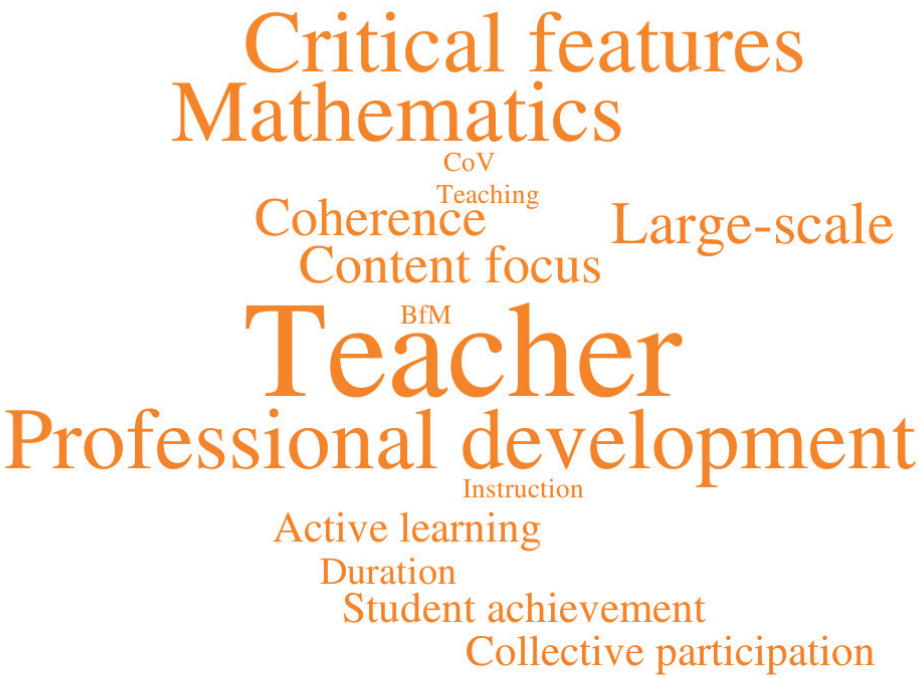
Mälardalen University Doctoral Dissertation 249

Critical features and impacts of mathematics teacher professional development programs

Comparing and characterizing programs implemented at scale

Jannika Lindvall



MÄLARDALEN UNIVERSITY
SWEDEN

Mälardalen University Press Dissertations
No. 249

CRITICAL FEATURES AND IMPACTS OF MATHEMATICS TEACHER PROFESSIONAL DEVELOPMENT PROGRAMS

COMPARING AND CHARACTERIZING PROGRAMS IMPLEMENTED AT SCALE

Jannika Lindvall

2017

School of Education, Culture and Communication

ISBN 978-91-7485-364-3
ISSN 1651-4238
Printed by E-Print AB, Stockholm, Sweden

Mälardalen University Press Dissertations
No. 249

CRITICAL FEATURES AND IMPACTS OF MATHEMATICS TEACHER PROFESSIONAL DEVELOPMENT PROGRAMS
COMPARING AND CHARACTERIZING PROGRAMS IMPLEMENTED AT SCALE

Jannika Lindvall

Akademisk avhandling

som för avläggande av filosofie doktorsexamen i didaktik vid Akademin för utbildning, kultur och kommunikation kommer att offentligen försvaras fredagen den 19 januari 2018, 13.15 i Zeta, Mälardalens högskola, Västerås.

Fakultetsopponent: Docent Torulf Palm, Umeå universitet

Akademin för utbildning, kultur och kommunikation

Abstract

The aim of this thesis is to contribute to the knowledge base on conceptualizations and impacts of teacher professional development (PD) programs. This is done by studying the case of two large-scale teacher PD programs. The first was mandatory for all public elementary school mathematics teachers in a larger Swedish municipality, and the second has been completed by 76% of all elementary school mathematics teachers in Sweden. In the municipality, and during the time frame in which this study was conducted, it was possible to make a comparison between teachers participating in different programs. Data on the programs' impacts on teachers, instruction, and student achievement were gathered both immediately and one year after the teachers' participation in one of them. In other words, the context of the study created an opportunity to respond to recent calls for studies that (a) examine the impacts of PD programs implemented on a larger scale, (b) adhere to PD programs' impacts on teachers, instruction, *and* student achievement, (c) examine the sustainability of PD programs' impacts, and (d) attend to variations within and between PD programs' impacts. The results show that the studied programs are highly similar if characterized according to established research frameworks on what constitutes critical features of teacher PD. At the same time, they demonstrate different impacts, both between the programs and within them. These results suggested an elaboration of two of the five critical features of teacher PD: Content Focus and Coherence. Through the development and application of a more finely grained tool to characterize the programs' Content Focus, differences between their characteristics were detected. Together with a systematic review of the PD research literature on Coherence, these results formed a basis for discussing plausible reasons for the difference in the programs' impacts as well as elaborating on the critical features of Content Focus and Coherence. In summary, the thesis contributes: (a) empirical results in relation to large-scale teacher PD programs' impacts; (b) methodological results in the form of tools for characterizing PD programs' Content Focus and Coherence; and (c) theoretical results, as it examines established frameworks for characterizing teacher PD programs by using them in practice and, in light of the results, also suggests an elaboration of them.

ISBN 978-91-7485-364-3
ISSN 1651-4238

To my family

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Lindvall, J. (2017). Two large-scale professional development programs for mathematics teachers and their impact on student achievement. *International Journal of Science and Mathematics Education, 15*(7), 1281-1301.

II Lindvall, J., Helenius, O., & Wiberg, M. (2017). Critical features of professional development programs: Comparing content focus and impact of two large-scale programs. *Teaching and Teacher Education*. Advance online publication. doi: 10.1016/j.tate.2017.11.013

III Lindvall, J. (2017). Large-scale professional development and its impact on mathematics instruction: Differences between primary and secondary grades. In J. Häggström, E. Norén, J. van Bommel, J. Sayers, O. Helenius & Y. Liljekvist (Eds.), *ICT in mathematics education: The future and the realities. Proceedings of MADIF10. The tenth research seminar of the Swedish Society for Research in Mathematics Education* (pp. 57–66). Gothenburg, Sweden: SMDF/NCM.

IV Lindvall, J. (2017). *Large-scale professional development and teacher change: The case of Boost for Mathematics*. Paper to be presented at the eleventh research seminar of the Swedish Society for Research in Mathematics Education, MADIF 11. Sweden: Karlstad, 23-24 January 2018.

V Lindvall, J., & Ryve A. (submitted). *Coherence and the positioning of teachers' in professional development: A systematic review*.

Preface

About five years ago, the political leadership in Västerås, one of the larger municipalities in Sweden, made the decision to invest a considerable amount of resources (1.6 million EUR) in the project Count on Västerås (CoV). The project included a year-long PD program for teachers, and was carried out in cooperation between Mälardalen University and the City of Västerås. Its main aims were to develop the mathematics teaching in the municipality's elementary schools and improve student achievement. At the same time as CoV was initiated, so was the national state-coordinated project Boost for Mathematics (BfM), run by the National Agency for Education. Like CoV, BfM also included a teacher professional development (PD) program. The program was offered to all teachers teaching mathematics in Sweden; and, in Västerås, participation was mandatory for all teachers teaching mathematics at a public elementary school. These circumstances created a rather unique situation in Västerås, as two large-scale PD programs with similar overarching aims and targeted at the same teachers were conducted within the same municipality and time frame.

Employed as both a doctoral student at the university and a mathematics mentor for the public elementary schools in Västerås, I had the opportunity to participate in the project group for CoV. This group, at the time, was led by Andreas Ryve, Professor of Mathematics Education, and Bodil Lövgren, mathematics developer for the City of Västerås. As a mathematics mentor, my main tasks were twofold. First, I led numerous collegial discussions among teachers participating in the project's teacher PD program. Secondly, I took on a large responsibility regarding the development and implementation of tools and procedures for evaluating the project. In the thesis, I have chosen to focus more deeply on selected parts of the collected data.

Initially, my doctoral studies were focused on how to support instructional improvement at scale. This focus resulted in a licentiate essay that attended to the role of two common approaches to supporting mathematics teachers' development of reform-based practices: teacher PD programs and curriculum materials. In my continuing doctoral studies,

however, I made the decision to look more deeply into the matter of teacher PD. The reasons for this were several. First, during the process of writing the licentiate essay, it became more and more apparent how unique the situation in Västerås was with regard to having two large-scale PD programs conducted within the same municipality and time frame. Second, the two (out of four) papers in my licentiate thesis that focused on teacher PD showed differences both between the programs' impacts as well as within them, even though they were largely similar if characterized according to established research frameworks. Third, within the project group we had access to a large amount of data on both of the projects, which could potentially support further analysis aiming to explore possible reasons for the detected differences. As a result, this doctoral thesis is a continuation of my licentiate essay (Lindvall, 2016) in which I focus on critical features and impacts of large-scale teacher PD programs in more depth. The reader should note that since this compilation thesis and the licentiate essay partially address the same subject, and partially include the same papers (Papers I and III), some of the introductory chapters (*kappa* in Swedish) in the thesis and the essay are similar. This applies particularly to specific parts of the methodology and results sections.

Contents

1 Introduction

Most teachers in high-income economies are currently involved in some kind of professional development (PD)[1] activity (OECD, 2016), and foundations and governments are investing large sums in the design and implementation of teacher PD programs (e.g., Swedish Ministry of Education, 2012; U.S. Department of Education, 2014). Despite this, research and empirical evidence related to teacher PD is a relatively new phenomenon and has only gained momentum mainly in the past three decades (Borko, 2004; Wilson & Berne, 1999). The concept has taken hold quickly, however. Today, the PD of teachers is argued to be a key to improving the quality of instruction and student achievement (e.g., Borko, 2004; Desimone, 2009; Kennedy, 2016; Timperley, Wilson, Barrar, & Fung, 2007).

In light of the increasing investments in teacher development, several reports on critical features of high-quality PD have emerged (e.g., Borko, Jacobs, & Koellner, 2010; Darling-Hammond & McLaughlin, 1995; Desimone, 2009; Garet, Porter, Desimone, Birman, & Yoon, 2001; Guskey, 2003). Features appearing in these reports can be related to either the PD content or the process and structure of the PD. Content features concern *Content Focus*, which means that the PD should focus on a combination of subject-specific matters and pedagogy, and *Coherence*, which means that the content should align with policy standards and teachers' knowledge and beliefs. Structural and process features, on the

[1] The term PD has been used to refer to a wide range of complex, interrelated, formal, and informal teacher learning opportunities (cf. Desimone, 2009; Lieberman & Miller, 2014). For example, Kriewaldt (2008) uses the terms professional learning and PD as generally synonymous, while others (e.g. Firestone & Mangin, 2014; Webster-Wright, 2009) distinguish between them. In this thesis, I build on Mayer and Lloyd's (2010) discussion of the concepts and regard PD as the planned activities teachers engage in to improve their practice, while professional learning refers to changed practices. This change can be a result of PD, but also of the informal learning that takes place in teachers' everyday practices. I acknowledge that the term teacher change can be interpreted in many ways, but the central focus of most current PD literature, as well as the view adopted in this thesis, most closely aligns with change as growth, development, and a process that involves learning (Clarke & Hollingsworth, 2002; Sowder, 2007).

other hand, relate to recommendations that the PD activities involve teachers working together with colleagues (*Collective Participation*), include multiple sessions spread over a longer period of time (*Duration*), and give teachers the opportunity to actively engage in tasks connected to their classroom practices (*Active Learning*). Some years ago, agreement about the five critical features mentioned above reached a level such that many in the field regarded it as consensus (e.g., Desimone, 2009; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Russell, Kleiman, Carey, & Douglas, 2009).

This consensus has been questioned lately, however. For example, scholars (Hill, Beisiegel, & Jacob, 2013; Kennedy, 2016) have argued that the core critical features are mainly derived from small-scale studies and do not address questions of sustainability, as data are collected only during, or immediately after, the PD programs' implementation. In addition, even programs incorporating the recommendations in research literature have shown to have difficulties in improving student achievement (Garet, Heppen, Walters, Smith, & Yang, 2016; Jacob, Hill, & Corey, 2017), and many of the terms used in the literature on teacher PD have been applied and defined in different ways (Sztajn, Campbell, & Yoon, 2009). This seems to be particularly widespread concerning critical features related to content (Salinas, 2010), namely Content Focus (Kennedy, 2016) and Coherence (Firestone, Mangin, Cecilia Martinez, & Polovsky, 2005). Furthermore, concerns are being raised that, although the studies conducted on teacher PD have greatly contributed to our understanding of the field, up to this point they have only shown a limited capacity to support practice and policy (Bryk, Gomez, Grunow, & LeMahieu, 2015; Cobb & Jackson, 2011; Guskey, 2014b). In other words, we still need to develop more knowledge regarding teacher PD offered on a larger scale, with non-volunteers and in multiple contexts (Cobb & Jackson, 2011; Goldsmith, Doerr, & Lewis, 2013; Kennedy, 2016; Marrongelle, Sztajn, & Smith, 2013; Wayne, Yoon, Zhu, Cronen, & Garet, 2008), and we need studies that explicitly examine best practices and critical features of high-quality PD (Dede, Jass Ketelhut, Whitehouse, Breit, & McCloskey, 2008; Desimone, 2009; Desimone & Garet, 2015; Hill et al., 2013). This thesis responds to both these calls.

Serving as a basis for the thesis, allowing for the investigation of phenomena connected to the impacts and critical features of teacher PD programs, are two large-scale instructional improvement efforts that include PD programs for mathematics teachers: *Boost for Mathematics* (BfM) and *Count on Västerås* (CoV). The two programs serve as interesting cases for several reasons. First of all, they were conducted in

collaboration with a large number of teachers, including non-volunteers. Such studies of PD programs have been called for by several scholars (e.g., Cobb & Jackson, 2011; Goldsmith et al., 2013; Marrongelle et al., 2013). Secondly, to be able to identify the specific features of PD programs that cause their impacts, studies comparing different programs with slightly different features are needed (Borko et al., 2010; Hill et al., 2013). The context for this study was unusual, as two large-scale PD programs were conducted within the same municipality and time frame. This offered a unique opportunity for a comparison between programs. Thirdly, scholars have called for studies examining the sustainable effects of PD programs, i.e. the effects that remain when the programs have come to an end (e.g., Desimone & Stuckey, 2014; Kennedy, 2016; Wayne et al., 2008). The design and implementation of the programs examined in this thesis also allowed for data collection a year after the teachers' participation in them. Fourthly, questions on "what works" always depend on where, when, and with whom (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Guskey, 2002a; 2014a). It is therefore recommended that studies of educational reforms not only look at the overall results of specific PD programs' impacts, but also especially attend to variations in the sample (Bryk et al., 2015; Desimone & Stuckey, 2014). BfM and CoV were both conducted with teachers from Grades 1–9, which allowed for comparisons between teachers teaching mathematics at different grade levels.

1.1 Aim

The overarching aim of this thesis is to contribute to knowledge on conceptualizations and impacts of teacher PD programs. This is done by characterizing two large-scale PD programs for mathematics teachers and studying their impacts on teachers, instructional practices, and student achievement (Papers I–IV), and by systematically reviewing the literature on teacher PD (Paper V). Drawing on the results of the papers, the aims of the thesis, more specifically, is to (a) merge the results in order to give a more comprehensive view of the two programs' impact and possible reasons for the variations between them, and (b) elaborate on two frequently mentioned critical features of high-quality PD, namely Content Focus and Coherence.

1.2 How to read this thesis

To make progress in the field of research on teacher PD, scholars have argued for the need of a consistent use of frameworks and models for describing different PD initiatives (Desimone, 2009; Scheerens, 2010; Schoenfeld, 2015; Sztajn et al., 2009). In light of this argument, this thesis draws on the core conceptual framework for studying teacher PD proposed by Desimone (2009). The framework, which is further elaborated on in later sections of the introductory chapters of the thesis (*kappa*[2] in Swedish), consists of a model of how a PD program is expected to affect teachers, instructional practices, and student outcomes, as well as a set of five core critical features of high-quality PD.

1.2.1 Connections between the papers

All the papers (I–V) in this thesis can be connected to Desimone's (2009) framework for studying teacher PD. This is illustrated in Figure 1.

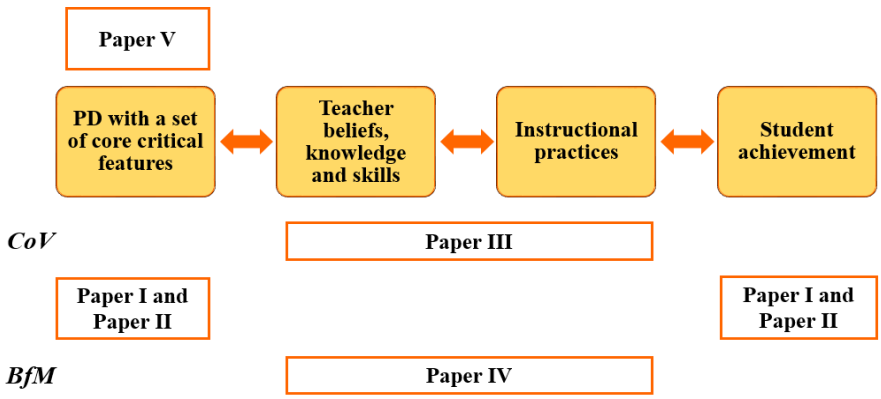


Figure 1. Connections between the papers based on Desimone's (2009) framework for studying teacher PD

Papers I-IV all use one, or both, of the PD programs BfM and CoV as cases to study large-scale teacher PD programs. Papers I and II attend to both programs, and focus on a characterization of them as well as their impacts on student achievement. In Paper I, the characterization is based on Desimone's (2009) core critical features of teacher PD, and the programs' impact on student achievement is measured just after the teachers' participation in one of the programs. The findings from Paper I

[2]In the absence of a concise English term for the introductory chapters of a compilation thesis, the Swedish term *kappa* is used in this thesis.

demonstrated differences in impacts on student achievement, both between the programs and between grade levels, even though the characterization suggested that they are largely similar. This motivated a need for additional studies to try to explain the reasons for the differences found. In Paper II, a finer characterization of the PD programs' Content Focus was therefore done by analyzing the PD materials they provided to teachers. In addition, an analysis of students' results again one year after the teachers' participation in the programs was conducted.

To get a more comprehensive view of the impacts of the two PD programs, Papers III and IV emphasize their effects on teachers and their instructional practices. Paper III examines how CoV supported and affected changes in the participating primary- and secondary-level teachers' reported mathematics instruction, respectively. Paper IV, on the other hand, focuses on the impact of BfM on fourth- and eighth-grade teachers' collaboration with colleagues, their confidence in carrying out mathematics instruction, and the support they received from school leadership.

The results from Papers I-IV raised several questions in regard to critical features of teacher PD programs mentioned in the research literature, especially the feature of Coherence. Therefore, Paper V explicitly adheres to core critical features of teacher PD programs, comprising a systematic review of the PD literature that emphasizes the critical feature of Coherence.

The kappa offers an opportunity to merge the results from Papers I–IV to give a more comprehensive view of the programs' characterizations and impacts. These results, together with the systematic literature review presented in Paper V and the kappa's literature review, offer an opportunity to reason about explanations for the differences found in the programs' impacts and to elaborate on critical features of high-quality PD.

1.2.2 Structure of the kappa

The kappa consists of seven chapters. This first gives a brief introduction, articulates the aims of the thesis, and describes the connections between the five papers.

In *Chapter 2*, different theories underlying teacher PD programs are described with a focus on the processes that are expected to happen in order to attain the program goals, in other words PD programs' change models (Chen, 2005, 2006, 2012). The change model in Desimone's (2009) framework is presented, but it is also compared to other relevant models within the field that will later support the discussion of the results.

Chapter 3 focuses on critical features of teacher PD. The chapter is organized according to Desimone's (2009) critical features of high-quality teacher PD, but also includes findings from other studies that both support and contradict her arguments.

Chapter 4 contains a description of the context of the study. First, the Swedish national educational context is described. Then, the PD programs in CoV and BfM are characterized using the critical features described in Chapter 3.

Chapter 5 accounts for the methodology. To begin with, the different methods and procedures used for collecting and analyzing data on the PD programs and the PD literature are described. Thereafter, the ethical considerations that have been taken throughout the study are discussed. Finally, the validity and reliability of the studies are reflected upon.

In *Chapter 6*, a summary of the five papers is provided, in which the results have the most prominent position.

In *Chapter 7*, the conclusions of the thesis are presented and discussed in relation to two themes connected to the thesis aims. The first theme considers the two PD programs' impacts and possible reasons for the variations between them, while the second considers an elaboration of two of Desimone's (2009) critical features of teacher PD: Content Focus and Coherence.

In *Chapter 8*, some concluding remarks are offered, related to possible limitations of the study and its contributions for research and practice in the context of teacher PD. Lastly, I reflect on the study's implications and offer some suggestions for future research related to impacts and critical features of teacher PD.

2 Theories underlying a professional development program

There are at least two central components in a conceptual framework for studying teacher PD (Desimone, 2009). One involves establishing how the PD program works to influence teacher and student outcomes, while the other concerns identifying the critical features that define effective PD. These reasonings can also be found in the broader perspective of program theories (cf. Chen, 2005, 2006, 2012). While engaging in the discussion of program theories is outside the scope of this thesis, two concepts that help clarify the above-mentioned central components of a conceptual framework for studying teacher PD are useful in the study. The first concept, referred to as a program's *change model*, concerns the causal processes that are expected to happen in order to attain the program's goals. For example, in the educational context, assumptions concerning students' poor achievement may be that it is the result of inadequate teaching methods. As a consequence, stakeholders might decide to invest resources in a PD program focusing on whatever teaching strategies are deemed effective. This concept can be connected to Desimone's (2009) argument that a conceptual framework for studying teacher PD needs to establish how the program works to influence teacher and student outcomes. The second concept concerns the design of the program and prescribes the components and activities, both within the reform itself and contextual factors, that are seen as necessary for its success. It dictates what components and activities will be needed to activate the change model, and is thus referred to as the program's *action model*. This resembles Desimone's (2009) notion of critical features that define effective PD.

