Article

Journal of Teacher Education
2018, Vol. 69(3) 307–320

DOI: 10.1177/0022487117704170
journals.sagepub.com/home/jte
SAGE

Effects of a Data-Based Decision-Making Intervention for Teachers on Students' Mathematical Achievement

Emmelien A. van der Scheer[1,2] and Adrie J. Visscher[1]

Abstract

Data-based decision making (DBDM) is an important element of educational policy in many countries, as it is assumed that student achievement will improve if teachers worked in a data-based way. However, studies that evaluate rigorously the effects of DBDM on student achievement are scarce. In this study, the effects of an intensive DBDM-intervention for Grade 4 teachers on students' mathematical achievement were investigated in a randomized controlled trial. Multilevel analyses showed that although no main effect on students' mathematical achievements was found, students who received "extended instruction" benefited significantly from the intervention. Based on the results, recommendations for the design of new DBDM-interventions and for their evaluation are presented.

Keywords

data-based decision making, professional development, mathematics

Introduction

Despite the worldwide emphasis in educational policy on data use to enhance student achievement, research has not shown convincingly that data use results in improved student achievement (Lai & Schildkamp, 2013; Marsh, 2012; Slavin, Cheung, Holmes, Madden, & Chamberlain, 2013). A possible explanation is that studies into the effects of data-based decision making (DBDM) generally do not evaluate interventions aimed to support *teachers* in the implementation of DBDM (Carlson, Borman, & Robinson, 2011; Coburn & Turner, 2012; Lai & Schildkamp, 2013; Slavin et al., 2013). Teachers training in DBDM might be crucial for improving student achievement, as teachers have a considerable impact on student learning (Hattie, 2009). Teachers need support in developing their knowledge and skills for using student achievement data to adapt classroom instruction to the students' needs (Datnow, Park, & Wohlstetter, 2007; Goertz, Oláh, & Riggan, 2009; Slavin et al., 2013).

This study aimed to provide evidence that data use can enhance student achievement if teachers are trained intensively in DBDM. Participating teachers were trained in multiple aspects of DBDM, with a strong emphasis on the implementation of DBDM in the classroom through coaching.

At the classroom level, Dunn, Airola, Lo, and Garrison (2013) defined DBDM as "the identification of patterns of performance that unveil students' strengths and weaknesses relative to students' learning goals as well as the selection and planning of instructional strategies and interventions to facilitate student achievement of learning goals" (p. 225). Research has shown that Dutch teachers can be trained effectively in interpreting and analyzing data (Staman, Visscher, & Luyten, 2014). However, working in a data-based way requires more than data analysis (Gummer & Mandinach, 2015; Mandinach & Gummer, 2013, 2016a, 2016b) An important prerequisite is that teachers provide instruction that suits each student's instructional needs (Means, Chen, DeBarger, & Padilla, 2011). In other words, this requires the acquisition of differentiation skills, whose mastery is rather complex (Tomlinson et al., 2003). Only 52% of Dutch primary school teachers master both the basic didactic skills (e.g., classroom management) and the more complex differentiation skills (Inspectorate of Education, 2014).

An effective way to support teachers in developing differentiation skills in the context of DBDM might be to coach them in their own classrooms. Currently, several reform programs use "coaching" to support teachers in implementing

[1]University of Twente, Enschede, The Netherlands
[2]University of Groningen, The Netherlands

Corresponding Author:
Emmelien A. van der Scheer, Groningen Institute for Educational Research, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, The Netherlands.
Email: e.a.van.der.scheer@rug.nl

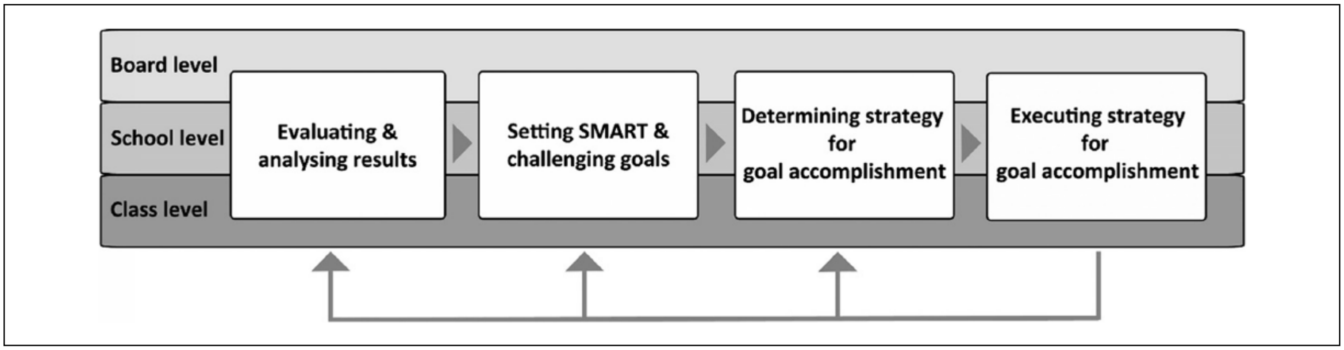

Figure 1. Four components and levels of DBDM (Keuning, Van Geel, Visscher, Fox, & Moolenaar, 2016).
Note. DBDM = data-based decision making.

DBDM in the classroom. Coaches offer on-site coaching, which is context-embedded and can promote reflection on teacher practice (Marsh, Sloan McCombs, & Martorell, 2010). Indeed, positive effects from coaching on both teacher practice and student achievement have been reported (Biancarosa, Bryk, & Dexter, 2010; Marsh et al., 2010; Van der Steeg & Gerritsen, 2013).

Based on the aforementioned DBDM principles, a DBDM-intervention was designed to promote DBDM implementation in the classroom: a combination of general meetings for groups of teachers (where teachers were trained in general DBDM-aspects, for example, data analysis) and coaching sessions during which teachers were coached individually with respect to how to implement DBDM in their classrooms. As the final goal of DBDM is improved student achievement, the following research question is addressed by means of a randomized controlled trial:

> **Research Question 1:** What is the effect of an intensive DBDM-intervention for Grade 4 teachers on students' mathematical achievement?

