Chapter 6

General discussion
In this chapter we analyze and discuss the four empirical studies that were presented in this dissertation, thereby answering our main research question: \textit{To what extent can a professional development program on data use, containing the three interrelated, evidence-informed components performance goals, data-analysis and instruction, enhance teacher change as well as students' mathematical achievement?}

This general discussion first considers the context and research aims of our study. After that, the main findings of the four empirical studies are recapitulated, integrated and related to the research question. Next, possible explanations for not finding proximal and distal outcomes, the study design, and issues concerning measurement and constructs are discussed. At the end of the chapter implications for future research and practice are addressed.

\textbf{6.1 Context and research aims}

For some years, there have been concerns about the level of mathematical proficiency of Dutch students. Some slight decreases took place over the last 15-20 years (Meelissen & Drent, 2008; Royal Netherlands Academy of Arts and Sciences, 2009; Scheltens et al., 2013). Further, the relatively low number of advanced or excellent students raises questions on whether these students are optimally challenged by their teachers. Given the importance of (early acquired) mathematical competence for students' future performance and opportunities (Aubrey et al., 2006; Claessens et al., 2009; Duncan et al., 2007; Reyna & Brainerd, 2007; Ritchie & Bates, 2013), these concerns are considered relevant, leading to national initiatives that attempt to enhance students' math proficiency.

Such recent initiatives have often been framed in the context of data-driven decision making (DDDM) or data use, as was done in the current study. Data use can be understood as “systematically analyzing existing data sources within the school, applying outcomes of analyses to innovate teaching, curricula, and school performance, and implementing (...) and evaluating these innovations” (Schildkamp & Kuiper, 2010, p. 482). This definition reflects a broad perception of data use as opposed to a narrower perception in which data are merely used for accountability purposes. Schildkamp and Kuiper’s broad conceptualization, in which data are used in a formative way to improve future teaching, was adopted in this dissertation and in the professional development program (PDP) on data use that we developed, conducted and evaluated.

The main aim of this dissertation was to evaluate a PDP on data use in terms of both its proximal (changes in teacher attitudes and teaching practices) and distal
outcomes (improved math achievement of the students). This PDP contained three components that reflect the core features of data use (Black & William, 2009; Visscher & Ehren, 2011): 1. deciding on the desired proficiency level by setting performance standards and goals, 2. establishing the actual proficiency level by using data-analyses and conducting diagnostic math interviews, 3. modifying instruction, using evidence-informed instructional methods (direct instruction and modelling) such that the gap between the desired and actual level can be closed. For an elaborate overview of the PDP, the reader is referred to Appendix 1.

The PDP was expected to foster teachers’ pedagogical data-literacy, meaning that teachers would be able to adequately interpret data and use this information to base meaningful instructional decisions and actions on (Mandinach, 2012). Consequently, changed teacher attitudes and teaching practices were assumed, resulting in improved math achievement of the students. We further expected positive effects given the empirical link of the three components to student outcomes (Locke & Latham, 2002; Fuchs et al, 1985; Lai et al., 2009; Timperley & Parr, 2009; Borman et al., 2003; d’Agostino, 2000; van Gog, 2013), and given the adherence of the PDP to the five critical program features of effective PD (Desimone, 2009) – focus on content, active or inquiry-based learning, coherence with educational policy, sufficient duration, and the use of collaboration.

6.2 Main findings

In the first study teachers were asked to set well-considered goals. Such well-considered goals were expected to be promoted by a step-by-step procedure in which initial teacher expectations were contrasted to performance data and team input. As such, the accuracy of teachers’ performance goals was assumed to improve, as these goals would be based on a more refined picture of students’ capacities (Good & Brophy, 2003). In this way, we expected that using a step-by-step procedure would reduce negative expectation bias or even promote some positive expectation bias.