In addition, all teacher PD programs aiming at improving student achievement rest on not one but two separate theories (cf. Desimone, 2009; Wayne et al., 2008). The first, the *theory of teacher change*, concerns the program's underlying assumptions about what causes increases in, for example, teachers' knowledge and/or practices and how this comes about. The second, the *theory of instruction*, concerns what in the changed practices leads to increases in student achievement as well as how this

happens. Both the theory of teacher change and theory of instruction include a change model and an action model. For example, a PD program's theory of teacher change incorporates both assumptions on the causal processes that are expected to happen in order for teachers to gain knowledge and/or change their practices, and assumptions on the critical features of the PD program that support these processes.

There is a substantial amount of literature related to instructional best practices (i.e. theories of instruction), but it is argued that studies identifying aspects of PD that reliably lead to teacher change have been more elusive (Desimone & Stuckey, 2014; Kennedy, 2016). Therefore, in this thesis the main focus is on theories of teacher change, with special attention to critical features of high-quality teacher PD programs. In other words, the main emphasis is on the action model of PD programs' theories of teacher change. Nevertheless, and as argued by Chen (2012), studies of educational reforms are studies of a package, and it is therefore impossible to completely distinguish all the components involved and the contexts in which the programs are implemented. For this reason, aspects concerning different change models connected to theories of teacher change will also be touched upon. Below, three different change models that are used in the thesis to varying degrees are presented. This is followed by an argument for why one of the models is used as the main basis for the thesis, and a description of how the other two complement it and thereby support the kappa's discussion.

2.1 Change models of teacher professional development programs

In the research field of teacher PD, different models have been used to describe the relationship between mainly four components: PD, teachers, instruction, and student outcomes. However, though many models include these four components or similar ones, they differ in how the components are assumed to relate to and affect one another.

Arguing that teacher change is a multifaceted process, some scholars (e.g., Clarke & Hollingsworth, 2002; Fishman, Marx, Best, & Tal, 2003) have presented complex and cyclical models. One of the most cited is probably Clarke and Hollingsworth's (2002) *Interconnected model of professional growth*, which suggests that change occurs through a mediating process of enactment and reflection in four domains: the *External Domain*, which includes external sources of information such as PD programs; the *Personal Domain*, including teachers' knowledge,

beliefs and attitudes; the *Domain of Practice*, for example teachers experimenting with new teaching strategies; and the *Domain of Consequence*, which comprises salient outcomes. This rather complex model, with multiple pathways and entry points, is argued to enable the identification of particular change sequences of teachers and thereby recognize the individual nature of teacher professional growth (Clarke & Hollingsworth, 2002).

Other models are linear, though very detailed, and therefore only focus on parts of the possible impacts of a reform. One example is *the Cognitive-Affective Model of Conceptual Change* (CAMCC) presented in Gregoire (2003). This model describes possible pathways between two of the four components, namely PD and teachers. It begins with a presentation of a reform message, which may be accepted or rejected by the teacher. An acceptance of the reform message can, according to the model, happen in two different ways depending on how the teacher interprets it. Firstly, a teacher can interpret the message as something that does not implicate the teacher herself, for example as something she is already doing. Such an interpretation is not accompanied by any discomfort and usually leads to a shallow processing of the reform message, which results in a superficial change. Secondly, the message can be interpreted as something that has a significant impact on the teacher herself, which in turn creates a stress appraisal. The teacher then has to decide whether she has the motivation and ability to go through with the reform. This process is affected by a range of factors, such as the teacher's self-efficacy, her beliefs about the value of the reform, and the support she has in the forms of time, tools, colleagues and school leadership. An evaluation that implementation of the reform is not wanted might, like the previous path, lead to a shallow processing of the reform message. If, however, the change demanded by the reform is perceived as worthwhile and achievable, the teacher accepts the challenge and formulates a goal of approaching the reform message. This approach leads to a systematic processing of the message, which results in true conceptual change.

The most commonly presented change model appearing in several studies, however, is a more simplistic and linear one. This model resembles the one presented in Figure 2 (e.g., Blank & de las Alas, 2009; Desimone, 2009; Scher & O'Reilly, 2009; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007), though some change the order of the components (e.g., Guskey,

2002b) to reflect the view that teacher change in beliefs may come last and as a function of improved student achievement.

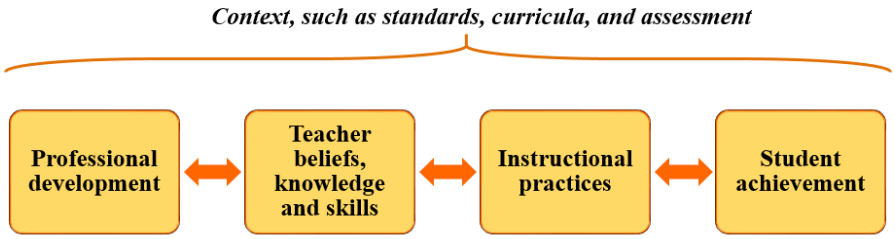


Figure 2. A linear change model for teacher professional development programs

Essentially, the model reflects that teachers' experiences from PD (with certain core critical features) should result in changes in their knowledge and beliefs, which in turn should foster changes in their instructional practices, which will finally contribute to increased student learning. The arrows between the components are usually double-ended, reflecting the interactive relationships between the components. Further, all this occurs within – and is also influenced by – the context, such as curriculum materials, student characteristics and the policy environment.

2.2 Theoretical point of departure

All models have their strengths and limitations, and I recognize that different models are more or less useful in different studies, depending on their research questions and available data. At the same time, a major obstacle to progress in the field of research on teacher PD is that there is no consistent use of frameworks and models for describing different PD initiatives, which makes it difficult to assemble the big picture (Desimone, 2009; Scheerens, 2010; Schoenfeld, 2015; Sztajn et al., 2009). In light of this, the current thesis adopts Desimone's (2009) framework for studying teacher PD as a theoretical point of departure. The framework consists of a linear change model resembling the one presented in Figure 2. It also includes an action model in terms of a set of five core critical features of high-quality PD, which is further elaborated on in Chapter 3 of the kappa. The reasons for choosing this framework are several. First of all, the article by Desimone (2009) is already widely used in studies on teacher PD, as proven by its many citations. Considering the argument that researchers within the field of teacher PD need to use a common language (e.g., concepts and frameworks) in reporting their findings (Schoenfeld, 2015; Sztajn et al., 2009), Desimone's framework could function as a common

base. In addition, similar models have been used in a range of other studies on teacher PD (e.g., Blank & de las Alas, 2009; Scher & O'Reilly, 2009; Yoon et al., 2007). Secondly, while it has not been exhaustively tested, the framework has received support from several studies linking PD programs with both teacher and student outcomes (e.g., Desimone, Porter, Garet, Yoon, & Birman, 2002; Penuel, Fishman, Yamaguchi, & Gallagher, 2007). Thirdly, as opposed to many other presented frameworks and models (e.g., Clarke & Hollingsworth, 2002; Gregoire, 2003; Guskey, 2002b), Desimone's framework includes both an operational theory for how the PD is assumed to work in influencing teachers, instruction and students (i.e. a change model) and a set of five core critical features of high-quality PD (i.e. an action model), for example Content Focus and Coherence. It is recommended that studies of educational reforms adhere to both (Chen, 2005, 2006, 2012).

As mentioned by Desimone (2009, p. 186) herself, however, the framework is basic, and "it is clear there are several potentially important components not included in the base model (…), as they have not yet been subject to much impact research". I acknowledge the simplistic viewpoint of the linear change model, but also contend that we need a common base that allows us to build on knowledge from different perspectives. Such a base needs to be simple in order to fit studies taking different approaches. In other words, adopting a more simplistic model enables a discussion of possible explanations of PD programs' impacts by drawing on selected parts of multiple change models. This is illustrated in this thesis. To begin with, Desimone's (2009) framework does not offer much of a description of what is included in the four components of the change model (PD, teachers, instruction and student achievement). It does, however, allow me to draw on Clarke and Hollingsworth's (2002) more detailed descriptions of the four components, which they refer to as the four dimensions of change. These descriptions function as support in explaining possible reasons for the differences found between the impacts of BfM and CoV on student achievement (see Section 7.1). Moreover, Gregoire's (2003) CAMCC can be interpreted as a more detailed account of the arrow between the first two components in the linear change model (PD and teachers), and complements it by revealing possible hindrances to the occurrence of true change. This extension of the linear change model, and the distinction between superficial and true conceptual change, supports the discussion of the critical feature of Coherence in relation to what kind of changes PD programs are aiming for (see Section 7.2). Consequently, the linear model serves as useful and basic support for studies on teacher PD. At the same time, it also has development potential and may need to

be complemented with other models depending on the aims of the specific studies.

3 Critical features of high-quality professional development

Apart from a change model, every intervention also rests on an action model concerning the design of the intervention program and prescribes the components and activities seen as necessary to its success (Chen, 2005; 2006; 2012). In the context of teacher PD, this can be related to the question of what constitutes critical features of high-quality teacher PD, which is also one of the kappa's two focuses.

In the past two decades, several scholars have presented reports on features of high-quality PD (e.g., Borko et al., 2010; Darling-Hammond & McLaughlin, 1995; Desimone, 2009; Guskey, 2003). These features can be sorted into either characteristics related to the content of the PD, or those related to its process and structure (Borko et al., 2010). One of the most cited reports is Desimone's (2009), in which it is argued that there is a consensus on five core critical features of teacher PD. First, the *Content Focus* of the PD should be on subject matter content and how students learn this content. Second, the teaching practices advocated in the PD should be *Coherent* with teachers' knowledge and beliefs, and aligned with school and state policies. Third, high-quality PD should engage teachers in *Active Learning* tasks, such as analyzing classroom videos or conducting mathematics lessons. Fourth and fifth, it should include sufficient *Duration* and *Collective Participation*, with teachers from the same school, grade or department all taking part in the in-service education and meeting regularly over an extended period of time.

In this section of the kappa, a brief overview of critical features of PD programs discussed in the research literature will be presented. This overview will be organized according to Desimone's (2009) content and structural features of high-quality teacher PD, but will also include findings from other studies within the field that both support and contradict her arguments.

3.1 Features concerning content

3.1.1 Content Focus

In an analysis of six reports on critical features of high-quality PD, Borko et al. (2010) concluded that features concerning PD content emphasized that it should be situated in practices and should focus on student learning. These two characteristics largely resemble the description of one of the most commonly occurring critical features of high-quality teacher PD cited in the literature — Content Focus (e.g., Desimone, 2009; Garet et al., 2001; Ingvarson, Meiers, & Beavis, 2005). This feature is described in terms of the assertion that the content of PD programs should focus on subject matter content, how students learn the content, and how to represent it in a meaningful way. The feature is further supported by empirical evidence, with a number of reviews and meta-analyses (e.g., Clewell, de Cohen, Campbell, & Perlman, 2005; Kennedy, 1998; Scher & O'Reilly, 2009; Slavin & Lake, 2008; Slavin, Lake, & Groff, 2009) suggesting that PD programs addressing both subject-specific content and pedagogy are the most effective ones when the aim is to improve student results. Still, Ball, Thames and Phelps (2008, p. 404) have stated that most subject matter courses for teachers, and teacher education courses in general, are "viewed by teachers, policy makers, and society at large as having little bearing on the day-to-day realities of teaching". Even in professional learning communities, it is argued that the literature on how students understand content and how to teach is neglected (Bausmith & Barry, 2011). This might explain why the critical feature of Content Focus is often described in terms of the need for the teacher PD to have a main focus on developing teachers' pedagogical content knowledge (PCK; Desimone, 2011; Knapp, 2003; Salinas, 2010).

In the area of mathematics educational research, the concept of PCK is highly influential. The most commonly used representation of the concept stems from either the definition by Shulman (1986, 1987) or the one by Ball et al. (2008) (Depaepe, Verschaffel, & Kelchtermans, 2013). Shulman (1986, p. 9) was the first to introduce the concept, and defined it as belonging "to the dimension of subject matter knowledge *for teaching*". Within this category of knowledge, he included knowledge of ways of representing subject content to make it comprehensible to others, as well as an understanding of what makes learning a specific subject easy or difficult. This type of knowledge is argued to be of special interest, as it is the type that is most likely to distinguish the understanding of a content specialist from that of a teacher (Shulman, 1987). Since 1986 several scholars (cf. Depaepe et al., 2013) have made attempts to refine

Schulman's conceptualization of PCK, one of the most influential of which is likely Ball et al.'s (2008) map containing different domains of mathematical knowledge for teaching. Within this map, Ball et al. (2008) differentiate between: (a) teachers' subject matter knowledge (SMK), which includes pure mathematical knowledge, specialized mathematical knowledge for teaching, and knowledge about how the mathematical topics in the curriculum are related; and (b) teachers' PCK, which includes knowledge about students' mathematical thinking, knowledge about the content and teaching, and knowledge of the curriculum and other instructional materials.

The concept of PCK is widely used in educational research, and there seems to be an emerging consensus that it should be emphasized in the PD content for teachers (Desimone, 2009; 2011). At the same time, there is also a lack of agreement on the concept's definition and interpretation (Depaepe et al., 2013). For instance, both Schulman's (1986) and Ball et al.'s (2008) conceptualizations of PCK have been criticized for mainly taking a perspective on teacher knowledge as something that can be controlled for, independent of the context in which it is used (Petrou & Goulding, 2011). For example, studies (Britt, Irwin, & Ritchie, 2001; McNeill & Knight, 2013) have shown that teachers participating in similar PD programs but at different grade levels experience unique challenges that need to be addressed for their specific contexts. Also, teachers themselves are asking for PD that is focused on both the content and the grade level they teach (Chval, Abell, Pareja, Musikul, & Ritzka, 2008), and it is proposed that future research needs to address the issue of whether PCK is different for teachers of different levels (Abell, 2008). In turn, the conceptualization of PCK as highly specific to context complicates the design of teacher PD programs (Van Driel & Berry, 2012), and can be linked to another commonly mentioned critical feature of teacher PD — Coherence.

3.1.2 Coherence

A critical feature mentioned in numerous studies within the research field of teacher PD (e.g., Desimone, 2009; Garet et al., 2001; Loucks-Horsley, Stiles, Mundry, Love, & Hewson, 2010) is that PD programs should be coherent. According to Desimone (2009), this feature is related to two different aspects. The first concerns the consistency of teachers' learning within the PD with their previous knowledge and beliefs. For example, the results from a study by Penuel et al. (2007) suggest that teachers' judgments about the coherence between the practices advocated in the PD

and their own goals for students' learning influence their implementation decisions. The critical feature of Coherence may thus explain the previous discussion regarding the possibility that PD programs with the same content may have different impacts on different teachers. In other words, teachers vary in their response to the same PD (Desimone & Stuckey, 2014).

The second aspect of Coherence concerns the alignment of school and state policies with what is taught in the PD. As mentioned by several scholars (e.g., Cobb & Jackson, 2011; Garet et al., 2001), the teaching practices advocated in PD programs, curriculum materials, national standards and assessments can facilitate teachers' efforts to improve their instruction if they provide a set of coherent goals. However, if the goals conflict, they may instead constrain teachers' efforts to develop their teaching practices in a consistent direction.

Besides the two aspects of Coherence mentioned by Desimone (2009), other studies present arguments that PD should be coherent with students' needs (e.g., Klinger, Volante, & Deluca, 2012; Robinson & Timperley, 2007) and teachers' practice (e.g., Brozo, 2014; Clark-Wilson, Hoyles, Noss, Vahey, & Roschelle, 2015). Further, some scholars describe Coherence more in line with structural features of high-quality PD, which is accounted for in the next part of the kappa.

3.2 Features concerning process and structure

3.2.1 Coherence

The critical feature of Coherence is mainly connected to the content of PD programs, but can also be described in relation to structural features. For example, both Firestone et al. (2005) and Villegas-Reimers (2003) stress that PD content should be delivered in a mode consistent with the pedagogy that is being promoted. Roth et al. (2011), on the other hand, argue that using just one main form of activity in a PD program (e.g., videocase-based analysis of practice) contributes to the coherence of the program.

3.2.2 Active Learning

High-quality PD should also include Active Learning. This means that teachers should have the opportunity to actively engage in a meaningful analysis of teaching and learning, for example by planning for mathematics instruction, reviewing student work, observing expert teachers, or being

observed (Desimone, 2009; Garet et al., 2001). In other words, it is recommended that PD providers treat teachers as active learners who construct their own understandings, and activities in which the preferred instructional strategies are modelled are particularly requested (Borko et al., 2010; Guskey, 2003).

The fact that teachers carry out activities does not, however, necessary imply that they are actively engaged in meaningful discussion or practice. For example, in reviewing the literature on teacher PD, Webster-Wright (2009) asserted that active engagement in experiences is not enough for professional learning, but that teachers also need to reflect upon these experiences. Moreover, in designing and trying out a questionnaire to measure characteristics of teacher PD, Soine and Lumpe (2014) found that items designed to measure Active Learning split into two separate components: *Active Learning in the Classroom* (e.g., items related to curriculum planning and instruction) and *Active Learning beyond the Classroom* (e.g., items related to presentation, coaching and observation). Of these two components, only the former showed a small correlation with teachers' use of new skills and knowledge. In other words, Active Learning alone is not enough; the content and context of the activities in which teachers are engaged must also be taken into account.

3.2.3 Duration

For the past 15 years, most research literature on teacher PD has called for programs that are intensive, ongoing, and sustained over time (Borko et al., 2010; Cohen & Hill, 2000; Garet et al., 2001; Guskey, 2003; Scher & O'Reilly, 2009; Timperley et al., 2007). These are the core aspects concerning the critical feature of Duration (Desimone, 2009). The duration of PD programs is suggested to be important for two reasons (Garet et al., 2001). Firstly, longer meetings give teachers greater opportunities to engage in in-depth discussions related to the PD content. Secondly, activities that are sustained over time (e.g., spread over one or several semesters) are more likely to give teachers a chance to try out new practices in their classrooms as well as obtain feedback on their instruction.

Even if the duration of teacher PD is generally considered as important, other factors have been shown to have greater influence on student results. Several studies (Kennedy, 1998, 2016; Scher & O'Reilly, 2009; Timperley et al., 2007) have demonstrated that the PD programs that provide the longest contact hours are not always the ones that show the greatest effects. In other words, the quality of the PD sessions is more important than the quantity. In addition, some changes are easier to elicit than others, and the

time needed will also depend on whether the PD program aims to develop teachers' content knowledge, procedural behaviors, or decision-making (Desimone & Stuckey, 2014). Yet, as mentioned by Timperley et al. (2007), comprehensive time frames are likely necessary in most studies of PD programs, as these programs involve substantive new learning that challenges teachers' already existing beliefs and practices.

3.2.4 Collective Participation

A critical feature of high-quality PD mentioned in several studies (e.g., Borko et al., 2010; Desimone, 2009; Garet et al., 2001; Guskey, 2003) is that programs should be organized around collaborative learning environments, and that groups of teachers from the same grade, subject, or school should participate in the PD together. PD programs that are designed for groups of teachers are believed to have several advantages in comparison with those designed for individual teachers (Garet et al., 2001). Firstly, these kinds of professional learning communities enable interaction and discourse among colleagues. Secondly, teachers from the same school or grade level are more likely to share common curriculum materials, course offerings, and assessment systems. Thirdly, teachers teaching the same students can engage in discussions around needs across grade levels and classes. Fourthly, the Collective Participation of teachers from the same school may support the development of a shared professional culture and a common understanding of instructional goals, which in turn may help sustain changes in practice over time.

At the same time, studies have shown that even in communities where teachers are given time to work together, significant gains in student achievement are not present (Kennedy, 2016; Timperley et al., 2007; Vescio, Ross, & Adams, 2008). *What* and *how* teachers discuss, and the support they are given, must also be taken into consideration. To begin with, the participating teachers need to be willing to openly discuss the issues they encounter in practice (Cobb & Jackson, 2011). Further, the discussions should be focused around investigating relationships between instructional practice and student learning (Vescio et al., 2008). Finally, these discussions often benefit from the support of external expertise, such as researchers within the field (Blank & de las Alas, 2009; Timperley et al., 2007).

3.2.5 Different methods for facilitating enactment

While these are not described as critical features of high-quality PD, some scholars (Gall & Vojtek, 1994; A. Kennedy, 2005; M. M. Kennedy, 2016) have classified and/or studied the impact of PD programs based on the programs' processes for communicating the PD content, namely their methods for facilitating teachers' enactment of the ideas presented in the PD. One of the most recent and extensive endeavors in this area is Kennedy's (2016) meta-analysis of the impacts of 28 PD programs on student achievement, in which she characterized the programs along a continuum of teacher autonomy. This framework deserves special attention in the kappa for several reasons. Firstly, it was formulated recently and is thereby based on current research on teacher PD. Secondly, the framework is empirically tested and used to characterize 28 PD programs and determine their impacts. Thirdly, Paper II makes use of the framework as a basis for characterizing the PD programs in focus in this thesis, and it thereby supports the kappa's discussion.

In short, Kennedy's (2016) framework is used to characterize PD programs according to four methods for facilitating teachers' enactment of the instructional strategies presented in the PD. The first category, *Prescription*, includes programs that offer teachers concrete descriptions of instructional activities or tasks to conduct with no explicit explanation of why. In other words, they are highly procedural and keep the amount of individual judgement by the teachers to a minimum. PD programs characterized into the second category, *Strategy*, are instead accompanied by a rationale to support teachers in why and when they should implement the strategies. These strategies can be as procedurally detailed as prescriptions, but are accompanied by a purpose and often include a multiple-choice quality. The third category, *Insight*, includes programs that rely on teachers' self-generated "aha moments", which can be stimulated by the PD providers raising provocative questions, forcing the teachers to reflect on familiar events differently. Insights are stated to alter teachers' behaviors in ways that cannot be prescribed by someone else, and thus rely on teachers to make their own decisions regarding how to respond. The fourth category for facilitating enactment is *Body of Knowledge*. These PD programs are inherently passive, and do not offer much in terms of stimulating particular teaching actions. Usually, teachers are presented with a number of concepts, for example in the form of a book, but with no additional support regarding whether or how to do anything with this knowledge. The results from Kennedy's (2016) meta-analysis indicate that PD programs relying on Strategies or Insights have a more positive impact

on student achievement compared to those characterized as Prescription or Body of Knowledge.