Conceptual Framework

In this section, we elaborate on what DBDM in the classroom entails and why working in a data-based way is expected to impact student achievement. Next, we discuss what the scientific literature describes concerning effectively preparing teachers for DBDM. Finally, the study hypotheses are presented and explained.

DBDM in the Classroom

Although DBDM is widely promoted, the concept varies somewhat between countries as educational policies and systems differ. Visscher and Ehren (2011) distinguished between three core DBDM components:

1. Determine the instructional needs of individual students and groups of students.
2. Define the SMART (Specific, Measurable, Attainable, Realistic, and Time-Bound) and challenging learning and performance goals.
3. Based on the information from Steps 1 and 2, teachers decide which instructional approach is most promising for accomplishing the goals.

Keuning, Van Geel, Visscher, Fox, and Moolenaar (2016) have introduced a fourth step: the execution of the planned strategy for goal accomplishment. The four components and the levels of DBDM are shown in Figure 1.

DBDM at the class level, which is the focus of this study, requires that teachers work systematically in addressing each DBDM component (van der Scheer, 2016). Teachers need to master each of the components; however, they need not necessarily execute them in the same order. For example, if a teacher notices during teaching that the instructional strategy does not fit the students' needs, the teacher could adapt the strategy (Component 3) without having to change the goals first (Component 2).

The DBDM components in Figure 1 relate closely to the framework of Ikemoto and Marsh (2007) and the Multi-Tiered Systems of Support (MTSS; Benner, Kutash, Nelson, & Fisher, 2013). However, Ikemoto and Marsh (2007) elaborated more on the collection of data, whereas in Figure 1 "goal setting" is included as a separate component of DBDM. This was done for two reasons. First, setting goals is necessary for determining whether students have made the expected (or required) progress. This way, data analysis and interpretation have a clear purpose and no time is spent on unnecessary analyses (Lai & Schildkamp, 2013). Second, working with well-defined goals generally improves performance (Locke & Latham, 2002). As presented in Figure 1, MTSS encompasses more than the approach to DBDM. It includes DBDM, just as analyzing students' behavior, and interventions aimed at improving their behavior. In this study, DBDM focuses predominantly on improving students' cognitive achievements, as this is aligned with the national goals for DBDM in the Netherlands (Inspectorate of Education, 2010b).

We will now describe in further detail all four DBDM model components and their relation with student achievement.

Evaluating and analyzing results. To gain insight into students' instructional needs, data have to be analyzed and interpreted. Student performance data can provide teachers with feedback on the results of their actions. Feedback can be an important mechanism for improving performance (e.g., Hattie & Timperley, 2007), as "data can help teachers to monitor their constantly changing environment, their functioning and to what extent curriculum aims are met, and react timely and in an evidence-based manner when problems need to be solved" (Schildkamp & Kuiper, 2010, p. 482).

Teachers have to combine the results of the data analysis with data from other sources (e.g., students' daily work) to adapt teaching in line with students' needs (Ingram, Louis, & Schroeder, 2004). Based on this information, learning and performance goals can be formulated for each student.

Setting SMART and challenging goals. The ability to set clear performance goals is not self-evident. Visscher and Ehren (2011) found that teachers formulate "action-oriented" goals like "we will strive for high scores." Locke and Latham (2002) found that setting SMART and *challenging* goals can improve performance considerably. When combined, goal setting and feedback synergize; feedback informs about the degree of goal accomplishment, thereby serving as an encouragement to try harder and/or explore alternative strategies for goal accomplishment (Locke & Latham, 2002).

Determining strategy for goal accomplishment. To decide on suitable instructional strategies for each student, the data analysis results and the goals have to be combined with instructional and curricular knowledge to place the data in a meaningful context; for example, (pedagogical) content knowledge (Mandinach & Gummer, 2016a), knowledge about learning progression, the structure of the curriculum, effective instructional strategies, and lesson planning (Visscher & Ehren, 2011).

In the case of Dutch DBDM, the use of an instructional plan is considered important to contemplate instruction carefully. Such a plan is supposed to include the student performance goals and the instructional strategies for each student. It is common practice in Dutch classrooms to distinguish between three instructional groups: a low-performing group of students (the "extended instruction group"), an average-performing group (the "basic instruction group"), and the high-performing group (the "shortened instruction group"). As whole-class instruction provided to the entire class is not sufficient for students in the *extended instruction group*, they receive *supplementary* instruction. It is important that these students are not withdrawn from whole-class instruction as this may weaken their performance (Braddock & Slavin, 1995; Houtveen & Van de Grift, 2012). *High performing* students might lose their interest when confronted with non-differentiated instruction (McIntosh, as cited in Tomlinson et al., 2003). Therefore, these students should be allowed to work independently as soon as the new subject-matter content is understood. As a result, they will have more time for additional, more difficult assignments and receive additional instruction at their own level (Inspectorate of Education, 2014).

Executing a strategy for goal accomplishment. Planning instructional strategies and formulating challenging and feasible goals are important prerequisites for effective differentiation (Tomlinson et al., 2003). However, although Dutch teachers write instructional plans, only a minority realizes these plans in the classroom (Inspectorate of Education, 2014). An explanation for this might be that differentiation is a complex skill to master (Van de Grift, Van der Wal, & Torenbeek, 2011), in that teachers need to work flexibly across instructional groups, vary instructional material and instruction as well as the pace of instruction in response to students' learning needs (Tomlinson et al., 2003). Moreover, the mere division of students into instruction groups is not effective; instruction has to address the needs of students within each group (Deunk, Doolaard, Smale-Jacobse, & Bosker, 2015).

Student Achievement

Each of the four components incorporates characteristics that are supposed to be associated with improving student achievement such as feedback, goal setting, and differentiation (Hattie & Timperley, 2007; Locke & Latham, 2002; Tomlinson et al., 2003). The components together are supposed to encourage teachers to tailor their instruction to student needs, which in turn is expected to positively affect student achievement (Campbell & Levin, 2009; Carlson et al., 2011; Lai & Schildkamp, 2013; Van Geel, Keuning, Fox, & Visscher, 2016).