The strong general relationship between (high) performance goals and student outcomes (with an effect size of $d=.80$) was further elaborated on by studying a) whether and how the step-by-step procedure affected the final goals and b) whether positive changes (higher final goals than the initial expectations) were related to students’ achievement. The teachers kept to their initial expectation for most students, but for more than a quarter of the students the step-by-step procedure resulted in a positive change. Investigating the relationship between these positive changes and student
achievement was assumed to do more justice to the various student characteristics that
teachers probably took into account (Martínez et al., 2009; Slavin, 1990). No general
relationship between the positive changes and student achievement was found, but they
did seem to foster better outcomes for the initially low-performers, whereas the initially
high-performers did not seem to profit from the positive changes: They tended to
perform worse than comparable high-performers for whom the initial expectation and
the final goals were the same. That the initially weak students profited from the positive
changes could either be explained by assuming that teachers had set more accurate goals
for them (reflecting decreased negative expectation bias) or by assuming that teachers
actually became more ambitious for them (reflecting increased positive expectation bias).
For the initially high-performing students teachers might also have had such increased
positive expectation bias, but given the negative relation between the positive change and
these students’ achievement, it was questioned whether teachers acted upon such
positive expectation biases.

In the second study we wanted to understand to what extent teachers who
participated in our PDP used differentiation practices in their lessons. Forty-three
teachers were observed during their math and reading comprehension lessons, using a
low-inference observation instrument. First, preconditions for differentiation, classroom
management and the use of extended instruction were studied, thereby investigating
whether teachers’ differentiation practices could be related to context factors like the
subject domain, the heterogeneity of the class, and the type of class (multi- or single-
grade). These context factors were found to play a role in the way teachers adapted
teaching to various students, confirming our idea that differentiation cannot be
straightforwardly studied: several class features should not be overlooked. Further, we
found that in a considerable number of classes teachers did not use small-group
instruction for either the weak or the advanced students. This might be considered a
missed opportunity, since during such additional instruction teachers could address
specific root problems, knowledge gaps, or misconceptions, or challenge the advanced
students in approaching more difficult tasks.

Next, we studied in more detail how teachers dealt with students of four different
achievement levels (the very weak, weak, average and advanced students) and found
that the very weak students were most often addressed by the teachers, mainly by means
of content-related utterances. Teachers did not tend to provide additional instructional
guidance to the advanced students and the length of the whole-class teaching was not shortened for them.

In the third study we investigated whether the PDP influenced the mathematical achievement of the students. For this purpose we compared the math achievement of students who were taught by a teacher who participated in our PDP to similar students who were taught by a teacher who did not participate in such a PDP. In order to optimize our impact analyses, we conducted two matching procedures, using propensity score matching (PSM). First, for reasons of sample preservation, we matched the students based on student and class characteristics. After that, we wanted to check the robustness of these findings by not only matching on student and class characteristics but also on teacher characteristics. In this way potential systematic differences between treated and control teachers were also controlled for. Both matching procedures showed that our PDP was not related to students’ math achievement. No differential effects were found either.

Having found no effects on the student outcomes, we attempted to identify whether the PDP influenced proximal outcomes in the fourth study. For this purpose we investigated the attitudes, the self-reported behavior and the observed teaching practices of 33 teachers who participated in our PDP. Using survey data that were filled in at the end of the PDP, we first compared teacher attitudes and self-reported behavior of the treated teachers to those of 31 control teachers. We found that both groups reported rather similar attitudes and behavior, except for the goal-directness and the focus on basic skills: the treated teachers had higher scores on these features.

Next, teacher profiles in terms of teachers’ data-mindedness were identified. Using latent class cluster analysis, we identified teachers who showed a higher level of data-mindedness and those who did so to a lesser extent. Using these teacher profiles, we then found that the teachers in the two profiles already differed at baseline measurement: the teachers who showed more data-mindedness at the end of the PDP already showed significantly more awareness of student outcomes and had better teaching practices at the start of the PDP. Investigating whether the teachers had changed during the PDP, showed that not many changes had taken place, except for a better acquaintance with performance standards (in both clusters) and greater use of explicit goals for students and a better overall lesson quality by the teachers who showed more data-mindedness. Implementation of the PDP by both clusters of teachers thus
seemed to be limited. Finally, we tried to identify whether students’ math achievement was related to the teacher profiles. Multilevel analyses, while controlling for relevant covariates, showed that the math achievement of students who were taught by the teachers who showed more data-mindedness did not differ from that of students whose teachers showed less data-mindedness.

Relating the studies to the general research question

The main research question, concerning the extent to which our PDP could enhance teaching practice and students’ math achievement, can be answered by the results of the four empirical studies. Apparently, the findings of study 3 and 4 showed that our PDP did not succeed in bringing about substantial changes in student outcomes, nor in teachers’ thinking and behavior. Before interpreting these findings in the next subsection, we first discuss how the findings of the 1st and 2nd study relate to the general research question.