3.3 A consensus on the core critical features of high-quality PD?

Numerous scholars (e.g., Borko et al., 2010; Desimone, 2009; Penuel et al., 2007) in the field of teacher PD have argued that the support for many of the critical features described above can be regarded as consensus. While, for example, Desimone (2009) argues that there is a consensus on five core critical features, Borko et al. (2010) conclude that the agreed-upon features are relatively extensive and broad in number. Moreover, and as also demonstrated in Papers I and II, it has happened that even programs incorporating the critical features mentioned in the research literature have not led to improvements in student achievement (Garet et al., 2016; Jacob et al., 2017; Kennedy, 2016). In addition, concerns have been raised that the research literature on teacher PD has not yet provided the kind of clear guidelines needed to steer PD investments (Cobb & Jackson, 2011; Guskey, 2014b; Hill et al., 2013; Wayne et al., 2008), partly because many of the terms used have been applied and defined in different ways (Sztajn et al., 2009). This seems particularly prevalent concerning critical features related to content (i.e. Content Focus and Coherence), and it is argued that few studies have investigated high-quality PD with a lens focusing on content rather than a structural lens (Salinas, 2010).

Regarding Content Focus and the argument that PD content should emphasize teachers' PCK (Desimone, 2009; 2011), first note that there is a lack of agreement on the definition of the concept (Depaepe et al., 2013). This in turn may have the result that scholars, who all define the programs they are studying as having a Content Focus, do not mean the same thing since they use different definitions of PCK. Secondly, later studies have begun to discuss, or even question, the argument that PD programs should focus on teachers' development of PCK. For example, based on the findings from recent U.S. research, Desimone and Garet (2015) suggest that PD designed to foster teachers' use of straightforward procedural routines (e.g., writing the goal for the lesson on the board at the beginning of each lesson) has been more successful, or perhaps traceable, than PD designed to change teachers' subject matter knowledge and more complex teaching behaviors (e.g., how to follow up on students' responses in order to help them develop knowledge of the mathematical content in focus in the lesson). Moreover, the findings from Kennedy's (2016) meta-analysis

suggest that programs focusing exclusively on content knowledge tend to have less effect on student achievement than those with a more generic focus, such as enlisting student participation. Actually, "when programs offering content knowledge were successful, the content was subsumed under a broader goal, such as helping teachers learn to expose student thinking" (Kennedy, 2016, p. 27).

Scholars have begun to question the aforementioned consensus in relation to Coherence as well. For example, Saylor and Johnson (2014) recently conducted a meta-analysis on the role of reflection within PD for science and mathematics teachers, which showed that only three of 21 PD programs were classified as being coherent. One of the studies included in the analysis even suggested that Coherence was not a helpful part of PD. Moreover, Firestone et al. (2005, p. 416) argue that "coherent professional development is often recommended but rarely clearly defined". This argument is further supported by the results of the systematic literature review in Paper V, which concludes that the concept of Coherence is not used in a consistent manner and that the consensus thereby appears to be illusionary.

The above arguments illustrate that the consensus on Content Focus and Coherence as critical features of teacher PD might not be as strong today as indicated by Desimone in 2009. Even if they appear together with the other critical features, they do not seem to guarantee that PD will result in positive impacts on teachers, instructional practices, and student achievement. Additional studies that explicitly adhere to critical features of high-quality PD are therefore needed in order to support future research and practice. This thesis responds to this call, as one of its aims is to deepen our knowledge of the critical features of Content Focus and Coherence by reviewing the literature on teacher PD and comparing the impacts of two large-scale PD programs conducted in the same context.

4 Context of the study

The role of context plays a major part in both mediating and moderating the effects of an intervention (e.g., Borko, 2004; Chen, 2012; Desimone, 2009; Timperley et al., 2007). For this reason, any effects of a PD program must be understood in light of the contextual factors surrounding it (e.g., curriculum, policy environment, teacher characteristics). Therefore, in order to interpret the results of Papers I–IV, and discuss the findings in relation to other studies within the field, it becomes important to describe the educational context in which they were conducted. Such descriptions are considered to strengthen the external validity of qualitative studies (Cohen et al., 2011).

The educational context can be described at many levels, from national to local. In this section, the national educational context in Sweden is described first and, thereafter, the local context is accounted for by means of a description of the PD programs BfM and CoV.

4.1 The Swedish educational context

In this section the Swedish educational context is described. The emphasis is on aspects that will later be of relevance in the interpretations of the findings from the papers, and aspects closely connected to the focus of the thesis (i.e. teacher PD programs).

4.1.1 The national curriculum and recent curriculum reforms

The Swedish education system is highly decentralized, and schools are either private or run by municipalities. However, all schools are required to follow a national curriculum, which states the content and goals of the education. Central authorities and various institutions ensure that the educational activities are implemented in line with the steering documents, for example by means of analysis of compulsory national exams as well as school visits.

Regarding the national curriculum, this has been rewritten several times in recent decades. The last two curriculum reforms for compulsory school in Sweden took place in 1994 and 2011, when new national standards were employed. The mathematics parts of the curriculum documents launched in 1994 by the National Agency for Education (2006), *Skolverket* in Swedish, were heavily influenced by the NCTM standards (cf. Boesen et al., 2014). The motive behind the reforms was to break with the dominating traditional approach to mathematics teaching that was mainly focused on procedural knowledge. Instead, the aim was to communicate a richer view of what doing mathematics means by also emphasizing broader competency goals related to mathematical reasoning, communication and problem-solving. Since 1994, however, there has been a shift in policy discussion from an emphasis on goals and teacher autonomy to one on results and control (Lundahl & Waldow, 2009; Morawski, 2010). For example, regarding the mathematics parts of the 1994 curriculum, several teachers considered them to be vague, and many had limited or non-existent knowledge of them (Boesen et al., 2014). Therefore, in the latest curriculum documents (Skolverket, 2011a), attempts were not made to change the mathematical orientation but rather to increase the degree of concreteness (Skolverket, 2011c). In response to this objective, the curriculum documents (Skolverket, 2011a) aim to more clearly draw attention to the specific competencies the mathematics teaching should give students opportunities to develop. These five competencies are expected to permeate all mathematical content, and are related to: (1) the formulation and solving of mathematical problems, (2) the use and analysis of mathematical concepts, (3) the selection of mathematics methods and procedures to solve routine tasks, (4) the application of mathematical reasoning, and (5) the use of mathematical forms of expression to communicate about calculations and conclusions. However, even if the goal of the latest curriculum reform was to concretize the mathematical competencies, a recent text analysis by Prytz (2015) suggests that the competency goals have not been clearly linked to the mathematical content to be addressed in instruction.

4.1.2 Mathematics classroom instruction

The fact that the latest curriculum documents (Skolverket, 2006, 2011a) place great emphasis on several mathematical competencies, such as problem-solving and reasoning, does not seem to be reflected in actual classroom teaching. This is visible in a quality review of the mathematics teaching in Swedish elementary schools, which was conducted in

cooperation between researchers and the Swedish Schools Inspectorate (*Skolinspektionen* in Swedish) (Bergqvist et al., 2009; Boesen et al., 2014). The study included interviews, questionnaires and classroom observations with nearly 200 Swedish elementary grade teachers, and showed that in a typical mathematics lesson, over 60% of the time was dedicated to the students solving procedural mathematical tasks by themselves or in small groups. This is supported by an even more recent examination by the Schools Inspectorate (Skolinspektionen, 2017). Their report on working methods used in Swedish classrooms demonstrates that, compared to other subjects, the mathematics instruction consists of more individual work, as well as less whole-class instruction and students working together in small groups.

There are, however, differences to be found in the mathematics instruction between different grade levels. In a more detailed analysis of the data material presented, Bergqvist et al. (2009) showed that students in Grades 4–9 usually spent almost 90% of the mathematics lessons on activities dedicated to procedural competency. For Grades 1–3, the corresponding figure was around 50%. Additionally, compared with the students at the upper levels, the primary-level students were given more opportunities to develop additional mathematical competencies besides procedural fluency. For example, 44% of all analyzed classroom situations in Grades 1–3 involved student communication, while the same figure for Grades 7–9 was 28%.

4.1.3 Mathematics teachers' education and professional development

As the Swedish teacher preparation programs have undergone a large number of reforms in recent decades, it is hard to describe a typical Swedish elementary school teacher's education. What can be established is that secondary-level (Grades 7–9) teachers are usually specialized in two to three subjects, while primary-level (Grades 1–6) teachers are expected to teach most of the subjects covered in the curriculum. This is similar to many other countries in the world, as most primary-school teachers are educated as generalists (Tatto, Lerman, & Novotná, 2009). Currently, to become a teacher in mathematics in Sweden for the primary levels, one semester of university studies in mathematics is required, while the corresponding time for the secondary level is generally three semesters (Axelsson, 2016).

After their pre-service teacher education, practicing teachers also have the opportunity to participate in different types of PD initiatives. Though

it is not legally required, teachers in Sweden are entitled to 13 days of PD each academic year (OECD, 2015). The design and implementation of PD for teachers in Sweden have looked different over the years, and some overviews and research studies of the PD for mathematics teachers have been published. One example is Emanuelsson (2001), who described and analyzed national PD initiatives in mathematics education during the period 1965–2000 that were initiated because of reasons such as new national curriculum documents, a decline in students' mathematics performance, and the introduction of new tools (e.g., electronic calculators). Another example is Kirsten and Wermke (2017) who, studying teacher PD initiatives in mathematics and standard language education during the period 1991–2016 in Sweden, demonstrated a tendency towards more centrally governed and evaluation-justified PD initiatives. This tendency is also visible in the previously described shifts in policy discussion from an emphasis on goals and teacher autonomy to one on results and control (Lundahl & Waldow, 2009; Morawski, 2010). Below, the three most recent PD initiatives for mathematics teachers in Swedish elementary schools are described in order to provide the reader with a contextual background relevant to the focus of the thesis.

The first initiative, the *local developer program*, was implemented in 2006–2008. Within this initiative, municipalities and other principal organizers were encouraged to designate teachers as local developers of mathematics teaching. All developers were supported by the National Agency for Education in terms of offers of PD activities such as conferences, university courses and regional center activities. An evaluation of the program (cf. Tengstrand, 2010) did, however, show large inequalities between schools and municipalities in how the program was implemented. Possible reasons for these inequalities were that the role of the local developers was not described in detail, and the municipalities did not have to establish specific preconditions in order for developers to participate in the initiatives. The municipalities were instead encouraged to discuss the authority of the developer together with the teachers. In other words, the municipalities and teachers were given a great deal of freedom in choosing the form and content of the mathematics development in the respective municipalities and schools. Some municipalities even chose not to take part in the program.

The freedom of municipalities and schools themselves to choose the content and form of the mathematics development is also apparent in the second PD initiative, *the mathematics initiative*. Within this initiative, municipalities and independent schools had the opportunity to apply for financial support to strengthen their local development efforts to improve

mathematics instruction. In total, the government granted schools 42 million EUR during the period 2009–2011. As for the local developer program, some municipalities and schools chose not to seek financial support for participating in the initiative. The themes for the development projects for which it was most common to seek financial support were laboratory material, information and communication technology, lesson studies, and assessment. Although the participating teachers largely experienced increased competence and altered teaching practices during the course of the projects, a national evaluation (Skolverket, 2012) showed that, in many cases, the changes were not permanent and that after the projects the teaching reverted to how it had been before. In the evaluation, possible explanations for the results are highlighted. One example is that in many cases the projects did not strive to develop actual school systems; instead, the focus was on developing individual teachers' skills or purchasing teaching materials.

The above two initiatives have contributed positive aspects to the development of mathematics instruction, but have also been criticized for not reaching out to all teachers, being too vague, and not being sustained (Skolverket, 2012; Tengstrand, 2010). In response to this criticism, in recent efforts agency engagement has become more prescriptive and elaborate (cf. Kirsten & Wermke, 2017). An example of this is the most recent PD initiative for mathematics teachers in Sweden, *Boost for Mathematics*. This initiative is further described in the next section.

4.2 Two professional development programs

Two similar large-scale PD programs for mathematics teachers, BfM and CoV, were used as cases to achieve the aims of the thesis. In this section I start by giving a brief overview of the two programs, which is then followed by a more detailed description of their action models concerning teacher change. This description is based on the critical features of high-quality PD, elaborated on in Section 3.

4.2.1 Boost for Mathematics

BfM is the most recent and extensive Swedish national improvement effort in mathematics, with a budget of 75 million EUR spread over four years (2012–2016). The project is run by the Swedish National Agency for Education, who collaborated with the National Center for Mathematics Education and several universities in developing the content and an

implementation plan. As formulated by the National Agency for Education (Skolverket, 2011b), the overarching goal of BfM was to improve student achievement in mathematics by strengthening the mathematics teaching, and by developing the mathematics teaching culture and the PD culture at the schools. In order to achieve this goal, the initiative included PD programs for teachers and principals as well as mathematics coaches. The project has been widely disseminated, and a recent evaluation (Skolverket, 2016a) shows that 76% of all elementary school teachers in Sweden teaching mathematics have participated in the year-long teacher PD program.

The teacher PD program was built around a digital platform containing PD materials organized into modules. The modules were structured around the core mathematical content in the national curriculum (Skolverket, 2011a) and grade levels (Grades 1–3, 4–6 and 7–9), for example *Understanding and Use of Numbers in Grades 1–3* or *Geometry in Grades 7–9*. In relation to the mathematical content, each module touched upon at least four different aspects: (1) Teaching for the mathematical competencies in the national curriculum (see Section 4.1.1), (2) Formative assessment and teaching in mathematics, (3) Routines and interactions in the classroom, and (4) Classroom norms. Teachers completed one module each semester. In the municipality of Västerås, where three of the studies in this thesis were conducted, the module *Understanding and Use of Numbers* was mandatory for all public elementary school teachers, while the choice of the second module was left to the respective schools. All modules followed a workflow consisting of eight rounds, starting with an individual preparation whereby teachers read texts and watched movies provided on the digital platform. This was followed by a collegial meeting where the teachers discussed the readings and planned for an activity (e.g., a lesson) with the support of a trained mathematics coach. Thereafter, the teachers carried out the activity, and finally reflected upon the round together with their colleagues and a mathematics coach.

At the time this text was written, several national evaluations of BfM were available (Ramböll, 2014, 2016; Österholm, Bergqvist, Liljekvist, & Van Bommel, 2016), and a number of research studies had recently been published (Boesen, Helenius, & Johansson, 2015; Hajer & Norén, 2017) or were awaiting confirmation of publication (e.g., Kaufmann, 2017; Van Steenbrugge, Larsson, Insulander, & Ryve, 2017). These evaluations and studies have, among other things, contributed information on the program's impacts on teachers and instruction, as well as possible reasons for these impacts. None of them, however, have focused on variation in the program's impacts between teachers from different grade levels or,

perhaps most noteworthy, the program's impacts on student achievement, which has been referred to as the bottom line of PD measurements (Guskey, 2002a; 2014a). This thesis contributes important complementary data to the already conducted evaluations and studies, as it comprises data on the program's impacts on student achievement, and emphasizes differences between teachers teaching mathematics at different grade levels.

4.2.2 Count on Västerås

CoV is a combined research and development project in cooperation between Mälardalen University and Västerås, a large municipality in Sweden. In designing the project, researchers from the university cooperated with teachers, principals, politicians and other central actors within the municipality. In total, the researchers involved in the project worked together with about 10,000 pupils, 450 teachers, heads of mathematics and principals at 40 elementary schools, mathematics developers, and politicians. The overarching aim of the project was to establish an effective mathematics education within the municipality that would offer students the opportunity to develop *all* the mathematical competencies mentioned in the mathematics parts of the national curriculum (Skolverket, 2011a). In order to achieve this aim, the project focused on multiple aspects at different levels of the educational system, from the district to teachers. For example, in efforts to institutionalize the reorganization of practices beyond the funding of the project, work has been done to establish both new positions (e.g., heads of mathematics at each school) and new routines (e.g., procedures for collecting, compiling and analyzing student results on mathematical tests; cf. Ryve, Hemmi, & Kornhall, 2016). In this thesis, the focus is on the year-long teacher PD program within the project.

More than 90% of the public elementary schools in Västerås have participated in the CoV teacher PD program during the period 2012–2017. The content of the program was directed at teaching for the five mathematical competencies set out in the national curriculum (see Section 4.1.1). In order to advance towards this focus, two main tracks within program were Formative assessment (cf. Wiliam, 2007, 2011) and Teaching mathematics through problem-solving (cf. Larsson, 2015; Smith & Stein, 2011). Moreover, as a way to ground the PD program in the teachers' real-world practice, special attention was also directed at the curriculum materials (teacher guides and student textbooks) teachers were already using in their instruction. All teachers, regardless of grade level,

followed the same program, which contained 19 sessions spread evenly throughout the school year. Most PD sessions began with teachers reflecting on a conducted activity, for example a lesson or an analysis of student results on a mathematical test. Thereafter followed a presentation and discussion of new content, and teachers planned for a new activity to conduct. A majority of the sessions were led by one of three doctoral students, including the author of the thesis, who guided the discussions and presented the new content. The doctoral students all had teacher degrees, and were also working as mathematics mentors within the municipality. The remaining sessions were led by the heads of mathematics at the respective schools. The heads of mathematics had all had PD of their own, and were further supported by PD materials provided in the project.

4.2.3 The professional development programs' action models

PD programs can be described in a number of different ways. In this thesis, the main focus is on studying action models concerning PD programs' theories of teacher change or, in other words, critical features of high-quality teacher PD programs. BfM and CoV are therefore described according to the critical features of high-quality PD, elaborated on in Section 3.

As illustrated in Paper I, BfM and CoV largely demonstrate the same characteristics when described according to the five critical features of high-quality PD, presented in Desimone (2009). A summary of the description is shown in Table 1.

Table 1. A description of the professional development programs based on critical features of high-quality professional development

Critical feature	**Operationalization within the PD program**	
	BfM	*CoV*
Content Focus	Pedagogical content knowledge in mathematics • Teaching for the mathematical competencies • Formative assessment • Classroom norms • Routines and interactions in the classroom	Pedagogical content knowledge in mathematics • Teaching for the mathematical competencies • Formative assessment • Teaching mathematics through problem-solving • Curriculum materials
Coherence	Emphasize mathematics teaching in line with the mathematics parts of the national curriculum	Emphasize mathematics teaching in line with the mathematics parts of the national curriculum
Active Learning	Teachers plan, conduct and discuss mathematics teaching practices	Teachers plan, conduct and discuss mathematics teaching practices
Duration	One year with one two-hour meeting every week	One year with one two-hour meeting every other week
Collective Participation	All teachers teaching mathematics are expected to participate	All teachers teaching mathematics are expected to participate

As illustrated in Table 1, both BfM and CoV include all the critical features of high-quality PD described in Desimone (2009). They focus on teachers' PCK in mathematics (*Content Focus*); in other words, both the subject matter and how to teach it. While the PD programs' *Coherence* with teachers' knowledge and beliefs cannot be established, they are coherent with school and state policies as attention is directed at teaching for the five mathematical competencies set out in the national curriculum (Skolverket, 2011a). Moreover, teachers are engaged in *Active Learning* tasks, such as planning for mathematics lessons or discussing video clips from classrooms. The programs also include *Duration* in terms of both intensity (teachers meet for two hours every, or every other, week), and endurance (the meetings are spread over one year). Finally, the programs are based on a design that emphasizes *Collective Participation* and collegial cooperation between teachers at each participating school. That is to say, all teachers teaching mathematics at the respective school are expected to take part in the PD programs; and during the sessions, which take place at the individual schools, time is devoted to joint discussion and instruction planning.

Furthermore, when described according to Kennedy's (2016) four methods of facilitating teachers' enactment of the ideas presented in the

PD, BfM and CoV are also very much alike. Both include a mix of either presenting teachers with *strategies* and a rationale for why to use them (e.g., a set of tasks to give to students, or a set of questions for teachers to ask students during problem-solving lessons), or relying on teachers' *insights* and that they themselves come up with the instructional design, with some support (e.g., a set of questions to discuss in regard to a text or the design of a lesson).

5 Methodology

This chapter consists of three subsections. In the first, the different methods used for data collection and analysis are accounted for in relation to the four components involved in teacher change models: PD, teachers, instruction, and student outcomes. The second subsection covers the ethical considerations made and, finally, the third contains a discussion of the studies' validity and reliability.

5.1 Data collection and analysis

The methods for collecting and analyzing data in the thesis can all be connected to one or several of the four components of teacher change models in the context of PD programs: (1) characteristics of the PD program itself, and its impacts on (2) teachers, (3) instruction, and (4) students (see figure 2 in section 2.1). A brief summary of the data collected for the papers and in connection to the four components is provided in Table 2.

Table 2. Data collected for the thesis

Component of change model	*Used to collect data on*	*Method of data collection*	*Time of data collection*	*Sample*	*Data used in Paper*
PD program	The PD literature's definitions of the critical feature of Coherence	Systematic literature review	Search conducted in June 2016 and papers analyzed from July 2016 to April 2017	Papers (n=95) of PD for in-service K–12 teachers that define or describe the critical feature of Coherence	V
	BfM's and CoV's Content Focus	Analysis of PD materials	Materials analyzed from December 2016 to March 2017	All PD materials provided to the teachers participating in BfM or CoV	II[a]
Teachers and/or instruction	CoV's impact on teachers' reported instruction as well as difficulties and support in carrying out the changes	Specially designed teacher pre- and post-questionnaires for CoV	The first week and one of the last weeks of the PD program in CoV 2013/2014 and 2014/2015	All teachers participating in the PD program in CoV 2013/2014 and 2014/2015 (n=135)	III
	BfM's impact on Teacher Confidence Teacher Collaboration and Support from School Leadership	TIMSS 2015 teacher questionnaires	Teachers completed the questionnaire in March or April 2015	Teachers teaching mathematics in Sweden in Grade 4 or 8 who did not participate in BfM during the spring term 2015 (n=284)	IV
Student achievement	BfM's and CoV's impact on student achievement	Test on number sense	At the beginning of the autumn term 2013, 2014, and 2015	All students studying at a public elementary school within the municipality who took the test (circa 8,000 students/year)	I and II

[a] The characterization of BfM and CoV in Paper I did not include a specific data collection procedure, as in Paper II. Paper I is therefore not included in this row of the table.