Now that DBDM at the classroom level has been defined in terms of its components and their relation with student achievement, we will briefly focus on the kind of professional development that might be required.

Professional Development of Teachers

Several reviews on the characteristics of effective professional development programs (PDPs) point to five features of such PDPs: "content focus," "active learning," "duration," "collective participation," and "coherence" (Desimone, 2009; Van Veen, Zwart, & Meirink, 2012). Research shows that the PDPs focusing on classroom practice (e.g., on the subject-matter *content* and/or how to teach it) have more impact on teacher cognition, classroom practice, and student achievement, than PDPs focusing on topics that are not directly related to classroom practice (Garet, Porter, Desimone, Birman, & Yoon, 2001; Van et al., 2012). In our

PDP, the *content focus* was incorporated, with each aspect of the intervention focusing on daily classroom practice.

Active learning only requires that the instructional method used in a PDP engages the participants; thus, it can take on many forms (Van et al., 2012). This intervention incorporated several forms of *active learning* such as coaching and making an instructional plan.

The *duration* of the PDP was one school year, consisting of approximately 45 hr. This way, teachers had multiple opportunities for in-depth discussion of the intervention-content to promote the integration of teacher knowledge and skills, and with multiple opportunities to implement new practices, receive feedback on them, and improve implementation (Garet et al., 2001).

According to Van et al. (2012), interaction and discussion between teachers and the exchange of feedback between them can be a powerful asset for the facilitation of learning. Therefore, *collective participation* is considered important. In our intervention, collective participation was assured by organizing central meetings.

An important aspect of a *coherent* PDP is that the PDP is in line with national as well as school policies, ensuring teachers' motivation to work on this professional development trajectory (Desimone, 2009; Van et al., 2012). Dutch schools are evaluated by the Inspectorate on whether they work in a data-based way, which implies that it was important for schools and teachers to work on DBDM.

Hypotheses for the Present Study

In this study, teachers were trained during one school year with respect to the knowledge and skills required to work in a data-based way in the classroom, in line with the four components. The PDP was designed in line with the five core features of effective PDPs and in line with current teaching practices in Dutch primary schools (to strengthen the latter as much as possible).

The PDP-effects on mathematical achievement were investigated based on the three hypotheses:

> **Hypothesis 1:** Mathematical achievement of students taught by teachers in the experimental group will be higher on the posttest than mathematical achievement of students in the control group.

Furthermore, teachers in the intervention were encouraged to work with three instruction groups, which, as previously indicated, is in line with most Dutch school policies. Based on data analysis, extended instruction group students received additional instruction matching their needs, while high-performing students could work more independently (shortened instruction group). Therefore, it was expected that the students in the extended and in the shortened instruction group would be the ones to benefit most from the intervention.

> **Hypothesis 2:** A positive interaction effect is expected for students who are in both the extended instruction group and in the experimental group.
> **Hypothesis 3:** A positive interaction effect is expected for students who are in both the shortened instruction group and in the experimental group.

Research Method

Initial Sample

Primary schools with a high percentage of low-SES (socioeconomic status) students were contacted by email to participate in the project.[1] It was decided to approach these schools, as the Inspectorate of Education showed that schools with a high percentage of low-SES students underperform more often (Inspectorate of Education, 2010a), and therefore require more support.

Contacted school leaders and teachers were informed about the design of the study, about what was expected from participating schools, and the content of the DBDM-intervention. Grade 4 (9-10 year old students) teachers of 60 different primary schools agreed to participate and were randomly allocated to an experimental group (30 classes, 39 teachers) or a control group (30 classes, 34 teachers). The experimental group received a 1-year DBDM-intervention during the school year 2013-2014; teachers in the control group (business as usual) only provided student achievement data and student and teacher background information. Control group teachers were not offered any support other than feedback from a student questionnaire filled out by their students and they were not involved in another DBDM-intervention during the intervention year. Control group teachers could participate in the intervention in the next school year (2014-2015).

In the case of part-time teachers, both teachers of a class had to participate if they were responsible for at least two out of five mathematics lessons a week. These teachers, responsible for the same student group, will be referred to as a "teacher pair" in this article.

Treatment fidelity. To retain a power of .80 ($\alpha = .05$) for detecting an effect of 0.22 of a standard deviation, we started with 60 classes (assuming an average class size of 20 students); power calculations were done with PINT2.12. Despite clear agreements beforehand regarding what participation would require from teachers prior to the intervention, 11 classes could not be included in the analysis. Three teachers dropped out of the intervention group at an early stage due to motivational difficulties. Four teachers were absent from school for more than 4 months during the intervention year (due to illness, other than a burn-out attributed to workload) and one teacher had participated in another intensive intervention

Table 1. Teacher Characteristics for the Experimental and the Control Group.

	Experimental group		Control group		Test for differences	
	n	%	*n*	%	χ^2 (*df*)	*p*
Classes	19		29			
Teachers	25		33			
Teacher pair	6		4			.13
Teacher gender						
Men	7	28.00	8	24.24		
Women	18	72.00	25	75.76	0.11 (1)	.75
Multigrade classroom	6	31.58	10	34.48	0.04 (1)	.84
	M	*SD*	*M*	*SD*	*t* (*df*)	*p*
Teaching experience in years	12.80	10.20	11.97	10.56	0.30 (56)	.77
Class size	18.79	4.04	19.28	4.07	−0.41 (46)	.69

Note. Degrees of freedom are in parentheses.
$^{**}p < .05$.

during the intervention year. In three classrooms, teacher training institute students took over the teaching activities during the second half of the year (for 3 or more days a week). As the reasons for exclusion were either unrelated to the intervention (e.g., considered irrelevant by teachers) or a common relevant teacher characteristic (e.g., poor teaching quality), we expected the included and excluded teachers to be similar. Indeed, no significant differences were found between the 19 experimental classes and the 11 excluded classes on any of the student and teacher background characteristics (teacher and student gender, teacher pair, teaching a multigrade classroom, class size, teaching experience, student age, and pretest scores). Thus, the remaining 19 experimental classes were deemed representative of the entire experimental group.

