Professional development in data use
Ritzema, Evelien

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 5

Teachers’ levels of implementation of a professional development program on data use and their relationship with students’ achievement in mathematics
Abstract

In this chapter it was investigated whether a professional development program on data use succeeded in bringing about changes in teachers' thinking and behavior. For this purpose we compared survey responses of the 33 teachers who had participated in the program to that of 31 control teachers. Next, we identified profiles of teachers' data-mindedness, operationalized in terms of their observed use of differentiation practices, their self-reported awareness of student outcomes and their self-reported level of goal-orientedness. Relating the teacher profiles to teachers' baseline measures and changes during the program it was found that early differences had remained and only very limited implementation had taken place. Moreover, no relation was found between the teacher profiles and students' math achievement. Possible explanations for the lack of program implementation as well as measurement issues are discussed.

5.1 Introduction

Concerns with the level of mathematics achievement due to a slight decrease in proficiency levels (Expert group Continuous Learning Progression, 2008; Royal Netherlands Academy of Arts and Sciences, 2009), have caused Dutch policymakers to focus on the basic skills and to promote the systematic use of performance data for making educational decisions. This emphasis on data-driven decision making (DDDM) is a tendency that is found in many western industrialized countries, like the USA (Datnow et al., 2007; Slavin et al., 2013; Wayman et al., 2012), New Zealand (Lai et al., 2009; Timperley & Parr, 2009), Flanders (Schildkamp et al., 2012; Verhaeghe et al., 2010), Great Britain (Downey & Kelly, 2013) and Canada (Anderson, Leithwood, & Strauss, 2010; Dunn et al., 2013). The idea is that by making use of different types of data, teachers are more aware of the proficiency levels of their students. This information supports making adequate instructional decisions that ensure better alignment between instruction and students' educational needs, reflecting a formative use of performance data. Such a better fit is supposed to lead to better student performance.

Several studies have shown that schools and teachers experience difficulties when implementing data use, or DDDM. First, the correct analyses and interpretations of data coming from digital student monitoring systems cause problems, due to insufficient levels of statistical knowledge and interpretative skills (Hellrung & Hartig, 2013; van der Kleij & Eggen, 2013; Verhaeghe et al., 2010). Second, using data in a formative way assumes that schools and teachers know the specific student outcomes they want to
attain. This assumes a goal-oriented way of working, reflecting awareness of the desired proficiency levels of students. Setting goals is considered essential in data use, but the use of explicit goals is not a common practice in the Netherlands and Flanders (Schildkamp et al., 2012; Verhaeghe et al., 2010). Third, teachers experience difficulties in translating the information provided by the data to instructional adjustments (Heritage et al., 2009; Goertz et al., 2009; Marsh et al., 2010). Given the difficulties schools and teachers experience with effectively using data, professional development in terms of DDDM is needed (Datnow et al., 2007; Mandinach, 2012; Vanhoof & Schildkamp, 2014; Wayman & Jimerson, 2014).

These considerations led us to develop and carry out a one-year teacher professional development program (PDP) in DDDM, targeted at teachers of grades 2 and 3.\(^2\) No significant effects on student achievement were found, despite the empirical evidence that the three program-elements of the PDP – goal setting, data analysis, and two effective instructional methods - were related to better student performance (e.g., Borman et al., 2003; Carlson et al., 2011; Fuchs et al., 1985; Lai et al., 2009; McNaughton et al., 2012; Muijs & Reynolds, 2011). The math performances of students that were taught by teachers who participated in the PDP did not significantly differ from those of comparable control students whose teachers were not involved in such a PDP (see chapter 4).

One could conclude that the current set-up of the PDP was not appropriate for yielding positive effects. An alternative explanation for this lack of finding significant effects might be related to the teachers’ implementation of data-informed teaching behavior and thinking. When considering teachers’ implementation, it might be the case that teachers incorporated the new way of working, at least to some extent, but that the adjusted attitudes and behavior did not yield better student achievement yet. Comparable suggestions have been put forth by Shaw and Wayman (2012), who suggested that it takes schools and teachers at least two years to fully incorporate the use of data. Moreover, the lack of finding an overall effect of the PDP on student achievement might have concealed differential effects of the PDP. The promoted way of working might just not have been adapted by all teachers, leaving room for the possibility that its effects on student achievement differed for teachers who were more data-minded and for those who were so to a lesser extent. The purpose of the current chapter is to investigate a) whether teachers who participated in the PDP had different behavior and attitudes at the

---

\(^2\) More specific information on the professional development program is provided in Appendix 1.
end of the intervention compared to teachers who were not engaged in such a training, b) whether it would be possible to identify different levels of teachers' data-mindedness at the end of the PDP, and c) whether different levels of data-mindedness were related to initial differences, teacher change, and student achievement.

5.1.1 Effective professional development programs

The effectiveness of educational reforms and policies for teaching practice can be mediated by professional development activities (Desimone, 2009). The rationale of professional development (PD) is that by offering new knowledge and/or supporting practical skills, teacher learning is fostered, resulting not only in increased knowledge, better skills, and changed attitudes, but also in improved teaching practice and, as a result, increased student performance. Regarding the content and set-up of such teacher learning programs, Desimone referred to a consensus in the PD field that five critical features are essential for effective and high quality PD programs (Desimone, 2009; Timperley et al., 2007): 1. focus on content, 2. active or inquiry-based learning, 3. coherence, 4. sufficient duration, and 5. collaboration. The first feature, focus on content, is considered most important (Garet et al., 2001; Hochberg & Desimone, 2010). Targeting classroom practices, by focusing PD activities explicitly on student learning and on content knowledge and pedagogical content knowledge, is not only found to be related to gains in teacher knowledge and skills, and teaching practice (Garet et al., 2001; Hochberg & Desimone, 2010), but also to improved student outcomes (Timperley et al., 2007). Active and inquiry-based learning means that professional development programs (PDPs) contain active learning, like interactive feedback based on lesson observations, or inquiry-based learning, in which teachers, for example, analyze and interpret student assessment data to base decisions for future teaching on (van Driel, Meirink, van Veen, & Zwart, 2012). The third feature, coherence, means that the PD content should be coherent with reform policies at school, district or national level, with teachers’ own knowledge and beliefs, and with problems teachers encounter in their daily classroom practice (Timperley et al., 2007). By being coherent, PDPs could solve teacher problems and therefore foster changes in teacher behavior. PDPs further are supposed to engage participants for a considerable amount of time (duration) and should be spread over a sufficient time span (Desimone, 2009). Finally, collaboration and collective participation offer teachers opportunities to interact with each other and work together, which stimulates their learning.
5.1.2 Teacher learning in professional development programs

Considerable PD initiatives do not succeed in bringing about real changes in teaching practices (Fullan, 2007). Fullan argued that it takes considerable time, 2 or 3 years, to implement changes after teachers made the decision to adopt the reform (the Initialization phase). However, teachers do not always decide to change, sometimes they reject the reforms. In their best evidence synthesis on effective professional learning and development Timperley and colleagues (2007) pointed out that relatively little is known about how professional learning opportunities, that are provided by PD programs, influence teaching practice. Unravelling this black box might help us to understand how the content of a PDP influences teachers’ interpretation and utilization of it, resulting in either changes in teaching practice or not (the so called possible ‘teacher outcomes’). For teachers’ actual implementation of the program content, context factors, like time and personnel resources, play an important role, as well as teachers’ motivation and emotions. For instance, teachers should consider the PD course practical and useful for daily practice (Timperley et al., 2007). This is important, not only for teachers’ motivation to adopt and implement the new information, but also for the promotion of positive emotions for change (Schmidt & Datnow, 2005).