In order to validate the PDP’s assumption that high teacher-set performance goals were indeed related to better student outcomes (Fuchs et al., 1985; Fuchs et al., 1989), we sought to investigate this goal-performance relationship. This would empirically justify the central role of such goals in our PDP. We found a strong general relationship between high performance goals and better student performance. However, when addressing changes in teacher goals during the goal setting procedure, more specifically by focusing on positive changes from initial expectations to final goals, the relationship between these positive changes and student achievement tended to be somewhat complex: The positive changes were related to increased achievement for the initially weak performers, but to decreased achievement for the initially high performers. We assumed that the translation of such goals to daily classroom practice inevitably would lead to the use of differentiation practices. The results on how teachers approached students of various math proficiency levels might support the first explanation, as we found that teachers did not provide additional instruction to their advanced students, whereas the very weak performers received most teacher attention.

Instructional adjustments for the (very) weak performers did, however, not lead to a differential effect in the general PD-effect analyses (study 3). The lack of finding a differential effect of the PDP might indicate that teachers did not modify their instruction in such a way that it really addressed the needs of the weak students. Moreover, in a substantial amount of classes we found no additional instructional moments for these weak students at all. It might also be the case that the treated teachers accurately
modified instruction for the weak students, but that the control teachers did the same, resulting in no differential program effects. Having had more information on teaching quality and on the quality of the additional instructional adjustments, ideally both for the treated and control teachers, would have further helped us in interpreting the findings. Nevertheless, the findings from the 1st and 2nd study correspond with the results from the TIMSS studies that there is a low spread in the math achievement of Dutch students and that the number of excellent Dutch math students is relatively low (Meelissen & Drent, 2008; Meelissen et al., 2012). In terms of our research question, studies 1 and 2 showed that (very) weak students received most teacher attention and they seemed to profit most from decreased negative or increased positive expectation bias. These findings might tentatively suggest that teachers adapted their teaching for a subgroup of their students, the weak performers, if they did so at all.

6.3 Interpreting the proximal and distal effects of the PDP

The studies 3 and 4 showed no distal and very limited proximal outcomes of our PDP. How can we explain that the PDP did not succeed in its aims, despite our positive expectations? In chapter 4, four possible explanations were put forward: the PDP 1) was effective, but we did not detect its effect, 2) was not effective yet, 3) was not effective at all, or 4) was not effective in this specific set-up. As was motivated, we consider the second and fourth explanation to be most probable. These will be reconsidered here.

Many PD programs, including programs on data use, do not succeed in bringing about (sufficient) changes in teachers and/or student outcomes (Fullan, 2007). Using Desimone’s core framework for effective PD we assume that potential distal outcomes result from preceding changes in teachers’ knowledge, skills, attitudes, and their teaching practices (Desimone, 2009). As such, expecting changes in student outcomes after a short period of time might just not be realistic. Desimone quotes in this respect Loucks-Horsley et al. (1998) who stated that "it is foolhardy to either expect or focus on measuring student learning when teachers have just begun to learn and experiment with new ideas and strategies” (p. 222). Borko (2004) stated that teacher learning is a complex and slow process, in which teacher change is found in different levels and paces. Moreover, “some elements of teachers’ knowledge and practice are more easily changed than others, e.g. it appears to be easier for teachers to incorporate strategies for eliciting students’ thinking into their teaching than to use what they hear from students to make instructional decisions” (Borko, 2004, p. 6). Hence, it seems difficult to change teachers’ behavior when it comes to translating the information coming from students (that is,
data) into meaningful adjustments. Some researchers also mentioned the perceived difficulties and complexities in terms of making teachers skilled data users through PD (Hubbard et al., 2014; Jimerson, 2014; Mandinach, 2012; Wayman & Jimerson, 2014). When considering the proximal changes of our PDP, we expected to find different levels of implementation (Borko, 2004). Yet the differences that were found between the two teacher profiles already existed, so the profiles could not be related to teacher change. Slightly optimistically we could interpret teachers’ enhanced goal-orientedness (in the strong data-mindedness-profile) as reflecting slow changes in teacher behavior.