In the control group, one teacher turned out not to teach Grade 4 students. The remaining 29 classes from the control group were included in the analyses.

Final Sample

Teacher characteristics for the experimental group and the control group are presented in Table 1. Chi-square tests (for teacher gender, teacher pair (Fisher's exact test), multigrade classroom) assessed for significant differences between the control and experimental group. The differences in class size and teaching experience were tested with an independent-samples *t* test.

Table 1 shows that no significant differences were found between the experimental and the control group with respect to teacher characteristics.

Student characteristics for the experimental group and the control group are presented in Table 2. A distinction was made between instruction groups: the extended instruction group, the basic instruction group, and the shortened instruction group. The control and experimental groups were compared using chi-square tests for the distribution of students within instruction groups and student gender. Furthermore, differences in student age and pretest scores were tested for with an independent-samples *t* test.

As expected, student pretest scores varied across instructional groups but not between the experimental and control groups. Table 2 shows that no systematic differences existed between students from the experimental and the control group prior to the intervention.

Intervention

Teachers in the experimental group followed a DBDM training course, which included seven meetings (or 36 hr contact time) and four coaching sessions on how to implement DBDM in the classroom. Teachers were divided into five groups based on their schools' location. The intervention was delivered by a trainer who had a university master's degree, was a primary school teacher, and is experienced in training teachers and schools in DBDM.

All meetings were attended by 68% of the teachers, 24% attended six meetings, and 8% (two teachers) attended five meetings. Generally, if teachers could not attend a meeting this was due to sickness and, in most cases, their teacher pair colleague was present to attend the meeting. All but one teacher were coached 4 times; one teacher was coached 3 times. After each meeting and coaching moment, teachers filled out an evaluation form; overall, they indicated the topics and assignments they worked on during the meetings were (very) useful. Furthermore, significant positive effects of the intervention were found on teacher efficacy and the use of instruction groups during the lessons (van der Scheer, Glas, & Visscher, in press; van der Scheer & Visscher, 2016).

During the seven meetings, teachers were trained to work in line with the four DBDM components (see Figure 1); an

Table 2. Student Characteristics for the Experimental and the Control Group.

	Experimental group		Control group		Test for differences	
	n	%	*n*	%	χ^2 (*df*)	*p*
Instruction group (IG)						
Total	269		404			
Extended IG	74	27.51	116	28.71		
Basic IG	115	42.75	183	45.30		
Shortened IG	80	29.74	105	25.99	1.15 (2)	.56
Student gender						
Boy	142	52.79	188	46.53		
Girl	127	47.21	216	53.46	2.53 (1)	.11
	M	*SD*	*M*	*SD*	*t* (*df*)	*p*
Pretest						
Total	77.20	14.40	75.66	14.30	1.36 (671)	.17
Extended IG	64.55	10.99	62.61	11.98	1.12 (188)	.26
Basic IG	76.38	9.44	76.28	10.12	0.08 (296)	.93
Shortened IG	90.07	12.01	89.00	9.14	0.69 (183)	.49
Student age						
Months	125.11	6.45	125.56	6.05	−0.93 (671)	.36

Note. Degrees of freedom are in parentheses.
**$p < .05$.

extensive description of how each component was addressed is presented now.

Evaluating and analyzing results. Dutch students are assessed by means of standardized assessments twice per school year (as described in more detail in the "Instruments" section). During the first, fifth, and seventh meeting, participating teachers analyzed the results of the most recent standardized mathematics test for their own classrooms by means of a student monitoring system (SMS) that was already in use by their schools. These analyses are incorporated in the SMS (for an overview, see Staman et al., 2014) and allow teachers to draw conclusions about a *student's* current ability level relative to students' prior ability level and the national benchmarks. In addition, *the class's* performance can be compared with that of prior classes from the same school and to national benchmarks. Moreover, detailed conclusions can be drawn regarding mathematical topics, which individual students or groups of students have not fully mastered yet.

Teachers were taught how to use their SMS and how to draw conclusions based on their findings. During the first meeting, the trainer showed and explained which kinds of analyses could be performed using the SMS and how to interpret the content of the resulting graphs and tables. Next, teachers were provided with an analysis protocol, including a step-by-step guide for retrieving the different graphs and tables, and requiring them to perform analyses while answering questions such as, "Please describe how your class performed at the last assessment in comparison with the previous assessment" and "Could you describe which students developed differently from the expected based on the relevant norm group?" Furthermore, they were asked to explain findings (e.g., "Could you explain, per student, why the student did not perform as expected?"). This way, teachers were encouraged to think more profoundly about the performance of their students. The trainer provided individual feedback, for example, on whether the conclusions drawn were correct. Research has shown that teachers can improve their SMS skills this way (Staman et al., 2014). Furthermore, teachers learned how to relate a student's ability level to the mathematics learning progression, that is, to determine the student's zone of proximal development. As the founders of the standardized tests provide a clear overview of how their tests relate to the learning progression, the trainer could easily explain this to the teachers (the learning progression was also extensively discussed in Meeting 3).

Setting SMART and challenging goals. Teachers had to make an instructional plan twice (during the second and fifth meeting when new student progress data were available). This plan had to include performance goals for all students, the allocation of students into instruction groups (extended, basic, or shortened instruction group), planned instructional strategies for the instruction groups and for individual students, and student assignment to special learning groups.

Teachers were taught how to set two types of goals; content-related goals that were directly related to the mathematical curriculum or curriculum-based tests and norm-based goals that were aligned with the standardized mathematics tests. As teachers were taught how to determine students' zones of proximal development, they had the required knowledge to set relevant curriculum goals. The trainer presented examples of these goals. Teachers could base their norm-related goals on benchmarks of average ability growth for mathematics in Grade 4 and for various categories of students.

