Appendices

- Appendix 1: Overview of the professional development program and specifications per meeting
  - 1. Overview of the PDP on data use
  - 2. Scheduling of the PDP’s meetings
  - 3. General set-up of each meeting
  - 4. Specifications of the content per meeting
- Appendix 2: The Dutch educational context in relation to the PDP’s three components
  - 1. Recent implementation of standards in the Netherlands
  - 2. The Cito LOVS standardized assessment system and student monitoring systems in the Netherlands
  - 3. General performance expectations, common instructional practices, and the curriculum
- Appendix 3: Overview of the cutscores and performance categories
- Appendix 4: Description of the data analyses used in the PDP
- Appendix 5: Lessons learned from the pilot study
- Appendix 6: Observation instrument
Appendix 1: Overview of the professional development program and specifications per meeting

The studies presented in this dissertation were conducted to evaluate the effectiveness of a teacher Professional Development Program (PDP) on data use. This PDP aimed to improve students’ reading comprehension and students’ mathematics due to performance concerns for both areas (Ministry of Education, 2010). All aspects of the program which have been reported in this dissertation in relation to mathematics (for example, setting standards and performance goals) have also been conducted for reading comprehension. The content of the Appendices will pertain to the PDP as it was delivered to the participants, and will thus pertain to both subject areas. The results of the program with respect to reading comprehension were discussed in the dissertation of van Kuijk (2014).

In this section, the reader can find detailed information on the overall rationale behind the PDP, its aims, and the way it has been conducted. This overview provides, for example, empirical support and practical arguments underlying the specifications of the program. In this way, we want to give a clear account of how our program was realized and provide other researchers with the necessary information to replicate our study. The following sections and paragraphs therefore mainly contain information of specific interest. First, we will provide an account of the entire PDP. Next, an overview of the program’s meetings is presented, and several general characteristics of these meetings - such as delivery format and duration – will be discussed. After this, each meeting is dealt with separately, describing its aim, content, the materials used, and the related homework assignments.

1. Overview of the multicomponent PDP

The teachers that participated in the PDP were supported in improving their practice with help of a three-component program on data use. These components were considered to reflect the core elements of data use (Visscher & Ehren, 2011; Black & William, 2009): 1) setting standards and performance goals for every student, 2) using data-analyses on students’ actual performance levels, and 3) acquiring relevant instructional skills and knowledge in reading comprehension and mathematics. All three components have shown to be positively related to student performance. The PDP was designed to foster student learning through teachers’ application of a multicomponent package. Given the
interrelatedness of its components, their integration into one synergetic package was supposed to foster more goal-oriented, focused, clear instruction, that better suited to students’ needs. In the paragraphs below, the information on each of the three components is briefly resumed (for more information, see Chapter 4 in this dissertation).

1.1. **Component 1: Establishing the desired level of students’ performance**

Goal setting was incorporated as the first component in the PDP since the insufficient results of Dutch students on both international and national assessments were attributed to the fact that, for schools and teachers, it was unclear what students should know and do at certain points in time (Expert group Continuous Learning Progression, 2008). Setting goals leads to an explicit notion of the desired performance level. Thereby, they make clear how success can be attained, they make teachers’ actions more meaningful, and enhance motivation to reach them (e.g., Fuchs et al., 1985; Locke & Latham, 1990). The goals were based on performance categories which had been identified by the participants using a *standard setting procedure* (see Deunk et al., 2014). As discussed in Chapter 2, we aimed to assist teachers in setting goals that were ‘challenging but realistic’ given their students’ capabilities. For this purpose, we developed a *step-by-step procedure* which incorporated performance data analysis and team discussion to help teachers reflect on and reconsider the appropriateness of their initial expectations before deciding on their final goals—following recommendations of the data use literature (e.g., Schildkamp & Kuiper, 2010).

1.2. **Component 2: Determining students’ actual level of performance**

In order to help the participating teachers set and attain the performance goals, it was important that they based their instructional decisions on performance data (e.g., Guskey, 2002). Using student performance data to adapt one’s teaching in order to better meet students’ needs is known as *formative assessment* (Black & Wiliam, 1998; Herman et al., 2010). The participants therefore received training in the use of the student monitoring system. Yet the concept of *performance data* not only pertains to the assessment results on standardized tests, but also, for example, to completed work book assignments or teacher observations of how students function in class (Lai & Schildkamp, 2013). Teachers’ reflection was targeted by focusing their attention on different sources of data they could use to enhance certainty on students’ actual levels of proficiency (Bennett, 2011) and therefore to help them make instructional decisions that would better suit different students’ needs. In the mathematics domain, this resulted in equipping teachers with knowledge and practice on how to conduct diagnostic math interviews, which would
allow them to get more detailed insight in students’ levels of mathematical understanding, their root problems, and misconceptions. Thus, by working with student-specific performance goals and monitoring performance in relation to these goals (i.e., components one and two), it was expected that teachers would attend more to different student needs, resulting in the use of differentiation. An important prerequisite, however, is that teachers translate the data into adjusted practices, a step which is not always guaranteed (Goertz, et al., 2009, Heritage et al., 2009). The third component of our PDP focused on how to take action after analyzing the data.

1.3. Component 3: Instructional modifications to close the gap between the desired and the actual level

In the PDP, next to setting the performance goals and monitoring the progress made toward them, it was important to help the teachers attain their own objectives by ensuring that they were sufficiently equipped with the most relevant instructional skills and knowledge about reading comprehension and mathematics. We targeted Direct Instruction (a teacher-centered model of instruction) and modeling (an instructional technique in which the teacher demonstrates how to apply a certain reading comprehension or mathematics strategy by "thinking aloud"; to show students which strategies are appropriate, how to pursue them and why). After having been introduced to these concepts, the teachers practiced and received feedback on the implementation of the instructional methods. In addition, the participants were informed on the set-up of the Cito-assessments for mathematics and reading comprehension in the grades under study. Here the degree of alignment between these assessments and the curricular text books used in their schools was discussed and tips were provided on how to bridge evident gaps. Specifically for reading comprehension, we discussed important determinants of reading performance and key concepts in the second- and third-grade reading comprehension curriculum.

In Figure 1, the interrelatedness of the components is illustrated. This graphical representation is given at the beginning of the detailed descriptions of the meetings (presented below). For each meeting, the most essential component is highlighted.
2. Scheduling of the PDP’s meetings

Throughout the school year, the time investment of the teachers was scheduled for 40 hours, including attending the nine after-school meetings and homework assignments. In Figure 2, a graphical overview of the nine meetings is presented. The three components of the PDP were addressed to comparatively the same extent. The standards and goal setting component (the first component) was addressed in meetings 1, 3 and 4, respectively. Information and training on data-analysis (our second component) was dealt with in meetings 2, 5, 6 and 9. Training in relevant instructional practices and information to improve teachers’ (content and curriculum) knowledge, the program’s third component, was targeted in meetings 6, 7, and 8. In the majority of the nine meetings, the subject areas of mathematics and reading comprehension were targeted simultaneously. In meeting 1 and 6, however, the emphasis was specifically on mathematics, while meeting 3 and 7 particularly addressed reading comprehension.
### 3. General set-up of each meeting

The standard setting meetings and the meetings on formative assessment and data-analyses (meeting 1, 2, 3, and 5) were set up as general gatherings which were held in a convention centre. These meetings were scheduled to last 2.5 hours. The other meetings (meeting 4, and 6 to 9) took place at the individual schools; in a few cases, the participants of two or three different schools joined together in one meeting. These meetings were scheduled to last 1.5 hours.
All meetings followed the same order: a) recapitulation of the last meeting, b) overview of the content to be discussed in the current meeting, c) presentation of information, d) recapitulation of the current meeting, and e) preview of the next meeting. The structure of these meetings resembled the elements in the Direct Instruction model, an effective instructional approach which was also discussed during the meetings.

Different delivery modes were used during the meetings. We provided short lectures using whole-group (power point) presentations. During several of these whole-group presentations, video-fragments were shown as illustrative material. In addition, during almost all meetings, the participants were asked to work on assignments “on the spot”. These assignments could be individual assignments, assignments requiring collaboration between colleagues from the same school, and assignments requiring collaboration between colleagues from other schools. Use of the digital student monitoring system (part of meetings 2 and 5) was practiced behind laptops. For the entire program, the majority of the hand-outs and other materials offered were self-developed. To organize the hand-outs and materials which were distributed during the meetings, we provided a binder to all participants.

At the beginning of each meeting, the attendance of the participating teachers, school principals and senior support coordinators was registered. At the end of each meeting, the participants were asked to fill in an evaluation form about how they had experienced the value and practicality of that particular meeting. During the standard setting meetings (meetings 1 and 3), specific questions were posed about the different rounds in the standard setting procedure and the degree to which the participants considered their own cutscores as well-considered. This was done because on-site evaluations by participants serve as an important check on the validity of the cutscores and the way in which they have been set (Cizek & Bunch, 2006; Hambleton & Pito, 2006).

4. Specifications of the content per meeting

In this section, the content of the nine meetings will be addressed separately. For each meeting the most vital component is highlighted in the figure next to the headings.
1: Introduction to the project and standard setting for mathematics

Summary of the first meeting: The goal of this meeting was a) to inform the participants on the set up of the PD, and b) to set cutscores and create performance categories for the second and third grade (end-of-the-school year) June-mathematics assessments using a standard setting procedure. This procedure was also conducted to facilitate and stimulate the teachers’ awareness of their own performance expectations and the instructional and curricular demands for the second- and third-grade mathematics.

Characteristics

<table>
<thead>
<tr>
<th>Format of meeting</th>
<th>General meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time span</td>
<td>2.5 hours</td>
</tr>
<tr>
<td>Delivery mode</td>
<td>Whole-group presentation, small-group discussion</td>
</tr>
<tr>
<td>Hand-outs</td>
<td>Ordered Item Booklet, standard setting forms, training materials for standard setting procedure, print-outs of powerpoint slides</td>
</tr>
</tbody>
</table>

Introduction to the project

At the beginning of the first meeting, a short outline of the project was presented to the participants. We briefly introduced the three components of the PDP as well as their timing throughout the school year. In addition, the participants were informed about certain data collection obligations related to their participation in the PD study (i.e., having a pre- and post-observation during the mathematics and reading comprehension lessons, filling in a questionnaire at the beginning and at the end of the PDP, and providing performance data to the researchers prior to several meetings).