We assume that the fourth explanation, an inadequate fit between the PDP’s set up and teacher needs, is more credible, especially in light of the findings from the joined PDP on reading comprehension, in which a small to medium effect on student achievement was found ($d=0.37$, van Kuijk et al., 2015). As was noted in the general introduction, the current study on math was part of a joint PDP on both reading and math. The treated teachers received one PDP in which both reading and math were targeted, addressing the same three components. Yet the subject-specific part was differently specified for reading and math. The math-specific part consisted of additional techniques to analyze root problems or misconceptions, thereby allowing teachers to better fit instruction to students’ needs. For math, we chose not to focus on the curriculum and to provide limited math instructional suggestions, because Dutch teachers tend to strongly stick to their curricular math textbooks. These math textbooks are rather structured, containing lesson plans using the direct instruction model, detailed learning lines and detailed recommendations for additional instruction and for variation in task-difficulty. This might be the reason why Dutch teachers generally feel confident in teaching math (Meelissen & Drent, 2008). The reading-specific part consisted of instructional as well as curricular information. In hindsight, we assume that enhancing teachers’ content knowledge and pedagogical content knowledge in reading strongly met their needs, more specifically since the curricular textbooks for reading comprehension tend to be bulky and rather unclear (Houtveen & van de Grift, 2012). Due to their awareness of limitations in their content knowledge and pedagogical content knowledge (Stoeldraijer & Vernooy, 2007), teachers were assumed to be more motivated for improving them, which probably fostered teacher learning (Timperley et al., 2007). For math there may have been less (perceived) need as regards the instructional support that was offered. Moreover, since change in multiple domains may be hard to accomplish, competition with the reading part of the program might have hampered potential improvements in math teaching. Thus although no effects were found on students’ math
achievement, the program in its essence might be effective. Hence, we will look at the content and study design in more detail in the next two sections.

**Program content**

In reflecting on the lack of distal and proximal outcomes we focused on the program’s set-up. Yet, although we had theoretical as well as empirical reasons for developing a PDP containing the three components *goal setting*, *data-analysis*, and *instruction*, our study did not specifically investigate whether the three components and their combination were effective. We only related the goal setting component to student achievement in order to seek evidence for their central role in the PDP. As our study findings might also be explained by inadequateness of the program itself, we acknowledge refraining from studying the details of our theory of action to be a limitation. Investigating the expected explicit relations between the program components (and their combination) and teacher attitudes, teaching practice, and student outcomes would have improved our understanding of the findings. We could have done this by measuring the three elements at a detailed level, by studying how the teachers actually used the data in their lesson preparation and in class. An alternative would have been to investigate and compare variations of our PDP on data use in order to understand what elements worked and which did not. The study of Heller, Daehler, Wong, Shinohara, and Miratrix (2012), in which different variations of PD in elementary science teaching were compared, provided such understanding in that it not only showed the effectiveness of raising content knowledge, but also that its integration with inquiry in student understanding seemed to be more effective than raised content knowledge alone or raised teachers’ metacognition. In our study, comparing our joint PDP with the other studies in the cluster might have provided such valuable knowledge, but was considered inappropriate given the large number of organizational and content differences between the studies that would hinder an adequate comparison.

**Study design**

The study had a pre-posttest control group design in which a comparable control group was sought using propensity score matching. By matching on student, class and teacher characteristics we assumed selection bias to be strongly reduced. Moreover, the possible pool of controls in our study on the PDP’s distal outcomes consisted of students in schools that were all involved in one of the four other PDPs on data use in our cluster of studies. It was assumed that, since all schools were willing to professionalize themselves in using data, the teachers in these schools already resembled each other to
some extent. By also matching on teacher attitudes towards data use we consider to have taken into account a relevant teacher characteristic, as teacher attitudes towards data have been related to teachers’ use of data in earlier studies (Bosker et al., 2007; Vanhoof, van Petegem, & De Maeyer, 2009). The use of a control group through matching allows the design of this quasi-experimental study to be considered rigorous (Guskey & Yoon, 2009). Yet we acknowledge that richer baseline information on teachers (for instance, their teaching quality, their actual use of data for instructional adaptations, their ability to correctly interpret data, their self-efficacy, or their willingness to professionalize), on classes (e.g., their average SES, their general level of student motivation), or students (for instance, their motivation or self-confidence) would have raised the quality of the matching, as the selection of relevant variables is crucial in PSM (Brookhart et al., 2006; Steiner et al., 2010).