Since teachers’ motivation, emotions, prior knowledge, and beliefs can influence implementation and, as a result of that, the effectiveness of PDPs, it is important to pay attention to them. Teacher motivation is known to be high if the PDP addresses topics that are meaningful to the teachers and when something is at stake for them as a professional (Hochberg & Desimone, 2010; Timperley et al., 2007). In a recent study, Wayman and Jimerson (2014) investigated the teacher needs for PD programs in data use. For this purpose, data were collected from educators coming from 3 school districts, having different positions in the schools. Interview and focus group results led Wayman and Jimerson to conclude that educators believe teachers to primarily need ‘immediately useful data-related professional learning’ (p. 31, italics from Wayman & Jimerson). In terms of how the PDP should be designed this meant that teachers should be facilitated to timely use data concerning their own students – meaning that the program is job-embedded and contextual. Further, teachers should ideally be able to collaborate in small groups relating to specific grade- or content-levels, and be actively engaged in hands-on activities. Educators further indicated that learning content should be provided by a credible expert, preferably an experienced classroom teacher, and should build upon prior knowledge of the teachers. Unfortunately, educators indicated that most PDP’s do not cover these needs, showing a gap between what is delivered in PDPs and what is
desired or needed by teachers. Since teachers’ motivation is found to be higher in cases where the content of the PDP is highly relevant to them (Timperley et al., 2007), it seems to be important that sufficient attention is paid to addressing the actual teacher needs while developing such programs (Wayman & Jimerson, 2014).

5.1.3 Teachers’ emotional states, beliefs and sense-making as mediating factors for implementation

In their study on how emotions had an impact on teachers’ sense-making of comprehensive school reforms, Schmidt and Datnow (2005) found that especially innovations related to changes in classroom practices elicited teacher emotions, both positive, like having feelings of enthusiasm, confidence and commitment, and negative, like experiencing feelings of resentment, frustration, worry or guilt. As emotions are found to be determinants for the implementation of innovations in several studies (Saunders, 2013; van Veen & Lasky, 2005), Schmidt & Datnow (2005) recommended that, in change processes, ample consideration should be given to teachers’ feelings of anxiety and worry, by making teachers knowledgeable about what the innovation amounts to, about the extent to which the innovation differs from teachers’ current teaching practices, and by providing teachers with support and tools to make the promoted changes. Further, Schmidt and Datnow (2005) stressed the importance of providing teachers with “the emotional support to take reasonable risks without concerning themselves about the possible consequences that might befall them personally” (p. 962).

Such considerations on teachers’ emotions and well-being are also be associated to PDPs on accountability topics, like the standards-based reform. One of the undesired consequences of accountability is that teachers perceive a lot of stress due to a strong focus on fast growth in student performance and the demands for teachers’ instructional skills. Hochberg and Desimone (2010) reflected on what accountability entails for effective PD. They argued that its main aims, rapid improvement of achievement for all students through considerable instructional changes, should have implications for PD initiatives. To support teachers in reaching these aims, Hochberg and Desimone (2010) specifically emphasized a content and long-term focus. The accountability context further requires that teacher beliefs are addressed sufficiently, since participation in the reform and therefore in PD initiatives is obligatory. Especially teachers’ epistemological and efficacy beliefs seem to play an important role in adapting to a more data-informed and outcome-concerned way of working. By fostering feelings of competence, by providing
teachers sufficient opportunity to practice and to incorporate the instructional modifications needed – which is often not common practice in PD initiatives in the United Stated (Hochberg & Desimone, 2010; Mandinach, 2012) -, and by showing the importance of the PD elements, discrepancies between the original and promoted efficacy beliefs can be dealt with. The same goes for teachers’ epistemological beliefs. These beliefs concern “the nature of knowledge and the process of knowing” (Gregoire, 2003, p. 172), like one’s belief that subject-matter knowledge is fixed or that it can be changed. Such epistemological beliefs have been related to teacher change (see for references Gregoire, 2003), thereby offering possible explanations for situations in which teachers do not implement the reforms as intended, even when they themselves assume that they do.

Taking teacher beliefs into account in innovation projects and PD initiatives was also stressed by other authors. Pajares (1992), in his review on teacher beliefs, stated that teacher beliefs might function as a filter through which new information is perceived. This filtering might cause the beliefs, in combination with teachers’ prior knowledge, to influence teachers’ interpretation processes in PDPs or reforms, leading to selective changes of only some aspects of teaching practice, while not implementing the full program or reform (Spillane, Reiser, & Reimer, 2002). Therefore, Spillane et al. (2002) stressed the need for focusing on teachers’ sense-making. When teachers make sense of, or comprehend, new policy or innovations, an active process of interpretation is assumed to take place “that draws on the individual’s rich knowledge base of understandings, beliefs, and attitudes” (Spillane et al., 2002, p. 391). Thus, effective PDPs should pay attention to such interpretation processes, making sure that all participants understand what the new way of working entails (cf. Cho & Wayman, 2014).

Especially in the context of data use, a prerequisite for effective PDPs seems to be that participants use the same definitions for the concepts and have mutual ideas about the aims and purposes of data use for their own school. Only when all educators in a school (or district) have common understandings, school teams will be able to constructively discuss and work together on data (Wayman & Stringfield, 2006; Kerr, Marsh, Schuyler Ikemoto, Darilek, & Barney, 2006; Young, 2006). In a recent study, Jimerson (2014) investigated the mental models of educators from a school district in Texas, using both survey and interview data. Jimerson examined the ways the educators considered the terms data and data use. The educators were found to have different perspectives on data use, with perspectives ranging from positive to negative; negative comments mainly coming from the teachers. Jimerson’s findings showed the importance
of spending time on co-constructing common understandings on data use during PD programs, meaning that participants (or school/regional teams) share the same meanings and perceptions, not only on what data are and how they can be used to support teaching practice, but also on (the purposes of) teaching and learning (Wayman et al., 2012).

Such co-construction and sense-making of the core concepts seem to play a central role in dealing with data use, especially given its connection to accountability. Jimerson (2014) found that, even in those schools where DDDM was primarily used for school/instructional improvement, accountability concerns were often present at some level. Although accountability might be most prominent in the United States, nowadays, it should be taken into account in other countries, like the Netherlands, as well (Ehren & Swanborn, 2012; Schildkamp et al., 2014).

5.1.4 Some recent examples of PD initiatives for data use

Although data use is one of the spearheads of educational policy, its full implementation in schools, leading to real, well-considered adaptations and instructional changes, is still rather limited (for instance, Goertz et al., 2009; Schildkamp et al., 2014; Vanhoof & Schildkamp, 2014). The availability of data does not automatically lead to its effective use: teachers should be trained to become data-literate, and should be supported in translating data into instructional actions (Heritage et al., 2009; Kerr et al., 2006; Mandinach, 2012; Means et al., 2010; Shaw & Wayman, 2012). Despite the apparent need for professional development in data use, PD initiatives do not always fulfill the needs and not always yield teachers that can be considered data-literate (Jimerson, 2014; Mandinach, 2012; Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009). In this section, we discuss some recent examples of PD initiatives on data use, teachers’ implementation of such programs, and their effectiveness for raising student achievement.

A recent study on whether observational data could improve teaching practice showed that discussing lesson observations fostered teachers to adapt their teaching in such a way that less time was spent on classroom management. Students were further found to be more on task and teachers spent more time on explicit instruction (Van den Hurk, Houtveen, Van de Grift, & Cras, 2014). In another study, Staman et al. (2014) described that their PDP on making school teams data-literate for instructional improvement purposes positively influenced teachers’ knowledge and skills regarding data analyses in the student monitoring system. However, when participants’ attitudes
regarding data and the student monitoring system collected after the training were compared to their attitudes before the PDP, no significant differences were found. Staman et al. (2014) argued that the high pretest scores on the attitudes made significant positive changes somewhat improbable. However, an alternative explanation might be that it is more difficult to change teacher beliefs than (part of) their knowledge and skills (Hochberg & Desimone, 2010).