Introducing performance standards

The participants were explained that specifying what students should be able to know and do would help them in making teaching more directed toward the performance of these desired objectives. We argued that having clear performance goals makes it easier to target instruction toward the performance of these outcomes (Fuchs, Fuchs, & Deno, 1985) and that this targeted instruction is expected to improve student results (Lauer et
al., 2005; Roeber, 1999)\textsuperscript{26}. For this purpose, the participants were asked to participate in the standard setting procedure, in order to establish performance categories. The benefit of working with goals that are based on performance categories was that teachers had reflected on desirable learning objectives for their students and that the performance of these goals would be easily established by conducting standardized assessments in class. Furthermore, we provided information on the current Dutch educational policy regarding the performance standards (as discussed in Appendix 2). By introducing them to this background, we hoped that the participants would recognize the value of working with performance goals, which would in turn positively influence their commitment to the PDP.

**Standard setting for mathematics**

Within the PDP, the Cito adaptation of the Bookmark procedure (see Van der Schoot, 2009) was used. Here, participants considered a selection of items from the Cito mathematics assessments and indicated which items they expected the students to answer correctly in the June-assessment. Note that for the remainder of this Appendix, the end-of-the-school year assessment is referred to as the *June-assessment* and the midway-of-the-school year assessment is referred to as the *January-assessment*. To facilitate the standard setting task, the items were ordered in such a way that they increased in difficulty and we presented them in a so-called *Ordered Item Booklet* (OIB)\textsuperscript{27}. The explication of performance expectations is done for students of different ability levels. In this PD, five performance categories were distinguished: below minimum, minimum, basic, proficient, and advanced. For each category, the participants decided at which item there was a suitable cutoff (between *below minimum* and *minimum*, between *minimum* and *basic*, etcetera). Thus, four cutoff points needed to be identified in total\textsuperscript{28}. Since all items can be converted into scores on the assessment scale, indicating a pupil’s level of proficiency, the cutoff points can also be converted into scores: these points are thus also referred to as *cutoff scores* or *cutscores*. The final cutscores were determined after three rounds. In the first round, the panelists individually studied

\textsuperscript{26} Positive findings are provided by the goal setting theory (Locke & Latham, 1990; 2002). For more information, see Chapter 1 of this dissertation.

\textsuperscript{27} Both the items and information on their difficulty (for which Item Response Theory was used) were provided by the Netherlands Institute for Educational Measurement, with whom we collaborated in this study.

\textsuperscript{28} Similar to the work of Roeleveld and Béguin (2009), four cutoff scores were set in the PDP.
the OIB and formulated cutoff scores based on their own opinion. In the second round, they came together in small groups (consisting of three to five people) and discussed their cutoff scores. Groups could reach consensus, but they did not have to. After this small group discussion each participant reset his/her scores. Based on these reset scores, the median cutscores were calculated ‘on the spot’, while the participants listened to a further explanation of the PD project and the three components. In the third round, the average cutscores of the group were presented and compared to the actual performance data of the student population (their “empirical equivalents”, see Chapter 3). These empirical data indicated the participants how realistic and ambitious their own cutoff scores were at that stage in the procedure. After this display, the panelists again reset their cutoff scores, after which the final ones were calculated by taking the median of the scores of the third round. This was done, however, at a later moment: the final scores were presented during the second PD meeting. After collecting the forms containing the final cutscores, the first meeting was brought to an end. More information on the standard setting procedure is provided in Deunk et al., 2014.

2. Formative assessment and data use

Summary of the second meeting: The goal of this meeting was to present the concept of formative assessment and data use. The participants also practiced different types of data analyses using the digital Cito student monitoring system (detailed further in Appendices 2 and 4).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>General meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of meeting</td>
<td>General meeting</td>
</tr>
<tr>
<td>Time span</td>
<td>2.5 hours</td>
</tr>
<tr>
<td>Delivery mode</td>
<td>Whole-group presentation, practice of data analysis behind laptops (in pairs)</td>
</tr>
<tr>
<td>Hand-outs</td>
<td>Data analysis booklets (containing navigational directions and screen shots of the student monitoring system), print-outs of power point slides, homework assignments for the teachers and the school principal/internal support coordinator</td>
</tr>
</tbody>
</table>

Recapitulation of meeting 1 and presentation of the final cutscores

At the beginning of the second meeting, the final mathematics cutscores were presented (their exact scores are presented in Appendix 3). The student performance at the level of
the cutscores was illustrated using several exemplary items. All performance categories (below minimum, minimum, basic, proficient, and advanced) and their accompanying test score intervals (also in Appendix 3) were discussed in order to explain to the participants how students’ June-assessment results would be converted into one of the five performance categories.

**A model for data use**

Next, the standards were coupled to the other two components of the PDP and these three components were briefly re-discussed. The second component focused on the use of assessment results, and other data on students, to improve instruction. The teachers were stimulated to use assessment data as feedback to help them modify their teaching activities in such a way that they met the individual students’ needs and thereby furthered their development. This practice is also known as *formative assessment* (Black & William, 1998). During the PD, the *evaluative cycle for data-driven teaching*, a model from Ledoux, Blok and Boogaard (2009) was used. This model was similar to models of formative assessment (Black & Wiliam, 2009; Carlson et al., 2011; Herman et al., 2010) and the well-known Plan-Do-Check-Act cycle (Deming, 1986). The evaluative cycle for data-driven decision making (DDDM), or data use, was used to help the teachers carrying out the three core elements of data use (establishing the desired performance level, determining the actual performance level, and conducting instructional adjustments in order to reach the goal) in a reflective and structured way. Teachers’ reflection was targeted in this model as well, by focusing their attention on their students’ performance in relation to their own teaching practice. The model consisted of five questions being:

1) **What do I want to accomplish with my students?**
2) **Which sources of information can I use to map out the proficiency of my students?**
3) **How are my students performing based on these different sources of information?**
4) **What do these performance results mean? Can I interpret them?**
5) **How have I been teaching my students and do I need to adapt my approach and/or goals?**

In order to “get acquainted” with this model, an introductory assignment was provided. The participants had to reconstruct the logic of the model by putting its elements (i.e., the five questions) in the right order. This activity was conducted in small groups consisting of participants from different schools. After the introductory assignment, the definition of “data-driven decision making” (a translation of the Dutch term “opbrengstgericht
werken”) was addressed. Empirical evidence for employing a data-driven way of teaching (e.g., Carlson et al., 2011) was presented as well. The experiences of several Dutch school teams who already worked in a data-driven manner were illustrated using video-material (Primary Education Council, 2009). Next, we elaborated on each of the five questions of the evaluative cycle for DDDM. With respect to the second question (Which sources of information can I use to map out the proficiency of my students?), we provided additional information on the difference between the Cito-assessments and the assessments of the curricular textbooks - also see Appendix 2 for more information on this topic. Additional information on the differences between the various student monitoring systems was provided as well (discussed in Appendix 2).

*Making use of the student monitoring system*

In the Netherlands, the use of student monitoring systems by teachers is still rather limited for analyzing problems and adapting instruction, and teachers who do use the student monitoring systems are often unaware of the possibilities for more sophisticated analyses (Ledoux et al., 2009; Meijer & Ledoux, 2011; Schildkamp & Kuiper, 2010; van der Kleij & Eggen, 2013). During this meeting, the participants were informed on three specific types of analysis which could be performed using data from the Cito assessments. These three types of analysis are called a) *the estimation of future performance*, b) *an overview of performance in the previous school year*, and c) *a progress report* (analysis of academic growth). For their exact content, see Appendix 4A, B, and C. These three analyses provided information on (estimates of) individual students’ performance and would be used in homework assignments later on in the PDP.

The participants practiced these analyses behind laptops (in pairs) while the researchers walked around to answer questions and give immediate feedback. The hand-outs for these exercises contained screen shots and directions on how to navigate through the student monitoring system. These exercises also contained additional information and questions to help the participants interpret and critically reflect on the output provided by the system.

*Homework assignment*

The participating teachers as well as the school principals and senior support coordinators - the latter two referred to as the *school management staff* - received booklets with specific homework assignments. The teachers were provided with a three-staged assignment, which focused on their own students’ performance in mathematics. This three-staged assignment has already been discussed as part of the step-by-step
procedure in Chapter 1; here, its content is recapitulated. First, we asked the teachers to make their initial expectations for their students’ future performance on the June-assessment explicit, using the performance category classification (below minimum, minimum, basic, proficient, and advanced). Second, the participants were asked to use the student monitoring system to compute the future performance estimation, which is based on prior achievements (see Appendix 4A). Third, we requested them to use their student monitoring system to present an overview of last year’s performance on the June-assessment (see Appendix 4B). Both the estimates and the performance scores could be converted to the performance category classification. The assignment booklet also contained questions about possible differences between the three assignments, like ‘Are there differences between your own intuitive prediction and the Cito future performance estimate? If so, can you explain them?’ These questions were developed to increase the teachers’ affinity with the performance categories and to stimulate critical reflection on their own performance expectations. The homework assignment was requested to be completed prior to the program’s fourth meeting, when the teachers would have to set performance goals for their students.

The school management staff was provided with a different assignment. The participants were asked to use the student monitoring system to compute the developmental growth (see Appendix 4C) of the current second- and third-grade students with respect to the prior school year, i.e. from the January-assessment to the June-assessment in grades 1 and 2. This assignment would offer more insight into how the students who were currently in second and third grades had developed the past school year. The assignment pertained to both the mathematics and the reading comprehension development of the students, thereby anticipating the subject area of interest in the next meeting. Similar to the homework assignment of the teachers, this task was requested to be completed before the fourth meeting.
3. Standard setting for reading comprehension

Summary of the third meeting: The goal of this meeting was to set cutscores and create performance categories for reading comprehension using the standard setting procedure, and to increase the participants’ awareness of the performance expectations and instructional demands with respect to second- and third-grade reading comprehension.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of meeting</td>
<td>General meeting</td>
</tr>
<tr>
<td>Time span</td>
<td>2.5 hours</td>
</tr>
<tr>
<td>Delivery mode</td>
<td>Whole-group presentation, small-group discussion</td>
</tr>
<tr>
<td>Handouts</td>
<td>Ordered Item Booklet, standard setting forms, training materials for standard setting procedure, print-outs of power point slides, homework assignments for teachers</td>
</tr>
</tbody>
</table>

Standard setting for reading comprehension

The cutscores for reading comprehension were set in the exact same way as those for mathematics. The only difference between both standard setting procedures pertained to the construction of the Ordered Item Booklet (OIB). The reading comprehension items increased in their difficulty, but the items that applied to the same text were grouped together to improve the booklet’s readability. As reading comprehension concerns answering text-related questions, the difficulty of the question is influenced by (the difficulty of) the text. For example, the question “To what word does *this* in sentence 3 refer?” might be easy or difficult to answer depending on the complexity of the text. In Appendix 3, the final cutscores as well as the related performance categories are presented.