Our PDP was deliberately carried out in a realistic setting, so that implementation would take place in schools’ daily practice. The schools all voluntarily participated, but no full commitment of the school principal or teachers was asked for. As a result, compliance to the treatment could vary among the participants. For instance, some teachers felt obliged to participate without the intrinsic motivation to change their own practice, some principals did not take part in the meetings and lacked any active contribution, and in some schools teachers felt they did not have time to prepare their homework. The situation in the schools was assumed to mirror the diversity in schools that is found in reality, making the study ecologically valid. As such, we conducted an effectiveness trial in which the levels of implementation were allowed to vary (Shadish, Cook, & Campbell, 2002). Alternatively we could have conducted an efficacy trial in which full implementation would be studied (Shadish et al., 2002). This would have provided relevant information on the potential of the program in an optimal setting. Yet our premise for ecological validity next to the current need for PD in data use (Hubbard et al., 2014; Mandinach, 2012) led us to refrain from this choice.

Measurement and conceptual issues

In our view, one of the valuable conclusions of our study is that it brought forward the current lack of clarity on conceptual and measurement issues involving data use and data-literacy. In this subsection we first address problems in measuring teachers’ actual use of data. Next, we discuss the lack of a clear definition of the construct of data-literacy and the current lack of clarity on PD in data use. Finally, it is debated how to best identify student learning as a result of teachers’ use of data.
The use of data is expected to guide educational and instructional decisions at school, class and the individual student level, being applicable to a wide range of knowledge and skills (Hubbard et al., 2014; Means et al., 2010). It encompasses a deliberate and consistent reflection and action on the aims and objectives of all educational activities. This comprehensiveness renders easy measurement difficult. In our study we tried to measure teachers’ data use through low- and high-inference observations and teachers’ responses to questionnaire items, and we related these to student outcomes. Given the assumption that data use implies the use of goals and differentiation, we focused on teachers’ use of differentiation, their goal-orientedness, and their awareness of student outcomes. These measures were assumed to provide information on crucial changes in teaching practice and, thus, on what happened to students in class. However, in all four empirical studies we encountered a measurement problem in that we did not fully grasp teacher actions. For instance, we did not measure whether teachers used data in deciding on organization and content in their lesson planning, we did not measure qualitative information (e.g., on teacher-student interaction or quality of instruction), and we did not measure nonverbal ways in which teachers are known to express their expectations to students (Rosenthal, 1994; Rubie, 2003). By no means were all intentions, planning, and actions of the teachers measured that they may have employed, having students’ data in their minds. A mixed-methods design in which teachers’ actions are not only measured in a quantitative but also in a qualitative way might improve understanding on what types and quality of decisions and actions teachers employ. This would entail a more in-depth study, using different instruments like teacher logs, interviews, and observation instruments.

Related to the measurement issue is the lack of conceptual clarity on the construct of data literacy (Mandinach, 2012). Mandinach stated that there is no consensus amongst researchers, educators, and policymakers on what it means to be data literate. As a result, there is no clear idea on how the concept should be operationalized either. The absence of a clear notion of the desired teacher abilities and behaviors severely complicates investigating data use in itself and evaluating PD in data use. Data literacy can, in a somewhat narrow view, be understood as being able to correctly interpret data and translate them into useful information (Vanhoof et al., 2009, Downey et al., 2013). But is such a narrow definition sufficient for PDPs in data use? Indeed, correctly interpreting data is an important precondition for using data to inform educational and instructional decisions, but in our view these data-informed instructional decisions for planning and carrying out modified teaching – pedagogical data literacy - are as crucial for an effective
use of data as the ability to adequately interpret the data. Only if teachers really act upon the data, the data can have a formative function to improve teaching (see also Bennett, 2011).