In a one year case study of a US primary school, Hubbard, Datnow, and Pruyn (2014) investigated whether and how teachers implemented data use across the curriculum and how data use was integrated in recent reform initiatives. Such reform initiatives, like teaching 21st century skills and the new Common Core Standards, demand students to learn skills like critical thinking and project-based working. Measuring such skills forces our ideas about data and data use to become broader than in current practice. In order to inform and optimize instruction, these recent initiatives demand an extended use of data across the whole curriculum, thereby not only using data to easily assess and monitor students’ achievement in basic skills. Based on observed grade-level meetings and interview data, Hubbard and colleagues found that such a broad use of data was limited: teachers used data mainly for improving instruction in math and English language arts, but such data use was not merged in other subject areas. Further, no real integration with other initiatives, like problem based learning, was found. Implementation of data use thus seemed to remain rather fragmented and isolated in the school. In this respect the researchers pointed to the complexities teachers perceive when uniting and implementing multiple initiatives.

Two recent studies concerned the same one year PDP (the current project), in which Dutch teachers were professionalized in using data for instructional improvement purposes. The studies focused on whether the PDP affected student achievement in reading comprehension (van Kuijk, et al., 2015) and mathematics (see chapter 4). The comprehensive joined program contained three interrelated, evidence-informed components (goal setting, data analysis, and instruction), and targeted teachers of grades 2 and 3, as well as school principals and senior support coordinators. This set-up allowed for collaboration in small school teams, while focusing on specific grade-levels. Since both studies were part of a joined PDP, most of the year-long program was the same for both studies; the main difference between the two programs was related to the domain-
Comparing the classes taught by teachers in the experimental condition to comparable classes in a control condition showed a significant, positive effect of the PDP \( (d=0.37) \) on reading comprehension. In the study on mathematics, students in the experimental condition were matched to comparable control students. These analyses did not yield any indication that the PDP positively affected the mathematics achievement of students.

5.1.5 Research questions

Research on PD’s and educational reforms, like DDDM, indicated that it is not straightforward that PDPs suit educators’ needs, foster sustainable changes in teachers’ knowledge, beliefs, skills, and practices, and ensure sufficient time for such changes. Recent examples of PDPs on data use further showed that these programs were not always effective, sometimes only partially, for teacher change or raising student achievement. This might be due to a lack of bringing about real changes in teachers’ beliefs, attitudes and behavior, or to a lack of implementation of all promoted teaching practices.

Several authors stressed the need for sufficiently taking into account teacher attitudes and beliefs when conducting PD programs and when studying their implementation by the teachers (Gregoire, 2003; Hochberg & Desimone, 2010; Jimerson, 2014; Pajares, 1992; Spillane et al., 2002). Researchers have related program implementation to mental models or sense-making of its participants, possibly causing only selective implementation of elements (Jimerson, 2014; Spillane et al., 2002), or causing such high levels of discrepancy with teachers’ own beliefs, attitudes and prior knowledge that hardly any implementation takes place. Teacher emotions of anxiety, worry or a lack of confidence could even cause teachers to (partly) reject the innovation. Levels of implementation might thus differ considerably between teachers, possibly leading to differential effects on student performance. Such considerations led to the following three research questions:

1. To what extent is attending a professional development program on data-driven decision making related to teachers’ data-minded attitudes and behavior?

---

Note that an important part of the instructional component targeted knowledge and skills in the two evidence-informed instructional methods – direct instruction and modeling –, that were related to both math and reading comprehension.
2. Is it possible to determine different profiles of trained teachers in terms of their data-mindedness?

If teacher profiles can be identified,

3a. what were the data-minded attitudes and behavior of the identified teacher profiles at the start of the PDP and to what extent had these attitudes and behaviors changed at the end of the PDP?

3b. to what extent are the teacher profiles related to the math performance of students?

These questions were answered using teacher and student data. The PDP was supposed to foster a) teachers’ awareness of what levels of student learning and achievement they strived for – teachers were thus assumed to have become more goal-oriented-, b) teachers’ ability to analyze and interpret data in order to determine students’ actual proficiency levels and c) teachers’ ability to close the gap between desired and actual levels, by making their lessons more structured, focused, and clear. Given the support, practice, and team discussions, we expected teachers who participated in the PDP to show higher levels of data-mindedness in their thinking and behavior than teachers who were not engaged in a PDP on data use (Hypothesis 1). Furthermore, we expected those teachers in the experimental condition, who reflected more data-mindedness at the end of the PDP, to have implemented program content to a higher extent than those teachers in the experimental condition, who showed a lower level of data-mindedness (Hypothesis 2). Moreover, teachers’ higher levels of data-mindedness were expected to influence their students’ math achievement in a positive way (Hypothesis 3).

5.2 Method

5.2.1 Design

Implementation of DDDM was studied in two ways. For studying the first question, a posttest-only control group design was used. Second, for investigating the teacher profiles, their relationship with baseline measurements, with teacher change, and with students’ math achievement, a post-hoc quasi-experimental pre-posttest design was used.
5.2.2 The current study

A PDP was developed to support teachers in using data in a formative way. The program consisted of three components: 1) setting performance standards and performance goals to establish students’ desired performance level, 2) analyzing and interpreting data to determine students’ actual performance level, 3) supporting teachers to use effective instructional methods in order to foster structured, clear and targeted lessons, thereby closing the gap between 1 and 2. These three components are considered to reflect the core features of DDDM (cf. Visscher & Ehren, 2011).

The PDP focused on data use in mathematics in grades 2 and 3 of primary school. It took place in 19 schools and lasted one school year. In total, the PDP consisted of nine meetings: four were held as a general assembly, five took place at the individual schools. Participation was on a voluntary basis. For more information on the PDP, the reader is referred to Appendix 1.

The PDP adhered to the five core features of effective PD (Desimone, 2009) and also to most of the features considered to meet teacher needs in PD on data use (Wayman & Jimerson, 2014). The researchers carried out all the meetings and the observations.
5.2.3 Participants

**Experimental condition**

Our PDP was conducted at 19 schools, targeting teachers from grades 2 and 3, senior support coordinators and school leaders. For the current study, data from 33 teachers were used. Criteria for including the teachers in the study were that a) teachers had filled in a questionnaire at the end of the project, b) a math lesson, given by the selected teacher, was observed, and c) data on both a standardized pre- and posttest in math were available for the students in their classes. For each class only one teacher was selected. Should two teachers share a position part-time and if both met the criteria, the teachers who taught the class most frequently were selected. The average years of experience of the 33 experimental teachers was 15 years (SD: 12). They taught 42 grade-specific classes and 471 students of grades 2 and 3 were taught by them.\(^{22}\)

**Control condition**

The PDP was part of a larger cluster of five PDPs, all targeting (aspects of) data use. Since all PDPs were directed at different grades, teachers of grades 2 and 3 coming from schools that took part in one of the other four interventions could be considered as possible controls: the current study was the only study that focused on grades 2 and 3. The current study was further the sole intervention in which senior support coordinators and school leaders were engaged. Therefore, contamination of the control grade 2- and grade 3-teachers by the other PDPs was assumed to be minimal.

The cluster of projects took place from 2010-2012. All five intervention studies lasted one school year: in our PDP, teachers participated in the school year 2011-2012, while the other four projects took place in either 2010-2011 or 2011-2012. Teachers could only serve as controls when a) they had filled in a questionnaire at the end of the project, b) complete mathematics assessment information (pre- and posttest) of their students was available, and c) the teachers and their students could be unambiguously linked to a specific class. Thirty-one teachers, teaching 48 classes, met these criteria and were selected as controls. Their average years of experience was 21 years (SD: 13). They taught 756 students of grades 2 and 3.

\(^{22}\) Next to single-grade classes, teachers taught multi-grade classes. In the multi-grade classes, the different grades were distinguished as separate classes, meaning that teachers who taught a multi-grade class containing both grade 2 and grade 3 were considered to teach two classes. Teachers who taught a multi-grade class containing both grade 3 and grade 4 were considered to teach one class.
5.2.4 Instruments

In this study, variables at teacher and student level were used. Data collection took place using three instruments: a) a questionnaire on aspects of data use, b) a high-inference observation instrument for observing teacher behavior during a mathematics lesson, and c) standardized mathematics assessments, providing information on student achievement.