Reading comprehension development

In this part of the meeting, the complexity of reading comprehension was discussed. In comprehension processes, individual differences (e.g., decoding skills, prior knowledge) interact with the text features (e.g., text difficulty) (Ozuru, Dempsey, & McNamara, 2009; Snow, 2002). This interrelatedness between reader and text affects the ease with which one can concretize performance expectations in this domain. In order to aid the participants in determining what could be expected from second and third-grade students, the guidelines of the Expertise Centre for the Dutch Language (2010) were
addressed (see Appendix 2 for their content). The Expertise Centre has identified seven skills that students in these grades are expected to master. All skills were briefly addressed, but the skill concerning students’ genre knowledge was discussed more elaborately as it was part of teachers’ homework assignment. The types of texts as discussed by the Expertise Centre for the Dutch Language (2010) were compared to those used in the Cito reading comprehension assessment. The types of questions used in these assessments were discussed as well, as some question formats were less common than those used in the curricular textbooks and workbooks (also in Appendix 2).

**Homework assignment**

The teachers were requested to identify (on a 3-point Likert scale) to what degree the text types and question formats in the Cito-assessment were similar to those in the curricular textbooks and workbooks as used in their classes. In the case of underexposure, the teachers were asked if they had ideas how to tackle this problem. We developed this homework assignment not to advocate teaching-to-the-test, but to improve teachers’ awareness on these differences in order to acquaint the students with these question formats (so that students would not be unnecessarily surprised when taking the Cito assessments). The homework assignment was requested to be completed prior to the program’s fourth meeting.

In addition, the teachers were asked to complete the same three-staged homework assignments for reading comprehension as was provided for mathematics. They had to 1) make their initial expectations for their students’ future performance on the June-assessment explicit, using the performance category classification - i.e., by selecting the below minimum, minimum, basic, proficient, or the advanced performance category -, 2) use the student monitoring system to compute the future performance estimation - see Appendix 4A -, and 3) present an overview of last year’s performance on the June-assessment - see Appendix 4B. This three-staged assignment was requested to be completed prior to the fourth meeting.
4. Goal setting for mathematics and reading comprehension

Summary of the fourth meeting: During this meeting the teachers had to set ‘challenging but realistic’ performance goals for each individual student in their classes, both for mathematics and reading comprehension. Furthermore, the teachers were stimulated to reflect critically on the curriculum alignment. Finally, for mathematics as well as for reading comprehension particular issues were addressed.

<table>
<thead>
<tr>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of meeting</td>
</tr>
<tr>
<td>Time span</td>
</tr>
<tr>
<td>Delivery mode</td>
</tr>
<tr>
<td>Hand-outs</td>
</tr>
</tbody>
</table>

Discussing the teacher expectations and performance estimates

During this meeting the homework assignments for mathematics and reading comprehension were discussed by the school team. The focus in this discussion was on explaining the differences between the teachers’ original expectations of the students and the estimates from the student monitoring system. By discussing performance data, school team members can learn from each other and thereby improve their knowledge of and skills in teaching (Datnow, Park, & Wohlstetter, 2007; Huffman & Kalnin, 2003; Schildkamp & Kuiper, 2010). For example, last year’s teacher might ask the current teacher: ‘Johnny was a rather good reader last year, and I would probably have expected him to perform on the proficient level. Why do you expect the basic level to be more suitable?’ By creating an open atmosphere in which the participants were willing to think along with one another and to offer (instructional) suggestions, the supportive role of the school team was increased. Furthermore, specific attention was paid to a comparison between the expectations for students’ performance in mathematics and in reading comprehension. The students’ ability growth (part of the school management staff’s homework assignment) was taken into account as well.
Setting performance goals

In this part of the meeting, the teachers were asked to set a performance goal for each individual student by reconsidering their initial expectations and taking into account all the information obtained during the group discussion. The performance goals were set for both mathematics and for reading comprehension. They were formulated using the performance category classification (below minimum, minimum, basic, proficient, and advanced) and pertained to each student’s desired performance on the June-assessment. The teachers were instructed to set ‘challenging but realistic’ goals given the students’ capabilities, as these have been proven to be the most effective goals (Erez & Zidon, 1984, in Locke & Latham, 1990).

The school level: vertical curriculum alignment, school goals, and mutual expectations

After setting performance goals at the student level several recommendations were provided like making explicit in school which knowledge and skills are taught to the students in the grades and how they are taught (Martone & Sireci, 2009; Webb, 1997). Such explicit knowledge would enhance curricular alignment in neighboring grades. To promote reflection on this issue, we asked questions such as: ‘Does the third-grade teacher know which skills are learned in grades two and four, and does he or she know in what ways these skills are taught to the students?’ Furthermore, we recommended the participants to openly discuss implicit expectations regarding instruction and the curriculum. These expectations could be either more general (‘In our school, we expect all our colleagues to have finished the curricular textbooks at the end of the year’) or content-specific (‘At the end of grade 2, we expect all students to master the multiplication tables 1-5’). In addition, we encouraged the participants to think of content-specific goals that would be in line with the individual performance goals that were just set (e.g. ‘At the end of grade 2, the basic/proficient and advanced students should be able to do automated additions up to 20’). All these suggestions were meant to promote a continuous learning progression within the school.
**Tips and practical suggestions on instruction**

As part of this meeting, we discussed known difficulties in relation to the Cito tests and provided several practical recommendations.

(i) **Mathematics: Word problems**

As the curricular textbooks do not address how to solve word problems in a systematic manner (see Appendix 2), students tend to get confused when confronted with (complex) word problems in the Cito assessments (Janssen, van der Schoot, & Hemker, 2005). In the PDP, the teachers were instructed on how to initiate mathematical questions during other parts of the school day, for example, when students share personal anecdotes at the start of the school day. By initiating these questions, teachers can demonstrate to their students that mathematics is an omnipresent phenomenon. Moreover, by elaborating on such questions, teachers have the opportunity to simultaneously demonstrate a systematic approach to solving such daily mathematical problems. As a result, the time spent on mathematics is extended. The participants were given an assignment in which they had to practice their ability to elicit mathematical thinking “on the spot”. They were asked to read a short story about a Christmas tree and broken decorations, and come up with questions that targeted different mathematical domains (these domains are discussed in Appendix 2).

(ii) **Reading comprehension: Text types and question formats**

During this part of the meeting, the outcomes of the previous homework assignment (pertaining to the reading comprehension textbooks) were discussed. After a brief recapitulation of these text types and question formats, we discussed their (under)exposure on the basis of questions such as ‘Do you have ideas on how to deal with certain underexposure?’ and ‘Which suggestions do your colleagues have?’ Specific ideas about how to develop exercises that appeal to students’ interests were mentioned. Furthermore, the teachers were supplied with a print-out with examples of different genres and question formats.
5. Data use and a formative use of assessments

**Summary of the fifth meeting**: The goal of this meeting was to inform the participants on how to implement data use at the levels of the school, the classroom and the lesson. Furthermore, they were given the opportunity to practice with different data analyses within the student monitoring system.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Type of meeting</th>
<th>General meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time span</td>
<td>2.5 hours</td>
<td></td>
</tr>
<tr>
<td>Delivery mode</td>
<td>Whole group presentation, video fragments, small group discussions, practice of data analysis behind laptops (in pairs)</td>
<td></td>
</tr>
<tr>
<td>Hand-outs</td>
<td>Data analysis booklets (with navigational directions and screen shots of the student monitoring system), print-outs of powerpoint slides, print-out with nine propositions, homework assignment for the teachers</td>
<td></td>
</tr>
</tbody>
</table>

**Data use at the school, classroom and lesson levels: introduction**

In our PDP, data use entailed, among other things, setting performance goals for each individual student and making instructional decisions based on data so that the students obtain the desired proficiency. Accordingly, our PDP mainly focused on helping teachers in enhancing their teaching at classroom level. However, data-informed decision making should pertain to the whole school. Data use therefore requires a change in school culture: important prerequisites for data-use include a clear vision of the school principal on issues such as the school’s goals and teacher collaboration (Schildkamp & Kuiper, 2010; Wayman, Midgley, & Stringfield, 2006). Within the PDP, the need for school and teacher change was illustrated by showing video-fragments (Primary Education Council, 2009) of teachers, senior support coordinators and school principals who explained in which ways DDDM had changed their way of working and their view on the school’s culture. These fragments also illustrated how these people changed their way of working as a result of their increasing awareness of the students’ performances and needs. After this general introduction, the participants split up into two groups: one groups with only teachers and one group with the school management staff (i.e., the school principal and senior support coordinator). The remainder of this meeting continued in parallel sessions which were conducted in separate rooms.
Data use at the classroom and lesson levels: teacher session

The teacher session started with a short recapitulation of why data use is important. Several characteristics of effective schools were introduced, such as having high expectations in regard to student abilities and frequently monitoring and evaluating performance (Sammons et al., 1995; Scheerens & Bosker, 1997). Next, the teachers’ reflection on their own data use and goal orientation was stimulated by having them conduct a small exercise. They had to score their own behavior by considering nine provoking propositions, among which “When I look at the students’ results I ask myself the following questions: do I see developments that I consider to be positive, do I see developments that I consider to be negative and are there any changes as regards the previous assessment results?” After rating their own behavior, teachers were asked to discuss their own behavior in pairs. In order to elicit vivid and informative conversations these pairs were designed to consist of participants with different scores (high and low; colored cards represented these scores and these pairs could thus easily be constructed).

In the next part of the meeting, we discussed the instructional model of Direct Instruction. This teacher-centered model was discussed as it has proven to be an effective instructional model (Borman et al., 2003; d'Agostino, 2000). We explained the model’s essential characteristics, while primarily focusing on the start and the end of the lesson (Leenders et al., 2010; Muijs & Reynolds, 2011). In addition, the topic of time-on-task was addressed. The amount of time teachers actually spend on teaching (instead of classroom managerial and organizational actions), and thus the actual time that students are learning determines student learning gains to an important extent (Houtveen, van de Grift, & Creemers, 2004). We provided tips to help the teachers identify their own time-on-task and classroom management behaviors.

(i) Teachers’ use of the student monitoring system

During this part of the teacher session, the participating teachers received training in making use of the student monitoring system (similar as in the second meeting of the PDP). Three types of analyses were discussed: a) a performance comparison (to make extreme high or low scores more tangible), b) identification of the average class performance for several consecutive years, and c) error analysis for mathematics – analyses are described in detail in Appendix 4D, E, and F respectively. Similar to the second meeting, the participants practiced these analyses behind laptops in pairs with the help of hand-outs containing step-by-step instructions, explanations and critical questions about the outcomes. One of the members of the research team was available for
assistance. As two schools had indicated beforehand that they were already quite familiar with the analyses that would be discussed during the meeting, they followed a slightly altered program during this part of the meeting. They were exempted from several computer assignments but they were asked about their actual data use by a second member of the research team, who gave them targeted information to improve their current data-analyses practices.\(^{29}\)

\textit{Data use at the school level: school management staff session}

The school management staff session was set up as follows. For schools wanting to work in a data-driven way, the school management staff plays an important role. They are responsible for several tasks, such as monitoring the school outcomes, formulating clear goals, and promoting teacher collaboration (Schildkamp & Kuiper, 2010; Young, 2006). In the PDP, we stressed the importance of these activities. The structure of this meeting was similar to that of the teacher session, as it started with the same recapitulation of characteristics of effective schools. Next the participants in this group were asked to respond to the nine propositions (comparable to the propositions the teachers received, but then formulated at the school level) on teachers’ goal-orientation and collaboration. Beliefs and practices were discussed in pairs (again, consisting of participants with a high and a low score) to boost awareness and generate practical ideas on how data use could be implemented within the school. After this exchange of ideas, the implications for the school and HR policies were discussed and illustrated by using video-material (Primary Education Council, 2009). It was further argued that the school management staff should play an important role in supporting the teachers’ professional development (also in Fullan, 2007) by regularly observing classroom lessons and providing them with constructive feedback to improve their instructional practices.