In our PDP, we supposed that, by supporting teachers in analyzing and bringing together relevant data, by providing them with evidence-informed instructional methods, and by using the joint knowledge in the PLC, teachers - as the experts on teaching - would more or less automatically make the instructional adjustments needed to deal with the needs of students with a specific proficiency level, with specific root problems or with certain misconceptions. We planned to evaluate this by investigating teachers’ use of differentiation practices. The appropriateness of these indirect measures for providing information on teachers’ data use can be disputed. In any case, analyzing the data of teachers’ own students in the PD meetings did not lead to a widespread use of differentiation. Moreover, teacher attitudes and behavior hardly changed during the PDP. In short, our hypothesis that teachers would be able to make data actionable was not confirmed and several authors reported the same experiences: teachers seem to need concrete suggestions for using data to inform instructional decisions (Heritage et al., 2009; Marsh et al., 2010; Faber & Visscher, 2014; Shaw & Wayman, 2012; Timperley & Parr, 2009). Next to offering such suggestions, changes in teaching practice might also be fostered by using a more adaptive program, that is not only based on teachers’ needs, but that is also designed in consultation with the teachers. Such a collaboration would help attaining a clearly specified notion of the skills and knowledge the teachers really need to effectively use data, which would help building the knowledge base on the core features of PD in data use. Knowledge on both precise definitions and aims (when is an educator (pedagogical) data-literate?) and on the core PD components, their ideal implementation, and evaluation is needed (Schildkamp et al., 2014). This is highly relevant given the widely recognized need for PD in data use (Mandinach, 2012; Wayman & Jimerson, 2014).

A final issue that complicates the evaluation of data use and teachers’ pedagogical data-literacy is that it is not fully clear how to adequately measure the effect of teacher actions on student learning. In our study, standardized math assessment scores of the students were used to determine distal outcomes. Yet if we stress the importance of teachers acting on the data in class, it is questionable whether formal assessment scores suffice and provide the most adequate instruments to measure student learning. Of course, we expect the long-term effects of teachers’ aligned, hence improved, instructional modifications to be visible in students’ standardized assessment scores.
However, additional studies on instructional adjustments in class, based on qualitative data, could yield a more elaborate picture of the effects of data use.

6.4 Implications and recommendations for future research and practice

Apart from the conceptual and measurement issues related to PD in data use, we would like to present some implications and recommendations based on our study. These will be discussed in terms of program improvement for practice, of the initial teacher training, and of research.

Improving the set-up and content of the program

Having found no substantial proximal and no distal effects of the program, it is doubtful whether teachers felt sufficiently supported to close the gap between students’ desired and actual proficiency level through the general instructional methods that were addressed and practiced. A different focus in the instructional part might be required. Since teachers have been found to experience difficulties in translating data into meaningful actions in class (Goertz et al., 2009; Heritage et al., 2009; Means et al., 2010; Marsh et al., 2010), more detailed guidance in such translational processes seems to be needed. For PDPs on data use this might entail that, instead of a focus on general instructional methods (cf. Slavin et al., 2013), enhancing teachers’ content knowledge and pedagogical content knowledge (Bennett, 2011; Timperley & Parr, 2009), including concrete suggestions for instructional adjustments, could be more effective. A clear focus on learning how to use data to derive meaningful actions for student learning is expected to better meet the needs of teachers. For attaining such an in-depth focus on data, content and instructional adjustments, we would suggest limiting a PDP to only one subject domain.

Our study on the proximal and distal effects of the PDP assumed that both outcomes could be changed in the period of one year. This assumption was too optimistic, more specifically since, in both the PD-literature and in the literature on data use, the long-term needs for bringing about substantial changes are put forth. Therefore, it would be preferred to develop PDPs in which schools are supported at least two years (cf. Shaw & Wayman, 2012; Staman, Visscher, & Luyten, 2014), maybe even longer (Lai et al., 2009; Slavin et al., 2013). Such a long-term set-up would, for instance, allow for a profound investigation of the problems raised by the schools (Schildkamp et al., 2014), such that the participants can really act upon the problems they experience themselves, leading to enhanced motivation. It would also clear the ways for sense-making and co-
construction of meaning, making it clear for all members in a professional learning community how they interpret data and data use and what they assume the main purposes of data use to be. Accountability issues should be raised in this respect too (Hochberg & Desimone, 2010; Wayman, Spikes, & Volonnino, 2013). Such awareness of the concepts and what they entail might support teachers’ understanding on how, why and to what extent they are supposed to change their own behavior. Ideally such a long-term PDP would be followed by a less intensive trajectory in which sustained incorporation is fostered.