Below, I describe in more detail the methods for collecting and analyzing data related to the four components of the change model.

5.1.1 Professional development programs and their characteristics

The first component of the linear change model of teacher PD programs (see Section 2.1) consists of the characteristics of the PD program itself. Thus, in measuring and studying educational improvement initiatives, it is not enough to measure simply the initiatives' impacts; data on the characteristics of the initiatives themselves (e.g., aims, content, and methods of delivery) must also be gathered and analyzed (Bryk et al., 2015; McKenney & Reeves, 2012). This thesis makes use of two different ways of collecting and analyzing data related to the PD program component, either systematically reviewing the PD literature on a specific critical feature of teacher PD (Paper V) or comparing and analyzing PD materials from two cases of large-scale teacher PD programs (Paper II).

5.1.1.1 Systematic literature review

Paper V is not specifically centered on the PD programs CoV and BfM. It takes a broader perspective and, through a systematic review (Gough, Oliver, & Thomas, 2013; Petticrew & Roberts, 2006) of the literature on teacher PD, focuses on how a specific critical feature of high-quality PD (Coherence) is defined and described in the research literature. The systematic review was conducted in four steps. The first step consisted of a literature search using four databases (ERIC, PsycInfo, Scopus, and Web of Science) and different combinations of the terms *Professional Development*, *Teacher* and *Coherence*, or synonyms. In the second step the literature selection, papers that concerned PD for in-service K–12 teachers and that in some way defined or described coherence (or synonyms) in relation to teacher PD were sorted out. Moreover, we examined the reference lists of the already extracted papers to ensure that we had not overlooked any important references in our initial literature search. The first two steps of the review resulted in a total of 95 papers. In the third step we extracted everything in the papers in relation to Coherence and, in the fourth step, we analyzed these parts of the literature.

The analysis was performed in an iterative and cyclic process in which the author of the thesis did the initial coding, discussed the results, and revised the coding scheme together with the co-author, and thereafter recoded the material. Though it is argued that Coherence is not clearly defined in the literature on teacher PD (Firestone & Mangin, 2014), it

seems to be an established concept in the research field of pre-service teacher education (e.g., Hammerness, 2006; Tatto, 1996), educational policy (e.g., Fuhrman, 1993; Honig & Hatch, 2004), and curricula (e.g., Muller, 2009; Schmidt, Wang, & McKnight, 2005). The coding scheme used for analyzing the papers therefore drew upon this literature, especially in regard to the discussion about coherence with external or internal factors, as well as Honig and Hatch's (2004) notion of coherence as objective alignment and coherence as craft. This resulted in a coding scheme with three main categories, with papers stating that PD programs should: (1) *be coherent* with something (e.g., standards, curriculum materials); (2) *be internally coherent*, for example that the activities within PD programs should be aligned; or (3) *create coherence*, which included papers treating coherence as something PD programs can or should create, rather than something they should possess. For each of the categories, we also summarized the papers' descriptions of coherence in relation to two questions: (a) *the "what" of coherence*, which included descriptions of what there should be coherence with and about; and (b) *how to achieve coherence*, for example which content to focus on or how to facilitate enactment. Note that since the review mainly took a configurative approach (cf. Gough et al., 2013), a single paper could be coded as belonging to several categories. The main goal was therefore more theory-building than theory-testing (Eisenhardt & Graebner, 2007), aiming to map how coherence can be defined. We did, however, also determine how many times the different definitions of coherence appeared as a way of revealing how prominent they are in the analyzed research literature.

5.1.1.2 Analyzing professional development materials

The PD program component of the BfM and CoV change models are explicitly addressed in Paper II. This is done by analyzing all the PD materials provided to the teachers participating in the programs. The resources for BfM consisted of the website materials[3] related to the module *Understanding and Use of Numbers* for Grades 1–3, 4–6 and 7–9, as this module was mandatory for the teachers in Västerås. The resources for CoV consisted of the materials presented to the teachers during the PD sessions, for example PowerPoint presentations showed at the PD sessions, batteries of questions for lesson plans, and task descriptions.

The analysis of the PD materials was performed by the author of the thesis and a co-author who, in several cycles, (a) coded a tenth of the materials together; (b) divided the materials between themselves and coded

[3] This material can be found at the following website: https://larportalen.skolverket.se/#/moduler/1-matematik/alla/alla

them separately, including a couple of materials that both coded for control; (c) met and discussed their coding to achieve agreement and to revise the framework, for example the inclusion of a new subcategory; and (d) went back to step (a) and recoded the materials according to the revised framework. After several of these cycles, the first and second authors met for a final meeting at which they went through more than half of the texts together to ensure that their coding had been performed equivalently. Two separate analyses of the PD materials were conducted, one focusing on the content and one on the methods for facilitating teachers' enactment of the promoted practices.

For the first analysis, a framework for coding the PD materials' content was developed by drawing on literature concerning what constitutes important knowledge for (mathematics) teachers (Ball et al., 2008; Cook & Brown, 1999). The final framework consisted of two main categories (referred to with uppercase Roman numerals), which in turn included subcategories (referred to with lowercase Roman numerals). The first main category, (I) *Teacher Knowledge*, included materials focusing on knowledge as something that teachers possess. These materials emphasized either (i) teachers' purely mathematical knowledge, and/or knowledge of students' mathematical thinking and learning, or (ii) teachers' knowledge of didactical phenomena. Two examples are materials emphasizing mathematical tasks for teachers to solve, or materials describing what is meant by formative assessment. The second main category, (II) *Teacher Practices*, included materials focusing on knowledge as something that is a part of action and emphasizing teachers' performance of the instructional practices. This could be done in three different ways: by (i) focusing on specific actions teachers do during instruction, such as questioning students with a specific purpose; (ii) emphasizing structural aspects of the lesson design, such as grouping of students or tasks to hand out; or (iii) taking a meta-perspective by supporting teachers' reflection on and planning of their own learning and development.

For the second analysis, Kennedy's (2016) four categories of PD programs' methods for facilitating teachers' enactment of the presented ideas (see Section 3.2.5) supported the coding. If the materials offered teachers concrete descriptions of instructional tasks to conduct with no explanation of why or their educational relevance, they were coded into the category *Prescription*. If, on the other hand, the materials contained several prescriptive elements accompanied by a rationale to support teachers in why and when the practices should be implemented, they were coded into the category *Strategy*. The third category, *Insight*, refers to teacher self-

generated "aha moments" that can be stimulated by PD providers raising provocative questions that force the teachers to reflect upon familiar events differently. Materials were coded into this category if they described an educational phenomenon and were accompanied by guidelines for discussing how they could affect teachers' instructional practice, but without explicit descriptions of lesson activities or teacher actions. Finally, materials were coded into the category *Body of Knowledge* if they gave an exhaustive description of a mathematical or educational phenomenon, but lacked any other support for why or how teachers should apply them to their instructional activities. A preliminary analysis of the PD materials revealed differences between the two programs' methods for facilitating enactment in relation to whether the coding focused on (i) teacher actions or (ii) lesson design. All materials were therefore coded twice with respect to the Kennedy-derived categories, once in regard to (i) teacher actions and once in regard to (ii) lesson design.

5.1.2 Professional development programs' impacts on teachers and instructional practices

Data on the teacher and instruction components of BfM's (Paper IV) and CoV's (Paper III) change models were addressed with the support of teacher questionnaires.

For CoV, pre- and post-questionnaires specifically designed for the project were used to collect data from teachers participating in the PD program in 2013/2014 or 2014/2015. Both questionnaires contained items related to teachers' views on, for example, curriculum materials, collegial support, and their classroom instruction. In addition, the post-questionnaire also included items regarding teachers' perceptions of the PD program. The questionnaire items were piloted by members of the CoV project group and/or retrieved from other teacher questionnaires used for studying large-scale reforms of instructional practice (Hamilton et al., 2003; Malzahn, 2002; Middle-school Mathematics and the Institutional Setting of Teaching, n.d.). For the study reported on in Paper III, the focus was primarily on items regarding teachers' reported instruction and they were asked to estimate, on a four-point Likert scale, how often they conducted certain activities in their mathematics classrooms. Paired sample t-tests were employed to assess whether the PD program had had any impact on teachers' reported instruction, with the categorical variable represented by the time at which the teachers completed the questionnaire (pre- or post-) and the continuous variable represented by the teachers' reported instruction. These analyses were run separately for the primary-and the

secondary-level teachers. Additionally, some items regarding the teachers' views of the PD program and its implementation were also analyzed. For example, on a four-point scale, teachers were asked to state whether they had experienced any difficulties (e.g., inadequate classroom resources) in trying to make changes to their mathematics instruction based on their experiences from the PD. To compare the primary- and secondary-level teachers' (i.e. two different groups of teachers) perceptions of the program, independent sample t-tests were employed.

For analyzing the impact of BfM on teachers and instruction, the Swedish data from the Trends In Mathematical and Science Study (TIMSS) 2015 fourth- and eighth-grade mathematics teacher questionnaires were used. These questionnaires ask teachers about a range of topics. In Paper IV, items connected to three constructs were analyzed: *Teacher Confidence* in carrying out mathematics instruction, *Teacher Collaboration* with colleagues, and teacher-reported *Support from School Leadership*. Each construct included four to nine Likert-scale items and their Cronbach alpha values were all above .86. To be able to determine the impacts of BfM as related to the constructs, two groups of teachers, participants in BfM and non-participants, were created for the fourth- and eighth-grade teachers, respectively. These groups were comparable in regard to both age and years of teaching experience. Independent sample t-tests were employed to assess whether there were any statistically significant differences between participants and non-participants in respect to the three constructs. The statistical analyses were conducted for both the fourth- and eighth-grade samples.

Regarding the statistical methods for analyzing data from the questionnaires used in Papers III and IV, one could argue that the Likert scales should be seen as ordinal instead of continuous, and that the non-parametric alternatives (Wilcoxon Signed Rank Test and Mann-Whitney U Test) should therefore have been used instead. For this reason, I either examined whether the data sets were normally distributed and thereafter ran the non-parametric tests for data that did not meet the assumption (Paper IV), or ran the non-parametric alternatives for all analyses (Paper III). These additional analyses showed the same results as the t-tests. This should also have been expected, as it has been shown that parametric tests are particularly robust and can usually be used with data from Likert scales without concern, even if the data are not normally distributed (De Winter & Dodou, 2010; Norman, 2010).

Lastly, a statistical significance does not necessarily imply an educational significance. In contrast to a simple significance test, the effect size tells us something about the size of the differences found (Cohen,

Manion, & Morrison, 2011). Therefore, to also assess the strength of the differences as well as facilitate comparison between groups (e.g., different programs or groups of teachers), it is recommended that studies on PD also report on the effect sizes (Blank & de las Alas, 2009; Clewell et al., 2005; Timperley et al., 2007). For the data obtained from the teacher questionnaires, the effect sizes were calculated with Cohen's d. For interpreting the results, the guidelines proposed by Cohen (1988) were used with .2 representing small effect, .5 representing moderate effect, and .8 representing large effect.

5.1.3 Professional development programs' impacts on student achievement

The last component of the linear change model, PD programs' impacts on students' learning outcomes, has been referred to as the bottom line of PD measurements (Guskey, 2002a; 2014a). In large-scale PD projects, it is recommended that such data be collected both during (or immediately after the end of) the program, as well as several years after its implementation (e.g., Avalos, 2011; Bryk et al., 2015; Desimone & Stuckey, 2014; Kennedy, 2016; Sowder, 2007; Wayne et al., 2008). For example, it is mostly preferred that changes to a program are made during the course of the program instead of after its completion. Immediate data are therefore needed in order to predict where the program is heading, to point to future directions, and to consider the demands of financial supporters, such as politicians and governments, who often require fast responses. However, changes taking place among teachers, instruction and students is not an event but rather a process, and thus collecting results too close to the time of implementation may be misleading in assessing a PD program's ultimate effects. Data collected several years after a program's implementation are therefore needed to assess its sustainability and long-term impacts. In this thesis, data on the impact of BfM and CoV on student mathematics achievement were collected both immediately (Paper I) and one year after (Paper II) the teachers' participation in one of the programs.

In order to measure the programs' potential impacts on students' mathematics achievement, a collection of mathematical tests (McIntosh, 2008) assessing students' number sense were used. Each grade (1–9) has its own test and, naturally, the items are not exactly the same since the degree of difficulty increases with the age of the students. Nevertheless, there are a number of anchor items that reoccur over several grades. Examples of test items include a multiple-choice item in which students are asked to estimate how many days they have lived, as well as an item in

which students are asked to give examples of and specify the amount of numbers between two decimals (e.g., 3.9 and 4.0) or two fractions (e.g., 2/5 and 3/5). For further examples of items, see McIntosh, Reys, Reys, Bana, and Farrell (1997) or McIntosh (2008). There were several reasons the McIntosh (2008) tests were chosen. Firstly, even before the PD programs' implementation, a procedure was in place that tests are administered annually to all elementary school students in the municipality at the beginning of each semester with established procedures for the administration and reporting of the results. After the students have taken the tests and their teachers have corrected them, the head of mathematics at each school sends in a report protocol regarding their results. Thus, data on students' results on each item as well as which grade and school they attend are collected each year. Without such established structures, it is hard to collect the type and amount of data needed to investigate large-scale PD programs' impact on student achievement. Secondly, the tests could serve as support for teachers in planning for the following instruction, since they are accompanied by a comprehensive teacher guide offering information on possible reasons for the difficulties detected among students, as well as suggestions for instructional activities. Thirdly, since the tests have been developed through several combinations of research and development projects, as well as through experiences of working together with teachers and students from Australia, Sweden, Taiwan and the United States (cf. McIntosh et al., 1997; Reys et al., 1999), they have been thoroughly tested. Fourthly, in collaboration with the Swedish National Center for Mathematics Education as well as researchers, teachers and students, the tests have been adjusted to the Swedish national context (cf. Emanuelsson, Johansson, Reys, & Reys, 1996). Fifthly, the tests and the accompanying teacher materials were included in both PD programs' content, either as readings on number sense (BfM) or as tools for formative assessment (CoV).

This thesis make use of students' results on the tests collected in 2013–2014 (Paper I) and 2013–2015 (Paper II). The statistical analysis in Paper I was conducted by the author of the thesis, and the statistical analysis in Paper II was conducted by the author of the thesis in cooperation with the third author of the paper. For both papers, the statistical analyses began by dividing the students into three groups depending on whether their teachers had participated in BfM, CoV, or no intervention. Note that students whose teachers had participated in both programs during the time frame when the studies were conducted were excluded from the sample in order to be able to compare the individual programs' impacts. Thereafter, some descriptive statistics, such as averages and standard deviations for students' results,

were produced for each grade, group, and year. Then more advanced methods of analysis were conducted, such as one-way ANOVAs and independent sample t-tests (Paper I), or two-way ANOVAs (Paper II). The ANOVAs were performed separately for each grade, and for all of them we sought to determine whether there was a significant interaction effect between the two categorical variables (Year and Intervention) on the continuous variable (Students' Test Results). The results for all grades, groups, and years were presented to get a more comprehensive view of the different projects' effects. The results that were emphasized, however, were those in which a significant interaction effect was found.

Finally, as in the analysis of the teacher questionnaires, the effect sizes were calculated in order to indicate the respective size of the PD programs' impacts, as well as the differences between them. In Paper I the effect sizes were calculated in terms of Cohen's d, and the guidelines proposed by Cohen (1988) – with .2 representing a small effect, .5 representing a moderate effect, and .8 representing a large effect – were used to interpret the results. In Paper II the effect sizes were calculated in terms of partial eta squared (η^2), and to interpret the results the guidelines proposed by Cohen (1969) – rounded so that .001 represents small effect, .06 represents moderate effect and .14 represents large effect – were used. It is important to note that when it comes to the impacts of PD programs on student results, the benchmarks for estimating the strength of effect sizes are debated. For example, Timperley et al. (2007) state that studies including large groups of students and that use control/comparison groups (like the studies included in this thesis) typically produce smaller effect sizes. Therefore, instead of merely relying on standardized benchmarks, one should also relate the results to studies conducted in similar contexts (Sink & Stroh, 2006). For example, in light of the results of her review of 28 PD programs' impacts on student achievement, Kennedy (2016) suggests that in studies of PD programs, effect sizes around .2 could be considered rather large. In addition, Hill, Bloom, Black, and Lipsey (2008) have shown that when studying the impacts of interventions in elementary schools using broad standardized tests, as is the case in Papers I and II, the effect sizes are most often very small.

5.2 Ethical considerations

The studies included in this thesis were guided by the legislation, ethical requirements and ethical recommendations discussed in a report by the Swedish Research Council (2011). According to this report, ethics within

the research community involve two aspects. The first, *research ethics*, mainly concerns ethical questions that apply to those participating in research as subjects or informants. The second aspect, *professional ethics*, mainly concerns the craft itself and the researcher's responsibility towards both research and the research community. Consequently, professional ethics includes, among other things, scientific quality such as the validity and reliability of the research, which is further elaborated on in Section 5.3. As Paper V is a systematic literature review, it does not include informants in the usual sense. Therefore, there are not many issues to discuss concerning research ethics in relation to the paper. What can be said is that responsibility towards publishers and the authors of the papers has been demonstrated by striving only to examine, and not evaluate or judge, the papers.

In this section of the thesis, the ethical considerations made in the included studies will be explained and discussed in relation to Stutchbury and Fox's (2009) tool for ethical analysis in educational research. This tool includes four different levels to consider concerning ethical questions: *External, Consequential, Deontological* and *Relational.* The different levels are interdependent of each other, and specific issues may thus be considered in more than one layer. However, to avoid repetition, I only mention the different issues once and in relation to one level.

5.2.1 External considerations

The external level of Stutchbury and Fox's (2009) tool brings to the fore that the context in which the research takes place may raise several ethical issues and that researchers have to consider their responsibilities and use of resources in relation to (a) the law and (b) society.

At the beginning of the project, the law on ethical review of research involving humans (SFS 2003:460) was taken into account. As the studies included in the thesis do not concern what can be termed sensitive data, it was judged that they are not the type that requires an application for formal ethical review by the central or regional ethical review boards.

As regards the responsibility for and use of resources in relation to society, this concerns issues such as efficient use of resources (e.g., people's time) and responsive communication including an awareness of the wishes of others (Stutchbury & Fox, 2009). A point of friction between research and development projects is "the generality that is valued in academic research and the specificity that is needed to guide practice improvement" (Bryk et al., 2015, p. 99). While academic research values long and detailed measures in an effort to measure each construct with high

reliability, practice improvement demands practical measurements whereby data collection needs to fit into, for example, teachers' often limited time frames and be collected on a regular basis. In order to deal with this friction and minimize additional workload for the teachers and students, the measurements and data used in this thesis were chosen to fit in with and be of support to the teachers' and students' instructional reality. For example, either the teachers were given time to complete the teacher questionnaires during the PD sessions (Paper III) or I made use of appropriate teacher data already collected in other studies, such as TIMSS 2015 (Paper IV). Another example is the choice to, instead of employing new tests specifically designed for the studies, make use of the McIntosh (2008) tests to collect data on student achievement (Papers I and II). These tests were already mandatory for all elementary school students in the municipality, and therefore did not create an additional workload for teachers and students. Moreover, as a researcher it is crucial to consider who will benefit from the research and what those who participate in the studies will get in return (Singer, 2008; Swedish Research Council, 2011). The data collected on the two PD programs were used not only to generate research papers, but also to constantly evaluate and improve the projects' designs and/or their respective implementation in Västerås. For example, each year, the project group for CoV gathered for several days to analyze the collected data in order to make informed decisions about revisions to the PD program and its implementation, so that it would better match the needs of the teachers in the municipality. For instance, the results from Papers I, II and III have been discussed in the project group and have contributed valuable information on important aspects to consider in the planning and adjustment of the project in forthcoming years.

5.2.2 Consequential considerations

Issues concerning the consequences of possible actions for society, groups, and individuals are related to Stutchbury and Fox's (2009) consequential level. This level involves aspects such as striving to conduct and report the research in such a way that it, to the greatest extent possible, avoids harm and is to the most benefit of the participants, society and the researcher. In the studies included in this thesis, multiple actions have been taken to minimize harm and strive to achieve the most benefit for all participating actors.

First of all, and as elaborated on in the section above (5.2.1), (a) no sensitive personal data about the respondents were collected, (b) efforts were made to minimize additional workload for teachers and students in

relation to the collection of data, and (c) the data were used to evaluate and improve the projects' designs and/or implementation in Västerås so that they would better match the needs of the municipality, teachers, and students.

Secondly, considerations were made in relation to the communication of results and how they might be interpreted by different actors. For example, if data show that teachers do not carry out the type of instruction advocated in the PD, it may be easy to jump to the conclusion that this is due to a lack of commitment and ability on the teachers' part. However, conducting mathematical instruction in which students are given the possibility to develop all the mathematical competencies is a major challenge, and does not only depend on the competence of the individual teacher. Previous studies, for example, have shown the importance of also attending to the support and resources teachers receive from curriculum materials, the political context, and their principal (Cobb & Jackson, 2011; Timperley et al., 2007). Throughout this thesis, I have tried to nuance the kinds of challenges faced by teachers in carrying out the teaching practices advocated in the PD programs, and thereby do them justice.