\textit{(ii) School management staff’s use of the student monitoring system}

During this part of the meeting, the school management staff received training in working with the so-called self-evaluation module of the Cito digital student monitoring system. The three types of analyses discussed were a) \textit{cross section of the performance levels of the middle school classes}, b) \textit{longitudinal development of the group mean score of one or more classes}, and c) \textit{the mean score of a specific grade} (e.g. grade three)

\(^{29}\) Since the teachers indicated that they only worked with error analyses for a few individual students, further information on whole group error analysis was provided using the output of a fictitious group.
throughout consecutive years. These analyses are elaborated in Appendix 4G, H, and I respectively. For the participants who were already more experienced in working with the digital student monitoring system, an additional type of analysis was offered, namely d) group analysis (elaborated in Appendix 4J), which examines the development of one class in terms of different subject areas and in several years. Again, the participants worked behind laptops in pairs, using hand-outs containing step-by-step instructions, explanations and critical questions about the outcomes. One of the members of the research team was available for questions and support.

Homework assignments
Both the teachers and the school management staff received a booklet with homework assignments. The teachers were given an assignment pertaining to the January-assessment for mathematics. They were asked to make two error analyses, one for the whole group and one for an individual student. The booklet contained instructions for executing the analyses in the digital student monitoring system. Guidelines on how to interpret the output were provided as well. The booklet also included questions to stimulate one’s reflection on the output, such as ‘Can you indicate mathematical domains for which a vast amount of students scored below or above expectation? If so, can you explain why these deviations occurred?’ The assignment was requested to be completed prior to the sixth meeting.

The management staff was provided with a different assignment. For the second time in the PDP, they were asked to compute the developmental growth for the current second- and third-grade students with the help of their digital student monitoring system (the progress report, discussed in Appendix 4C). This time, they had to calculate the developmental growth from the June-assessment in the previous year until the current January-assessment. Similar to the previous progress report assignment, the developmental growth of both mathematics and reading comprehension had to be analyzed. The assignment was requested to be completed prior to the sixth meeting.

6. Mathematics instruction and curriculum
Summary of the sixth meeting: The goal of this meeting was to evaluate the January-assessment results for mathematics, to interpret the mathematics error analysis, to re-examine the teachers’ performance goals for the June-assessment (and adjust them if
desired) and to extend teachers’ knowledge of the provision of explicit instruction on mathematical word problems.

### Characteristics

<table>
<thead>
<tr>
<th>Format of meeting</th>
<th>Meeting at school</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time span</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Delivery mode</td>
<td>Whole-group presentation, group discussion</td>
</tr>
<tr>
<td>Hand-outs</td>
<td>Overview of student achievements on January-assessment and performance estimates for June-assessment, overview of exercises in terms of different math domains, print-outs of powerpoint slides, homework assignment for teachers</td>
</tr>
</tbody>
</table>

**Start of the meeting**

At the start of this meeting, an overview was handed out to each class, consisting of the performance goals that teachers had set in terms of the five performance categories (below minimum, minimum, basic, proficient, and advanced) during the fourth meeting. The aim was twofold: to provide a clear overview of the performance goals at the class level and to remind the teachers of the goals they had set earlier.

**Discussion of performances on the January-assessment**

In the Netherlands, the predominant analyses that are conducted in the digital student monitoring systems focus on the development of individual students. Analyses of the proficiency development of an entire class are less frequently executed. Furthermore, adjustments to teaching practice on the basis of the data-analyses are still relatively uncommon (Inspectorate of Education, 2010; Ledoux et al., 2009; Schildkamp & Kuiper, 2010). During this meeting, attention was paid to both these aspects by discussing the output of the error analysis of the January-assessment on the basis of group performance and by investigating how the analysis could be used for data-informed instructional decisions. For this purpose, the homework assignments distributed during the fifth meeting were discussed in detail. As most of the participants were not familiar with the error analyses and it was only briefly discussed in the fifth meeting, its purpose and possibilities were explained in more detail. The output from a fictitious class was used for illustrative purposes. With respect to this fictitious example, we also provided suggestions on why and how the teacher of this class could adapt his or her whole-class teaching. When discussing the error analyses for their class as a whole, the teachers were requested to explain the over- or underachievement within specific domains. For example, if a large percentage of the class scored relatively low in the “time domain” (one
of the subdomains for mathematics, in Appendix 2), teachers were asked how and how
often this topic was addressed in the curricular textbooks that were used in school and
whether the teachers felt that they had sufficiently targeted this domain during
instruction. Practical tips on how to tackle such issues were provided.

**Mathematical subdomains and the diagnostic math interview**

As the error analysis indicates students’ performance on the different mathematical
subdomains (see Appendix 4F), the teachers are supposed to have sufficient content
knowledge to distinguish the characteristics of such subdomains. To enhance or refresh
teachers’ knowledge on the specific mathematical domains, this information was
provided using the categorization of domains employed in the Cito assessment system
(Cito, 2003). Following this information, the teachers were asked to classify several
mathematical exercises into the corresponding subdomains. The correct classifications
were discussed during the meeting, and a handout with additional examples per domain
was distributed among the participants.

The error analysis only indicates students’ relative strong and weak achievements on
particular mathematical domains (i.e. what goes right or wrong) but a more detailed
analysis is needed to determine why students make certain mistakes. Information about
students’ problems in understanding or their misconceptions is crucial in this respect
(Van Groenenstijn, Borghouts, & Janssen, 2011; Goertz et al., 2009). This information
can be attained using “diagnostic math interviews”. During this meeting, we stressed the
importance of teacher-student interactions as these interactions help to attain valuable
information on a student’s knowledge and skills. When practicing a diagnostic math
interview, we recommended teachers to focus on the students’ problem solving
capabilities, because the mathematics textbooks do not pay specific attention to this
problem solving skill while it is an essential part of the Cito-assessments (Jacobse &
Harskamp, 2011). In the problem solving process, several stages can be distinguished30: a)
the approach phase, which refers to the steps of reading, exploring and planning, b) the
calculation phase, which refers to the step in which the computational plan is
implemented by the student, and c) the evaluation phase, which refers to the step in
which the student evaluates the process and verifies the answer. Information on students’

30 Problem solving is a complex activity. It requires several steps to be taken (Schoenfeld, 1992):
read, analyze, explore, plan, implement and verify. In the PDP, these steps were reduced to three
stages.
skills and understanding in these different phases would enable the teachers to adapt the instruction in such a way that it meets students’ needs.

*Resetting the performance goals*

At the end of the meeting, the teachers were asked to re-examine the performance goals that they had set for the individual students in the fourth meeting. For this purpose, the teachers used the overview handed out to them earlier and discussed its content with their colleagues. On the basis of this information the teachers could either reset (some of) their goals or leave them unaltered.

*Homework assignment*

In order to encourage the teachers to further investigate the output of the error analysis and to enhance their awareness of the nature of the students’ mistakes - i.e., focusing on the *why*. For this purpose, the teachers were asked to interview a student about the mathematical subdomain on which he or she had scored weakly. By discussing exercises with an individual student, the teachers were asked to assess the student’s ways of thinking, level of understanding, and possible misconceptions, thereby classifying problems into the three phase(s) of approach, calculation, or evaluation. The outcomes of the diagnostic math interview could be reported on a form. This form contained additional space for writing down how the teacher planned to adapt instruction to meet this particular student’s needs. The homework assignment was requested to be completed before the eighth meeting.
7. Reading comprehension instruction and curriculum

Summary of the seventh meeting: The goal of this meeting was to analyze the January-assessment results, and to extend the participants’ knowledge of reading comprehension acquisition and the curriculum. During this session, the teachers were also asked to re-examine and, if desired, adjust their performance goals for the June-assessment.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of meeting</td>
<td>Meeting at school</td>
</tr>
<tr>
<td>Time span</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Delivery mode</td>
<td>Whole-group presentation, group discussion</td>
</tr>
<tr>
<td>Hand-outs</td>
<td>Overview of the student test results on the January-assessment and the performance estimates for the June-assessment, overview of the reading strategies as discussed in various instructional and curricular materials, print-outs of powerpoint slides, additional reading materials, homework assignment for the school principal/internal support coordinator</td>
</tr>
</tbody>
</table>

Discussion of performances on the January-assessment

The performances on the reading comprehension assessment were addressed in a similar way as in the discussion of the mathematics results during the sixth meeting. However, the digital student monitoring system does not provide error analyses for reading comprehension, as in statistical analysis this skill is found to be unidimensional; an analysis of performance on (theoretical) subdomains is therefore not plausible (Feenstra et al., 2010).

Reading comprehension; interplay among reader characteristics, text characteristics, and reading goal

In training the teachers in how to make informed instructional decisions, they were informed about the important determinants of reading comprehension achievement as comprehension stems from an active and interactive process between the reader (with a specific level of e.g., decoding skills, vocabulary and motivation), the specific text (with certain characteristics in regard to e.g., text genre, audience appropriateness and coherence), and the goal a reader has for that specific text (Snow, 2002; Sweet & Snow, 2003). Different reading goals require different reading strategies. We discussed these three types of characteristics and referred to observations made in the participating
teachers’ classrooms. For example, when discussing different genres, statements could be made such as “I saw you addressing the distinction among different types of texts by asking your students “what is the difference between a narrative and an expository text?”.