Determining strategy for goal accomplishment. To accomplish the goals, teachers were required to formulate instructional strategies. Teachers were encouraged to form instruction groups. They divided students over these groups based on mathematical achievement (according to the results of the standardized test) and student performance ratings in the classroom, for example, based on daily work and curriculum-based tests. In addition, teachers formed special learning groups to plan extra support for few students on specific topics; it was advised that students from the shortened instruction group would be one of these groups. Teachers planned this additional support mostly once or twice a week during their mathematics lessons, for example, in lessons when students have to work independently. The instructional plan was an important aspect of the intervention as planning is an important prerequisite for effective differentiation (Lou et al., 1996; Tomlinson et al., 2003).

After all teachers had made their instructional plans, they were asked to evaluate and discuss each other's plans in pairs. Each teacher identified similarities and differences between the two plans, wrote down questions regarding its practical feasibility, and planned execution. Then, these questions were discussed critically to improve their instructional plans. This evaluation resulted in not only very practical questions (e.g., "How do you organize this in your classroom?"), but also in more critical questions regarding the feasibility of the instructional plan (e.g., "Are you able to work with four special learning groups, when you work in a multigrade classroom?"). Moreover, the trainer provided each teacher with individual feedback on the instructional plan, the match between the analysis and the instructional plan, and on the appropriateness of the selected instructional strategies. In the fifth meeting, teachers were encouraged to evaluate their own instructional plan, for example, by evaluating whether and why goals were met or not. In this meeting, teachers also adapted their instructional plans based on the new data analysis.

Executing a strategy for goal accomplishment. The implementation of the instructional plan in the classroom was encouraged by means of coaching, reflection on one's own teaching practices, and peer feedback. Coaching offers teachers the opportunity to actively apply and integrate new knowledge into practice (Knight, 2009). Coaching interventions can differ, for example, in who the coach is (a peer teacher or an expert teacher) or in the focus of the coaching activities (Cornett & Knight, 2009; McKenna & Walpole, 2008).

In this study, an external coach supported the execution of the instructional plan in the classroom, especially working with three instruction groups. Thus, teachers were encouraged to provide their instruction, while allowing shortened instruction group students to work independently soon after the start, and providing additional instruction to students in the extended instruction group.

During four individual coaching sessions, each teacher received feedback from the trainer mainly on how he or she implemented the instructional plan in his or her classroom, and on the extent to which instruction was in line with the varying instructional needs of students. After the observed lesson, the trainer and teacher discussed the various lesson phases: the introduction, the formulation of the lesson goal, the presentation of new subject matter, independent work by students, and the evaluation of the lesson goal. Teachers first presented their opinions regarding their strengths and weaknesses for each lesson phase. Subsequently, the trainer provided his own opinion on the phases. Finally, the teacher and trainer discussed how the teacher could improve the implementation of DBDM. The trainer indicated that the content of the discussions varied considerably between teachers and over the course of the intervention. Especially at the beginning of the intervention, the discussions were focused on organizational aspects (how to organize a classroom that allows additional instruction for a small group of students). During the last few coaching sessions of the intervention, the discussions focused more on mathematical content, for example, on the kind of instructional strategy that would have suited the lesson or an individual student better. Each of the lessons observed by the trainer were recorded and teachers were able to re-watch these lessons online.

In the fourth and sixth meeting, teachers received and provided peer feedback on their own videotaped lessons (reduced to approximately 20 min of the original lesson) and on a selection of the recorded lessons. The lesson was evaluated and discussed based on a feedback form either in a teacher pair (Meeting 4) or with all attending teachers (Meeting 6). Guiding questions were formulated for each lesson phase, for example, "What are the strong aspects?" and "What could be improved?" Despite of initially reserved attitudes among most teachers, all teachers indicated on an evaluation form that they experienced these meetings as very useful. The trainer indicated that for most teachers, these activities led to more awareness about their own teaching behavior. It appeared to be rather difficult to have critical, content-focused discussions based on a short video fragment, but teachers exchanged tips and tricks on how to approach instructional topics and deal with organizational difficulties in the classroom.

An overview of the content of the intervention is provided in Table 3.

As shown in Table 3, the first three meetings were held during the first 2 months of the school year to ensure that teachers could start working in a data-based way almost immediately.

Table 3. Intervention Content.

Meeting	Month	Duration	Content
1	September	8 hr	Background of DBDM Explanation Cito assessments Explanation use and interpretation SMS Analyze and interpret own data with a SMS by means of a protocol Individualized feedback on the SMS-protocol
2	September	8 hr	Drawing up an instructional plan Formulating SMART goals Individualized feedback on instructional plan
3	October	4 hr	Explanation Dutch core goals Learning progression mathematics Instructional models Drawing up personal improvement goals
Coaching Session 1			
4	November	4 hr	Observation mastery lessons Observation videotaped lessons in teacher pairs Peer teachers provide feedback on videotaped lessons Feedback focused on how to differentiate instruction in the classroom
Coaching Session 2			
5	February	4 hr	As Meetings 1 and 2
Coaching Session 3			
6	April	4 hr	As Meeting 4
Coaching Session 4			
7	June	4 hr	As Meeting 1

Note. DBDM = data-based decision making; Cito = (Dutch central institute for test development), SMS = student monitoring system; SMART = Specific, Measurable, Attainable, Realistic, and Time-Bound.

Instruments

Student achievement was measured by means of the standardized mathematics tests from the Dutch Central Institute for Test Development (Cito), used by about 90% of Dutch primary schools (Ledoux, Blok, & Boogaard, 2009). The Dutch COTAN (responsible for assessing the quality of the tests used in the Netherlands) assessed the quality of the mathematics Cito test (e.g., reliability and validity) for six criteria as "good," and for one criterion as "not applicable" (ToetsGids, 2014). Test results are converted into a student ability score, depicted on a vertical ability scale (ability scores can be compared across grades).