Teacher questionnaire

Self-reported data, containing information on teacher attitudes and behavior related to data use, were measured using a questionnaire. The questionnaire was filled in by the experimental and control teachers, both at the start and at the end of the cluster of projects. The controls filled in the questionnaire in June 2010 and in June 2012, the teachers in the experimental condition in July/August 2011 and in June 2012. As we wanted to avoid having different time-frames between pre- and posttest, question 1 will be answered using only data from the posttest (June 2012). Since the questionnaire was expensive and since we intended to minimize teachers’ time and effort needed for filling it in, teachers were randomly assigned to one of two versions, consistent of a mutual set of questions and selective of additional questions. In the current study, we limited the use of questionnaire items to those that were provided to the teachers in both versions. The scales and items that were identified are listed in Section 5.2.5, and are summarized in Appendix A. Appendix A further contains examples of the questions that were posed.

Observation instrument

All 33 teachers in the experimental condition were observed during two mathematics lessons, at the start of the PDP and at its end. The observation instrument consisted of two parts: a low-inference time sampling part, capturing the set-up and flow of the lesson (see chapter 3), and a general high inference part. The latter part indicated whether various teaching practices were used or not. In the current study only the high inference part was used. The high inference observation instrument consisted of 16 dichotomous items – yes indicating that the teacher showed the (instructional) behavior, no indicating that the teacher did not. Taken together, these items were considered to provide a rough indication of teaching quality. The main purpose of these observations was to provide teachers with immediate feedback on their use of instructional elements that were discussed in the PDP, like direct instruction, modeling and differentiation. As such the observations were part of the PDP’s content. Given the formative role of the
observations in the PDP, observation data were only collected for the teachers in the experimental condition. No observations were carried out in the math lessons of the teachers in the control condition.

The observations were conducted by two researchers and two research-assistants. After being trained by means of video material, inter-rater reliability was considered sufficient (Cohen’s kappa = .85). The identified scales and variables are discussed in Section 5.2.5.

**Standardized mathematics assessment scores**

The standardized assessment for mathematics is part of the Cito LOVS assessment system which is used in 85 percent of Dutch schools and by all the participating schools. The tests are used in primary school (grades 1-6), and students are assessed twice a year; mid-term assessment in January and end-term assessment in June. Test scores (from grade 1 to 6) are linked to a single proficiency scale, ranging from 0 to 169. The tests have a good overall reliability in all grades (Cronbach’s alpha is at least .91, Janssen et al., 2010).

**5.2.5 Variables**

The teacher characteristics were derived both from the questionnaire and from the observed math lessons, denoted by (q) and (o), respectively. First, the three reliable unidimensional composite scales that were identified are discussed; these are denoted by an asterisk. These three reliable unidimensional composite scales were used for answering question 1, 2 and 3a. Next, the separate items, averaged scores, as well as the sum scores are discussed. Separate items or averaged scores, from the teacher questionnaire (q) were used for answering both question 1 and 3a. Separate items or sum scores from the observations (o) were only used for answering question 3a.

**Teacher characteristics**

*Attitude towards basic skills (q)*

This scale, $\alpha=.67$, determined teachers’ attitudes towards the recent educational policy to focus on the basic skills (math and reading). The scale was based on three items. It reflects teachers’ awareness of the importance of students’ basic skills for future learning and development (e.g., Bodovsky, et al., 2011). Teachers who strongly agree with a focus on basic skills are assumed to also aim for maximizing students’ math and reading scores. A higher score on this scale is therefore considered to reflect stronger
awareness of measurable outcomes. The scale ranged from 1 (complete disagreement with an emphasis on basic skills) to 5 (complete agreement).

**Goal-oriented working (q)***

This scale, $\alpha = .89$, measured whether teachers consider it relevant to share information with their students on how they perform, on what they should minimally be able to master and on what they are expected to learn in the near future. Sharing this information assumes that teachers explicitly know what levels of performance they want their students to reach, meaning that they used clear goals. The scale ranged from 1 (complete disagreement) to 5 (complete agreement).

**Differentiation practices (o)**

This scale, $\alpha = .82$, identified the differentiation practices teachers used during the observed math lessons. These were scored using three items: a) the provision of extended instruction by the teacher to a subgroup of the students, b) the assignment of easier tasks to low performers, and c) the provision of more difficult tasks for the high performers. Information on teachers’ use of differentiation practices is considered valuable for assessing their implementation of data use, as, especially in highly heterogeneous classes, differentiation can be considered a consequence of such data use. Scores went from 0 (no observed differentiation practices) to 3 (three differentiation practices observed).

**Attitude towards standardized assessments (q)**

This item reflected the relative level of importance that teachers attribute to information from standardized assessments compared to information coming from a) their own knowledge and impressions, b) observations in class, c) students’ daily work on math tasks and d) assessments in the curricular textbooks. Five categories were distinguished, ranging from 1 (the standardized assessment is least informative) to 5 (the standardized assessment is most informative).

**Attitude towards differentiation (q)**

DDDGM assumes that teachers adapt their teaching to students’ instructional needs. This entails differentiation practices. Knowing teachers’ intentions to, for example, focus only on the low or on both the low and high performers, provides information on how teachers think they should adapt their teaching practices to

---

23 Interpretation of this score is rather ambiguous. A high score could indicate that teachers use information from analyses on the standardized assessment often and/or thoroughly, but during the PDP it was argued that standardized assessments are only one source of information on students’ proficiency; daily work, observations and textbook assessments should also be taken into account.
differences between students. The variable was a sum score of four separate items, transformed in a 3-point scale.

*Use of within-class grouping over all basic skills (q)*
This score summarized the number of ability groups teachers tended to distinguish in the lessons focusing on basic skills, averaged over lessons in reading, reading comprehension, language arts, and mathematics.

*Use of within-class grouping during math lessons (q)*
This score represented the number of ability groups that teachers distinguish in their math lessons.

*Use of explicit performance levels for students in mathematics (q)*
Teachers were asked whether they made explicit what specific math performance levels should be reached by the students in their class. Four categories were distinguished - 1: no, 2: I know what the levels are, 3: I actively determine these levels, either alone or with the senior support coordinator, 4: these levels and expectations are openly discussed in the school team.

*Acquaintance with performance standards (q)*
Performance standards reflect the level of performance at which students master content (Hambleton & Pitoniak, 2006). They are supposed to provide clear goals and therefore are assumed to focus teachers’ instruction on what is considered to be most important. Two categories were distinguished (1: no, 2: yes).

*Ambitious mind-set (q)*
Information on the teachers’ willingness to constantly strive for bringing about better student achievement, even in case of good performance, might be an indication of how ambitious they are. Answers ranged from 1 (negative attitude) to 5 (positive attitude).

*Direct Instruction (o)*
Knowledge and implementation of the direct instruction model (DIM - Borman, et al., 2003), was part of the PDP. Teachers' use of DIM in math lessons was scored using four items. These items measured whether the teacher a) activated prior knowledge of students at the lesson start, b) addressed key lesson objectives at the start of the lesson, c) discussed these key lesson objectives at the end of the lesson, and d) looked ahead at the next lessons – thereby clarifying relations between concepts and lessons. Dummy scores were used to indicate whether the DIM elements were present (0=no, 1=yes) and were summed up.
Modeling (o)
Modeling is an effective teaching method (van Gog, 2013) in which teachers explicitly demonstrate their use of strategies to let students see or hear how they should deal with a given task or problem. In the PDP, modeling was discussed and all participants had the opportunity to practice the skill in a safe setting. Dummy scores were used to indicate whether the behavior was observed in class or not (0=no, 1=yes).

Informal diagnostic assessment (o)
During their lessons teachers have ample opportunities to informally assess students’ levels of understanding. This information allows teachers to be able to fit future teaching to students’ actual needs. In the observation instrument, three items that were related to these informal assessment tools were measured: a) teachers asking their students to demonstrate the strategies used while performing a written task, and teachers asking students to orally elaborate on b) a right or c) a wrong answer. Based on whether the three behaviors were observed or not (dummy variables; 0:no, 1:yes), a sum score was calculated.