Thirdly, several actions have been taken in order to ensure the confidentiality of the teachers and students participating in the studies in Papers I–IV. To begin with, information that could reveal students' identities was removed before the data on their achievement were transferred from the municipality's database. Moreover, teachers participating in the studies in Paper III were asked not to put their names on the questionnaires. At the same time, however, data on which teacher had completed each questionnaire was needed in order to follow the progression from pre- to post-questionnaire; therefore, code keys were used. There is, however, no possibility to connect the codes with the names of the respective teachers since only the teachers themselves know their codes. The data from the TIMSS 2015 teacher questionnaires were obtained from the National Agency for Education, which had already anonymized the data. Complete anonymity is rare, however, and may not even be achievable in statistical studies (Moore, McCabe, & Craig, 2009; Singer, 2008). Considering the era of advanced technology and all the information stored in different databases, "a clever computer search of several data bases might be able, by combining information, to identify you and learn a great deal about you even if your name and other identification have been removed from the data available for search" (Moore et al., 2009, p. 228). Still, as we kept the written questionnaires in locked cabinets and used codes instead of names, combined with the large number of participants and the fact that the schools are not named in the studies, the

possibility for an unauthorized individual to identify individual teachers or students is regarded as minimal.

5.2.3 Deontological considerations

Deontological considerations are connected to issues regarding one's duty and consideration of possible actions. In contrast to the previously described level, it is concerned with the way things are done rather than the consequences of doing them (Stutchbury & Fox, 2009). Examples of "duties" are keeping promises (e.g., ensuring confidentiality, as described in Section 5.2.2) and informing respondents about the research and their rights. This last issue can be connected to the requirement of informed consent (Swedish Research Council, 2011).

Regarding informed consent and the teacher questionnaires used in Papers III and IV, before the questionnaires were distributed all participating teachers were informed about the purpose of the questionnaires, that the information gathered could be used in research studies, how they might get involved in these studies, and how the results would be utilized and to what end. This information was given at information meetings (Paper III), and/or was summarized on the cover sheets of the questionnaires (Papers III and IV). Before the CoV questionnaires were distributed, the teachers were also informed that completing the questionnaires was optional, and a few of them (less than 1%) chose not to complete them. In light of the information given to the teachers, the response to and handing in of the questionnaires can be seen as provided consent. As for the TIMSS 2015 teacher questionnaires, these are also intended to be optional for teachers to hand in. Then again, based on the documents provided on the website for the Swedish National Agency for Education, it cannot be ascertained whether or how this was communicated to the teachers. Concerning data on student achievement (Papers I and II), these were administrated by the municipality and would have been gathered independent of the research study. Thus, the situation for the teachers and students did not differ from other municipality evaluations and, therefore, consent was not requested from the students in relation to the tests.

5.2.4 Relational considerations

The relational level is concerned with issues such as (a) respecting the autonomy of the respondents (cf. informed consent in Section 5.2.3), (b)

the relationship between respondents and researcher, and (c) confirmation of the findings (Stutchbury & Fox, 2009).

It is important to note, in relation to (b) the relationship between respondents and researcher, that the author of the thesis also worked as one of the three mathematics mentors who guided more than half of the PD sessions within CoV. This role can be seen as not only an advantage (as the researcher has had a broad and deep view of the project) but also a disadvantage (as the researcher's objectivity might be questioned). To reduce any negative aspects, measures have been taken to make the research process as transparent as possible, and thus also open to critique. Throughout this thesis I have strived to clearly describe the research process, from the basic assumptions and theoretical frameworks adopted to how the data collection and analysis were carried out. All the data material from the studies is kept in locked cabinets, making it possible to retrieve it to run the analyses again. In addition, all the studies included in the thesis have undergone, or will undergo, peer-review processes and have been presented and discussed at, for example, research conferences. They have thereby also been made open to critique. Moreover, as mathematical tests and teacher questionnaires were used to gather data on CoV, the author's influence on the respondents was reduced compared to if it had been gathered in interviews (Schwarz, Knäuper, Oyserman, & Stich, 2008). Finally, in the analysis of the PD materials within CoV, the author conducted the analysis together with a co-author who was not involved in the project.

Issues concerning (c) the confirmation of findings are highly related to questions about the reliability and validity of the study (Stutchbury & Fox, 2009). This is further discussed in the next section.

5.3 Validity and reliability

General definitions of the concepts of validity and reliability – for example that a study's validity concerns the issue of whether an indicator actually measures what it is intended to measure, or that reliability refers to the repeatability and consistency of the findings – are very broad. In fact, there are many different types of validity and reliability, and which ones should be paid more careful attention to depends very much on the research tradition and the method in which the researcher is working (Golafshani, 2003; Messick, 1995; Swedish Research Council, 2011). Therefore, in this section, issues related to validity and reliability will be discussed in three parts, depending on the methods used for data collection and analysis: (1)

content analysis and/or systematic review, as in Papers II and V; (2) questionnaires, as in Papers III and IV; and (3) tests, as in Papers I and II.

5.3.1 Validity and reliability in systematic reviews and content analysis

Two of the papers in the thesis include qualitative content analyses, in terms of either an analysis of the content of PD materials (Paper II) or an analysis of the content of research papers (Paper V). Weber (1990) describes three types of notions mainly associated with reliability in such analyses: stability, reproducibility, and accuracy. *Stability* can be determined when the same coder codes the content more than once. As the coding procedures in both Papers II and V were conducted in multiple cycles whereby the coders recoded all, or selected parts of, the materials in each cycle, the stability was enhanced. *Reproducibility*, sometimes also referred to as inter-coder reliability, is regarded as a stronger form of reliability as it concerns the consistency of the results when the same content is coded by more than one coder. In Paper II, the author of the thesis conducted the coding together with a co-author to ensure reproducibility. More than half of the PD materials were either jointly coded by the authors, or first coded individually by each author and thereafter compared and discussed until agreement was reached. In Paper V, on the other hand, the author of the thesis conducted all the coding. Reproducibility was reached in certain terms, however, as the results of the coding were discussed with the second author between every coding cycle. Finally, *accuracy* concerns the extent to which the coding of certain content corresponds to a standard or norm. As (to the authors knowledge) there exist no standards for coding PD materials' Content Focus, or coding research papers in relation to critical features of teacher PD, the conditions for achieving accuracy could not be met in Papers II and V.

Validity in relation to qualitative research (Cohen et al., 2011), and especially content analysis (Weber, 1990), can be clarified using two distinctions: (1) *internal validity*, which refers to the correspondence between two sets of things, such as concepts and data; and (2) *external validity*, which refers to the generalizability of the results. The internal validity of the studies is strengthened by the fact that the coding schemes used in Papers II and V are grounded in established definitions and frameworks. Additionally, the process of coding and re-coding, as already discussed in relation to the studies' reliability, is also argued to strengthen the internal validity of the studies. Finally, in relation to systematic reviews, studies' internal validity can be strengthened by taking measures

to ensure that all relevant studies are included in the literature search (Petticrew & Roberts, 2006). Examples of such measures taken in Paper V include the choice to search for relevant literature in *several* databases, and to also examine the reference lists of the relevant papers found in the initial literature search.

As for external validity in qualitative research, this can be interpreted in terms of the comparability and transferability of the findings (Cohen et al., 2011). An aspect that can be considered to strengthen the external validity of the study in Paper II is the inclusion of two PD programs that have been implemented in either a whole country (BfM) or a large municipality (CoV). In addition, in Paper V we have aimed at providing the reader with detailed descriptions of the national contexts in which the PD programs were implemented, and in Papers II and V a thorough description of the analysis processes is provided. Such thick descriptions are considered to strengthen the external validity of qualitative studies (Cohen et al., 2011).

5.3.2 Validity and reliability in teacher questionnaires

Regarding validity and reliability in studies using questionnaires, this can be seen from two perspectives (Cohen et al., 2011). Firstly, whether those who did not respond to the questionnaire would have given the same answers as those who did respond. Secondly, whether those who did respond did so accurately, honestly, and correctly. To deal with these two perspectives in Papers III and IV, they were considered with respect to the types of errors that can occur in relation to the four cornerstones of survey research: *Coverage*, *Sampling*, *Response* and *Measurement* (Hox, de Leeuw, & Dillman, 2008).

Errors regarding *Coverage* and *Sampling* occur in connection with the selection process and the fact that the entire population does not answer the survey (Hox et al., 2008). In Paper III this did not cause any problems, since all the teachers participating in the CoV PD program were asked to answer the questionnaires. Paper IV, on the other hand, made use of the Swedish data from the TIMSS 2015 teacher questionnaire. TIMSS employs a rigorously developed methodology to obtain school and student samples that are representative of the participating countries (Martin, Mullis, & Hooper, 2016), and for Sweden the school coverage was 100% and the number of ineligible school excluded from the sample was less than 3%. It should be noted, however, that even though the selection is representative of the students, it is not necessarily representative of the teachers (Skolverket, 2016b). Generalizations of the results should therefore be made with caution.

Errors in relation to *Response* can occur when a teacher refuses to answer the questionnaire (unit nonresponse), or skips a question in the survey (item nonresponse; Hox et al., 2008). For the study reported on in Paper III, this was not a problem. More than 99% of the teachers who attended the PD sessions where the questionnaires were distributed chose to answer them. The unit nonresponse was thus very low. Further, the item nonresponse for the items analyzed for the paper was, at most, only a few percent. In relation to Paper IV, the number of teachers whose answers regarding participation in BfM were missing in the TIMSS 2015 teacher questionnaire or not administrated was also merely a few percent.

Errors associated with *Measurement* arise when the survey questions are (a) inappropriately formulated, or (b) if a respondent, more or less deliberately, provides false information (Hox et al., 2008). In order to prevent these issues, it is important to pose clear questions. Campanelli (2008) therefore recommends that researchers pilot their questionnaires in order to refine, for instance, their content and wording to make them appropriate for the targeted sample. Regarding the questionnaires used in CoV (Paper III), several of the items had previously been used in studies with similar objectives (Hamilton et al., 2003; Malzahn, 2002; Middle-school Mathematics and the Institutional Setting of Teaching, n.d.; Stein & Kaufman, 2010). This is a relatively common procedure within the scientific community, simply because the items have been shown to be useful (Harkness, 2008). At the same time, as noted by Harkness (2008), survey items taken from other studies must be adapted to the specific context in which the current study is being conducted. This was also done for the questionnaires used in Paper III which were, several times, discussed and refined in the project group as well as piloted with teachers in the schools. The field tests, however, were not conducted according to all recommendations in the literature, for example Campanelli (2008). Nevertheless, as stated by the mentioned author, in many cases these kinds of extensive field tests are impossible to conduct due to insufficient resources in terms of both time and money. In such cases, she recommends that researchers at least test the surveys themselves, and thereafter conduct a small field test with an associated briefing report together with the respondents. These actions have all been taken in developing the questionnaires used in the study presented in Paper III and led to some minor changes in the original questionnaires. The changes were mainly related to choice of words and sentences structures in questionnaire items that teachers stated were hard to interpret.

Regarding the TIMSS 2015 teacher questionnaires used to collect data for Paper IV, these have gone through an extensive development process

to ensure their validity and reliability. This process is stated to include both discussions between staff at the TIMSS & PIRLS International Study Center and meetings with reviewing committees, as well as field tests (cf. Martin et al., 2016). Moreover, if multiple questions are used to create constructs, like in Paper IV, risks of measurement errors in relation to the reliability of the constructs can be detected by means of calculating Cronbach's alpha (Hox, 2008). For the constructs used in Paper IV, all alpha coefficient values were above .8, which is regarded as highly reliable (Cohen et al., 2011).

Finally, measurement errors in relation to (b), in addition to misinterpretation of the question, can occur when teachers deliberately provide false information. This is a particularly great risk when it comes to items that may be considered sensitive. It has been shown, however, that using self-administered questionnaires, similar to those used in Papers III and IV, minimizes this risk since respondents often experience these as more secure than interviews when it comes to confidentiality (Schwarz et al., 2008). Furthermore, when the questionnaires were distributed, it was stressed that the results would not be traceable to individual teachers, and that the teachers themselves would benefit from the questionnaires since the results would be used to improve CoV (Paper III) or contribute to valuable knowledge for improving mathematics instruction both nationally and internationally (Paper IV).

5.3.3 Validity and reliability in tests

Concerning reliability in tests, for example those used in Papers I and II, Cohen et al. (2011) discuss threats with respect to several aspects, such as examiners and markers as well as students' motivation and previous knowledge. To assure reliability in relation to examiners and markers, all teachers in the municipality were given the same written instructions for how to mark the tests. Furthermore, since most of the test items are either multiple-choice or short-answer questions, they are considered relatively easy to correct. With reference to aspects such as students' previous knowledge and motivation to take the test, these were not a large issue in this thesis. This is because the purpose of using the tests was not to measure students' mathematics abilities per se, but rather the effects of the PD programs on their achievement. For example, there is no reason to suspect that the motivation of the students taking the test in 2013 would be any different from that of those taking it in 2014 and 2015.

As for the validity of the tests, this concerns whether the test items actually measure what they are intended to measure. In the case of Papers

I and II, this regards the question of whether the tests used actually measure students' number sense. Given the amount of research and field testing (cf. Emanuelsson et al., 1996; McIntosh et al., 1997; Reys et al., 1999) undertaken to develop the tests in McIntosh (2008), it is argued that they are valid.

Finally, in relation to the discussion of the reliability and validity of the studies reported on in Papers I and II, it is also important to consider whether one could claim that it is the PD programs that have contributed to the differences seen in students' results. As mentioned by Cohen et al. (2011), there may be other confounding variables that affect the results of an experiment. Therefore, to increase the validity and reliability of the studies, a control group (schools not participating in a PD program) was included in the analyses of interaction effects between year and intervention.

6 Results of the papers

In this chapter, a summary of the five papers included in the thesis is provided. As both theory and methodology have been elaborated on in previous chapters, the summary is mainly focused on the studies' results. These results will be used in the next chapter of the kappa to discuss plausible reasons for the difference in the impacts of BfM and CoV, as well as elaborations of the critical features of Content Focus and Coherence.

6.1 Summary of Paper I

Paper I reports on two things. Firstly, it contains a description of BfM and CoV based on Desimone's (2009) five critical features of high-quality PD. The description shows that, regarding the five core critical features, the programs are very similar (see Section 4.2.3). The greatest difference lies in their respective content. Both programs emphasize teachers' development of PCK, but do so with slightly different focuses.

Secondly, the paper reports on the respective impacts of BfM and CoV on student achievement immediately after the teachers' participation in one of the two programs. Despite the programs' similarities, the results suggest that there are differences in their impacts on student achievement. Differences can also be seen for the same PD program, but within different grade levels. Data on students' results on the McIntosh (2008) tests in 2013 (n=8882) and 2014 (n=8848) were analyzed, and significant interaction effects between year and intervention were found for Grades 2, 8 and 9. When the descriptive statistics were examined, the mean changes in students' test scores for Grades 2–5 pointed at slight improvements for students whose teachers participated in CoV, and a small deterioration for students whose teachers participated in BfM. In contrast, the changes in mean scores for students in Grades 7–9 mainly indicated a declining trend for CoV, while the same results for BfM pointed at improvements. Considering that both programs were implemented within the same municipality and are highly similar when it comes to several of the critical features of high-quality PD, as identified in the research literature, these

variances in results are noteworthy. Therefore, the paper ends with a discussion on possible explanations for the variations found by taking into account the differences in the programs' Content Focus, the Swedish educational context (Bergqvist et al., 2009; Boesen et al., 2014), and results from teacher questionnaires (Ramböll, 2014; Paper III). Based on the results from the study, and the discussion of possible explanations for the differences found, future studies are encouraged to look into what kinds of PCK in PD programs for mathematics teachers have the strongest impact on student achievement.

6.2 Summary of Paper II

In view of the findings from Paper I, Paper II set out to contribute to research on the critical feature of Content Focus. It did so by means of a closer examination of the content of BfM and CoV and their sustained impacts on student achievement.

As in Paper II, the McIntosh (2008) tests were used to examine the impacts of BfM and CoV on student achievement. However, this time we examined their impacts one year *after* the teachers had finished the respective program. Considering the calls for more studies on the impacts of PD programs several years after their implementation (e.g., Desimone, 2009; Kennedy, 2016; Sowder, 2007; Wayne et al., 2008), this analysis was both a method for increasing the study's reliability and an important result in its own right. The results from the ANOVAs showed significant interaction effects between year and intervention for Grades 8 and 9, with the descriptive statistics suggesting a somewhat more positive performance development for students whose teachers participated in BfM, compared to those whose teachers participated in CoV. It is worth noting that, for Grade 2, the variations in students' results between pre- and post-measurements differed significantly between students whose teachers had participated in CoV and those whose teachers had not participated in an intervention. A closer examination of the descriptive statistics indicated that students whose teachers had participated in CoV showed more positive development compared to those whose teachers had not participated in an intervention. However, no significant interaction effect for Grade 2 was detected when students whose teachers had participated in CoV were compared to those whose teachers had participated in BfM. Nevertheless the numbers did indicate that it was close ($p=.053$), with the descriptive statistics suggesting a more positive development for students whose teachers had participated in CoV.

Regarding the analysis of the PD programs' respective content, we developed a framework (see Section 5.1.1.2) for examining the PD materials provided to the teachers in terms of two aspects: (a) their content, and (b) their methods for facilitating teachers' enactment of the presented ideas. This framework allowed us to detect distinct differences between the programs. In relation to analysis (a) the programs' content, the results showed that CoV heavily emphasized Teacher Practices (i.e. knowledge as part of action), while BfM placed more weight on Teacher Knowledge (i.e. knowledge as something that is possessed). Moreover, considering Teacher Practices, BfM mainly emphasized practices related to the lesson design (e.g., grouping of students, tasks to hand out), while CoV placed the primary focus on teacher actions (e.g., asking students questions with a specific purpose). Differences were also found between the programs in regard to analysis (b) the programs' methods for facilitating enactment. In relation to teachers' actions, BfM heavily emphasized Insight and thus provided teachers with more agency, or less support, in regard to their decisions on how to act during instructional activities. In contrast, the rounds for CoV had a fairly equal balance between Insight and Strategy, and provided teachers with more detailed descriptions of how to act during the instructional activities. The results regarding PD materials focusing on lesson design, on the contrary, showed that BfM placed a strong emphasis on Strategy and provided teachers with multiple samples of lesson activities to conduct with students. In contrast, CoV placed more emphasis on Insight and relied on teachers to come up with the lesson designs themselves, together with the support of questions to discuss among their colleagues when planning the activity.

The results of the analyses of the PD materials provided explanatory value in discussing reasons for the PD programs' differing impacts on student achievement. The conclusion that was supported by the available data, and that was offered as the most plausible, was the following: If a PD program is to lead to higher student achievement after one to two years, the discrepancies between teachers' current knowledge and practice on the one hand and the knowledge and practice promoted in the PD on the other should be moderate. For an elaboration of the conclusion and the arguments that support it, the reader is referred to Paper II and Section 7 of the kappa.

6.3 Summary of Paper III

In light of the results regarding the differences found between the grade levels in the impact of CoV on student achievement, the aim of Paper III was to examine and compare how the particular PD program had affected the participating primary (n=104) and secondary teachers' (n=31) reported mathematics instruction, respectively. To achieve this aim, data from teacher pre- and post-questionnaires (see Section 5.1.2) were analyzed.

Compared to the primary levels, the results demonstrate that the secondary-level teachers reported having made greater changes in their mathematics instruction. For example, statistically significant differences could be seen for both the primary- and secondary-level teachers in statements on how common whole-class discussions and individual work in the textbook were during instruction. Still, while the changes made by the primary-level teacher were all considered to be small, as measured by effect size, those made by the secondary-level teachers were mainly moderate or great. The same also applied to questions regarding how much emphasis the teachers reported having placed on different mathematical competencies in their mathematics instruction. Furthermore, compared to the primary-level teachers, those at the secondary level expressed to a greater extent that they had difficulty (due to, e.g., insufficient time or classroom resources) in conducting changes in their mathematics instruction based on their experiences from the PD program.

Based on the results of the study, it is suggested that the coherence between the Content Focus of the PD program and teachers' initial practice, instead of merely their knowledge and beliefs, is important to consider in designing the program. As the Content Focus of the PD program in CoV seemed to be less coherent with the secondary-level teachers' initial practices, and thereby also more challenging, it may be even more important to consider how to support these teachers.

6.4 Summary of Paper IV

While Paper III focused on CoV, Paper IV examined and compared the respective impacts of BfM on fourth- and eighth-grade teachers. More specifically, data from the TIMSS 2015 teacher questionnaire were used to study the program's impacts on Teacher Confidence, Teacher Collaboration, and Support from School Leadership.

The results from the statistical analyses indicated that, for both the fourth- and eighth-grade samples, BfM did not have a sustained impact on the teachers' experiences of school leadership support or their confidence

in teaching mathematics to students. Regarding Teacher Collaboration with colleagues, however, a significant difference between participants and non-participants was detected for the eighth-grade teachers. A subsequent examination of the descriptive statistics showed a more positive value for BfM participants compared to non-participants. For the fourth-grade teachers, no significant difference between participants and non-participants was established. The numbers did indicate that it was close ($p=.54$), however, with the descriptive statistics pointing to non-participants reporting on more collaboration with colleagues compared to participants.

The results of Paper IV are somewhat contradictory to those presented in national evaluations of BfM (Ramböll, 2016; Österholm et al., 2016). Potential explanations for the differences were therefore discussed, resulting in a recommendation that future studies of PD explicitly adhere to variations in impact.

6.5 Summary of Paper V

In Paper V, a systematic review of the literature on K–12 teacher PD focusing on the critical feature of Coherence was undertaken. In total, 95 papers that in some way defined or elaborated on Coherence were analyzed using a framework containing three categories derived from the research literature (see Section 5.1.1.1).

The results showed that more than 90% of the papers could be categorized into the first category, *Be coherent with*. These papers described coherence in terms of the notion that PD should be coherent with constructs outside itself, for example standards or assessments. A vast majority of the papers in the first category treated coherence as a desired feature of teacher PD by stating, and sometimes also arguing for, different things PD ought to be coherent with (e.g., standards, assessments) and/or what there should be coherence about (e.g., goals for student learning). However, 17 papers did discuss whether coherence is always possible to achieve and/or desired, for example stressing that it is the diversity of the skills and perspectives within PD programs and their participants that drives the growth of a community. The second category (*Be internally coherent*), which included 24% of the papers, treated coherence as an internal property of PD programs themselves, for example stressing that activities within the programs should be aligned. Finally, 25% of the papers fell into the third category (*Create coherence*), treating coherence as something PD can or should create, rather than something it should

possess. Most of these papers assumed that there is a "right" answer to the alignment — a predetermined goal stating what PD should create coherence about, or among what. Three papers, on the other hand, described coherence as something that is negotiated between different actors, amongst them teachers, over the course of PD. In summary, the results from the analysis showed that there were great variations in the papers' descriptions of what there should be coherence about and/or among what. Moreover, suggestions for how to achieve coherence were also scattered. Among the papers, 30% did not mention it, while the rest offered various suggestions regarding both methods for facilitating enactment and PD content.

