relation to the fundamental idea of formative assessment. In addition, by using this framework it was possible to show how the teachers built on and developed their previous use of formative assessment.

By using this conceptualisation it was possible to define high quality formative assessment and the three-dimensional possibility to expand the use of formative assessment in classroom practice. In other words, we could define the kind of formative assessment for teachers to strive for and possible directions for development. Other quality indicators and dimensions for development may be possible, but those proposed in this thesis are central in formative assessment.
In the next section I clarify the value of characterising high quality formative assessment classroom practice based on formative assessment as a unity of integrated formative assessment strategies.

**7.3 The value of characterising high quality formative assessment classroom practice**

In this thesis, high quality formative assessment is defined as a use of formative assessment activities following the underlying principles of the five key strategies, with a fidelity to the fundamental idea in formative assessment. The opposite (low quality) would be a use of formative assessment activities without following those general principles and the fundamental idea in formative assessment. This view of high and low qualitative formative assessment is very much consistent with the notions of using formative assessment “by the spirit” for an integrated formative practice into the flow of lessons and “by the letter” for an instrumental use of formative assessment (James & McCormish, 2009; Marshall & Drummond, 2006). The difference in our definition of high quality formative assessment lies within how we conceptualise formative assessment.

As described earlier, (1) there is a large amount of studies showing the impact of formative assessment practices on student achievement, (2) there is a variation in conceptualisation of formative assessment, (3) these conceptualisations have a common core in the fundamental idea, and (4) implementation of formative assessment has caused controversies about the available evidence for large achievement gains from formative assessment and experiences of difficulties in identifying best practices related to formative assessment.

Thus, when referring to the effects of formative assessment it is most important to be clear about the way formative assessment is conceptualized. Similarly, for purposes of gaining valuable insights about high quality formative assessment it is important that implementations of formative assessment that in scientific studies are empirically linked to student achievement are carefully analyzed and described. These analyses may provide information about specifics of such practice as well as how these specific characteristics may have functioned as part of an enhanced learning process.

In this thesis I have emphasized the importance of the fidelity to the underlying principles and fundamental idea in formative assessment and the use of integrated strategies in line with those principles as indicators of high
quality formative assessment. Focusing on the fundamental general principles, there were several other indicators of qualitative use of formative assessment that were not systematically investigated in the studies, for example the reliability and validity in the assessment process. At a more detailed level, the quality in teachers’ use of the formative assessment activities is of course very important. For example, concerning the quality of the learning intentions; the tasks and questions; and the interaction between the agents (teacher, learner, peers) in the classroom. We did have information about the way teachers’ used the formative assessment activities, but not always at a very detailed level. For example we did not collect information about how the teachers formulated their tasks. Thus, I describe high qualitative formative assessment at a macro level rather than a micro level, but without saying that the micro level is not important.

So far, I have discussed the kind of formative assessment to include in professional development programs. Next I discuss other important aspects to consider in the design of professional development programs in formative assessment.

7.4 Important characteristics of professional development programs in formative assessment

The design of the professional development program set up as a part of the intervention study described in this thesis was inspired by previous research (see Subchapters 3.2, 3.3 and 3.4). The sub studies presented show that the professional development was successful in motivating the teachers to make changes in their classroom practice to an extent that enhanced the student achievement in mathematics. The program afforded the teachers ground for the strengthening of the teachers’ value beliefs of this practice and expectancy of success in implementation. Teachers’ feelings and beliefs are known to affect their actions and consequently needs to be considered in design of professional development in formative assessment (e.g. Black & Wiliam, 1998b; Marchall & Drummond, 2006; Cooper & Cowie, 2010).