Lesson quality (o)
The lesson quality of the observed math lesson was measured by taking the sum score of all 16 variables used in the high inference observation instrument. This score therefore also included teachers’ scores on direct instruction, modeling, informal diagnostic assessment and on the scale differentiation practices. The other 5 items (all dummy-coded) related to whether the teacher 1) promoted students’ automated skills, 2) asked students who finished their tasks to do other math work (and not work on other subjects or doing something for themselves), 3) used sufficient waiting time, providing students time to think, 4) repeated the answers given, so that all students in class could hear the right answer, and 5) praised students.
**Student and class characteristics**

**Students’ mathematics performance**

In the outcome analyses students’ prior math achievements were taken into account as a covariate, using the pretest scores on the standardized mathematics assessment (see 5.2.4). For the grades 2 and 3, the assessment scores at the end of grades 1 and 2 were used, respectively. Scores on the standardized math assessment a year later, at the end of the grades 2 and 3, were used as the posttest.

**Sex**

Ganley et al. (2013) referred to various studies in which a (small) sex difference in mathematics test performance was found, with boys outperforming girls in assessments. Sex, dummy-coded with boys as reference group, was used in the outcome analysis.

**Class size**

To account for possible influences of class size on student achievement (Brühwiler & Blatchford, 2011) or on the amount of individual teacher attention, time on tasks and students’ active engagement (Blatchford et al., 2011; Doolaard & Bosker, 2006), class size was controlled for in the outcome analyses. Class size was defined as the total number of students in one grade. According to this definition, multi-grade classes that contained both grade 2 and grade 3 were considered to consist of two classes. In the other multi-grade classes, the grade 1 or grade 4 students were not included.

**5.3 Analyses**

Differences between experimental and control teachers regarding their attitudes and self-reported behavior (question 1), were investigated using descriptive analyses. For answering question 2, teacher profiles were constructed in order to investigate whether teachers who had received the PDP differed in terms of their data-mindedness. For constructing the profiles the three identified scales were used. The scores on the scales Goal-oriented working and Attitude towards basic skills, both originally a 5-point scale, were transformed into a 3-point scale in order to reduce the number of parameters that had to be estimated.

For comparing the initial differences between the identified teacher profiles and their levels of implementation, question 3a, nonparametric tests were used, respectively the Mann Whitney U test and the Wilcoxon rank sum test. Here, data from the teacher questionnaire and from the observed math lessons were used.
5.3.1 Missing data

One criterion for including teachers in the study was that teachers had filled in the posttest of the questionnaire. However, not all teachers had filled in all questions. Some missing data remained for the teachers. Of the teachers in the experimental condition, the percentage of missings on items in the scales *Attitude towards basic skills* and *Goal-oriented working* was 3% and 6.1% respectively. As we wanted the teacher profiles to be based on the information from all selected teachers in the experimental condition, we imputed the missing data in these two scales. Given the simple structure of the data – that is, questionnaire items at teacher level - and the low number of missing data, multiple imputation in SPSS (Fully Conditional Specification) was used. Five datasets were imputed, after 10 iterations.

5.3.2 Teacher profiles

Teacher profiles were identified using latent class cluster analysis (LCA). In LCA, class membership is modeled statistically by basing such class membership on probability models. LCA was conducted using Latent Gold 4.5 (Vermunt & Magidson, 2005). The advantage of this program is that it allows for multiply imputed data, for which pooled estimations are provided.

For the 33 teachers who participated in our PDP we tried to identify clusters in a latent cluster model. This was done using the three scales, *Attitude towards basic skills, Goal-oriented working* and the observed *Differentiation practices*. Models with one through four latent classes were fitted to the data. For assessing the models, the absolute model fit was based on the likelihood-ratio statistic $G^2$, while taking into account considerations on parsimony. The models were further assessed in terms of the interpretability of the grouping.

The model in which two latent classes were estimated fitted the data best. These two profiles described levels of teachers’ use of differentiation, goal-orientedness and a focus on basic skills, teachers’ so-called *DGB-behavior*. The profiles were used in the subsequent analyses. Higher average cluster scores were interpreted as more use of DGB-behavior. The teachers in the profile with the highest cluster means were expected to be more aware of their explicit performance aims and of educational outcomes and were expected to act upon differences between students. Since these are considered important identifiers of data use, DGB-behavior is assumed to reflect teachers’ data-mindedness.
5.3.3 Impact analyses

The question whether teachers’ stronger use of DGB-behavior was related to better math achievement of the students, was investigated using multilevel multiple regression analyses in MLwiN (Rasbash et al., 2012). In the multilevel analyses, we regressed the math posttest scores of students on the profile of the teachers. We first considered an empty model. To control for relevant student or class characteristics, all available covariates, including type of class (i.e., single-/multi-grade), grade, average class score on math pretest, and class heterogeneity (SD of math pretest), were used in constructing a covariate model. For reasons of sparseness, the covariate model presented in Section 5.4 only contains significant predictors. The final model that was estimated contains the teacher profiles as explanatory variable predicting the math posttest scores. In this Effect Model, the teachers who showed DGB-behavior to a relatively low extent, the cluster 1-teachers, were considered the reference group. Outcome analyses were conducted for all five datasets that derived from the missing-data imputation. The resulting five regression coefficients and standard errors were then pooled (Snijders & Bosker, 2012). The multilevel models presented in Section 5.4 contain these pooled estimates.

5.4 Results

First, we considered whether attitudes and behavior in terms of teachers’ data-mindedness were related to their attendance to a PDP on data use. This question was investigated in an explorative manner by considering the teachers’ scores on items from the questionnaire and comparing them to the item scores of the control teachers. Descriptives for both the experimental and control group are depicted in Table 1, as well as Effect sizes (Cohen’s d) and 90%-Confidence Intervals. Considering teachers’ scores at the end of the PDP, we found that both the treated and control teachers responded in a rather positive way: They were aware of the relevance of basic skills, they were at least somewhat aware of students’ desired performance levels, although they did not usually made these explicit, they used within-class grouping, they thought it desirable to spread their attention between students having different performance levels, and they were rather ambitious.

Differences between the treated and control conditions were primarily found in terms of the use and awareness of goals, that is, determining an explicit performance level for each student and acquaintance with performance standards: teachers in the experimental condition scored significantly higher on these items. The teachers who were
engaged in the PDP further valued a focus on basic skills more than the controls. On the other hand, teachers in the control condition tended to make more use of within-class ability grouping during the math lessons and tended to value the information provided by standardized assessments on students’ development more. Further, both groups scored relatively high on ambitious mind-set. Overall, Table 1 indicates that the attitudes and self-reported practices of teachers who participated in a PDP on data use did not strongly differ from teachers who did not, despite the tendency that the treated teachers were more aware of the desired performance level of their students, a core feature of DDDM.