Paper V ends with a discussion of the results in relation to the teacher positioning that the various definitions of coherence bring to bear, and suggestions for future research are made. For example, the fact that few papers questioned whether coherence is possible to achieve and/or desired raised questions in relation to how teachers are presented as learners. There is likely no point in discussing tensions and struggle if it is assumed that all things are consistent. Several models of teacher learning (e.g., Gregoire, 2003; Mevarech, 1995) do, however, suggest that for true conceptual changes to occur, in contrast to mere assimilation, cognitive struggle is likely necessary. Moreover, in light of the results from the analysis, it is recommended that future studies in the area of teacher PD define what type of coherence is referred to. Furthermore, more research regarding how to achieve coherence, as well as studies conducted in contexts other than the US, are needed to move our understanding forward.

7 Conclusions and discussion

In this section, conclusions from the results of the papers are discussed in relation to previous research. The discussion is organized into two themes connected to the aims of the thesis. The first theme considers plausible reasons for the differences between the impacts of BfM and CoV on teachers, instruction, and student achievement, while the second considers an elaboration of two of Desimone's (2009) critical features of high-quality PD: Content Focus and Coherence.

7.1 Differences within and between the professional development programs' impacts

The results from Papers I–IV indicate that BfM and CoV have had complex differential impacts on teachers, instructional practices, and student achievement, of which the last has been referred to as the bottom line in education and teacher PD (Guskey, 2002a, 2014a). The differences in impact are seen both between the programs and within them. For example, Papers I and II indicate that BfM has mainly had a positive impact on student achievement for Grades 8 and 9. The results for CoV, on the other hand, point at a positive impact on student achievement for Grade 2. The differences within and between the two programs' impacts are noteworthy, as they are highly similar when described according to established frameworks of what constitutes high-quality PD (see Paper I and Section 4.2.3), and were also conducted in the same municipality and during the same time frame.

In the complex setting of educational systems, it is difficult to draw firm conclusions as to the causes of PD programs' differences in impact on student achievement. Then again, it is important that such attempts be made in order to advance our understanding. Considering that (a) BfM and CoV have been conducted in the same context, (b) data on student achievement were gathered on more than one occasion (Papers I and II), (c) additional teacher data are available (Paper III; Paper IV; Ramböll, 2014; 2016; Österholm et al., 2016), and (d) a thorough examination of the

programs' Content Focus and methods for facilitating enactment has been done (Paper II), I argue that this thesis offers an exceptional opportunity to discuss plausible explanations for variations found between and within the programs' impacts. To organize this section of the thesis, I draw upon Clarke and Hollingsworth's (2002) four analytic change domains of professional growth: *the Personal Domain, the Domain of Practice, the External Domain,* and *the Domain of Consequence*. These domains can be seen as more detailed descriptions of the four components (PD, teachers, instruction, and student achievement) in the linear change model. In this discussion, *the Domain of Consequence* (i.e. the salient outcomes) is seen as the programs' impacts on student achievement, while the pre-conditions and changes in the other three domains are used to explain the changes in this particular domain.

7.1.1 The Personal Domain

The available data indicate that the differential impacts of BfM and CoV may be due to differences in the Personal Domain, namely teachers' previous knowledge. More precisely, the conclusion is that the PD programs have had a more positive impact on student achievement when the teachers, before the start of the programs, were already somewhat familiar with the content. This conclusion holds for both BfM and CoV.

If we first consider BfM, note that the results from Paper II show that the program focuses mainly on teachers' knowledge, especially in relation to mathematical content. Considering that primary-level teachers in Sweden generally take fewer mathematical courses than secondary-level teachers, the latter should be more familiar with the program's content. It is also among these teachers that BfM has had a positive impact on student achievement. This suggestion is further supported by other studies within the field of teacher PD (e.g., Britt et al., 2001; McNeill & Knight, 2013), which have demonstrated that teachers at different grade levels respond differently to the same PD program. The authors have proposed that this may be due to differences in teachers' knowledge of the subject content and/or instructional practices. In a recent article, Covay Minor, Desimone, Lee, and Hochberg (2016) examine this issue more deeply by comparing various PD programs' impacts on teachers with different levels of content knowledge (CK). There results show that if a PD program focuses on both CK and PCK, like BfM does, teachers with strong initial CK will mainly develop their PCK, while those with weak initial CK will mainly develop their CK. The authors suggest that, in practice, this means that teachers with high CK will make more use of the PD program, as they can focus on

the pedagogical ideas and why they are implementing the strategies. Teachers with low CK, instead, tend to make shallow instructional changes and see the PD materials as scripts, as they do not detect the big ideas behind the intervention.

Regarding CoV, the results from Paper II show that the program places almost no emphasis on developing teachers' knowledge of mathematical content. The focus is instead on Teacher Practices (i.e. knowledge as part of action), especially in relation to teachers' actions in the classroom. Recall, from Section 4.1, that a previous study (Bergqvist et al., 2009) has shown that Swedish primary-level teachers' instructional practices give students more opportunities to develop additional mathematical competencies besides procedural fluency. In other words, before the start of the program, the instructional practices promoted in CoV were more familiar to the primary-level teachers than to those at the secondary level. This statement is further supported by data from the pre-questionnaires used in Paper III, as well as the reports from another recently conducted study on teacher PD in Sweden focusing on formative assessment practices (Boström, 2017). This study showed that, compared to primary-level teachers, teachers in the secondary grades experienced a more pressured working situation in implementing the promoted instructional changes and found the implementation of the latest national curriculum to be greater barriers to change. The fact that the primary-level students whose teachers participated in CoV have demonstrated more positive achievement gains compared to the students in the secondary grades can thus be explained with the same reasoning that was conducted for BfM.

In summary, teachers with strong initial knowledge of the PD content may be able to make more use of the PD program, as they can focus on the pedagogical ideas and why they are implementing the strategies. Teachers with low initial knowledge of the same content, on the other hand, tend to make shallow instructional changes and see the PD materials as scripts, as they do not detect the big ideas behind the intervention.

7.1.2 The Domain of Practice

The differences in the programs' impact may also be due to issues related to the Domain of Practice. This domain includes aspects related to classroom instruction, but also other dimensions of teachers' professional practices, such as collegial collaboration. In this thesis, the discussion in relation to the Domain of Practice includes issues concerning: a) teachers' initial practices, b) students' previous instructional experiences, and c) teacher collaboration. The first issue, a) teachers' initial practices, is

regarded as being a part of teacher knowledge and has already been discussed in relation to the Personal Domain.

In relation to b) students' previous instructional experiences, the results indicate that BfM and CoV have had a more positive impact on student achievement when the students, before the teachers' participation in the programs, were already somewhat familiar with the promoted content/and or instructional practices. In other words, BfM and CoV have had a more positive impact on student achievement if the instructional changes for the students have been moderate. For example, just as the secondary-level teachers are typically unfamiliar with the teaching practices advocated in CoV, so are their students. The results from Paper III also show that both the primary- and secondary-level teachers report having made changes in the mathematics instruction. The changes were larger, however, for the secondary-level teachers than for those at the primary level. As opposed to CoV, Paper II shows that BfM does not place such a strong emphasis on teachers' instructional practices but instead stresses the mathematical content. Compared to CoV, it should therefore not result in equally high demands for instructional changes for the secondary-level teachers and students. This claim is supported by the descriptive statistics from Ramböll (2014) which indicate that, compared to teachers in the primary grades, the mathematics teachers in Grades 7–9 state to a lesser extent that the project has contributed to their changing how they teach. In addition, the same teachers report having conducted the lesson activities suggested in the PD program less often than do those in the primary grades.

Considering c) teacher collaboration, the results indicate that PD programs that bring about an increase in this aspect also have a positive impact on student achievement. First of all, Paper IV shows that BfM had an impact on teacher collaboration for the eighth-grade teachers, with the descriptive statistics pointing at more collaboration for participants than for non-participants. The same impact was not detected for the fourth-grade teachers. In a similar way, though not reported on in the paper, additional analyses of the teacher questionnaires used in Paper III indicate that CoV had a stronger positive impact on teacher collaboration for the primary-grade teachers than for those in the secondary grades. In other words, the grades in which a particular PD program had a positive impact on teacher collaboration were also those in which student achievement for teachers participating in the program increased. Note, however, that increased teacher collaboration is likely not enough to increase student achievement on its own. How, and about what, teachers collaborate must also be taken into consideration (Cobb & Jackson, 2011; Timperley et al., 2007; Vescio et al., 2008). Studies have suggested that teachers take

advantage of teacher collaboration to varying degrees depending on their initial knowledge and practices. One example of this is Penuel, Sun, Frank, and Gallagher (2012), who propose that when teachers have limited experience of implementing a PD program's promoted practices, they mostly need organized and sustained PD as the gap is so great between their current practice and the promoted practices. On the other hand, teachers who are already familiar with the promoted practices benefit more from collegial cooperation as their "primary learning goals are to develop more nuanced, differentiated approaches to teaching that are facilitated by interacting with colleagues about their practices" (Penuel et al., 2012, p. 110). Put differently, for BfM and CoV, teacher collaboration seems to be beneficial for increasing student achievement when there is a moderate gap between teachers' initial knowledge and practices on the one hand and the PD program's promoted knowledge and practice on the other.

7.1.3 The External Domain

The differences in the impacts of BfM and CoV on student achievement can also be explained from the perspective of the External Domain, which includes external sources of support (e.g., from principals) and stimuli (e.g., a teacher PD program). Within this domain, the results indicate that the differences in the PD programs' impacts are affected by issues related to either the external support in terms of available time to conduct the instructional changes, or the PD programs' design in terms of their methods for facilitating enactment.

Considering the external support for conducting the instructional changes, several scholars have argued that, for improving instructional practices at scale, merely providing teachers with PD opportunities is not enough (e.g., Bryk et al., 2015; Cobb & Jackson, 2011; Darling-Hammond & McLaughlin, 1995; Knapp, 2003). In other words, instructional change at scale is not only dependent on teacher learning, but also requires an organization that supports this learning and the instructional changes the teachers are expected to implement. These arguments are supported by the results from Paper III, which show that the secondary-level teachers participating in CoV reported receiving less support in regard to available time to implement the promoted practices compared to the primary-level teachers. In contrast, the descriptive statistics in Ramböll (2014) point at the opposite relationship for teachers who participated in BfM. Students whose teachers experienced having available time for planning and implementing the practices promoted in the PD as a larger problem were also those whose results did not improve. Note that time alone is not

enough to promote instructional improvement in student achievement (Kennedy, 1998, 2016; Scher & O'Reilly, 2009; Timperley et al., 2007). The results in this thesis show that teachers participating in the same PD program report different needs of support in terms of time. In other words, the level of support provided to teachers must be considered in relation to the demands of the changes requested.

Similar reasoning can be applied to another aspect of the External Domain, namely the design of PD programs in terms of their methods for facilitating enactment. The findings from Kennedy's (2016) meta-analysis indicate that PD programs focusing on providing teachers with strategies or relying on teacher insights have the most positive impact on student achievement. However, the results from Paper II suggest that the effects of the two methods for facilitating enactment are relative to the nature and demands of the changes requested. In short, the paper's results demonstrate that relying on teacher insights may be more beneficial when the teachers are somewhat knowledgeable about the promoted content and/or practices. If, however, teachers are novices to the content and practices encouraged in the program, providing them with more prescriptive strategies may be more helpful. Similar reasoning is offered by Penuel et al. (2012). In light of the results from their study, the authors conclude that teachers not familiar with a PD program's promoted practice may be more likely to change their instructions when they participate in more organized PD. By contrast, teachers who are already familiar with the promoted practices may benefit more from interaction with colleagues.

7.2 Elaborating on critical features of teacher professional development

The above discussion regarding the results from Papers I–IV, the results from the systematical review of the literature on teacher PD (Paper V), Gregoire's (2003) distinction between superficial and true conceptual change, and the studies presented in Section 3 of the kappa raises several questions regarding the definitions of the critical features of Content Focus and Coherence. Drawing on the available data, I suggest an elaboration of Desimone's (2009) definitions of these features.

7.2.1 Content Focus

Content Focus is suggested, by several scholars (Desimone, 2009; Ingvarson et al., 2005; Kennedy, 1998), to be one of the most influential

critical features of PD programs' impacts on teachers, instructional practices, and student achievement. Particularly programs that address both the subject-specific content and the pedagogy (i.e. PCK) have gained a great deal of support in the literature (e.g., Clewell et al., 2005; Desimone, 2011; Salinas, 2010; Scher & O'Reilly, 2009; Slavin & Lake, 2008; Slavin et al., 2009). At the same time, though, the results from this thesis demonstrate that BfM and CoV have had very different impacts on student achievement. This is the case even though the programs are highly similar when described according to Desimone's (2009) five critical features and both emphasize teachers' PCK in mathematics. Also within the programs, the impacts on students (cf. Papers I and II), teachers, and instructional practices (cf. Papers III and IV) differ between the various grade levels. Recently, other studies within the field have also shown that PD programs incorporating the five critical features and emphasizing teachers' development of PCK do not necessary lead to improvements in student achievement (Garet et al., 2016; Jacob et al., 2017; Kennedy, 2016). For example, Kennedy's (2016) meta-analysis did not detect any particular PD content as more beneficial than another. In a recent article (Desimone & Garet, 2015), Desimone herself also offered evidence from studies indicating that PD programs emphasizing teachers' performance of routine procedural tasks, as opposed to enhancing their knowledge of the content and pedagogical principles, may be more beneficial in changing instructional practices and improving student achievement. In light of these results, one could question whether it should be recommended that teacher PD programs focus on developing teachers' PCK.

To be able to respond to this issue, one has to consider the kind of instructional practices a PD program is striving to attain. As indicated by the results from Paper II, and as argued by other scholars (Desimone & Garet, 2015; Desimone & Stuckey, 2014; Santagata, 2009), it is easier for PD to change teachers' procedural routines than their skills with reflective practices. Nevertheless, to prepare students for the demands of today's society, both scholars (e.g., Clewell et al., 2005; Goldsmith et al., 2013; Schoenfeld, 2014) and national curricula (cf. Boesen et al., 2014; Marrongelle et al., 2013) call for more conceptual and inquiry-oriented instructional practices. This is also reflected in the content of BfM and CoV, as they strive to support teachers' implementation of the most recent national curriculum (Skolverket, 2011a). In order to implement such practices with quality (e.g., elicit student thinking and foster higher-order discussions), Desimone and Garet (2015) argue that teachers need to develop their PCK and not simply adopt certain procedures (e.g., increases the amount of time spent on certain activities). Additionally, teachers' PCK

has been shown to have greater predictive power for student achievement in mathematics than their content knowledge (Baumert et al., 2010). In other words, if a PD program aims at implementing more inquiry-based instructional practices, it is not enough to focus on teachers' implementation of procedures or the development of their mathematical knowledge; it likely must also emphasize the development of their PCK.

In summary, the results from Papers I–IV, as well as Kennedy's (2016) recently conducted meta-analysis, have not been able to detect any particular PD content as more beneficial than another in improving student achievement. Instead, the results from the papers in this thesis and other studies (e.g., Covay Minor et al., 2016; McNeill & Knight, 2013; Penuel et al., 2012) indicate that different content works to differing degrees for teachers depending on their previous knowledge and/or practices. The inference is, thus, that in determining the Content Focus of a particular PD program, one has to consider: a) the nature of the content (e.g., striving for the implementation of procedural or conceptual knowledge) and b) the coherence between the program's Content Focus and teachers' current knowledge and practice. At the same time, it is impossible to discuss a PD program's coherence without paying attention to its Content Focus, as one needs to know what there should be coherence about. The critical features of Content Focus and Coherence are therefore highly intertwined.

7.2.2 Coherence

The results from Paper V demonstrate that the consensus on Coherence as a critical feature of teacher PD seems to be an illusion. Several scholars argue for the need of coherent PD programs, but what there should be coherence about, and among whom or what, differs between the studies. In Desimone's (2009) framework, Coherence is defined as including two parts: (1) the PD content being aligned with school, district and state policies; and (2) the extent to which teachers' learning is consistent with their knowledge and beliefs. As shown in Paper V, the first part, which is related to Coherence with Clarke and Hollingsworth's (2002) *External domain*, seems to be unproblematic. Most scholars argue that the PD content should be coherent with policy, standards, curriculum, and/or assessments. However, about what (e.g., goals for student learning) is not clearly defined. For the second part, which relates to Coherence with teachers' *Personal domain*, the results from Paper V and the discussion of explanations for the differences found between the differential impacts of BfM and CoV, show a more complicated view. In addition, Desimone (2009) does not mention aspects of Coherence related to teachers' practice

and classroom work (i.e., *the Domain of Practice*), though the results from Paper V show that this is included in several other scholars' definition of the concept. As a consequence, aspects of Coherence in regard to the Personal Domain and the Domain of Practice will be elaborated on below.

As stated in Section 7.1.1, the impact of teacher PD programs depends on aspects connected to the Personal Domain; more precisely, teachers' previous knowledge of the PD content. Note that this does not imply that teachers' previous knowledge and practices should be the same as those raised in the PD. Instead, it is about ensuring that the design of PD programs must adhere to the notion that the distance between teachers' current knowledge and the PD content should be kept at a reasonable level. Opfer and Pedder (2011, p. 378) describe this phenomenon by referring to the Goldilocks principle, which suggests that "the relationships between variables are often curvilinear—too little and learning will not occur, too much and it is counterproductive or negative". This implies that PD programs should be differentiated to match teachers' previous knowledge and practices. Therefore, it is critical to assess teachers' understanding of the content before the program design is settled. On the other hand, providing differentiated PD may collide with the critical feature of Collective Participation. A pressing challenge for the design of future PD programs is thus "how to successfully merge the idea of collective participation with differentiated PD" (Desimone & Garet, 2015, p. 255). The two features are not incommensurable, however. Scholars have, for example, suggested that teachers can be asked to take on different roles during learning activities (Santagata et al., 2010), or that additional support can be added for certain types of teachers (Desimone & Stuckey, 2014). Moreover, Opfer and Pedder (2010) state that the Goldilocks principle also has an application on teacher collaboration, since too much collaboration may emphasize conformity to group norms at the expense of creativity and initiative. This argument is further supported by some of the papers identified in the systematic literature review (Paper V), which stress that it is the diversity of the skills and perspectives within PD programs and their participants that drives the growth of a community.

In relation to the Domain of Practice, Desimone (2009) does not state coherence with teachers' instructional practice and classroom work as an aspect of the critical feature of Coherence. However, the results from Paper V do demonstrate that it is a significant aspect of the feature in the PD literature. More than 20% of the papers included in the study emphasized that the PD content should be coherent with teachers' practices and/or classroom work. In addition, scholars (Penuel et al., 2012; Spillane & Zeuli, 1999) have suggested that if the gap is too great between teachers'

current practice and the PD targeted practice, they are likely to integrate the new ideas into their existing frameworks for teaching rather than transforming them. In relation to this, one can question whether instructional improvement efforts should strive to reorganize practices (cf. Cobb & Jackson, 2012; Cobb, Jackson, Smith, Sorum, & Henrick, 2013) or support existing ones. If one strives for all things being consistent at the beginning, there is likely no point in discussing tensions and struggle. However, Gregoire's (2003) model of teacher change, as well as others (Mevarech, 1995), suggest that for true conceptual changes to occur, as opposed to mere assimilation, cognitive struggle is probably necessary. For example, in the case of BfM and CoV, more substantive changes are likely needed if one considers the instructional changes the programs are aiming for (see Section 4.2) in relation to current mathematics teaching in Sweden (Bergqvist et al., 2009; Boesen et al., 2014). Coherence regarding teachers' instructional practices is therefore not an argument for uniformity. Rather, it is an argument that this type of coherence is an important aspect to consider when designing teacher PD programs in relation to what kind of support one is able to provide. Two aspects of the support provided to teachers are important to consider. Firstly, one has to consider the amount of support, such as time, opportunities for collaboration, and artifacts (Spillane, 1999). Secondly, one must adhere to the type of support that is provided. As indicated by the results from Papers I–IV, and as argued by Peneul et al. (2012), teachers who are further away from the promoted instructional practices benefit more from prescriptive strategies. This is not to say that PD programs should only focus on teachers' procedural behaviors and not adhere to the epistemological regularities of instruction (i.e., conceptions of knowledge in the classroom; Spillane & Zeuli, 1999). Then again, even if the final goal of PD is to achieve transformative changes, superficial changes focusing on procedures may be a necessary first step (Kazemi & Hubbard, 2008; Santagata, 2009). In other words, routine instructional activities may function as tools that give teachers the opportunity to focus on specific aspects of their practice instead of everything at once. Or, as expressed by Kazemi and Hubbard (2008, p. 438), "the recipe you must at first rely heavily on is what later allows you to experiment with and respond to the specific conditions of your enactment".

8 Concluding remarks

Although I would have preferred to tell a straightforward story, the collected data and the complex settings of educational systems do not allow it. Instead, the results have shown a complex story and I have provided possible interpretations. In this final section of the thesis, I reflect on the limitations of the study, but also on its contributions and implications for research and practice. In doing so, I aim to offer some helpful suggestions for future work and to those devoted to supporting the development of mathematics instruction at scale.