Article 3 provided the six characteristics of the professional development program that the teachers experienced as most important for their development of a formative classroom practice. The teachers’ responses accentuated the wholeness of the program, the iterative processes of theoretical input, experiments and reflection, together with the formative approach of the program. The teachers appreciated that the program was continuously adjusted to meet their questions and needs.
Theoretically the formative assessment approach of the professional development program is indicated from the use of information about teachers’ use of formative assessment before the professional development program (Study I, but also continuously during the implementation process) as a baseline process (establishing where they are) to make adjustment in the design of the professional development program. Similarly, the directional process (establishing where they are going) in formative assessment is indicated in the definition of high quality formative assessment as a competent use of the five key strategies and an adherence to the fundamental idea. The conceptualisation of formative assessment used in our research studies afforded viewing development of formative assessment in three dimensions, which can be related to the process of movement in the right direction (establishing what needs to be done to get where the learner is going). As described in a previous chapter (2.2.3), formative assessment is theoretically based on these three processes.

In practice this means that the teachers already before the in-service training have developed their individual way of teaching. Thus, the implementation of new formative assessment activities should be individually adapted to that previous classroom practice. The professional development program encouraged the teachers to find their own way for implementation of formative assessment. In addition, the program supported the teachers’ implementation of new formative assessment activities through defining formative assessment as a fundamental idea and five key strategies. As argued above (Chapter 7.3), this conceptualisation has the potential to define high quality formative assessment practice. It is likely that this conceptualisation helped the teachers to identify what a high quality formative assessment classroom means. In our second study (Article II) we saw that the teachers implemented new formative assessment activities (took the next step) in three ways (dimensions): by strengthening any of the fundamental processes in teaching and learning, by involving all agents in the classroom, and by shortening the adjustment cycle. Any such next step in developing the use of formative assessment in classroom practice must be individually adapted to previous classroom practice guided by a goal of a high quality formative assessment practice.

As in student learning, individualisation does not mean that teachers should not work together. On the contrary, teacher learning communities have been promoted to be especially supportive for teachers’ learning about formative assessment (Wiliam, 2007a). In the professional development program described in this thesis, the teachers from 22 different schools formed a teacher learning community. Thus, it was possible for the teachers to function as resources for each other in the professional development.
From this theoretical and practical point of view, I have discussed what a formative professional development program might mean. Next I will discuss how the process-oriented aspect of the program was dependent on the other five important characteristics identified.

During the professional development program the teachers functioned as resources for each other in the work and discussions in whole group and smaller group constellations. They discussed literature, lectures and experiences from practice. The connection between theory and practice was indicated as an important aspect by the teachers. Throughout the professional development program the teachers made connections of theories about formative assessment both to their previous practical experiences of being a teacher and their experiences of using new formative activities in their classroom practice. Consequently it was important to provide activities directly useable in classrooms to create such situations in which they could generate such new experiences of using formative assessment activities. The teachers experienced the discussions after the weekly experiments as a very important driving force, but also as an opportunity for teamwork to overcome barriers in using any of the formative activities. The conceptualisation of formative assessment as a fundamental idea and five key strategies can be seen as facilitating the connection between theory and practice. As mentioned above, the framework, with clear implication for practical use, can support the teachers to apply theory of high quality formative assessment into practice.

The project leader brought knowledgeable support by providing the theoretical foundations of formative assessment, but also by organizing the possibilities to connect to practice and arrange learning situations that promoted cooperative problem solving that supported the teachers in the process of implementation. In addition, the project leader lead the discussions to focus on main factors in formative assessment, and at the same time listened for and focused on the needs of the teacher. Time is an essential aspect to make all these other important aspects in the program possible - both actual time to, for example, plan, discuss and reflect, but also duration of time that afford possibilities to make iterative experiments to discuss and reflect on.

A weakness in the professional development program was that the work in the learning community ended when the professional development program ended, and lack of support to create new teacher learning communities at the schools after the professional development program ended. After the program was finished it was hard for the teachers to meet since they worked at different schools spread out in the municipality. A strength of the program
was the affordance for the teachers to actively engage in formative assessment activities to develop their skills, understanding and beliefs about formative assessment as self-regulated learners in a collaborative and supportive environment.