In this study, a student's ability score at the end of Grade 3 (administered between May 15, 2013, and August 1, 2013) was used as a pretest score, the mid-grade 4 assessment (administered between January 1, 2014, and March 1, 2014) as the midtest. The end-of-grade-4 test (administered between May 15, 2014, and August 1, 2014) was used as a posttest score. This entails that the pretest scores were mostly administered by teachers other than those participating in this study, as the students were not taught by the participating teachers during third grade. An average score was calculated for those students who underwent multiple assessments during the specific time period.[2] The data from the schools' SMSs provided us with both students' achievement scores and students' demographic characteristics (e.g., gender).

Demographic data concerning teachers (e.g., age) were obtained via an online questionnaire. Teachers were also asked to provide information on the instruction group in which students had been placed (extended, basic, or shortened). They were asked to give this information for the first half and for the second half of the school year 2013-2014. This resulted in 123 missing values for the first half of the year (out of 673 students) and 50 missing values for the second half of the year (despite multiple requests to provide us with the information). For the statistical analyses, the instruction group of the second half of the year was used. For the 50 students with an unknown instruction group for the second half-year, the instruction group from the beginning of the year was used. This was considered reasonable as it was established from teacher reports at both time periods that 75% of the students remained in the same instruction group. However, for four students, it proved to be impossible to find out in which instruction group they had been taught, as such they were left out from the analyses.

Teachers from both the experimental and the control group identified a total of 46 students as "having their own learning progression" (5% of students in the experimental group and 7% of students in the control group). These students neither follow daily classroom instruction (although they are in that classroom) nor belong to any of the three instruction groups. They underperform strongly and work according to their own individual plan. They are predominantly supported by teaching assistants outside the classroom. As these students do not participate in the daily

Table 4. Average Ability Scores at the Pretest, the Midtest, and the Posttest, and Ability Growth of the Experimental Group and Control Group.

		Pretest		Midtest		Posttest		Growth	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experimental group									
Total	269	77.20	14.40	82.72	12.76	89.25	11.96	12.05	9.13
Extended IG	74	64.55	10.99	70.02	6.93	79.32	8.42	14.77	9.43
Basic IG	115	76.38	9.44	81.93	7.94	88.24	9.27	11.86	8.13
Shortened IG	80	90.07	12.01	95.61	9.84	99.88	9.36	9.81	9.64
Control group									
Total	404	75.66	14.30	81.13	15.19	87.40	12.98	11.74	8.81
Extended IG	116	62.61	11.98	66.95	12.10	75.33	11.82	12.72	10.91
Basic IG	183	76.28	10.12	81.52	9.87	88.17	8.77	11.89	7.82
Shortened IG	105	89.00	9.14	96.10	10.55	99.40	7.37	10.40	7.84

Note. IG = instruction groups.

mathematics routine, they also did not participate in the DBDM-intervention and were therefore excluded from the subsequent analyses.

Analysis

Grade 4 students from participating classes were included in the analyses, if at least their pretest and posttest scores were available. Grade 3 (114), Grade 5 (69), and Grade 6 (3) students in multigrade classrooms were left out of the analyses as the intervention was aimed at Grade 4 students.

As students are nested within classes, two multilevel regression models were estimated with student ability scores at the midtest and the posttest as the dependent variables, respectively. School was not considered as a level as all classes were from different schools. The intraclass correlation (ICC) was calculated to assess whether multilevel modeling was required (using the following formula: design effect = 1 + [*n* classes − 1] × ICC), that is, whether variation at the class level was large enough. This resulted in an ICC of .19 for the model with the midtest scores as dependent variable (design effect of 10.13) and an ICC of .17 (design effect of 8.99) for the model with the posttest scores as dependent variable. These results show that multilevel modeling was necessary (Peugh, 2010).

To check for the effect of student and teacher background variables, the following variables were first added to the model: student SES (based on student weight), gender, age (standardized), extended instruction group (yes/no), shortened instruction group (yes/no), pretest scores, teacher gender, number of years of teacher experience (standardized), multigrade classroom (yes/no), teacher pair (yes/no), and class size (defined as the total number of students in the classroom). In the case of teacher pairs, the background characteristics of the teacher who taught the class most of the time were used. Significant background variables ($p < .05$) stayed in the model, and this model was estimated again to ensure that the included background variables were truly significant. The remaining significant background variables are presented in the "context model," as shown in the "Results" section. Next, the intervention effect was added to the model to test the first hypothesis (the "intervention model"). In the last model, the interaction effects Intervention × Extended Instruction and Intervention × Shortened Instruction were added ("instruction group model"). The models were estimated with SPSS version 22, using maximum likelihood (ML) estimation to allow for model fit comparison.

Results

Before presenting the results of the multilevel analyses, an overview of the average ability scores on the pretest, midtest, and the posttest, as well as the growth across the pretest and posttest is provided in Table 4.

Table 4 shows that average ability growth is in favor of the experimental group (12.05), although the difference is limited: 0.31 ability score points. Students receiving extended instruction in the experimental group showed most growth (14.77).

Multilevel Regression Analysis

First, the effects of the DBDM-intervention during the first half of the school year were estimated. The results are presented in Table 5.

In the "intervention model," participation in the intervention hardly influences the scores on the midtest. The "instruction group model" shows that both interaction variables (participating in the intervention and being in the shortened or extended instruction group) do not have a statistically significant effect.

The results from the multilevel regression analysis with the posttest as the dependent variable are presented in Table 6.

Table 5. Multilevel Regression Analyses Predicting the Midtest Results.

	Empty model		Context model		Intervention model		Instruction group model	
Fixed part	Coefficients	*SE*	Coefficients	*SE*	Coefficients	*SE*	Coefficients	*SE*
Intercept	82.30**	1.04	39.79**	2.13	39.93**	2.22	40.06**	2.24
Participation in intervention					−0.35	1.52	−0.54	1.60
Pretest			0.55**	0.03	0.55**	0.03	0.54**	0.03
Extended instruction			−6.40**	0.67	−6.41**	0.67	−6.91**	0.82
Shortened instruction			7.07**	0.66	7.07**	0.66	7.32**	0.81
Participation × Extended							1.27	1.16
Participation × Shortened							−0.52	1.15
Multigrade			3.12*	1.60	3.11**	1.60	3.16*	1.60
Random								
Level: Class	37.75	10.32	23.67	5.43	23.65	5.43	23.43	5.38
Level: Students	165.15	9.33	33.18	1.88	33.18	1.88	33.08	1.87
−2 × log likelihood	5,413.24		4,377.99		4,377.94		4,375.77	

*$p < .10$. **$p < .05$.