In addition, we discussed the two goals of formal reading comprehension instruction. These are 1) developing students’ knowledge and vocabulary, and 2) teaching students how to control their own reading processes by applying a number of reading strategies (van de Mortel & Förrer, 2010). During this meeting, the topic of reading strategies was given ample attention, as these strategies are crucial tools in helping the reader to understand a text (National Reading Panel, 2000; Pressley, 1998). The participants were provided with a handout containing five different overviews of reading strategies\footnote{1) Overview of ‘evidence-based’ strategies by Vernooy & Stoeldraaijer, ‘Algemene leerlijnen begrijpend lezen’, \url{www.taalpilots.nl} 2) overview of strategies as used by the curricular textbook ‘Nieuwsbegrip’, 3) overview of strategies according to Filipiak (2006), 4) overview of strategies as provided by (van de Mortel & Förrer, 2010), 5) overview of strategies as used by the curricular textbook ‘Kidsweek’.} - some more elaborate than others – and they were asked to compare these overviews and reflect on their content. To assist teachers in this task, teachers were asked which strategies they valued, which ones they used themselves and which ones they regarded as appropriate for the grades under study. The seven skills identified as appropriate for students in the second and third grades – addressed in Appendix 2 – had been translated into exemplary questions and linked to the different strategies provided in the overviews. Next, the strategies used in the teacher’s own textbooks and workbooks (as analyzed by the one of the research team members) were considered and compared to those in the overviews and those mentioned in the guidelines of the Expertise Centre. The results of the homework assignment from the third meeting (determining the degree to which the students encountered the text types and question formats used in the Cito assessments) were also taken into account in this part of the meeting. Last, the importance of explicit instruction in reading comprehension strategies, i.e., explaining what they entail, why they are used and how they are used (Guthrie et al., 2004; Pressley, 1998) was addressed as it is commonly found that explicit instruction in reading comprehension is lacking (also in Appendix 2). These issues would be discussed in more detail in the next meeting.
At the end of the meeting, additional materials were provided to the participants to advance their knowledge in these areas.

**Homework assignment – school management staff**

The school management staff was asked to observe a lesson in mathematics or reading comprehension given by the participating teachers, during which they had to fill in a simple observation form which targeted teachers’ implementation of Direct Instruction (we specifically focused on the elements of this model pertaining to the start and end of the lesson). In this way, the school management staff would obtain an impression of the extent to which this model was implemented within their school’s lessons. The school principal and senior support coordinator were asked to provide the teachers with constructive feedback with regard to their implementation of this practice. The observation instrument used was a simplified version of the high-inference instrument employed by the researchers (see Chapter 5). As the observations would be discussed during the next meeting, the assignment was requested to be completed in time.

---

**8. Effective instruction**

**Summary of the eighth meeting:** The goal of this meeting was to re-activate the teachers’ awareness of their own performance goals for student performance and to train the teachers in instructional practices which would help them to attain these goals. This was done by reflecting on their experiences with the diagnostic math interviews, by further elaborating on explicit problem solving instruction, and by explaining and practicing how to model the application of reading comprehension strategies.

---

32 The additional reading material contained an analysis of the curricular text books as used in their classes (developed by Projectbureau Kwaliteit; [www.schoolaanzet.nl](http://www.schoolaanzet.nl)), a special issue on reading comprehension and instruction (Loman & Marreveld, 2010), and an overview of reading strategies as provided by the curricular method Nieuwsbegrip ([www.nieuwsbegrip.nl](http://www.nieuwsbegrip.nl)).
Characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format of meeting</td>
<td>Meeting at school</td>
</tr>
<tr>
<td>Time span</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Delivery mode</td>
<td>Whole-group presentation, group discussion, practical modeling exercises</td>
</tr>
<tr>
<td>Hand-outs</td>
<td>Print-outs of powerpoint slides, overviews of goals set for mathematics and reading comprehension, modeling examples</td>
</tr>
</tbody>
</table>

Data use: discussion of differentiation practices
At the start of this meeting, we provided the teachers with two overviews (one for reading comprehension and one for mathematics) of the ‘revised’ performance goals as set in the sixth and seventh meeting. The aim of discussing this overview was threefold. First, it was used to stimulate the teachers’ awareness of their performance goals. Second, the information was used to focus teachers’ attention to the performance expectations for their whole class and to draw their attention to how they could divide their class into subgroups for instruction. Third, in line with the second aim, the teachers were asked to compare the classifications in the overviews with their current ability grouping practices. They also were asked to compare their within-class grouping practice for mathematics to that for reading comprehension. This exercise aimed at tackling the teachers’ hesitance on using analyses of their students’ capabilities for adapting their instructional routines (recommended in, e.g., Ledoux et al., 2009; Schildkamp, Visscher, & Luyten, 2009).

Instruction on problem solving and automation
As part of the homework assignment provided in the sixth meeting, the teachers were asked to conduct a diagnostic math interview with at least one of their students. During the current meeting, the information about this students’ level of knowledge and skills was discussed in terms of the three problem solving phases as discussed during the sixth meeting (being approach, calculation and evaluation). With respect to the first phase (the approach-phase), guidelines were given on explicit instruction on how to approach contextual problems in a structured way. The teachers were shown practical steps which students can take when facing a word problem, such as trying to reformulate the question in their own words, underlining the relevant elements in the text, or drawing a picture representing the situation at hand (Fuchs et al., 2008; Griffin & Jitendra, 2009). With respect to the second phase (the calculation-phase) the importance of spending time on automation within the mathematics lessons was addressed. Automated knowledge of basic skills in the long-term memory facilitates more complex mathematical operations.
as it provides room in the working memory to conduct non-automated calculations (Ruijsenaars, Van Luit, & Van Lieshout, 2004). Especially interactive automation seems to be a fruitful way of working, since active participation of the students increases the effectiveness of the exercises (Inspectorate of education, 2011). The researchers provided practical examples for such exercises. Last, the importance of the third phase (the evaluation-phase) was emphasized. Again, practical suggestions on how to address task-evaluation more effectively was provided.

**Modeling strategies**

Next, the concept of *modeling* was explained to the participants. Modeling is an instructional technique in which the teacher demonstrates to the students how to apply a certain reading comprehension or mathematics strategy by "thinking aloud". This approach is considered to be the primary method of showing students how they can interact with a text (e.g., Taylor & Pearson, 2002 in Fischer, Frey & Lapp, 2009) and has proven to be an effective instructional technique (Fisher et al., 2008; National Reading Panel, 2000; Pressley & Harris, 1990). Teachers in Dutch primary schools are still rather unfamiliar with modeling, although this instructional approach has received some attention in journals targeting teachers and schools rather recently (e.g., Filipiak, 2006; Loman & Marrevel, 2010) and in other reading improvement PDPs (for example, Droop et al., 2012). In order to practice modeling, a hand-out was distributed in which a number of texts (from the Cito assessments of the grades under study) had been tagged with questions suitable for modeling “on the spot”. We modeled three questions, after which the participants were asked to demonstrate how they would model the other questions in their classrooms. Immediate constructive feedback was provided to the participants on their implementation of this instructional technique. The appropriateness of using the modeling practices during math instruction was further specifically stressed.

**Homework assignments**

To practice modeling strategies, teachers were asked to implement modeling during their book reading practices (i.e., when the teacher reads a story out loud to the class; a common practice in primary schools). The teachers were asked to prepare a section of the book that they were currently reading in their class and choose one or two reading strategies that they wanted to model. As part of the homework assignment, they were asked to adopt these strategies in the classroom and report back on their experiences. Using the same story, the teachers also had to formulate two mathematical problems
which the students had to solve in pairs. In this way, the teachers could both create additional time for practicing word problems, elaborate on students’ problem solving approaches, and stimulate cooperative working (Slavin & Lake, 2008). The teachers were requested to complete both parts of the assignment prior to the ninth meeting.

The second assignment concerned “near future behavior”. The teachers were given a postcard containing a list of six teacher activities that were discussed during the PDP, namely: a) modeling strategies, b) interactive automation exercises, c) discussing the differences between genres, d) conducting a diagnostic math interview, e) initiating mathematical questions (following anecdotes or other situations that would lend themselves for the construction of a mathematical word problem), and f) addressing reasons for using (reading) strategies and relating the use of these strategies to topics important to the students. From these six items, the teachers were asked to mark those activities that they intended to carry out until the summer holidays. The postcard was sent to them three weeks later to remind them of these intentions.

The third assignment had a similar purpose as the second one, but focused on “long term behavior”. The participants were requested to write a letter to themselves about aspects of the program that they wanted to remind themselves of after the PDP had finished. We recommended the teachers to make this assignment while taking a close look at the information which was discussed throughout the school year (which they had collected in the binder). The participants were asked to hand in their letter in a closed envelope during the ninth meeting. The letters were sent to them at the beginning of the following school year to re-activate the skills and knowledge they had acquired during the PDP and to remind them of their aims.

Discussion of teacher observations – school management staff only
At this point the teachers were requested to leave the meeting. Now the school management staff was asked to report on their classroom observations, their conversations with the teachers, and the insights they had gained from their impressions and from filling in the observation instrument. These experiences were compared to our own observations of these teachers’ instructional practices.

Homework assignment – school management staff only
This assignment focused on the connection between the data and the actual teaching practices in class, building on the classroom observations that were made. The school staff was requested to analyze the longitudinal development of the group means for the
classes observed (discussed in appendix 4H) and to compare this analysis to their perceptions during the lesson observations. As the longitudinal development of the group mean would, at least partly, give an indication of the educational quality, this assignment might be of use in helping to explain the developments in student performance. For instance, if the group mean development’s learning curve was less steep than the national average, the lesson observations might help to gain understanding on aspects of teaching practice that might be in need for improvement. In the reverse case, a very steep learning curve of the group mean might indicate “high-quality teaching” and the instructional practices of the teacher responsible for this result could then be discussed in the entire school team to stimulate other teachers to implement such high-quality teaching.

9. Evaluation of the PDP

Summary of the ninth meeting: The goal of this meeting was to connect students’ performance on the June-assessment to the performance goals (in order to establish to what extent teachers had attained their goals) and to evaluate the PDP as a whole.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Format of meeting</td>
<td>Meeting at school</td>
</tr>
<tr>
<td>Time span</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Delivery mode</td>
<td>Whole-group presentation, group discussion</td>
</tr>
<tr>
<td>Hand-outs</td>
<td>Print-outs of powerpoint slides, June-assessment scores, analyses of longitudinal group mean developments and cross-sections, overview of student performance in relation to performance goals</td>
</tr>
</tbody>
</table>

Discussion of the homework assignments

First the homework assignments of the previous meeting were briefly discussed. The teachers shared their experiences with modeling reading strategies and the formulation of mathematical exercises. Also the execution of teachers’ instructional intentions (as marked on the postcards during the previous meeting) was addressed.

Data use: student achievement in relation to the performance goals

The results of the June-assessment were discussed, as these assessments had been administered to the students prior to the current meeting. Data from the different steps
of the multistep procedure as well as later meetings on student performance were combined into one overview. These overviews were handed out and discussed, while we pointed out visible patterns of over- or underperformance in relation to the teacher-set goals. The results were linked to the teachers’ (assumed) level of ambition: some teachers indicated that they had been somewhat more “cautious and reserved” than other teachers in their goal setting practices. After discussing the results at the classroom level, the participants received comparisons between the performance goals and the actual achievements at the individual student level. As the focus of this meeting was particularly on evaluation at the group level, these accounts of individual students’ performances were only briefly dealt with.