Even though formative assessment might be promoted in policy documents and in line with reform initiatives as in Sweden (see Chapter 1.4) implementation of formative assessment cannot be expected to be straightforward. Teachers not entering a professional development program from own initiative may not see their teaching as being of low quality and causing problems. In addition, many teachers experience time pressured and high workload. Thus, entering the professional development all teachers may not have the expectancy and value beliefs required for sufficient engagement in the program and be hesitating engagement because of high cost expectancies. The importance of long duration has been pointed out as particularly true for professional development programs in formative (Schneider & Randel, 2010, p. 271). Since formative assessment is complex the teachers would need much time to learn, make mistakes and plan for implementation. Therefore, time must be provided to the teachers so that they can engage in the formative assessment activities by which they may develop the necessary expectancy and value beliefs. The expert can support the teachers when they have doubts about the value of the new practice, when they feel they do not have sufficient understanding of the concept and practice of formative assessment, when they experience difficulties in implementation, and when they feel pressured by other responsibilities. Such support may have to take into consideration content issues, as well as social and emotional issues. Consistent with formative assessment practice, the support may also be most useful when it is based on knowledge about individual’s needs, and therefore information about these needs must be recurrently gathered.

Next the implications for educational practice are discussed.

7.5 Implications for educational practice

The education system calls for research-based ways to improve teaching and learning. Formative assessment has received attention for its potential and many reform initiatives in formative assessment are taken. The research findings presented in this thesis can be most valuable in such reforms by its implication for what formative assessment to include in professional development programs in formative assessment and important aspect to consider in the design of the programs.
Thus, schools and teachers can benefit from improved professional development programs, meaning programs that provide appropriate support to the teachers. The research presented in this thesis can also contribute to the content level in such a program by providing terms, concepts and models for teachers to use in talking and thinking about the complex formative assessment classroom practice. I found that formative assessment to a high degree coincided with the teachers desired classroom practice (Study III), but that the teachers lacked words to use, guidance and practical tools. The framework used in this thesis can support teachers by providing words to use, and structured guidance about how to think about the classroom practice as well as what formative assessment activities to use. For example, it would be possible for any individual teacher to, based on their current use of formative assessment, identify in which dimension(s) of formative assessment their classroom practice could be developed. Based on what formative assessment activities already are used, the teacher can add new formative assessment activities guided by the underlying ideas and principles in formative assessment (structured from the fundamental idea and five key strategies).

Characterising the changes the teachers made in their classroom practice due to the professional development program also visualized the complexity of this practice. Consequently, it is justifiable to believe that developments in classroom practice would require major changes in most teachers’ practice, and significant support in professional development initiatives.

Next, implications for the educational research field of formative assessment are discussed.

7.6 Implications for the educational research field of formative assessment

There has been a request about consensus in the terminology of formative assessment and a sound research-validated framework for best practices in formative assessment (e.g. Dunn & Mulvenon, 2009). Much work has already been done to reach such consensus and arriving at a common framework (e.g. Klenowski, 2011), although some disagreements still remain (e.g. Wiliam, 2011a). In this research project we contribute to the research field by the empirical use of a framework that conceptualises formative assessment as a unity of integrated formative assessment strategies. Few studies have used such a conceptualisation. Building on other researchers’ work, by using a framework that put together different formative assessment aspects of the complex classroom practice, it was possible to contribute to
Moreover, studying the effect from a professional development program on teachers' classroom practice as well as on student achievement, made it possible to empirically link the formative assessment used in the classroom to increased student learning. These kinds of studies are extremely valuable in characterising high quality formative assessment. To be able to link in-service training to impact on classroom practice and student achievement, such studies need to provide detailed analysis and descriptions of what formative assessment is used in the classroom practice as well as clear descriptions of the support provided in the professional development program (Schneider & Randel, 2010). The intention of this thesis has been to do that.

Few studies have directly addressed the issue of understanding the processes involved in changing teaching practice (Timperley et al., 2007). The expectancy-value theory used in Study III functioned well to explain why the teachers participating in the professional development program made changes in their classroom practices, and therefore can be advantageous to use in similar interventions studies.

In summary, by building on previous empirical and theoretical research, it has been possible to meet some important requests from the research field and hopefully the generated knowledge can be used to build upon.

Next, my suggestions for further research work are provided.

### 7.7 Future research

The ideas for future research provided here concern characterisation of high quality formative assessment and dimensions for development and teachers' professional development seen from perspectives of self-regulated learning and co-regulated learning.