For answering the second question, we first attempted to identify latent clusters for the treated teachers. These clusters were based on the three ordinal scales for which we could construct reliable unidimensional composite scales. As was mentioned earlier, a two cluster model fitted the data best. In Table 2 shows both the latent class prevalence in the sample and the item-response probabilities are presented for the two clusters. More teachers were assigned to cluster 1 than to cluster 2, 55% and 45% respectively. Many teachers in cluster 1 did not use differentiation in the math lessons and used goal-oriented working at a medium level. Almost half of the cluster 2-teachers used differentiation to a strong degree and these teachers were categorized as strong goal-oriented workers. Regarding the third scale, attitude towards basic skills, both clusters of teachers mainly had item-response probabilities in the medium category.
Table 1: Descriptives of Item/Scale Scores, Treated and Control Teachers (Questionnaire, Posttest)

<table>
<thead>
<tr>
<th>Item/scale</th>
<th>Control condition</th>
<th>Experimental condition</th>
<th>90%-CI</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude towards basic skills *</td>
<td>29</td>
<td>32</td>
<td>3.52 .54</td>
<td>3.89 .48</td>
</tr>
<tr>
<td>Attitude towards standardized assessments</td>
<td>29</td>
<td>30</td>
<td>3.59 1.59</td>
<td>3.17 1.5</td>
</tr>
<tr>
<td>Goal-oriented working *</td>
<td>29</td>
<td>32</td>
<td>3.95 .72</td>
<td>4.18 .39</td>
</tr>
<tr>
<td>Use of explicit performance levels for students (in math)</td>
<td>29</td>
<td>31</td>
<td>2.48 .91</td>
<td>2.90 .87</td>
</tr>
<tr>
<td>Acquaintance with performance standards</td>
<td>30</td>
<td>32</td>
<td>1.60 .50</td>
<td>2.00 .00</td>
</tr>
<tr>
<td>Use of within grouping during math lessons</td>
<td>31</td>
<td>33</td>
<td>2.26 .68</td>
<td>2.03 .73</td>
</tr>
<tr>
<td>Use of within-class grouping over all subjects</td>
<td>31</td>
<td>32</td>
<td>1.77 .80</td>
<td>1.78 .83</td>
</tr>
<tr>
<td>Attitude towards differentiation</td>
<td>29</td>
<td>32</td>
<td>2.07 .80</td>
<td>2.06 .76</td>
</tr>
<tr>
<td>Ambitious mind-set</td>
<td>29</td>
<td>32</td>
<td>4.17 .47</td>
<td>4.03 .65</td>
</tr>
<tr>
<td>*: Identified scales</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Latent Class Prevalence and Item-response Probabilities of the Identified Clusters

<table>
<thead>
<tr>
<th>Latent class prevalence</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>.55</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Item-response probabilities**

**Differentiation practices**

<table>
<thead>
<tr>
<th>No differentiation</th>
<th>.69</th>
<th>.24</th>
<th>.49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Little use</td>
<td>.24</td>
<td>.24</td>
<td>.24</td>
</tr>
<tr>
<td>Medium use</td>
<td>.02</td>
<td>.05</td>
<td>.03</td>
</tr>
<tr>
<td>strong use</td>
<td>.06</td>
<td>.47</td>
<td>.24</td>
</tr>
</tbody>
</table>

**Goal-oriented working**

| Low         | .08 | .00 | .04 |
| Medium      | .83 | .27 | .58 |
| High        | .10 | .73 | .38 |

**Attitude towards basic skills**

| Low | .34 | .11 | .24 |
| Medium | .62 | .73 | .68 |
| High | .04 | .16 | .09 |
We investigated whether all the three scales were related to variation in allocation in different categories per cluster and identified that only the variable \textit{differentiation practices} discriminated significantly, using a critical value of 0.10 (Wald test, \( p=.08 \)). For descriptive purposes, we then studied whether and how the two identified groups of teachers differed on each of the three scales. We investigated their ranked orders on the three scales using the Wilcoxon Ranked Sum test. Since the ordinal scores reflected directionality, we tested one-sided. Teachers in cluster 2 (pooled \( Mdn=22.11 \)) used significantly more differentiation practices than the teachers in cluster 1 (pooled \( Mdn=12.94 \)), \( U=59.5, Z=-2.92, p<.01 \). Teachers in cluster 2 (pooled \( Mdn=25.03 \)) worked also more goal-oriented than the cluster 1-teachers (pooled \( Mdn=10.62 \)), \( U=17.0, Z=-4.90, p<.01 \). Finally, teachers in cluster 2 (pooled \( Mdn=19.73 \)) were more focused on basic skills than the teachers in cluster 1 (pooled \( Mdn=14.84 \)), \( U=94.3, Z=-1.75, p=.04 \).

The descriptive analyses thus indicated that cluster 2-teachers used more differentiation practices, worked more goal-oriented, and focused more on basic skills, thereby showing more so-called DGB-behavior. Assuming that DGB-behavior reflects a certain level of data-mindedness and given the directionality of the scales (higher scores on the scales representing a better profile), teachers in cluster 2 were considered to be more data-minded than teachers in cluster 1.

After having described the two groups of teachers in terms of their relatively weak or strong DGB-behavior, we investigated their attitudes, self-reported behavior, and observed teaching practices at the start of the PDP in order to determine whether there were initial differences between the two clusters of teachers. Table 3 presents pretest descriptives on the questionnaire items that were filled in by the teachers before the PDP started, and on measures in the observation instrument. Given the small sample size we used the non-paramatric Mann Whitney \( U \) test for determining pretreatment differences between both clusters. One of the teachers was not observed at the start of the PDP.

From the data presented in Table 3 it can be derived that, in general the teachers in the second cluster already had higher mean scores on almost all items before the PDP started. In terms of significant differences, the cluster 2-teachers used differentiation practices significantly more and their lesson quality was also significantly higher than that of the teachers in cluster 1. The teachers in cluster 2 also tended to score higher on attitudes towards the basic skills than their cluster 1 counterparts (\( p=.07 \)). Notably, the teachers in cluster 1 significantly used more modeling than the cluster 2-teachers (who did not use it at all), although only few cluster 1-teachers were found to use modeling.
Table 3: Pre-Treatment Differences between Teachers in Cluster 1 and 2

<table>
<thead>
<tr>
<th>Item/scale</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Man-Whitney U-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>Mdn</td>
</tr>
<tr>
<td><strong>Questionnaire</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude towards basic skills</td>
<td>16</td>
<td>3.76</td>
<td>4.00</td>
</tr>
<tr>
<td>Attitude towards standardized assessments</td>
<td>15</td>
<td>2.93</td>
<td>3.00</td>
</tr>
<tr>
<td>Use of explicit performance levels for students</td>
<td>14</td>
<td>2.64</td>
<td>2.00</td>
</tr>
<tr>
<td>Acquaintance with performance standards</td>
<td>16</td>
<td>1.56</td>
<td>2.00</td>
</tr>
<tr>
<td>Use of within-class grouping during math lessons</td>
<td>16</td>
<td>1.88</td>
<td>2.00</td>
</tr>
<tr>
<td>Use of within-class grouping over all subjects</td>
<td>16</td>
<td>1.38</td>
<td>1.00</td>
</tr>
<tr>
<td>Differentiation attitudes</td>
<td>16</td>
<td>2.38</td>
<td>3.00</td>
</tr>
<tr>
<td>Ambitious mind-set</td>
<td>16</td>
<td>3.94</td>
<td>4.00</td>
</tr>
<tr>
<td><strong>Observed math lesson</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct Instruction</td>
<td>18</td>
<td>1.94</td>
<td>2.00</td>
</tr>
<tr>
<td>Modeling</td>
<td>18</td>
<td>.22</td>
<td>.00</td>
</tr>
<tr>
<td>Informal diagnostic assessment</td>
<td>18</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Differentiation</td>
<td>18</td>
<td>.72</td>
<td>1.00</td>
</tr>
<tr>
<td>Lesson quality</td>
<td>18</td>
<td>7.94</td>
<td>8.00</td>
</tr>
</tbody>
</table>

*: p<.05, #: p<.10; #: scales; 1#: Since the median scores were mostly the same for both groups, the pooled mean scores are also presented as the central tendency measure.

Next, the level of implementation was studied. For this purpose we compared pre- and posttest data within the two clusters on both the questionnaire and the lesson observations, using the Wilcoxon rank sum test (see Table 4). Considering teacher change in cluster 1 first, it can be derived from Table 4 that these teachers had become significantly more aware of performance standards, reported to differentiate significantly more over all subjects, and had less positive attitudes towards differentiation. The cluster

24 No information on the scale goal-oriented working is provided by Table 3, since the teachers did not fill in questions on their goal-oriented working on the pretest.
2-teachers also became more aware of performance standards. Moreover, they were found to make significantly more use of explicit performance levels in math, they used modeling significantly more (albeit still not much), and their overall lesson quality had improved somewhat \( (p=.09) \). Overall, it can be concluded that on most items no noticeable changes had taken place during the PDP, both in cluster 1 and cluster 2. In many respects, the posttest scores of the teachers in cluster 2 were still higher than the scores of the teachers in cluster 1, meaning that they both had developed in a rather similar way. Both before and after the teacher training, the teachers in both clusters did not reach the maximum scores.
Finally, we investigated whether relatively strong DGB-behavior by the teachers was associated with higher student achievement in math, as was hypothesized. For this purpose, multilevel models were estimated that were based on data of 471 students. Averaged over the 5 imputed datasets, teachers in cluster 1 taught 224 students. Although the prevalence of cluster 2-teachers (45%) was lower than that of the cluster 1-teachers (55%), they taught a larger number of students: 247 students.