Table 6. Multilevel Regression Analyses Predicting the Posttest Results.

	Empty model		Context model		Intervention model		Instruction group model	
Fixed part	Coefficients	*SE*	Coefficients	*SE*	Coefficients	*SE*	Coefficients	*SE*
Intercept	88.55**	0.89	50.37**	2.17	50.11**	2.25	50.73**	2.26
Participation in intervention					0.65	1.45	−0.33	1.54
Pretest			0.49**	0.03	0.49**	0.03	0.49**	0.03
Extended instruction			−5.48**	0.69	−5.48**	0.69	−6.61**	0.85
Shortened instruction			4.94**	0.68	4.94**	0.68	4.64**	0.84
Participation × Extended							2.74**	1.19
Participation × Shortened							0.89	1.19
Multigrade			2.85*	1.60	2.88**	1.53	3.01**	1.53
Random								
Level: Class	27.22	8.00	21.09	5.09	21.00	5.07	20.81	5.03
Level: Students	132.69	7.52	35.71	2.02	35.71	2.02	35.43	2.01
−2 × log likelihood	5,262.20		4,419.48		4,419.28		4,414.01	

*$p < .10$. **$p < .05$.

Table 6 shows that the main intervention effect is positive in the "intervention model"; however, this effect is not statistically significant. The anticipated difference between the experimental and the control group was not found; therefore, the first hypothesis was rejected.

In the "instruction group model," clear differences in results are visible between Tables 5 and 6. In Table 6, the interaction of participating in the intervention as a student and being in the extended instruction group shows a statistically significant positive effect on the posttest score. This implies that the posttest scores in the extended instruction group of the experimental group gain an additional 2.74 ability score points, and that their posttest scores are on average 2.41 ability score points higher than those of the control group with extended instruction. Model fit improved significantly (from 4,419.28 to 4,414.01) with the inclusion of the interaction effect ($\chi^2 = 4.72$, $df = 1$, $p < .05$).[3] The effect size of the interaction effect is $d = .19$ (calculated by adding the intervention effect and the interaction effect Participation × Extended Instruction group, divided by the pooled standard deviation of the posttest score). Therefore, the second hypothesis was confirmed.

The third hypothesis that participation in the intervention as a student while being in the shortened instruction group leads to improved student achievement could not be confirmed based on our analysis.

Instruction groups are formed subjectively, as the teacher decides in which instruction group a student belongs to. To check whether the intervention effect also held for objectively lower performing and objectively higher performing students, the multilevel analysis was repeated by replacing the extended instruction group variable with the weakest 28% of students in each classroom (the percentage of students in the extended instruction group in this study). Furthermore, the shortened instruction group variable was replaced with the 28% best performing students from each classroom. The results showed similar outcomes, but the interaction effects in this case were not statistically significant. In other words, the intervention was not effective for the objectively lowest performing students, but for those students who, from the teacher's perspective, were most in need of extra instruction. The influence of the teacher proves to be very important here.

Discussion and Conclusion

The main aim of this study was to provide evidence, by means of a randomized controlled trial, that DBDM can enhance students' mathematical achievement. Although no intervention effect on mathematical achievement was found across all students, the students in the extended instruction group benefited considerably from the DBDM-intervention (effect size of .19 for 74 students). These students had been placed in the extended instruction group (based on their mathematical achievement and their teacher's assessment of the student during daily classroom practice), and received more instruction adapted to their needs than the students in the shortened or the basic instruction groups.

At the posttest, extended instruction group students in the experimental group scored on average 2.41 ability score points higher than students in the extended instruction group of the control group. In an entire school year (10 months), Grade 4 students normally gain about 12 ability score points in mathematics. Thus, the intervention students in the extended instruction group of the experimental group grew approximately two additional months in mathematics relative to their control counterpart.

It is not possible to describe exactly which part of the intervention caused this effect: the extended *instruction time* or the deeper understanding and execution of the instructional strategies required for these students. However, effective differentiation requires not only additional instruction time, it also requires that the content of instruction matches students' instructional needs (cf. Deunk et al., 2015; Tomlinson et al., 2003).

The goal of the intervention was that teachers provide instruction that would suit the needs of all students better, but the intended student achievement effects were only found for the students in the extended instruction group. This could be due to the features of the PDP.

During the PDP, extra support for students in the extended instruction group was emphasized more in both the *design* of the instructional plan that teachers had to make and in its *execution* in the classroom (more than providing extra support to students in the basic, and in the shortened instruction group). In their instructional plan, teachers *planned* extra support for the extended instruction group as part of their daily mathematics lessons, while extra support for a few of the basic and shortened instruction group students was planned less frequently by means of special learning groups. During the coaching sessions, in which the *execution* of the instructional plan was promoted, the trainer could only provide feedback on how the teacher gave extra support for the extended instruction group as the special learning groups were generally not scheduled for these lessons. It may be that teachers did not provide as much support to the students in the special learning groups as intended and planned. As both the *amount* of extra support and the *execution* of this support for students in the basic and the shortened instruction group were emphasized less during the PDP, this might explain why these students did not benefit as much from the intervention. This is also supported by the results of the multilevel analyses, which assessed whether the effect for students in the extended instruction group also held using a more objective measure (the weakest 28% of the students). This did not appear to be the case, which implies that the intervention was indeed helpful for the students allocated to the extended instruction group and, as a result, treated as one group of students in need of extra support.

For the interpretation of the effect size of .19 (the intervention effect for students in the extended instruction group), the context of the study is important (Lipsey et al., 2012), because the effect size standards as developed by Cohen (1988) are based on interventions aimed at *individual* participants. In this study, the PDP was aimed at influencing the behavior of teachers in their classrooms and, thereby, student achievement (thus, not directly at students). Lipsey et al. (2012) found that, on average, studies using a randomized controlled trial aimed at the *classroom level* reported effect sizes of .18 on student achievement.