I do not say that the characterisation of high quality formative assessment put forward in this research is the one and only. Similarly, there might be other ways (dimensions) for development of formative assessment classroom practice not investigated in the project. Accordingly, future research could examine such quality aspects and dimensions for development. For example, computer support could be viewed as adding a fourth agent (or resource) into the classroom. Quality criteria related to the use of the computer as a
resource would be needed to guide the way the computer should be used and not be used.

Previous research (e.g. Timperley et al., 2007) as well as the research presented in this thesis implicate that studies of teachers’ development of formative assessment could use a teacher self-regulation perspective when studying the teacher development processes as well as a teacher co-regulation perspective studying teachers as resources for each other. Moreover, studying teacher learning communities in formative assessment could have a focus on the problem-solving process used in the groups of teachers, to study how teachers overcome obstacles in the implementation process to reach a status of using formative assessment activities as new fully implemented habits of mind. Lastly, research is also needed about what kind of teacher learning communities are most effective. For example, is it possible to say that school-based professional development is always the best?

7.8 Summary and final words

The research presented in this thesis contributes to the limited number of studies of professional development programs in formative assessment using a framework that integrates the strategies in formative assessment to a unity and to studies of formative assessment classroom practice empirically linked to student achievement accompanied by detailed descriptions of the actual teaching underlying this impact. A major contribution comes from the analysis of the characteristics of the new classroom practice based on formative assessment as a unity of integrated formative assessment strategies, and examples of how this practice affords new learning opportunities. The characteristics visualize the complexity of this practice, and thus why such developments would require major changes in most teachers’ practice and significant support in professional development initiatives.
Acknowledgements

This research process is finally coming to its end and with contradicting feelings I am finishing my doctoral studies. These five years are the most stimulating and interesting years in my carrier. They are also the most academically demanding, in the way it should be. I am finishing the work with sadness, joy, and gratefulness. I owe thanks to many persons who, in different ways, have contributed. First of all, the teachers participating in the research, who generously let me visit them in their classrooms and shared their stories and experiences with me. When listening to you I recognised how complex and complicated the work of teachers is.

Fortunately, during my years as a doctoral student, I have had the opportunity to meet and collaborate with many researchers and graduate students from whom I have learnt a lot, become inspired and been supported. In addition, you cared about me on travels that often became more adventurous than planned.

I would like to extend my warmest thanks to my first supervisor, docent Torulf Palm. I should need lot of space in this thesis to outline all kind of knowledge you shared with me. I would like to use the words brilliant and caring to characterise your support. I would also thank my other two supervisors docent Magnus Österholm and docent Mikael Winberg. Special thanks to Carina Granberg, who supported me in the end of completing this thesis.

I would also specially thank Lärarhögskolan (School of Education) in Umeå that gave me several opportunities to participate in courses and seminars. I learnt a lot that I do not know how I could have learnt about otherwise. Certainly those opportunities raised the quality of my doctoral studies by broadening my knowledge and giving me several new perspectives to think about educational research.

I also want to thank Professor Mary Hill at Auckland University who hosted me in the end of 2013. You shared with me your extreme amount of knowledge about formative assessment and experiences of applying formative assessment into classroom practice. You also put me in contact with other knowledgeable people and made my journey worth it all.

Especially warm thanks to my roommates and all staff at the Department of Science and Mathematics Education in Umeå.
I dedicate this thesis to my dog Lester who kept me company the last months of writing.

Umeå, March 2015

Catarina Andersson
References


Regeringsuppdrag. (U2013/6845/S). *Uppdrag att svara för genomförandet av kartläggningar av forskningsresultat med relevans för praktiskt arbete i skolväsendet* [Assignment to be responsible for the implementation of surveys of research relevant to practical work in the school system]. Stockholm: Utbildningsdepartementet.


Skolinspektionen. (2012). Framgång i undervisningen. En sammanställning av forskningsresultat som stöd för granskning på vetenskaplig grund [Success in teaching. A summary of the research findings that support the examination on a scientific basis]. Stockholm: Skolinspektionen.


112