To estimate whether relatively strong use of DGB-behavior positively predicted the math posttest, we first estimated an Empty model (see Table 5). In the Covariate Model, the significant predictors math pretest, gender, and class size were taken into account. The Covariate model showed that, all being equal, students with higher prior

---

\[\text{In Table 4, abbreviations are used to denote the item names. See the Methods section and Appendix A for their full name, a short description, and examples of the items used.}\]
math achievement performed better on the posttest than students with lower prior math achievement, boys outperformed girls, and students in small classes performed better than those in larger classes. Deviance testing showed that the Covariate Model fitted the data significantly better than the Empty model: the deviance decreased by 374.34. This exceeds the critical value (16.27) in a chi-square distribution with $df=3$ for $p=.001$. As a next step, we added the identified teacher profiles in the Cluster Model. Here, a pooled model was estimated. From the data presented in Table 5 it can be derived that, while controlling for the covariates, students who were taught by teachers in cluster 2 did not significantly perform better on the math posttest than students who were taught by teachers in cluster 1 ($\chi^2=0.25$, $df=1$, $p=.62$). The Cluster Model thus did not fit the data better than the Covariate Model, indicating that the teacher profiles were not significantly related to posttest scores. If anything, a negative average effect size of $d=-.02$ was found. According to Cohen’s classification this would amount to a negligible negative effect of relatively strong DGB-behavior.

Note that for the analyses presented in Table 5 students were considered to be nested in classes. Analyses in which students were nested in classes and teachers, and in which students were nested in teachers, were also carried out. These showed comparable results: relatively strong data use was not significantly associated with the math achievement of students.
Table 5: Multilevel Analyses Predicting Students’ Math achievement, Pooled Estimates

<table>
<thead>
<tr>
<th></th>
<th>Empty model</th>
<th></th>
<th>Covariate model</th>
<th></th>
<th>Cluster model - pooled</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>SE</td>
<td>( \beta )</td>
<td>SE</td>
<td>( \beta )</td>
<td>SE</td>
</tr>
<tr>
<td>Response: Math posttest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Part</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>73.10</td>
<td>1.44</td>
<td>36.54</td>
<td>2.39</td>
<td>34.42</td>
<td>2.08</td>
</tr>
<tr>
<td>Math pretest</td>
<td>.74*</td>
<td>.03</td>
<td>.74*</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>-3.73*</td>
<td>.98</td>
<td>-3.53*</td>
<td>.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class size</td>
<td>-.18*</td>
<td>.08</td>
<td>-.17</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher profile (cluster 1=reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Part</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student level</td>
<td>218.35</td>
<td>14.88</td>
<td>106.72</td>
<td>7.23</td>
<td>106.78</td>
<td>7.24</td>
</tr>
<tr>
<td>Class level</td>
<td>60.83</td>
<td>18.74</td>
<td>5.32</td>
<td>3.32</td>
<td>5.13</td>
<td>3.27</td>
</tr>
<tr>
<td>-2*loglikelihood</td>
<td>3928.28</td>
<td></td>
<td>3553.94</td>
<td></td>
<td>3553.69</td>
<td></td>
</tr>
<tr>
<td>Number of classes</td>
<td>42</td>
<td></td>
<td>42</td>
<td></td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Number of students</td>
<td>471</td>
<td></td>
<td>471</td>
<td></td>
<td>471</td>
<td></td>
</tr>
</tbody>
</table>

*: \( p < 0.05 \)

5.5 Discussion

5.5.1 Main findings

The purpose of the current chapter was to gain a better insight in whether and how professionalisation of primary school teachers was related to their attitudes and behavior. More specifically, we focused on teachers’ levels of data-mindedness and their relation to students’ math achievement.

Our first line of investigation was directed at identifying whether attitudes and self-reported behavior in terms of data-mindedness were more positive for teachers who participated in the PDP than for teachers who did not (Hypothesis 1). The explorative, descriptive analyses showed that the attitudes and self-reported behavior of teachers who were involved in the PDP resembled those of the control teachers in most respects, although we found teachers in the experimental condition to be more aware of measurable outcomes and of explicit desired levels for their students’ math proficiency. Our first hypothesis was therefore partially confirmed.

Our second line of inquiry concentrated on whether different teacher profiles could be identified in terms of their data-mindedness. For estimating the profiles, we restricted ourselves in basing them on three scales: Goal-oriented working and Attitude...
towards basic skills, both derived from the questionnaire, and teachers' Differentiation practices, based on observed math lessons. Using latent class cluster analysis, two profiles were distinguished. Investigating whether all three scales were related to variation in allocation in different categories per cluster showed that only the scale differentiation practices discriminated significantly. Yet univariate analyses, using Wilcoxon Ranked Sum tests, showed that the two clusters differed on all three scales: Teachers in cluster 2 were a) more aware of the desired levels of students’ math proficiency, b) more aware of measurable outcomes, and c) acted more upon differences between students. The groups were therefore described in terms of teachers’ use of so-called DGB-behavior (where DGB represented differentiation, goals, basic skills). Levels of teachers’ DGB-behavior were assumed to reflect their levels of data-mindedness.

The third line of investigation focused on whether the identified teacher profiles were related to initial differences, to implementation of the PDP, and to students’ math performance. When relating the teacher profiles to teachers’ prior levels of data use and teaching quality, we found that there were already differences between the two clusters at baseline measurement. The teachers in cluster 2 had higher scores on almost all items, with two significant differences: These teachers already were more aware of their educational outcomes and their teaching quality in terms of differentiation and overall lesson quality was already better. When looking at teacher change, comparing teachers’ scores at the start and at the end of the PDP, we saw that the teachers in both clusters scored significantly higher on their acquaintance with performance standards at the end of the PDP. The teachers in cluster 2 also had significantly higher scores on their use of explicit math performance levels for their students, and on their overall lesson quality at the end of the PDP, but maximum scores on both the questionnaire and the observation items were not found. For most items, teachers did not change significantly. Beforehand, we expected teachers who showed more data-mindedness at the end of the PDP, that is, those teachers who showed more DGB-behavior, to have implemented the program to a higher extent (Hypothesis 2). Our findings did not confirm this hypothesis: teachers in both clusters did not seem to have sufficiently implemented the program content as intended by the researchers and most initial differences between the two clusters had remained at the end of the PDP.

We further expected that being taught by teachers with relatively high levels of DGB-behavior was related to better math achievement of their students (Hypothesis 3). Relating the two identified teacher profiles to students’ math achievement, while controlling for relevant covariates, showed that the math achievement of students who
were taught by teachers who used more DGB-behavior did not significantly differ from the math achievement of students who were taught by teachers who showed little DGB-behavior.

5.5.2 Interpreting the study findings

Our PDP on data use adhered to the five core features of effective PD (Desimone, 2009): the PDP addressed content (effective instructional methods and math specific recommendations), it contained inquiry-based learning (data discussions, feedback on observations), it was coherent with one of the spearheads of educational policy (data use), it had sufficient duration (it lasted one year and teachers spent approximately 40 hours), and teachers collaborated in professional learning communities. Nevertheless, the PDP only yielded very limited changes in teacher behavior and no change in teacher attitudes. In this discussion we try to understand our findings by discussing four possible explanations.