8.1 Limitations

The results presented in this thesis should, as in all studies, be interpreted in light of their limitations. One possible limitation is related to the use of questionnaires to measure instruction. Both questionnaires and interviews as well as classroom observations are frequently used to measure instructional practices, but studies using questionnaires seem to receive the most criticism in educational research (cf. Desimone, 2009). Researchers are often seen making excuses for the use of this method of data collection (e.g., Fishman et al., 2003; Hamilton et al., 2003), stating that they acknowledge that such measures are weak and that the questionnaire items may be subject to inaccurate responses. At the same time, using interviews and observations to collect data is considerably more expensive and is therefore not always appropriate for large-scale studies (Hamilton et al., 2003). As stated by Mayer (1999, p. 44), we cannot continue to solely "rely on in-depth studies of a small number of classrooms because they are of little value for assessing state and national instructional trends". In addition, the bias regarding the use of teacher questionnaires to measure instruction might not be entirely fair. First of all, the questionnaires used in Swan (2006) showed satisfactory consistency between classroom observations and teacher surveys. Secondly, Mayer's (1999) results also demonstrated that teacher questionnaires are valid and reliable in broad

terms when asking if teachers are carrying out certain instructional activities (e.g., asking students specific types of questions). Thirdly, based on her review of the literature, Desimone (2009) argues that the bias towards surveys is often based on early studies, many of which, by current standards, would be considered to contain fatal flaws. She concludes that when survey questions ask teachers *what* they did and *how often* (as in this thesis), in contrast to *how well* they did it, observations and surveys have been shown to produce much the same information. Hence, though questionnaires may not yield depth in the data collected, they have been used to identify broad patterns of impact when exploring differences in practice among teachers and schools (e.g., Stecher & Borko, 2002). This is not to say that the thesis would not have benefited from also including additional interview and observational data. Future studies should therefore explore the presented results using other methodological approaches.

The above discussion is an argument that teacher questionnaires have been proven useful in measuring the effects of teacher PD on instructional practices. Note, however, that they still have their shortcomings. For example, sometimes it is not enough to know how often teachers do a specific kind of instructional action; data on the quality of the instruction are also needed. For example, two teachers could both be asking students open-ended questions and requiring them to explain their answers, and thereby score high on a specific questionnaire item. One of the teachers, however, might ignore the students' answers while the other uses them and engages the students in productive discussions about differences and similarities. In other words, "more work needs to be done to devise surveys that better distinguish between teachers who perfunctorily use reform practices and those who use them effectively" (Mayer, 1999, p. 42).

To overcome the shortcomings mentioned above and increase measurement quality, researchers could take different actions. One example brought up in Mayer (1999) is related to the idea that measurement quality is increased if additional items measuring the same characteristic are combined into one indicator of that characteristic. Such action was taken in Paper IV, as several items were combined to measure specific constructs. In Paper III, on the other hand, single items were used. Note, however, that no conclusions were drawn based on a single item. Instead, several items were used to draw conclusions about where the results were pointing, for example that primary teachers have made less changes to their instruction and that secondary-level teachers experienced less support. The items used to draw the conclusions were all pointing in the same direction.

Another possible limitation of the study is related to the choice to use secondary data (i.e. data not specifically designed to measure the PD programs' impacts) in Papers I, II and IV, Firstly, the TIMSS questionnaires were used to measure BfM's impacts on teachers and instruction, though they are not specifically designed for that purpose. Secondly, McIntosh's (2008) tests were used to measure student achievement. These tests place sole emphasis on number sense and thereby neglect other mathematical content areas, such as Geometry or Algebra. In relation to the aims of CoV and BfM (see Section 4.2), one could thereby argue that there is a lack of alignment between the interventions and the outcome measures. However, a test that measures all the mathematical content competencies and content in the national curriculum (Skolverket, 2011a) would take a long time to carry out, and time is a highly limited resource for teachers and students (Bryk et al., 2015). Therefore, it is most likely that large-scale evaluations will have to continue using secondary data that are not completely designed for their purposes (Hamilton et al., 2003). Such data also have an advantage as they may detect possible unintended outcomes of the PD (Guskey, 2002a; 2014a). Moreover, and as discussed in Paper IV, the three constructs analyzed in the TIMSS questionnaires were closely connected to BfM's overarching goals. In addition, even though the McIntosh (2008) tests were not specifically designed for the PD programs, various elements of them can be found within the projects. Parts of McIntosh's book (2008) were integrated into both programs, either as teacher PD materials on number sense (BfM) or as examples of tools for formative assessment (CoV). In addition, the teaching of numbers should not be seen as solely concentrating on computation. It also involves other aspects, such as the communication and understanding of central mathematical concepts (National Research Council, 2001; Reys et al., 1999), an approach that can also be increasingly detected throughout the history of the mathematics parts of the Swedish national curriculum (Prytz, 2015). Finally, though number sense does not cover all the mathematical content, it has been argued that it is a central foundation for further mathematical studies, and the learning of concepts associated with numbers has been more thoroughly investigated than other content areas (National Research Council, 2001). Several of these areas are also highly intertwined with number concepts. Admittedly, apart from the McIntosh (2008) tests, national exams designed to test selected parts of the national curriculum were taken annually by all students in Grades 3, 6 and 9. Data on student achievement from these tests were, initially, also analyzed for Papers I and II. These data were excluded from the final results, however, as the specific parts of the national curriculum that are

tested differ from year to year. In other words, the national tests lack a common trait and scale. It is recommended that such tests be avoided when measuring changes (Linn & Slide, 1977), as is the case in this thesis.

Finally, one could question whether it is even possible to measure the impacts of teacher PD programs. Studies of PD take place in real-world settings, and the relationships between PD, teacher, instruction and student achievement are far too complex and include far too many intervening variables to allow for simple causal inferences (Bryk et al., 2015; Guskey, 2002a). As mentioned by Guskey (2002a) and Schoenfeld (2000), research in the educational context is often not a matter of conclusively *proving* something. What we can do, however, is to collect *evidence*; and if done in a cumulative process, we can hopefully move towards a conclusion that is regarded to be beyond reasonable doubt. This thesis should be viewed as a contribution to this cumulative process.

8.2 Contributions

The contributions of this thesis are empirical, methodological and theoretical. To begin with, Papers I–IV contribute important *empirical* results in relation to recent calls for studies investigating the impacts of PD programs on teachers, instructional practices, and student achievement. Scholars have argued that there is a lack of empirical studies on improvement efforts (a) being implemented on a large scale (Marrongelle et al., 2013; Wayne et al., 2008), (b) that include non-volunteers (Desimone, 2009; Wayne et al., 2008), (c) in contexts other than the US (Clewell et al., 2005), (d) that follow the results on student achievement over time (Desimone & Stuckey, 2014; Kennedy, 2016; Wayne et al., 2008), and (e) that explicitly attend to variations between teachers (Bryk et al., 2015; Desimone & Stuckey, 2014). Papers I–IV adhere to all these calls. The results should be interesting, particularly in the Swedish context, for all those involved in decisions regarding the choice and design of PD initiatives. In an international perspective, the thesis can be seen as a comprehensive case study that could support the discussion and knowledge base regarding what constitutes high-quality PD for mathematics teachers.

Moreover, Paper V contributes empirical results related to calls regarding the need to clarify the language and frameworks used in the field of teacher PD in order to support future research and practice (Schoenfeld, 2015; Sztajn et al., 2009). It does this by describing how a commonly mentioned critical feature of teacher PD, Coherence, is defined in the PD research literature and how it is suggested that Coherence can be achieved.

Moreover, it demonstrates how common or unusual the different definitions and suggestions are. These results should be interesting for both researchers working in the field of teacher PD as well as those involved in decisions regarding the investment in and the design and/or implementation of PD programs for K–12 teachers.

In regard to *methodological* contributions, scholars have reported on the lack of appropriate tools designed for measuring teacher PD (Desimone, 2009; Hill et al., 2013; Wayne et al., 2008). In particular, there is a need for measurements designed to test specific features of teacher PD (Desimone, 2009; Wayne et al., 2008). The frameworks developed for Papers II and V respond to this call. These frameworks can function as valuable tools for researchers, PD providers, and PD designers when determining and/or assessing PD programs' Content Focus, Coherence, and methods for facilitating enactment.

Considering the overall *theoretical* contribution, the thesis shows that Desimone's (2009) framework is useful for characterizing teacher PD programs, supports the design of studies aiming to investigate the particular effects of PD programs, and aids the interpretation of the results. However, as discussed in the kappa and the papers, and as noted by other scholars (Kennedy, 2016; Opfer & Pedder, 2011) and Desimone herself (Desimone & Garet, 2015), the framework also has its limitations. For example, the results illustrate that the framework is not always helpful in detecting the finer points of details between different PD programs and explaining why the same program may have a different impact on different teachers and in various contexts. Moreover, although it has been stated that there is consensus on the five core critical features (Desimone, 2009; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Russell, Kleiman, Carey, & Douglas, 2009), the results from the thesis show that this consensus appears to be illusionary. First of all, throughout the research literature, the features are not always defined in a consistent way. Secondly, even PD programs incorporating all the features have shown to have notorious difficulties in improving student achievement. Guided by the results from the papers, as well as previous research, this thesis has discussed and suggested the development of parts of Desimone's (2009) framework for studying teacher PD. It has thereby contributed to the ongoing efforts and discussions (cf. Desimone & Garet, 2015; Kennedy, 2016; Schoenfeld, 2015; Sztajn et al., 2009) concerning developing a more refined framework to better support both practice and future research.

8.3 Implications and future research

The results of this thesis and the different conceptualizations of critical features of teacher PD on which they have a bearing raise several issues regarding implications for practice and future research. These issues can be summarized in four recommendations.

The first recommendation for researchers within the field of teacher PD is: *Be explicit in how Content Focus and Coherence are defined in the study.* As shown and discussed in this thesis, these two features are broadly and/or vaguely defined in the research literature. This, in turn, makes it hard for politicians, developers, principals and teachers to make practical use of research findings when designing PD programs. Note that I do not advocate any single view of the features as better than another; they all contribute to important insights about teacher PD. What I do question is that the same concepts are used to refer to different, and sometimes even inconsistent, aspects of high-quality PD. Such an inconsistency of use also makes it complicated for researchers to build on each other's studies and thereby assemble a larger picture of what constitutes high-quality teacher PD.

The second recommendation strengthens the already mentioned argument that researchers and others involved in the measurement of teacher PD should explicitly *Adhere to variations*. The results show, for example, that no specific PD content is more valuable than another in a one-size-fits-all manner. Instead, different PD programs work to varying degrees for different teachers depending on their previous knowledge and beliefs and the contextual factors in which they are working. Moreover, without the separate analyses conducted for the different grade levels in Papers I–IV, the impacts on CoV and BfM on teachers, instructional practices and student achievement would often have been blurred. Future studies of PD that adhere to variations between teachers and contexts are therefore needed in order to shed further light on questions related not only to "what works", but also to when, how and for whom. Thus, in order to understand how to better design and target PD so that it is effective for as many teachers as possible, it is important that we have a richer understanding of the variation both between and within programs. For which groups of teachers is the PD effective? Why do certain groups of teachers have higher implementation than others? How can we expand the range in which the PD works?

The third recommendation is that, in determining the content and methods for facilitating teachers' enactment of the promoted practices, one should *assess the Coherence between the PD program and teachers' initial knowledge and instructional practices before the program starts*. As

discussed in Section 7.2.2, this recommendation is supported by the results presented in the thesis. The argument regarding assessing Coherence with teachers' content knowledge is further strengthened by the empirical results from Covay Minor et al. (2016), while the argument regarding Coherence with teachers' initial practices is supported by the empirical results from Penuel et al. (2012). Still, to be able to determine PD programs' Coherence with teacher knowledge and practice, methodological tools for making the evaluations are needed. The frameworks presented in Papers II and V can be seen as primary efforts to develop such tools. Future studies that examine PD programs in different contexts, and that collect observational and interview data from teachers are needed, however, to refine and further develop the frameworks. In addition, more empirical studies that could strengthen the given recommendation ought to be conducted. Future studies might, for example, compare the impacts on groups of teachers participating in the same PD program with different levels of coherence between the teachers' current knowledge and practices on the one hand and the knowledge and practices promoted in the PD on the other.

The final recommendation relates to the nature of the changes requested, and it is recommended that designers and providers of teacher PD *Tailor the support to match the requested changes*. PD programs focusing on supporting already existing practices or changing teachers' procedural routines are easier to come by than programs striving to reorganize practice and accomplish more difficult-to-reach dimensions of teaching. At the same time, at least in the Swedish educational context, substantive changes are likely needed if one wishes to achieve an instruction that offers students the opportunity to develop all the mathematical competencies mentioned in the national curriculum. As demonstrated in Gregoire's (2003) model of conceptual change, such change is dependent not only on teachers' Personal Domain but also on the External Domain in terms of the support they have in the forms of time, tools, colleagues, and school leadership. The results presented in this thesis have shown that the cases in which BfM or CoV did not lead to improvements in student achievement were also those in which teachers felt they had the least support, for example in terms of time to conduct the changes. If these teachers had been given more support, the results may have been different. However, it is probably naïve to expect that the PD program itself will provide teachers with all the support they need. Such a view of teacher PD is argued to underplay the complexity of the problem of instructional change, and "leads to focus on the micro context (individual teachers or individual activities or programs) to the exclusion of influences from meso (institutional) and macro (school

system) contexts" (Opfer & Pedder, 2011, p. 378f). In other words, the impacts of teacher PD are not only dependent on theories of instruction and theories of teacher change, but also on theories of schooling (cf. Hatch, 2002). For practice this implies that, on the one hand, researchers and PD providers have to be more explicit about what it takes to make the required changes, for both individual teachers and the school as a whole. On the other hand, politicians, municipality leaders, principals and teachers must develop a realistic understanding of the time and effort involved in making the required changes.

Sammanfattning på svenska

I denna avhandling har två storskaliga kompetensutvecklingsprogram (KUP) för lärare använts som fall för att bidra till kunskap om kritiska aspekter av sådana program samt deras inverkan på lärare, undervisning och elevresultat. Det första programmet, *Räkna med Västerås* (RäV), var obligatoriskt för alla kommunala grundskolelärare som undervisade i matematik i kommunen. Det andra programmet, *Matematiklyftet*, har genomförts av 76 % av alla grundskolelärare i Sverige som undervisar i matematik. Under den tidsperiod som studien genomfördes var det möjligt att göra en jämförelse mellan programmen, då det vid tillfället fanns lärare i kommunen som antingen deltagit i RäV, Matematiklyftet, eller inget program alls. Data gällande programmens inverkan på lärare, undervisning och elevresultat samlades in både direkt och ett år efter lärarnas deltagande. Med andra ord gjorde studiens kontext det möjligt att svara upp mot nyliga uppmaningar om att det behövs fler studier som studerar: (a) inverkan av KUP som genomförts i större skala, (b) KUPs inverkan på lärare, undervisning *och* elevresultat, (c) KUPs långsiktiga inverkan, och (d) variationen mellan olika KUPs inverkan. Resultaten visar att de studerade programmen är ytterst lika om de karaktäriseras enligt etablerade ramverk om vad som utgör kritiska faktorer av KUP för lärare, samtidigt som de visar på olika inverkan på lärare, undervisning och elevresultat. I ljuset av dessa resultat föreslås en utveckling av två kritiska aspekter gällande KUP för lärare: *innehållsfokus* och *samstämmighet*. Genom att utveckla ett mer finkänsligt verktyg för att karaktärisera programmens innehållsfokus upptäcktes också skillnader mellan dem. Tillsammans med en systematisk litteraturöversikt av forskningslitteraturen kring ”samstämmighet” utgjorde dessa resultat en bas för att diskutera möjliga förklaringar till skillnaderna i programmens inverkan, samt en utveckling av de kritiska aspekterna innehållsfokus och samstämmighet. Sammantaget bidrar denna avhandling med: (a) empiriska resultat i förhållande till storskaliga KUPs inverkan, (b) metodologiska resultat i form av verktyg för att karaktärisera KUPs innehållsfokus och samstämmighet, och (c) teoretiska resultat då etablerade ramverk för att karaktärisera KUPs har studerats genom praktisk användning och i ljuset av resultaten föreslås även utvecklingar av dem.

Acknowledgements

First and foremost, I could not have wished for better supervisors! My critical friend, sounding board, mentor and main supervisor Professor Andreas Ryve: He always allowed this thesis to be my own work, but challenged me to think a step further and examine things from a different point of view. Your remarkable expertise, inexhaustible positive driving force, sharing of contacts, and not the least faith in me, have made me grow as a researcher more than anything else. My statistical advisor, role model and supervisor Professor Marie Wiberg: Her statistical expertise, encouraging comments and external perspective (i.e. non-didactical) have been invaluable in my doctoral studies. Furthermore, your curiosity, professionalism and warmth have served (and still do) as a great source of inspiration. I would also like to express my gratitude to Professor Kirsti Hemmi for her valuable support as a supervisor during my first year as a doctoral student. Your energy and enthusiasm were always inspiring.

I am sincerely grateful to the 90% licentiate seminar opponent and, later, also colleague and co-author, Ola Helenius. Your theoretical and philosophical knowledge challenged me in our discussions and always kept me on my feet.

Many thanks to all of the reviewers (Peter Gustafsson, Niclas Månsson, Laila Niklasson, Wieland Wermke, Magnus Österholm) for their thorough reviews, including encouraging comments as well as helpful critique and suggestions for improvement.

The City of Västerås and Mälardalen University: Without their financial support, this dissertation would have been an impossibility. The same applies to all teachers who set aside time to answer the questionnaires.

My sincere gratitude to colleagues and co-workers within the municipality: Bodil Lövgren, Patrik Gustafsson, Per Kornhall and Majvor Mårts. Our cooperation within the CoV project always reminded me of the complex reality in the educational context. A special thank you to Anna Östman: I have really appreciated our discussions about both theoretical and practical matters. Your genuine kindness and caring thoughts have meant more than you can imagine. Moreover, all my colleagues at the university (M-TERM, VR18, the Division of Physics and

Mathematics/Natural Science with Didactics, fellow doctoral students): it has been both fun and stimulating sharing workplace with you for the last couple of years!

Last, but by no means least: my family. Dad, who read all of my articles and always cheered on me, although I doubt you grasped all of the content. Mom and grandmother, the two strongest people I know: your unconditional love and care has made me the courage and faith that I can do whatever I set my mind to if I am prepared to work hard for it. Jelena and Nelly: your genuine interest and keen eyes have helped me express the results to a broader audience. Finally, Mattias… sometimes words are not enough.

References

Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education, 30*(10), 1405–1416.

Avalos, B. (2011). Teacher professional development in Teaching and Teacher Education over ten years. *Teaching and Teacher Education, 27*(1), 10–20.

Axelsson, M. (2016). Sweden. In I. V. S. Mullis, M. O. Martin, S. Goh, & K. Cotter (Eds.), *TIMSS 2015 encyclopedia: Education policy and curriculum in mathematics and science*. Retrieved from http://timssandpirls.bc.edu/timss2015/encyclopedia/

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching what makes it special? *Journal of Teacher Education, 59*(5), 389–407.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., … Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180.

Bausmith, J. M., & Barry, C. (2011). Revisiting professional learning communities to increase college readiness: The importance of pedagogical content knowledge. *Educational Researcher, 40*(4), 175–178.

Bergqvist, E., Bergqvist, T., Boesen, J., Helenius, O., Lithner, J., Palm, T. & Palmberg, B. (2009). *Matematikutbildningens mål och undervisningens ändamålsenlighet* [The goal of mathematical education and the purposefulness of teaching practices]. Gothenburg, Sweden: National Center for Mathematics Education, Gothenburg University.

Blank, R. K., & de las Alas, N. (2009). *Effects of teacher professional development on gains in student achievement: How meta-analysis provides scientific evidence useful to education leaders*. Washington, DC: The Council of Chief State School Officers.

Boesen, J., Helenius, O., Bergqvist, E., Bergqvist, T., Lithner, J., Palm, T., & Palmberg, B. (2014). Developing mathematical competence: From the intended to the enacted curriculum. *The Journal of Mathematical Behavior, 33*, 72–87.

Boesen, J., Helenius, O., & Johansson, B. (2015). National-scale professional development in Sweden: Theory, policy, practice. *ZDM, 47*(1), 129–141.

Borko, H. (2004). Professional development and teacher learning: Mapping the Terrain. *Educational Researcher, 33*(8), 3–15.

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In P. L. Peterson, E. Baker, & B. McGaw (Eds.),

Third international encyclopedia of education (Vol. 7, pp. 548–556). Amsterdam, The Netherlands: Elsevier.

Boström, E. (2017). *Formativ bedömning: En enkel match eller en svår utmaning? Effekter av en kompetensutvecklingssatsning på lärarnas praktik och på elevernas prestationer i matematik* [Formative assessment: A piece of cake or a difficult challenge? Effects of a professional development program on teachers' classroom practice and students' achievement in mathematics] (Doctoral dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Aumu%3Adiva-135038

Britt, M. S., Irwin, K. C., & Ritchie, G. (2001). Professional conversations and professional growth. *Journal of Mathematics Teacher Education, 4*(1), 29–53.

Brozo, W. G. (2014). Investing in youth by investing in teachers: Transforming adolescent literacy through professional development. In L. E. Martin, S. Kragler, D. J. Quatroche, & K. L. Bauserman (Eds.), *Handbook of professional development in education: Successful models and practices, preK–12* (pp. 265–283). New York, NY: The Guilford Press.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Campanelli, P. (2008). Testing survey questions. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 176–200). New York, NY: Lawrence Erlbaum Associates.

Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage.

Chen, H. T. (2006). A theory-driven evaluation perspective on mixed methods research. *Research in the Schools, 13*(1), 75–83.

Chen, H. T. (2012). Theory-driven evaluation: Conceptual framework, application and advancement. In R. Strobl, O. Lobermeier, & W. Heitmeyer (Eds.), *Evaluation von programmen und projekten für eine demokratische kultur* (pp. 17–40). Wiesbaden, Germany: Springer Fachmedien Wiesbaden.

Chval, K., Abell, S., Pareja, E., Musikul, K., & Ritzka, G. (2008). Science and mathematics teachers' experiences, needs, and expectations regarding professional development. *Eurasia Journal of Mathematics, Science & Technology Education, 4*(1), 31–43.

Clark-Wilson, A., Hoyles, C., Noss, R., Vahey, P., & Roschelle, J. (2015). Scaling a technology-based innovation: Windows on the evolution of mathematics teachers' practices. *ZDM, 47*(1), 79–92.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 947–967.