A final recommendation for improving the program is that more support for teachers is needed in implementing its content. Coaches can guide teachers in how they can use the knowledge coming from several data sources in planning a lesson period or specific lessons, by discussing and questioning teachers’ intentions and their decisions based on the data. Such coaching in the planning phase can be a one-to-one activity, but it could also be done in a professional learning community, thereby profiting not only from the expertise of the coach but also of the colleagues. Subsequent reflection and feedback on the teachers’ actual behavior in class is expected to make teachers more knowledgeable of their own instructional behavior and how this could be enhanced. Video-taping teachers’ behavior, combined with constructive and concrete feedback from a coach who questions and addresses their teaching practices, was found to be a promising way of altering teachers’ behavior (van den Hurk et al., 2014).

**Recommendations for initial teacher training programs**

If pedagogical data-literacy is considered an objective for teachers, then student teachers ought to be provided with suitable learning experiences in the initial teacher training program. Teacher training programs in the Netherlands seem to contain at least some aspects of data use, like learning how to use the digital student monitoring system. But data use is comprehensive in nature. As such, teacher training courses should be carefully programmed, covering and showing how data pertain to education at all levels. Student teachers should not only learn how to analyze data, but their levels of content knowledge and pedagogical content knowledge should be adequate, as well as their skills to conduct professional learning conversations (Datnow, Park, & Kennedy-Lewis, 2013; Earl & Timperley, 2009; Timperley, 2009).

We further suggest initial teacher training programs to focus more on goal setting, preferably within a team setting. We strongly advocate students’ enhanced knowledge on the importance of goals and high teacher expectations for improving student
achievement (Fuchs et al., 1985; Rubie-Davies, 2007; Scheerens & Bosker, 1997). In our study on goals we stressed the need for teachers to act upon their goals. Learning and practicing the strategies and activities that high-expectation teachers employ, like flexible grouping, creating a positive classroom climate, using goal setting and feedback, and encouraging student motivation (McDonald et al., 2014), is expected to further enhance student teachers’ future teaching quality and effectiveness.

**Recommendations for research**

According to Borman, a prescribed one-size-fits-all design is preferable in light of clarity and replicability (Borman, 2005). However, when dealing with the data on students’ (cognitive) development in schools, some adjustment of the program to the context seems to be inevitable. As McNaughton et al. (2012) stated that joint problem solving by both the participants and the researchers positively and sustainably affects student outcomes (McNaughton et al., 2012), such specific inquiry in the *(teaching)* problems of a school seems to be a promising starting point for a PDP. Yet it probably results in various needs, for which a program fit to the local context is needed. McNaughton and colleagues (2012) showed an example of such a process-oriented intervention that allowed for a contextual fit of the program. It contained both a specified process part (identifying that a program on data use was integrated with fidelity) and a content part that was contextually bound. Such a process-oriented approach can be a promising and realistic approach that needs further exploration.

A second recommendation for future research concerns the need for clarity on important constructs in data use. Not only in terms of the construct of data-literacy and its operationalization (Mandinach, 2012), as was discussed earlier, but the boundaries of data use are also unclear. Data use is considered to be comprehensive and not only related to using standardized assessments for monitoring purposes. We can consider “data use as nested in a broader context of inquiry and continuous improvement” (Jimerson, 2014, p.7). And, specifically when combining data use with other recent reforms, like project-based learning, it should be concluded that “(w)hat counts as data will become increasingly wide-ranging…. We are now not only asking teachers to use data to inform decision making, but also to use more complex forms of data and to implement new instructional strategies” (Hubbard et al., 2014, p. 54). Such broad approaches for data use might complicate a uniform definition of data use, hindering its investigation and comparisons of study findings. Our study provided an example in which this need for more conceptual knowledge was demonstrated.
Concluding remarks

The main aim of our study was to investigate the effectiveness of a PDP on data use for raising students’ math achievement. We found no evidence for its proximal and distal outcomes. Yet the study offered valuable information. It not only showed the inappropriateness of the PDP’s specific set-up for enhancing teachers’ acting on data in the domain of mathematics, but also the lack of clarity on the construct of data-literacy. This definitional issue causes the design, implementation, measurement and evaluation of PDPs on data use to be complicated, as combining different study results (and drawing conclusions from them) is not straightforward. As data use is internationally one of the spearheads of educational policy, policy makers should be aware that its evaluation results will be difficult to interpret.