A first explanation for not finding adequate teacher change might be that teachers need sufficient time to adopt and fully implement new teaching practices (Fullan, 2007; Hochberg and Desimone, 2010). In this respect, the limited implementation of our PDP might indicate that it could have taken more time for teachers to incorporate and implement the full concept of data use (Desimone, 2009; Shaw & Wayman, 2012; Wayne, Yoon, Zhu, Cronen, & Garet, 2008).

A second explanation for finding very limited implementation of the program might be that several contextual features were not taken into account. For instance, we did not consider relevant features like the school culture or the principal’s educational leadership, both known to be facilitators for effective data use (Anderson et al., 2010; Earl & Fullan, 2003; Park & Datnow, 2009; Park et al., 2013). Such facilitating factors at school level could have influenced teachers’ levels of implementation. Confounding variables at class level were also not discussed. For instance, the specific math curricular textbooks that were used at the school might have affected program implementation, because some offered more elaborate instructional suggestions than others. The same holds for class size, since it might be easier for teachers in smaller classes to adapt teaching to differences between students. Checks were carried out to see whether these class level variables acted as confounders showing no relationship to both the clustering and implementation, except for the variable class size, which was controlled for in the outcome analysis.
A third explanation might be that teachers adapted the information provided by the PDP, leading to only partial implementation of the promoted way of working (Pajares, 1992; Spillane et al., 2002, Timperley, et al., 2007). Teachers might thus have been selective in which elements of data use to adopt. At the end of the PDP, teachers who had participated in our PDP differed from those in the control condition in their goal-orientedness, a core feature of DDDM. Moreover, the cluster 2-teachers had become more goal-oriented than the teachers in cluster 1. Being more goal-oriented fits the data-informed way of working that is promoted by both the Dutch educational inspectorate (Educational Inspectorate, 2010) and setting goals seems to be relatively easy (compared to really acting upon them by tailoring instruction). Teachers might therefore have been motivated to implement the goal setting part, but changing their instructional practices could have been somewhat more difficult to achieve. Especially when the new knowledge and practices resembled what they already did to some extent, this might have caused a situation in which teachers did not feel the need to change. Although, beforehand, the teachers had not reached the maximum scores in terms of teaching practices like direct instruction, modeling, or differentiation, our PDP apparently did not trigger the teachers to improve their instruction to the upper limit.

A fourth explanation for the very limited program implementation might be that the obtained information was not suitable for measuring the concept of data use or data-mindedness, as the measures used were both too limited and indirect. Our measures of data-mindedness mainly concentrated on teachers’ goal-orientedness, differentiation practices, and awareness of measurable outcomes. When comparing the two clusters for determining baseline differences and levels of implementation we also looked at instructional methods that were addressed in the PDP, like direct instruction and modeling. These all measured attitudes or behavior that are considered to be strongly related to data use, but they did not provide evidence on whether teachers actually used the data to inform their decisions. The concept of data use is further far more wide-ranging than our items conveyed. At the class level, teachers’ knowledge and skills in interpreting analyses from the student monitoring system (Staman et al., 2014; van der Kleij & Eggen, 2013), teachers’ skills in bringing together performance information from formal and informal assessments, teachers’ adequate determination of students’ root problems (Goertz, et al., 2009), and their application of content knowledge and pedagogical content knowledge to deliberately and effectively plan and carry out instructional adjustments based on the data (Timperley & Parr, 2009), should ideally all be taken into account in assessing teachers’ use of data. We acknowledge that the items
used in this study were by no means comprehensive. As a result, we might not have measured the core features from which implementation of data use, or data-mindedness, could be derived.

5.5.3 Limitations

Several limitations to the current study should be addressed. First, as was mentioned above, the study did not take all aspects of data use into account, even if the teacher characteristics used reflect aspects of data-mindedness and consequent behavior. The set-up of the assignment procedure of the teacher questionnaire, having two versions randomly assigned, prohibited us from taking into account teacher information on issues like the types of analyses used by the teachers, teachers’ current collaborative behavior, their reflective activities—based on proficiency information—, and their willingness to develop themselves professionally. Using these data would have caused us to severely reduce our sample size, leading to a considerable loss of power. Should we have been able to take such information into account, the question would still remain what PDPs on a formative use of data, their implementation, and their evaluation should look like (Schildkamp et al., 2014). Clarity in this respect is considered crucial given the widely recognized need for PD in data use (Mandinach, 2012; Datnow, et al., 2007; Wayman & Jimerson, 2014). Future research should therefore strongly focus on gaining knowledge on the core features of effective PDP in data use and on how the effects of such PDPs can be measured.

A second limitation is that we only measured whether the attitudes and practices of teachers in the experimental condition changed, when comparing their pre- and posttest data. However, we have no indications why the implementation was only very limited. Teacher emotional states, like not feeling confident enough to make data actionable, might have played a role (Hochberg & Desimone, 2010; Schmidt & Datnow, 2005; Saunders, 2013). Motivational aspects could also have influenced implementation (Timperley et al., 2007). Since the PDP was not explicitly based on problems that were felt by the teachers, the teachers might not have considered the PD-content to be relevant for their daily practice (Timperley, et al., 2007). Moreover, the rationale and the main purpose of the PDP might also not have been sufficiently clear to the participants, since no explicit focus on sense-making and co-construction of the meaning and purpose of data use was used (Jimerson, 2014; Wayman et al., 2012; Spillane, et al., 2002). These issues might have been addressed too casually in our PDP. Such co-construction would also have addressed the prevalent association of data use with accountability pressures.
(Jimerson & Wayman, 2014). Additional knowledge on these issues would have helped us to better understand why only very limited teacher change took place.

The lack of using qualitative information on teachers’ (instructional) practices can be considered a third limitation. Combining questionnaire and observational information on teacher characteristics gave us some indication on teachers’ data-mindedness and teaching practices, but the dichotomous items in our observations did not provide us with elaborate information on whether or how teachers tailored their instruction, questioning, and materials to students’ needs. In studies on differentiation, more specifically on grouping, such instructional modification has been found to be the core factor for its effectiveness (Lou, et al., 1996). Future studies should focus more on understanding the relationship between data and the actual instructional changes that teachers make by using different sorts of data. Not only the presence of specific teaching practices should be taken into account, but also more detailed and qualitative information on teacher-student interaction and on teachers’ intentions and planning. A mixed-methods approach is considered to extend our knowledge in this respect.
## Appendix A: Items and scales used in the questionnaire

<table>
<thead>
<tr>
<th>Questionnaire item</th>
<th>Example</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic skills *</td>
<td>‘I think it is important to mainly focus on math.’</td>
<td>Scale</td>
</tr>
<tr>
<td>Goal-oriented working *</td>
<td>‘I think it’s proper to tell students what I expect from them and in what topics they still have to improve.’</td>
<td>Scale</td>
</tr>
<tr>
<td>Items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude towards Standardized assessments</td>
<td>‘How important is standardized assessment information for receiving information on students?’</td>
<td>Separate item</td>
</tr>
<tr>
<td>Attitude towards differentiation</td>
<td>‘I think that low performers should get most teacher attention’</td>
<td>Averaged score</td>
</tr>
<tr>
<td>Use of within-class grouping over all basic skills</td>
<td>This is the average number of subgroups mentioned by teachers over all subgroups</td>
<td>Averaged score</td>
</tr>
<tr>
<td>Use of within-class grouping during math lessons</td>
<td>‘How many subgroups do you distinguish in your math lessons?’</td>
<td>Separate item</td>
</tr>
<tr>
<td>Use of explicit performance levels for students in math</td>
<td>‘Did you define the desired level of math performance for each student, for the end of the year?’</td>
<td>Separate item</td>
</tr>
<tr>
<td>Acquaintance with performance standards</td>
<td>‘Are you aware of performance standards?’</td>
<td>Separate item</td>
</tr>
<tr>
<td>Ambitious mind-set</td>
<td>‘I always strive for a higher proficiency level of the students, even in cases that they already perform well’</td>
<td>Separate item</td>
</tr>
</tbody>
</table>

*: identified scales