Clewell, B. C., de Cohen, C. C., Campbell, P. B., & Perlman, L. (2005). *Review of evaluation studies of mathematics and science curricula and professional development models.* Washington, DC: Urban Institute.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

Cobb, P., & Jackson, K. (2011). Towards an empirically grounded theory of action for improving the quality of mathematics teaching at scale. *Mathematics Teacher Education and Development, 13*(1), 6–33.

Cobb, P., & Jackson, K. (2012). Analyzing educational policies: A learning design perspective. *Journal of the Learning Sciences, 21*(4), 487–521.

Cobb, P., Jackson, K., Smith, T., Sorum, M., & Henrick, E. (2013). Design research with educational systems: Investigating and supporting improvements in the quality of mathematics teaching and learning at scale. *National Society for the Study of Education Yearbook, 112*(2), 320–349.

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*(2), 294–343.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7 ed.). London, UK: Routledge.

Cook, S. D., & Brown, J. S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization science, 10*(4), 381–400.

Covay Minor, E., Desimone, L. M., Lee, J. C., & Hochberg, E. D. (2016). Insights on how to shape teacher learning policy: The role of teacher content knowledge in explaining differential effects of professional development. *Education Policy Analysis Archives, 24*(61), 1–34.

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 76*(8), 597–604.

De Winter, J. C., & Dodou, D. (2010). Five-point likert items: t test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation, 15*(11), 1–12.

Dede, C., Jass Ketelhut, D., Whitehouse, P., Breit, L., & McCloskey, E. M. (2008). A research agenda for online teacher professional development. *Journal of Teacher Education, 60*(1), 8–19.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34*, 12–25.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199.

Desimone, L. M. (2011). Outcomes: Content-focused learning improves teacher practice and student results. *Journal of Staff Development, 32*(4), 63–69.

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society, & Education, 7*(3), 252–263.

Desimone, L. M., & Stuckey, D. (2014). Sustaining teacher professional development. In L. E. Martin, S. Kragler, D. J. Quatroche, & K. L. Bauserman (Eds.), *Handbook of professional development in education: Successful models and practices, preK–12*. New York, NY: The Guilford Press.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal, 50*(1), 25–32.

Emanuelsson, G. (2001). *Svårt att lära - lätt att undervisa? Om kompetensutvecklingsinsatser för lärare i matematik 1965–2000* [Difficult to learn – easy to teach? About professional development initiatives for mathematics teachers 1965–2000]. Gothenburg, Sweden: National Center for Mathematics Education, Gothenburg University.

Emanuelsson, G., Johansson, B., Reys, B., & Reys, R. (1996). Using a journal to engage teachers in developmental work. *Nordic Studies in Mathematics Education, 4*(4), 61–73.

Firestone, W. A., & Mangin, M. M. (2014). Leading professional learning in districts with a student learning culture. In L. E. Martin, S. Kragler, D. J. Quatroche, & K. L. Bauserman (Eds.), *Handbook of professional development in education: Successful models and practices, preK–12* (pp. 319–338). New York, NY: The Guilford Press.

Firestone, W. A., Mangin, M. M., Cecilia Martinez, M., & Polovsky, T. (2005). Leading coherent professional development: A comparison of three districts. *Educational Administration Quarterly, 41*(3), 413–448.

Fishman, B. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*(6), 643–658.

Fuhrman, S. H. (1993). *Designing coherent education policy: Improving the system.* San Francisco, CA: Jossey-Bass.

Gall, M. D., & Vojtek, R. O. (1994). *Planning for effective staff development. Six research-based models*. Eugene, OR: ERIC Clearinghouse on Educational Management

Garet, M. S., Heppen, J., Walters, K., Smith, T., & Yang, R. (2016). *Does content-focused teacher professional development work? Findings from three Institute of Education Sciences studies*. Retrieved from https://ies.ed.gov/ncee/pubs/20174010/pdf/20174010.pdf

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597–606.

Goldsmith, L. T., Doerr, H. M., & Lewis, C. C. (2013). Mathematics teachers' learning: A conceptual framework and synthesis of research. *Journal of Mathematics Teacher Education, 17*(1), 5–36.

Gough, D., Oliver, S., & Thomas, J. (2013) *Learning from research: Systematic reviews for informing policy decisions: A quick guide.* A paper for the Alliance for Useful Evidence. London, UK: Nesta.

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review, 15*(2), 147–179.

Guskey, T. R. (2002a). Does it make a difference? *Educational leadership, 59*(6), 45–51.

Guskey, T. R. (2002b). Professional development and teacher change. *Teachers and Teaching, 8*(3), 381–391.

Guskey, T. R. (2003). Analyzing lists of the characteristics of effective professional development to promote visionary leadership. *NASSP Bulletin, 87*(637), 4–20.
Guskey, T. R. (2014a). Measuring effectiveness of professional development. In L. E. Martin, S. Kragler, D. J. Quatroche, & K. L. Bauserman (Eds.), *Handbook of professional development in education: Successful models and practices, preK–12* (pp. 447–466). New York, NY: The Guilford Press.
Guskey, T. R. (2014b). Planning professional learning. *Educational leadership, 71*(8), 10–16.
Hajer, M., & Norén, E. (2017). Teachers' knowledge about language in mathematics professional development courses: From an intended curriculum to a curriculum in action. *Eurasia Journal of Mathematics, Science & Technology Education, 13*(7b), 4087–4114.
Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis, 25*(1), 1–29.
Hammerness, K. (2006). From coherence in theory to coherence in practice. *Teachers College Record, 108*(7), 1241–1265.
Harkness, J. A. (2008). Comparative survey research: Goal and challenges. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of survey methodology* (pp. 56–77). New York, NY: Lawrence Erlbaum Associates.
Hatch, T. (2002). When improvement programs collide. *Phi Delta Kappan, 83*(8), 626–639.
Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177.
Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher, 42*(9), 476–487.
Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher, 33*(8), 16–30.
Hox, J. J. (2008). Accommodating measurement errors. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 387–402). New York, NY: Lawrence Erlbaum Associates.
Hox, J. J., de Leeuw, E. D., & Dillman, D. A. (2008). The cornerstones of survey research. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 1–17). New York, NY: Lawrence Erlbaum Associates.
Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *Education Policy Analysis Archives, 13*(10), 1–28.
Jacob, R., Hill, H., & Corey, D. (2017). The impact of a professional development program on teachers' mathematical knowledge for teaching, instruction, and student achievement. *Journal of Research on Educational Effectiveness, 10*(2), 379–407.
Kaufmann, O. T. (2017). *Teachers' attention to student thinking, mathematical content, teachers' role and lesson management*. Paper presented at the 2017

NORMA Conference. Retrieved September, 2017 from http://www.mnd.su.se/om-oss/evenemang/norma-17/programme/paper-session-3/kaufmann

Kazemi, E., & Hubbard, A. (2008). New directions for the design and study of professional development: Attending to the coevolution of teachers' participation across contexts. *Journal of Teacher Education, 59*(5), 428–441.

Kennedy, A. (2005). Models of continuing professional development: A framework for analysis. *Journal of In-service Education, 31*(2), 235–250.

Kennedy, M. M. (1998). *Form and substance in inservice teacher education (Research monograph No. 13).* Madison, WI: National Institute for Science Education, University of Wisconsin, Madison.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980.

Kirsten, N., & Wermke, W. (2017). Governing teachers by professional development: State programmes for continuing professional development in Sweden since 1991. *Journal of Curriculum Studies, 49*(3), 391–411.

Klinger, D. A., Volante, L., & Deluca, C. (2012). Building teacher capacity within the evolving assessment culture in Canadian education. *Policy Futures in Education, 10*(4), 447–460.

Knapp, M. S. (2003). Professional development as a policy pathway. *Review of research in education, 27*(1), 109–157.

Kriewaldt, J. (2008). *Research into relationships between teacher professional learning and teaching standards: Reviewing the literature*. Paper presented at the 2008 AARE Conference. Retrieved March 5, 2017 from http://www.aare.edu.au/data/publications/2008/kri08759.pdf

Larsson, M. (2015). *Orchestrating mathematical whole-class discussions in the problem-solving classroom: Theorizing challenges and support for teachers* (Doctoral dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Amdh%3Adiva-29409

Lindvall, J. (2016). *Supporting instructional improvement at scale – The role of teacher professional development programs and curriculum materials* (Licentiate thesis). Retrieved from http://www.diva-portal.org/smash/get/diva2:921655/FULLTEXT02.pdf

Linn, R., & Slinde, J. (1977). The determination of the significance of change between pre- and post-testing periods. *Review of Educational Research, 47*(1), 121–150.

Loucks-Horsley, S., Stiles, K., Mundry, S., Love, N., & Hewson, P. (2009). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Sage Publications.

Lundahl, C., & Waldow, F. (2009). Standardisation and 'quick languages': The shape-shifting of standardised measurement of pupil achievement in Sweden and Germany. *Comparative Education, 45*(3), 365–385.

Malzahn, K. A. (2002). *The 2000 national survey of science and mathematics education: Status of elementary school mathematics teaching.* Chapel Hill, NC: Horizon Research, Inc.

Marrongelle, K., Sztajn, P., & Smith, M. (2013). Scaling up professional development in an era of common state standards. *Journal of Teacher Education, 64*(3), 202–211.

Martín, E. (2015). Pathways that converge in teacher professional development: Are they present in Spain? *Psychology, Society and Education, 7*(3), 327–342.
Martin, M. O., Mullis, I. V. S., & Hooper, M. (Eds.). (2016). *Methods and procedures in TIMSS 2015*. Retrieved from http://timssandpirls.bc.edu/publications/timss/2015-methods.html
Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*(1), 29–45.
Mayer, D., & Lloyd, M. (2011). *Professional learning: An introduction to the research literature*. Melbourne, Australia: Australian Institute for Teaching and School Leadership.
McIntosh, A. (2008). *Förstå och använda tal: En handbok* [Understand and use numbers: A handbook]. Gothenburg, Sweden: National Center for Mathematics Education.
McIntosh, A., Reys, B. J., Reys, R. E., Bana, J. & Farrell, B. (1997). *Number sense in school mathematics: Student performance in four countries* (MASTEC Monograph Series No. 5). Perth, Western Australia: Edith Cowan University.
McKenney, S. E., & Reeves, T. C. (2012). *Conducting educational design research*. New York, NY: Routledge.
McNeill, K. L., & Knight, A. M. (2013). Teachers' pedagogical content knowledge of scientific argumentation: The impact of professional development on k–12 teachers. *Science Education, 97*(6), 936–972.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.
Mevarech, Z. R. (1995). Teachers' paths on the way to and from the professional development forum. In T. R. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 151–170). New York, NY: Teachers College Press.
Middle-school Mathematics and the Institutional Setting of Teaching (n.d). *Mathematics and the institutional setting of teaching. Vanderbilt University. Teacher Survey*. Retrieved from http://peabody.vanderbilt.edu/docs/pdf/tl/Generic_Teacher_Survey_2010_Plus_Sources_100803.pdf.
Moore, D. S., McCabe, G. P., & Craig, B. A. (2009). *Introduction to the practice of statistics*. New York, NY: W. H. Freeman.
Morawski, J. (2010). *Mellan frihet och kontroll: Om läroplanskonstruktioner i svensk skola* [Between freedom and control: Constructions of curricula in Swedish schools] (Doctoral dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Adu-6366
Muller, J. (2009). Forms of knowledge and curriculum coherence. *Journal of Education and Work, 22*(3), 205–226.
National Research Council (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press
Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*(5), 625–632.
OECD (2015). *Improving schools in Sweden: An OECD perspective*. Paris, France: OECD Publishing

OECD (2016). *PISA 2015 results (Volume II): Policies and practices for successful schools*. Paris, France: OECD Publishing.

Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research, 81*(3), 376–407.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44*(4), 921–958.

Penuel, W. R., Sun, M., Frank, K. A., & Gallagher, H. A. (2012). Using social network analysis to study how collegial interactions can augment teacher learning from external professional development. *American Journal of Education, 119*(1), 103–136.

Petrou, M., & Goulding, M. (2011). Conceptualising teachers' mathematical knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 9–25). Dordrecht, the Netherlands: Springer.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell Publishing.

Prytz, J. (2015). Swedish mathematics curricula, 1850–2014. An overview. In K. Bjarnadóttir, F. Furinghetti, J. Prytz, & G. Schubring (Eds.), *"Dig where you stand" 3. Proceedings of the third international conference on the history of mathematics education* (pp. 309–325). Uppsala, Sweden: Uppsala University.

Ramböll (2014). *Delutvärdering: Matematiklyftets första år* [Intermediate evaluation: The Boost for Mathematics' first year]. Retrieved March 3, 2017, from https://www.skolverket.se/om-skolverket/publikationer/visa-enskild-publikation?_xurl_=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fskolbok%2Fwpubext%2Fbilaga%2FBlob%2Fpdf599.pdf%3Fk%3D599

Ramböll (2016). *Slututvärdering: Utvärderingen av Matematiklyftet 2013–2016* [Final evaluation: Evaluation of the Boost for Mathematics 2013–2016]. Retrieved March 3, 2017, from http://www.skolverket.se/publikationer?id=3705

Reys, R., Reys, B., Emanuelsson, G., Johansson, B., McIntosh, A., & Yang, D. C. (1999). Assessing number sense of students in Australia, Sweden, Taiwan, and the United States. *School Science and Mathematics, 99*(2), 61–70.

Robinson, V. M., & Timperley, H. S. (2007). The leadership of the improvement of teaching and learning: Lessons from initiatives with positive outcomes for students. *Australian Journal of Education, 51*(3), 247–262.

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching, 48*(2), 117–148.

Russell, M., Kleiman, G., Carey, R., & Douglas, J. (2009). Comparing self-paced and cohort-based online courses for teachers. *Journal of Research on Technology in Education, 41*(4), 443–466.

Ryve, A., Hemmi, K., & Kornhall, P. (2016). *Skola på vetenskaplig grund* [School on a scientific basis]. Stockholm: Natur & Kultur.

Salinas, A. (2010). *Investing in our teachers: What focus of professional development leads to the highest student gains in mathematics achievement?* (Doctoral Dissertation). Retrieved from http://scholarlyrepository.miami.edu/oa_dissertations/393

Santagata, R. (2009). Designing video-based professional development for mathematics teachers in low-performing schools. *Journal of Teacher Education, 60*(1), 38–51.

Santagata, R., Kersting, N., Givvin, K. B., & Stigler, J. W. (2010). Problem implementation as a lever for change: An experimental study of the effects of a professional development program on students' mathematics learning. *Journal of Research on Educational Effectiveness, 4*(1), 1–24.

Scheerens, J. (Ed.) (2010). *Teachers' professional development: Europe in international comparison. An analysis of teachers' professional development based on the OECD's teaching and learning international survey (TALIS*). Luxembourg: Office for Official Publications of the European Union.

Scher, L., & O'Reilly, F. (2009). Professional development for k–12 math and science teachers: What do we really know? *Journal of Research on Educational Effectiveness, 2*(3), 209–249.

Schmidt, W. H., Wang, H. C., & McKnight, C. C. (2005). Curriculum coherence: An examination of US mathematics and science content standards from an international perspective. *Journal of Curriculum Studies, 37*(5), 525–559.

Schoenfeld, A. H. (2000). Purposes and methods of research in mathematics education. *Notices of the American Mathematical Society, 47*(6), 641–649.

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? A story of research and practice, productively intertwined. *Educational Researcher, 43*(8), 404–412.

Schoenfeld, A. H. (2015). Thoughts on scale. *ZDM, 47*(1), 161–169.

Schwarz, N., Knäuper, B., Oyserman, D., & Stich, C. (2008). The psychology of asking questions. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 18–34). New York, NY: Lawrence Erlbaum Associates.

SFS 2003:460. *Lag (2003:460) om etikprövning av forskning som avser människor* [Law (2003:460) on the ethical review of research involving humans]. Retrieved from http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/lag-2003460-om-etikprovning-av-forskning-som_sfs-2003-460

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–23.

Singer, E. (2008). Ethics issues in surveys. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 78–96). New York, NY: Lawrence Erlbaum Associates.

Sink, C., & Stroh, H. (2006). Practical significance: The use of effect sizes in school counseling research. *Professional School Counseling, 9*(4), 401–411.

Skolinspektionen (2017). *Arbetsformer och lärarstöd i grundskolan: Resultat efter Skolinspektionens oanmälda besök på 60 grundskolor den 18 oktober 2016* [Working methods and teacher support in elementary school: Outcomes after the Schools Inspectorate's unannounced visits to 60 elementary schools on October 18, 2016]. Retrieved from https://www.skolinspektionen.se/globalassets/publikationssok/granskningsrapporter/flygande-inspektioner/2017/rapport-arbetsfo-lararst-grsk-2017-mars.pdf

Skolverket (2006). *Curriculum for the compulsory school system, the pre-school class and the leisure-time centre Lpo 94*. Stockholm, Sweden: Skolverket.
Skolverket (2011a). *Curriculum for the compulsory school, preschool class and the recreation centre, 2011*. Stockholm: Skolverket.
Skolverket (2011b). *Delredovisning av uppdrag om att stärka undervisningen i matematik, naturvetenskap och teknik* [Interim report of the commission to strengthen teaching in mathematics, natural science and technology]. Dnr 2011:643. Retrieved from https://www.skolverket.se/publikationer?id=2651
Skolverket (2011c). *Kommentarmaterial till kursplanen i matematik* [Commentary material to the mathematics curriculum]. Stockholm, Sweden: Skolverket.
Skolverket (2012). *Tid för matematik — Erfarenheter från Matematiksatsningen 2009–2011* [Time for mathematics — Experiences from the Mathematics Initiative 2009–2011]. Stockholm, Sweden: Skolverket.
Skolverket (2016a). *Slutredovisning: Uppdrag att svara för utbildning* [Final report: Commission to provide professional development]. Dnr: 2011:643. Retrieved from https://www.skolverket.se/publikationer?id=3701
Skolverket (2016b). *TIMSS 2015. Svenska grundskoleelevers kunskaper i matematik och naturvetenskap i ett internationellt perspektiv* [TIMSS 2015. Swedish elementary school students' knowledge of mathematics and natural sciences in an international perspective]. Retrieved from https://www.skolverket.se/publikationer?id=3707
Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427–515.
Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research, 79*(2), 839–911.
Smith, M. S., & Stein, M. K. (2011). *5 practices for orchestrating productive mathematics discussions*. Reston, VA: National Council of Teachers of Mathematics.
Soine, K. M., & Lumpe, A. (2014). Measuring characteristics of teacher professional development. *Teacher Development, 18*(3), 303–333.
Sowder, J. T. (2007). The mathematical education and development of teachers. In F. K. Lester (Ed.), *Second handbook on research of mathematics teaching and learning* (pp. 157–223). Reston, VA: National Council of Teachers of Mathematics.
Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies, 31*(2), 143–175.
Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis, 21*(1), 1–27.
Stein, M. K., & Kaufman, J. H. (2010). Selecting and supporting the use of mathematics curricula at scale. *American Educational Research Journal, 47*(3), 663–693.
Stutchbury, K., & Fox, A. (2009). Ethics in educational research: Introducing a methodological tool for effective ethical analysis. *Cambridge Journal of Education, 39*(4), 489–504.

Swan, M. (2006). Designing and using research instruments to describe the beliefs and practices of mathematics teachers. *Research in Education 75*(1), 58–70

Swedish Ministry of Education (2012). *Regeringsbeslut* [Cabinet decision] I:44. Stockholm, Sweden: Ministry of Education.

Swedish Research Council (2011). *Good research practice*. Stockholm, Sweden: Swedish Research Council.

Sztajn, P., Campbell, M., & Yoon, K. S. (2009). Conceptualizing professional development in mathematics: Elements of a model. In M. Tzekaki, M. Kaldrimidou, & H. Sakonidis (Eds.), *Proceedings of 33rd annual conference of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 209–216). Thessaloniki, Greece: PME.

Tatto, M. T. (1996). Examining values and beliefs about teaching diverse students: Understanding the challenges for teacher education. *Educational Evaluation and Policy Analysis, 18*(2), 155-180.

Tatto, M. T., Lerman, S., & Novotná, J. (2009). Overview of teacher education systems across the world. In R. Even & D. L. Ball (Eds.), *The professional education and development of teachers of mathematics* (pp. 15–23). New York, NY: Springer.

Tengstrand, A. (2010). *Kommunala matematikutvecklare: En av förutsättningarna för en hållbar utveckling av svensk matematikundervisning* [Municipal mathematics developers: One of the prerequisites for a sustainable development of Swedish mathematics education]. Gothenburg, Sweden: National Center for Mathematics Education.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration [BES]*. Wellington, NZ: Ministry of Education.

U.S. Department of Education (2014). *Fiscal year 2014 budget summary and background information*. Washington, DC: U.S. Department of Education.

Van Driel, J. H., & Berry, A. (2012). Teacher professional development focusing on pedagogical content knowledge. *Educational Researcher, 41*(1), 26–28.

Van Steenbrugge, H., Larsson, M., Insulander, E., & Ryve, A. (2017). *Curriculum support for teachers' negotiation of meaning: A collective perspective*. Manuscript submitted for publication.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher, 37*(8), 469–479.

Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Review of Educational Research, 79*(2), 702–739.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*(1), 80–91.

Wiliam, D. (2007). Keeping learning on track. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1053–1098). Reston, VA: National Council of Teachers of Mathematics.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Villegas-Reimers, E. (2003). *Teacher professional development: An international review of the literature*. Paris, France: International Institute for Educational Planning.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education, 24*(1), 173–209.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Österholm, M., Bergqvist, T., Liljekvist, Y., & van Bommel, J. (2016). *Utvärdering av Matematiklyftets resultat: Slutrapport* [Evaluation of the Boost for Mathematics' results: Final report]. Retrieved March 3, 2017, from http://www.skolverket.se/publikationer?id=3706

Address: P.O. Box 883, SE-721 23 Västerås. Sweden
Address: P.O. Box 325, SE-631 05 Eskilstuna. Sweden
E-mail: info@mdh.se **Web:** www.mdh.se

ISBN 978-91-7485-364-3
ISSN 1651-4238