Data use: learning gains at the class level

Prior to the meeting, the school management staff was requested to deliver the output of the longitudinal developments of the group mean scores (see appendix 4H) as well as cross-sections (see appendix 4G) pertaining to the second- and third-grade students. The results of these data analyses were discussed with the participants. We provided support to the participants in their interpretation of the output. The development of the group mean was considered indicative of the educational quality in the different classes and facilitated the comparison of parallel classes. These analyses could be further explained by comparing them to the cross-sections, as they depict the distribution of achievement within a class. In addition, the results of the school management staff’s homework assignment (in which students’ performance data was connected to the teachers’ observations) were discussed. The aim of combining this information with the former analyses was to foster the participants’ insight into issues such as student growth, school performance, and the important role of effective teaching therein.

General remarks with respect to the lesson observations

As part of the PDP, the researchers observed one mathematics and one reading comprehension lesson of each participating teacher, both at the start of the PDP and at the end. During this evaluation meeting, general constructive feedback was given.

Evaluation of the PDP

At the end of this last meeting, the participation in the project as a whole was evaluated. The researchers asked the participants for comments and suggestions for improvement. In addition, they summarized the PDP’s most important aspects of which they hoped that they would be maintained within the school team. These were: a) a continuing focus on goals and performance, b) team discussions about data and teaching in an open
atmosphere, c) teacher reflection on their own instruction practices, and d) a continuing focus on the provision of high-quality instruction in mathematics and reading comprehension. Finally, the researchers expressed their gratitude with regard to the participants’ collaboration in the program.
Appendix 2: The Dutch educational context in relation to the PDP’s three components

The studies presented in this dissertation were conducted to evaluate the effectiveness of a teacher Professional Development Program (PDP). This PDP aimed to improve students’ reading comprehension and students’ mathematics due to performance concerns for both areas (Ministry of Education, 2010). All aspects of the program which have been reported in this dissertation in relation to mathematics (for example, setting standards and performance goals) have also been conducted for reading comprehension. The content of the Appendices will pertain to the PDP as it was delivered to the participants, and will thus pertain to both subject areas. The results of the program with respect to reading comprehension were discussed further in the dissertation of van Kuijk (2014).

In this section, several aspects of the Dutch educational context are discussed in relation to the three components of the PDP, being 1) setting standards and performance goals for every student, 2) data-analysis, and 3) instruction in both reading comprehension and mathematics. We will start by addressing the recent implementation of national standards in the Netherlands, which is appropriate in light of the program’s first component. After this, we discuss the Cito LOVS standardized assessment system as well as the different digital student monitoring systems which are used in the Netherlands. This information is relevant in light of the program’s second component. In regard to the third component on relevant instructional skills and knowledge in reading comprehension and mathematics, we will give an outline of the general second- and third-grade performance expectations in these areas and address difficulties in the field of instruction which are commonly identified. General information on the curricular textbooks and alignment of these textbooks to the Cito assessments will be discussed as well.

Recent implementation of standards in the Netherlands

A standard setting procedure was conducted as part of our PDP. Commonly, this procedure is conducted in countries that employ standards-based education (such as in the United States, see Cizek & Bunch, 2007) in order to identify performance categories on, for example, state-wide or national tests (Hambleton & Pitoniak, 2006). Our reason for conducting this procedure was slightly different: the participating teachers would
formulate their student-specific performance goals on the basis of the performance categories they established themselves.

Recently, the political climate in the Netherlands has become more oriented toward working with standards in education. By doing so, the Netherlands have followed in the footsteps of other countries, such as the United States, England, Germany, New Zealand, Australia and South Africa, where standards-based education has already been introduced (Fuhrman, 2001; Klieme & Maag Merki, 2008; OECD, 1995; Taylor, 2009). Commonly, two types of standards are identified. Content standards define what should be taught and what students should learn. Performance standards provide descriptions and examples of what students have to know and do, to demonstrate proficiency in the knowledge and skills framed by the content standards (Ravitch, 1995). In the Netherlands, particularly the implementation of performance standards is a relatively recent development. Content standards, known as the core objectives\(^{33}\), were already actuated in 1993. In 2010, performance standards have been defined for the end of primary school in grade six and for the end of each academic track in secondary school (the Netherlands has a tracked secondary school system). By having performance standards for these points in time, it is expected that they will help boost current achievement levels as well as facilitate the transition from primary to secondary and from secondary to upper-secondary and higher education. The Ministry has identified two performance categories, namely basic (a performance level currently attainable for 75 percent of the student population) and proficient (a performance level currently attainable for 50 percent of the student population)\(^{34}\). These standards have been mandated by law for grade six since August 2010 (Ministry of Education, 2010) yet changes still need to be made in the Cito LOVS assessment system before it can be measured at which level a student is performing (e.g., ‘basic’ or ‘proficient’ when using the performance category classification). These assessment changes are scheduled for 2014-2015 (Ministry of Education, 2011). In short, the Dutch national performance standards are not yet fully operational and for the grades under study in this dissertation (grades two and three), no performance standards have or will be set nationally.

\(^{33}\) in Dutch: kerndoelen

\(^{34}\) For designing these standards a committee was installed. The standards have not yet been connected to test scores, but will be in the near future.
The Cito LOVS standardized assessment system and student monitoring systems used in the Netherlands

The Cito LOVS standardized assessment system
The Cito LOVS standardized assessment system, developed by the Netherlands Institute for Educational Measurement, is the most widely used assessment system in the Netherlands: it is employed in approximately 85 percent of the Dutch primary schools (Inspectorate of Education, 2010b). It provides standardized assessments throughout primary school, in different subject areas. These assessments will be referred to as the Cito assessments. For most subject areas, there are two yearly assessments. The midway-of-the-school year assessment is conducted in January, and the end-of-the-school year assessment is conducted in June. In the remainder of the Appendices, these assessments will be referred to as the January-assessment and the June-assessment.

The Cito assessments for reading comprehension and mathematics cover a broad content and have a reliability of above $\alpha = 0.89$ and $\alpha = 0.91$ respectively (Feenstra et al., 2010; Janssen et al., 2010). Both assessments have been approved by the Dutch National Committee of Tests and Testing, responsible for the review of tests (COTAN). After conducting these assessments, students receive a test score which is indicative of their proficiency in either reading comprehension or mathematics. Specifically for mathematics, performance on different dimensions, or subdomains, is identified as well. These mathematical domains are slightly different across second and third grade. In second grade, three domains are distinguished. These three domains are 1) numerical knowledge, 2) arithmetics, and 3) geometry, time, and money. In third grade, four domains are distinguished being 1) numerical knowledge, 2) arithmetics, 3) geometry, and 4) time and money. For reading comprehension, no distinction in subdomains is made as psychometric analyses of performance on the reading comprehension assessments have indicated that reading comprehension is a unidimensional skill.

Student monitoring systems in the Netherlands
Test results of Cito-assessments can be registered and analyzed in administrative digital student monitoring systems. In the Netherlands, the use of student monitoring systems by teachers is still rather limited for analyzing problems and adapting instruction, and teachers who do use the student monitoring systems are often unaware of the possibilities for more sophisticated analyses (Ledoux et al., 2009; Meijer & Ledoux, 2011; Schildkamp & Kuiper, 2010; van der Kleij & Eggen, 2013). The three most commonly used digital student monitoring systems are the Cito student monitoring system,
Parnassys, and ESIS (Meijer & Ledoux, 2011). In the Netherlands, the Cito student monitoring system is the most frequently used student monitoring tool. In our PDP, we used the Cito student monitoring system when training the participants in their use of such systems: the majority of the participants made use of this system in their school. For schools working with Parnassys and ESIS, we explained the similarities and differences between these three systems as often as possible in order to support the transfer of newly acquired knowledge and skills toward these two student monitoring systems: Parnassys and ESIS contain a majority of the analyses which are available in the Cito student monitoring systems but there are some differences in how the results are acquired and how the output should be interpreted. In realizing this transfer of explanations and assignments from the Cito student monitoring systems to Parnassys and ESIS, we collaborated with the publishers of these three systems. One school worked with the Magister system. For this system, we could not provide instructions for transfer. However, as this school had previously worked with the Cito student monitoring system, they could easily navigate through the Magister system themselves. For the remainder of the Appendices (for instance, in our account of the meetings on data use as reported in Appendix 2) we will exclusively refer to the Cito student monitoring system.

**General performance expectations, common instructional practices, and the curriculum**

**Reading comprehension: general performance expectations**

In the Netherlands, primary school starts with two years of kindergarten, followed by grades 1-6. Formal reading instruction starts in first grade. During this grade, the focus is on learning and improving the students’ decoding skills. Although reading comprehension commonly starts at the beginning or halfway the second grade, there has been debate about the best time to start teaching this subject (Stoeldraijer & Vernooy, 2007). With respect to the performance expectations for second- and third-grade reading comprehension, the Expertise Centre for Dutch Language (2010) has developed guidelines on what should be expected of students during the grades under study. Throughout the PDP, these guidelines were used to help the teachers in their thinking about reading comprehension performance. The seven skills that students in the second and third grades are expected to master (as explicated in the guidelines) are the following:
1. Being able to identify the main topic of a text and to activate one’s own prior knowledge on this topic.

2. Being able to connect an anaphor (a word or group of words referring back or forward to another word or group of words) to its referential tie.

3. Knowing what to do to succeed in comprehending a difficult sentence or sentences, for example, by re-reading the same section of words more slowly or looking up the meaning of a difficult word in the dictionary.

4. Being able to predict future information/content in a text.

5. Being able to process information provided in a text as well as “read between the lines”.

6. Being able to distinguish among different genres, for example a narrative, expository, directive, descriptive, or argumentative text.

7. Being able to recognize the structure of a narrative text (i.e., begin, middle, and end, including the introduction of the main characters, the plot of the story, and its ending).

Reading comprehension: common instructional practices and the curriculum

In Dutch primary schools, reading comprehension lessons generally take the following sequence. First, the students read a text either out loud or in silence. Next, a few questions about the text are discussed with the whole class, after which the students have to answer the remaining questions independently. Last, the correct answers are discussed with the whole class (Aarnoutse, 1992). During these lessons little explicit instruction is given (e.g., Van Elsäcker, 2002) and there is little differentiation between students (Van Berkel et al., 2007; Van Elsäcker, 2002). These are aspects which call for improvement. According to Collins-Block and Pressley (2002 in Houtveen, 2002), teachers do not provide instruction in reading comprehension because they are unaware that this may improve comprehension. Instead, students are expected to master this skill on their accord via immersion. Moreover, Stoeldraaijer & Forrer (2012) hypothesize that teachers in the Netherlands find reading comprehension a difficult subject to teach, given the complexity of the different reading comprehension skills and the curricular textbooks used, which are not always as clear as they should be. These textbooks have been criticized as being “more bulky than necessary, containing a substantial amount of material that has little or nothing to do with learning to read” (Houtveen & Van de Grift, 2012, p. 88). They also contain a large number of reading strategies, but not all of these
strategies which are presented as “effective” can be supported by empirical evidence (Droop et al., 2012; Stoeldraijer & Forrer, 2012). The inadequacy of the curriculum is considered to be problematic as it is known that teachers in the Netherlands follow the content of the curricular textbooks to a very large extent in their lessons (Meelissen et al., 2012).

The content of the curricular textbooks used in second- and third-grade differ somewhat in comparison to the Cito assessment for reading comprehension in terms of text types and question types. As these textbooks differ per publisher (and there is a large number of publishers and textbooks available for teaching reading comprehension), we will only discuss characteristics of the Cito assessment here. For the grades under study, the Cito assessments contain narrative, expository, directive, and argumentative texts (Feenstra, Krom, & van Berkel, 2007a; 2007b). Especially the latter two types are relatively uncommon in the Dutch student textbooks for the grades under study. In addition, the Cito assessments make use of the following question formats: 1) multiple choice questions, 2) multiple choice items requiring the student to replace missing words in a text (i.e., a cloze test), and 3) items requiring the student to identify the first or last sentence of a short narrative text in which the sentence order has been mixed up (Feenstra, Krom, & van Berkel, 2007a; 2007b). Especially the latter two are relatively uncommon in the Dutch student workbooks.

**Mathematics: general performance expectations**

The formal mathematics education starts in first grade. In 2006, eleven content standards were formulated to determine the content that should be offered throughout primary school, including grade-specific elaborations (Buijs et al., 2008). The aim of these content standards is to achieve continuity in the learning progression.35 For the grades under study (grades two and three), the following knowledge and skills were formulated (and required to be mastered):36

1. **Mathematical knowledge and skills**

---

35 For other references to the learning progression in the Netherlands, see van de Craats (2007) and the “rekenlijn”-website (Freudenthal Institute, SLO, & KPC, 2010).

36 The description of the knowledge and skills required for the grades under study is rather rough and is not meant to provide a refined specification (for more information, see Buijs, Klep, Noteboom, & Klein Tank, 2008; van den Craats, 2007).
a. Knowing the mathematical language and being able to solve practical mathematical problems (word problems).

2. Numerical knowledge
a. Knowing and understanding the (structure of the) number system up to 1000 (and further).

3. Arithmetics:
   a. Being able to carry out additions and subtractions up to 100 (grade 4) and up to 1000 and further (fifth grade).
   b. Being able to carry out multiplications up to 10 and corresponding divisions (starting multiplications in fourth grade, automation in fifth grade and divisions mainly in fifth grade).

4. Geometry, time, and money:
   a. Getting acquainted with natural measures (length, weight, volume).
   b. Knowing the metric system of length (and being able to perform calculations with it).
   c. Being able to tell the time (on digital as well as on analogous clocks).
   d. Knowing the money system and being able to perform calculations with money.

**Mathematics: common instructional practices and the curriculum**
The Dutch curricular textbooks all contain grade-specific information on learning progression. Dutch teachers base their lessons mostly on their curricular textbooks (Meelissen & Drent, 2008). These textbooks offer suggestions concerning issues such as lesson structure and differentiation aspects (such as task difficulty, pacing, and concrete recommendations for providing extended, small group instruction). In Dutch primary schools, the mathematics lessons generally take the following sequence (as suggested by the textbook). First, the lesson begins with a short activating exercise. Next, the teacher provides whole-class instruction. After that, the students do seatwork which entails making exercises which are based on the preceding instruction, while the teacher gives small-group instruction to the weak performing students (also referred to as extended instruction). Then, all children do seatwork. During mathematics lessons, students thus work quite some time without instructional support of the teacher (Harskamp, 2010). Some curricular textbooks dictate that in each week, two teacher- and three student-centered lessons should be provided; in the latter, the instructive role of the teacher is less dominant. As already mentioned, curricular textbooks support the adaptation of
instruction to different cognitive levels of students by providing an array of tasks, ranging
in difficulty, and by providing explicit recommendations for extended instruction, such as
simplification of the subject matter, using different materials to enhance the students’
understanding. However, there are some doubts about the effectiveness of how these
differential practices are applied in classrooms as it is unclear whether teachers base
their decisions on formal or informal assessment data (Harskamp, 2010). Not using such
information leads to instructional practices that still do not really fit the students’ needs
(Inspectorate of Education, 2010a).

In the Netherlands, the learning of mathematics is nowadays more focused on
contextualized items as opposed to the traditional approach which was purely based on
solving arithmetic calculations. Although the modern mathematical textbooks make use
of real-world questions and contexts to elicit solution strategies for improving the
students’ understanding, Dutch teachers’ interpretation and implementation of this
intended curriculum has been found to differ from the way it was meant to be
implemented (Royal Netherlands Academy of Arts and Sciences, 2009). Furthermore,
the contextual exercises presented in these curricular textbooks seem to focus primarily
on introducing a new type of problem and are designed in such a way that they facilitate
the nature of the calculation which has to be made. These exercises focus mainly on
improving the students’ arithmetic skills and on solving simple, superficial contextual
problems (Royal Netherlands Academy of Arts and Sciences, 2009). Yet the question
items in the Cito assessments for mathematics focus on mathematical understanding,
requiring both computational skills as well as skills to translate the real-world problem
into a mathematical problem. In the curricular textbooks, students are hardly expected to
transfer word problems into actual sums (Meelissen & Drent, 2008). Since there is no
emphasis on explicitly instructing students how to systematically solve word problems,
they are bound to get easily confused when confronted with more complex word
problems (Janssen et al., 2005).
Appendix 3: Overview of the cutscores and performance categories

The studies presented in this dissertation were conducted to evaluate the effectiveness of a teacher Professional Development Program (PDP) in regard to mathematics. The teacher PDP targeted this subject area as well as the subject area of reading comprehension due to performance concerns for both areas (Ministry of Education, 2009; 2010). The results of the program with respect to reading comprehension were discussed in the dissertation of van Kuijk (2014).

Below, the cutscores and the accompanying performance categories are presented for both mathematics and reading comprehension. These cutscores were set in meeting 1 and 3.

Table 1: Overview of the Cutscores for Mathematics and Reading Comprehension

<table>
<thead>
<tr>
<th>Cutscore</th>
<th>Mathematics Second grade</th>
<th>Mathematics Third grade</th>
<th>Reading Comprehension Second grade</th>
<th>Reading Comprehension Third grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>37</td>
<td>57</td>
<td>-7</td>
<td>12</td>
</tr>
<tr>
<td>Basic</td>
<td>48</td>
<td>67</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Proficient</td>
<td>61</td>
<td>77</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>Advanced</td>
<td>71</td>
<td>92</td>
<td>27</td>
<td>45</td>
</tr>
</tbody>
</table>

Table 2: Overview of the Performance Categories and the Associated Range of Assessment Scores for Mathematics and Reading Comprehension

<table>
<thead>
<tr>
<th>Performance category</th>
<th>Mathematics Second grade</th>
<th>Mathematics Third grade</th>
<th>Reading Comprehension Second grade</th>
<th>Reading Comprehension Third grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below minimum</td>
<td>≤ 36</td>
<td>≤ 56</td>
<td>≤ -8</td>
<td>≤ 11</td>
</tr>
<tr>
<td>Minimum</td>
<td>37 – 47</td>
<td>57 – 66</td>
<td>-7 – 0</td>
<td>12 – 19</td>
</tr>
<tr>
<td>Basic</td>
<td>48 – 60</td>
<td>67 – 76</td>
<td>1 – 11</td>
<td>20 – 29</td>
</tr>
<tr>
<td>Proficient</td>
<td>61 – 70</td>
<td>77 – 91</td>
<td>12 – 26</td>
<td>30 – 44</td>
</tr>
<tr>
<td>Advanced</td>
<td>≥ 71</td>
<td>≥ 92</td>
<td>≥ 27</td>
<td>≥ 45</td>
</tr>
</tbody>
</table>
Appendix 4: Description of the data analyses used in the PDP

In this appendix, reference is made to several analyses provided by the digital student monitoring system, below referred to as the Cito student monitoring system. The participants received training in filling in and interpreting this system. Below, their analyses are briefly described.

During the second meeting, the participants were trained in performing the following analyses:

A. *An estimation of future performance*

Based on the prior achievements on the Cito assessments in a particular subject area, the Cito student monitoring system can provide an estimate for each student’s “expected future performance”. Estimates are provided with respect to two subsequent Cito assessments.

B. *A performance overview (in the PDP: from the previous school year)*

The Cito student monitoring system can list the performances of students on the January and June-assessments for each school year.

C. *A progress report (‘ability growth report’)*

The Cito student monitoring system can calculate the difference between a current and a prior assessment score to see whether or not a student has improved his or her performance over a certain time period.

During the fifth meeting, the teachers and management staff received training in performing analyses relevant to their function. The teachers were trained in the following analyses:

D. *The alternative student report (a performance comparison which makes extreme high or low scores more tangible)*

This analysis gives information about the actual performance level of students with extreme high or low scores. As the proficiency scale for the Cito-assessments is the same for grades one to six, this analysis can indicate the actual level in terms of a grade mean.
that is (far) below or above the student’s current grade. For example, a second-grade student with a very high math score might perform similar to the average performance of Dutch students in fourth grade (which is two years above the mean).

E. The average performance of the class

The Cito student monitoring system can produce a graph representing a particular class’ average test performance for several consecutive school years. Salient patterns signal a need for further investigation.

F. Error analysis for the mathematics assessment (‘category analysis’)

The Cito-assessments for mathematics can be further analyzed by calculating students’ weighed performance on the different mathematical domains. The analysis compares a student’s actual performance on a specific mathematical domain to his/her expected performance, relative to his/her overall performance level. This comparison is made for all domains assessed, resulting in an overview of the student scores on each mathematical domain relative to his overall performance level. The analysis delivers two tables. The first one shows how the actual performance deviates from the expected performance per student. This leads to three categories, for which different signs are given ‘non-salient’, ‘salient’, and ‘very salient’. These categories provide information on whether and how much the actual performances deviate significantly from the expected achievements. In the second table, group information is summarized as regards positive and/or negative deviations of the group’s expected performance on the different mathematical domains. On both levels - individual and whole group - the teachers are thus informed about students’ strengths and weaknesses in domain-specific knowledge and skills.

During the fifth meeting, the school management staff was trained in performing the following analyses:

G. Cross-section of performance levels

Students’ performance on the Cito assessments can be classified in terms of ability level indicators (A to E, and I to V): we refer to the A to E distribution as it is still the most common classification in Dutch primary schools. The top 25 percent of the performance distribution (thus, the best performing students) are given an ‘A’. The next 25 percent of the performance distribution (thus, the ‘second best’ performing students) receive a ‘B’.
The next 25 percent of the performance distribution (thus, the ‘third-best’ performing students) receive a ‘C’. The 15 percent of students performing below that C receive a ‘D’ (15%), and the 10 percent of the lowest performing students receive an ‘E’. This distribution is computed by the Cito student monitoring system. The cross-section analysis has two options for reporting its results: 1) cross-sections for one or more classes with respect to one subject area, and 2) cross-sections for one or more classes with respect to different subject areas.

**H. Longitudinal developments of the group’s mean score(s) of one or more classes (‘trend analysis’).**

The Cito student monitoring system provides two options for looking at the development of student performance throughout primary school using means: 1) the cross-sectional development of the group mean, and 2) the longitudinal development of the group mean. The cross-sectional development of the group mean shows the scores of several grades (e.g., grades one, two, and three) on a specific subject area throughout different years. Here the performance stability of different groups of students is compared. The longitudinal development of the group mean indicates the stability of a specific group of students’ mean growth during primary school (grade 1-6). Here the mean growth of one or more classes in one year is compared to its/their growth in other school years.

**I. Mean score of a specific grade throughout several years**

The student monitoring system can also show the stability of the mean scores of a particular grade throughout several years.

**J. Group analyses**

In this analysis, data from the cross section and the trend analysis are combined, revealing the development pattern of a class with respect to different subject areas over several years. It provides an overview of a class’s mean score development, based on the five ability levels as identified in the Cito LOVS assessment system. This analysis helps to gain an insight into the developmental patterns of a class in different subject areas, again showing deviations if there are any.
Appendix 5: Lessons learned from the pilot

The PDP was conducted in the school year of 2011-2012, after the PD meetings and the associated materials had been piloted in the previous school year of 2010-2011. The experiences of the pilot study led to three relatively large modifications of the PDP’s set-up. The first one pertained to the simultaneous focus on the subject areas mathematics and reading comprehension. During the pilot study, there was a five-month period for reaching the goals for reading comprehension, followed by a comparable five-month period for mathematics. However, we experienced that this five month time frame per subject area was “too short” for the teachers to be able to work toward the full attainment of their goals. Therefore, it was decided to focus on both subject domains simultaneously in the 2011-2012 PDP. After the standard setting procedures for both subjects were completed, the teachers set their goals in November/December 2011 for the June 2012 assessment, which gave them more time to work toward their goal attainment.

The second modification in the 2011-2012 PDP’s set-up applies to the role of the school management staff. During the pilot study, teachers, school principals and internal support coordinators participated in the meetings, whereas the focus was mostly on the teachers’ knowledge, practices and assignments. The school management staff from the pilot study schools indicated that they would have liked to receive more information and assignments targeted at their specific function in the school. Furthermore, it was found that in those schools in the pilot study where the school management staff was particularly dedicated to the project, the teachers were too. This is in line with the findings of, for example, research on data use (Schildkamp & Kuiper, 2010, Park & Datnow, 2009). In the 2011-2012 PDP, the content of the performance data analyses was therefore adapted in such a way that the school management staff was targeted more actively (resulting in different data analysis assignments and the use of parallel sessions during the fifth meeting). Furthermore, in the 2011-2012 PDP, the school management staff was asked to conduct a lesson observation and provide teachers with constructive feedback on the implementation of instructional practices.

The third modification concerned the instructional practices addressed during the meetings to help the teachers meet their achievement goals. In the pilot study, both modeling and the Direct Instruction model were discussed around February. However, the researchers experienced that the participants failed to recall these instructional
practices later on and that their actual implementation was insufficient. Throughout the 2011-2012 PDP, both the content and value of the Direct Instruction approach and explicit strategy-instruction (modeling) were therefore more explicitly and repeatedly discussed. In addition, the teachers were given feedback on both their modeling practices (provided by the researchers) and on their implementation of Direct Instruction (provided by the researchers as well as the school management staff).

All modifications to the PDP’s set-up following the pilot study aimed to better fit the needs of the participants in order to increase student achievement.
Appendix 6: Observation instrument

The time-sampling instrument consisted of several aspects of teachers’ and students’ activities which were coded every two minutes for the entire lesson (maximum duration: 60 minutes). With help of this instrument, we coded the phase of the lesson as well as several aspects of teacher behavior, namely teacher position, teacher talk, and the student at whom this teacher talk is directed. These categories are elaborated in Table 1. Furthermore, we selected four specific students prior to the lesson observation on the basis of their academic performance and coded the activity in which they were participating. These four students, being the same at pre- and postmeasurement, were a) a very weak performing student, b) a weak performing student, c) an average performing student, and d) an advanced student. The teachers were not informed that we were interested in students’ activities nor were they informed which students were selected for the purpose of the observation.

In Table 2, we present the scheme we used for conducting the time-sampling data collection. In light of parsimony, only the first fifteen two-minute intervals (i.e., the first 30 minutes) and the last interval (at the 60th minute) are depicted in this table.

In Table 3, the high-inference measure is presented. This measure contained 16 items pertaining to different aspects of teachers’ behavior which were filled in directly after the lesson was observed. All items had dichotomous response options, with a 0 (no) or a 1 (yes) in the case that implementation of these practices was observed.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classroom organization (i.e., Phase)</strong></td>
<td>1. Whole-class instruction</td>
<td>The teacher provides whole-class instruction.</td>
</tr>
<tr>
<td></td>
<td>2. Extended instruction</td>
<td>One or more students receive extended instruction, while the other students in class do seatwork.</td>
</tr>
<tr>
<td></td>
<td>3. Seatwork</td>
<td>All students are working on exercises (individually, in pairs or small groups).</td>
</tr>
<tr>
<td><strong>Position of the teacher</strong></td>
<td>1. In front of the class</td>
<td>The teacher is standing or sitting in front of the students</td>
</tr>
<tr>
<td></td>
<td>2. At a student’s table or group of tables</td>
<td>The teacher is standing or sitting at a student’s table or a group of tables</td>
</tr>
<tr>
<td></td>
<td>3. Walking around</td>
<td>The teacher is walking around the class</td>
</tr>
<tr>
<td></td>
<td>4. At the desk</td>
<td>The teacher sits at the desk</td>
</tr>
<tr>
<td></td>
<td>5. Other</td>
<td>Teacher position is not options 1 - 4 (e.g., outside the class)</td>
</tr>
<tr>
<td><strong>Teacher talk</strong></td>
<td>1. Task at hand</td>
<td>The teacher refers to the task at hand (e.g. ’we will start with exercise 1, page 14’, ’now, we are going to do some automation exercises’)</td>
</tr>
<tr>
<td></td>
<td>2A. Explanation pertaining to content</td>
<td>The teacher provides information on the task, strategies and solutions (e.g. ‘18 times 6. To solve this, you can take two steps. First, you calculate 10<em>6 and then 8</em>6’)</td>
</tr>
<tr>
<td></td>
<td>2B. Content-related questioning</td>
<td>The teacher asks for information on the task, strategies and solutions (e.g. ’How much is 6*8?’ or ’How did you answer that question?’)</td>
</tr>
<tr>
<td></td>
<td>3. Organization</td>
<td>The teacher refers to the general sequence of the lesson or to conditions for working (e.g. ’Maria, please pay attention’ or ’you can come to me after the whole class instruction’)</td>
</tr>
<tr>
<td></td>
<td>4. Other</td>
<td>Teacher talk is not options 1 - 3 (e.g., ’well done’).</td>
</tr>
<tr>
<td><strong>Student who is addressed during teacher talk</strong></td>
<td>1. Very weak achieving student</td>
<td>Selected student, minimum level (10% lowest performers)</td>
</tr>
<tr>
<td></td>
<td>2. Weak achieving student</td>
<td>Selected student, basic level (25% lowest performers)</td>
</tr>
<tr>
<td></td>
<td>3. Average achieving student</td>
<td>Selected student, proficient level (average performer)</td>
</tr>
<tr>
<td></td>
<td>4. Advanced student</td>
<td>Selected student, advanced level (25% highest performers)</td>
</tr>
<tr>
<td></td>
<td>5. Other student</td>
<td>Non-selected student in the same class</td>
</tr>
<tr>
<td></td>
<td>6. Group of students or whole class</td>
<td>The whole class or a (small) group</td>
</tr>
<tr>
<td></td>
<td>7. Other</td>
<td>Other, e.g., a colleague who comes into the class or students of the other grade in a multi-grade class</td>
</tr>
<tr>
<td><strong>Activity very weak student (the same for the weak, average, and advanced student)</strong></td>
<td>1. Whole-class teaching</td>
<td>The student is engaged in whole-class teaching</td>
</tr>
<tr>
<td></td>
<td>2. Extended instruction</td>
<td>The student receives additional instruction in a small group</td>
</tr>
<tr>
<td></td>
<td>3. Individual teacher instruction</td>
<td>The student receives additional, individual instruction or is working individually with the teacher</td>
</tr>
<tr>
<td></td>
<td>4. Working independently</td>
<td>The student works on his own or with (a) peer(s)</td>
</tr>
<tr>
<td></td>
<td>5. Other</td>
<td>The student is outside the classroom or is working on exercises from a different subject area.</td>
</tr>
</tbody>
</table>
Table 2: Time-sampling scheme

<table>
<thead>
<tr>
<th>Class</th>
<th>Teacher</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Phase</td>
<td>Position</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>..</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: High-inference measure

<table>
<thead>
<tr>
<th>The teacher ....</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ... summarizes the content of the prior lesson or activates relevant prior knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 ... explicates the learning goal, content and/or topic of that lesson</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 ... starts the mathematics lesson with an automation exercise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 ... clarifies in which way the assignments are accomplished in a satisfactory manner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 ... provides extended instruction to a student or group of students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 ... differentiates for the weaker students in the assignments these students are expected to complete</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 ... differentiates for well achieving students in the assignments that these students are expected to complete</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 8 ... lets students that have completed their assignments ...  
- for math: work on more advanced mathematical materials  
- for reading: read for themselves |    |     |
| 9 ... gives the students time to answer the question (+ 3 seconds). |    |     |
| 10 ... repeats the right answer. |    |     |
| 11 ... compliments students when they have answered the question correctly. |    |     |
| 12 ... returns to the learning goal of that lesson and/or the new skill and/or knowledge that have been addressed at the end of the lesson |    |     |
| 13 ... connects the content of the current lesson to the following lesson |    |     |
| 14 ... goes into depth in regard to the approach and used strategies after right answers are provided |    |     |
| 15 ... goes into depth in regard to the approach and used strategies after wrong answers are provided |    |     |
| 16 .. models his/her application of knowledge, skill or strategy by thinking aloud |    |     |