Moreover, student achievement in this research was not measured by means of a researcher-developed test but a standardized test. In the case of self-developed tests, the content of the test is often closely aligned with the specific subject-matter content covered in the intervention. Standardized tests have been developed independently from a specific intervention and cover subject-matter content more broadly. Consequently, it is harder to find an effect using a standardized test. Lipsey et al. (2012) found an average effect size of .08 for randomized studies using a standardized test in elementary education.

Thus, the effect size found for students in the extended instruction group is of a comparable magnitude to those from other classroom-level studies, but larger than the effect sizes in studies using standardized tests.

Although it is hard to compare the results of this study to the effects of other DBDM effectiveness studies due to differences in subject matter, research design, and intervention target group, the results of such studies may give an idea of what effects are possible.

Other DBDM studies evaluating the effects of interventions on students' mathematical achievement showed a small effect (Carlson et al., 2011) or no effects (Slavin et al., 2013). However, these interventions were aimed at districts and school leaders rather than directly at teachers. The intervention used in the study of Van Kuijk (2014) was directly aimed at teachers but was meant to improve students' *reading comprehension*. A positive effect ($d = .37$) on student achievement was found. Slavin et al. (2013), after four intervention years aimed at reading, found a significant positive effect (reported effect size of .50). Seemingly, reading interventions show stronger effects than interventions aimed at improving students' mathematics' achievement.

Desimone (2009) stressed that improved student achievement can only be expected when it is preceded by changes in teachers' behavior within the classroom. Studies by Lockwood, Sloan McCombs, and Marsh (2010) and Biancarosa et al. (2010) show that it may take longer than one school year to find coaching effects for all students. Support for the latter is found in the study of Slavin et al. (2013) and in the results of this study: No effects were found in the first half-year (from pretest to midtest), while the effect for students in the extended instruction group appeared after the entire intervention year. As already mentioned, the trainer indicated that the discussions during the coaching sessions at the beginning of the school year were mainly focused on organizational aspects, but later on these discussions became more about differentiation of the content of mathematics lessons. As the emphasis was more on students from the extended instruction group, this explains why an effect was found for these students. Due to the complexity of DBDM and the skills it requires, teachers may need more time to fully learn to implement the newly acquired skills and, ultimately, improve the achievement of all students.

Future Research

Due to the costs of intensive PDPs such as the one reported in this study (approximately US$2,500 per teacher for a 1-year intervention), it may be a challenge to support and coach all teachers in school in this way. A practically more feasible approach may be to train one or a few teachers from a school intensively, who can then coach their colleagues. This will be cheaper and teachers will have more time to learn to implement DBDM gradually. This way, intervention effects can be studied more protractedly and longitudinally. A similar approach was followed in Florida, where internal coaches worked at a school to support teachers in working in a data-based way (Lockwood et al., 2010; Marsh et al., 2010). Success for All Schools also work with school internal "program facilitators" who support teaches in their schools to work in the intended way (Slavin, Madden, Chambers, & Haxby, 2009).

Another way to promote the implementation of DBDM in schools may be to make DBDM part of the curriculum of teacher training institutions. Currently, DBDM is rarely included in preservice training (Mandinach & Gummer, 2013).

Overall, this study indicates that DBDM can positively influence student achievement, while highlighting that this is not easy to accomplish and that the average teacher cannot be expected to implement DBDM on his or her own. On the contrary, teachers have to adapt their activities considerably to work in a DBDM-way. Changing old habits is difficult, especially when it requires the development of new complex professional skills. There is still much to learn about how data-informed, deliberate teaching can be promoted further. The results of this study can serve as a starting point for deliberately designed and rigorously evaluated DBDM-interventions.[4]

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Dutch Ministry of Education, Culture and Science, the Ministry of Economic Affairs and the Ministry of Finance with 450.000 euros.

Notes

1. The socioeconomic status (SES) of students is based on the parent's level of education. A "weight" is given to students (0, 0.3, or 1.2: The latter two weights indicate that the parents' level of education is low). Schools receive extra funding for each "weight"-student. Schools with an average student weight of 0.15 were contacted for participation in the study.
2. The test results are not reliable for students who perform way above or below the national average; schools can therefore decide to administer a test from another grade. Hence, a small number of students have multiple ability scores for a specific time period.
3. To calculate this effect, the model was re-estimated without the non-significant interaction effect of Participation × Shortened Instruction. The ModelFit −2 × log likelihood changed to 4,414.56, showing that the increased model fit is almost completely due to the interaction effect Participation × Extended Instruction.
4. The authors would like to express their gratitude to Professor Dr. R. J. Bosker and Dr. H. Luyten for their statistical advice.

Reference

Benner, G. J., Kutash, K., Nelson, J. R., & Fisher, M. B. (2013). Closing the achievement gap of youth with emotional and behavioral disorders through multi-tiered systems of support. *Education & Treatment of Children*, *36*(3), 15-29.

Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *The Elementary School Journal*, *111*(1), 7-34. doi:10.1086/653468

Braddock, J. H., & Slavin, R. E. (1995). Why ability grouping must end: Achieving excellence and equity in American education. In H. Pool & J. A. Page (Eds.), *Beyond tracking, finding success in inclusive schools*. Bloomington, IN: Phi Delta Kappa Education Foundation.

Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability*, *21*, 47-65.

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, *33*(3), 378-398. doi:10.3102/0162373711412765

Coburn, C. E., & Turner, E. O. (2012). The practice of data use: An introduction. American Journal of Education, *118*, 99-111.

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

Cornett, J., & Knight, J. (2009). Research on coaching. In J. Knight (Ed.), *Coaching approaches and perspectives* (pp. 192-216). Thousand Oakes, CA: Corwin Press.

Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high-performing school systems use data to improve instruction for elementary students*. Los Angeles: Center on Educational Governance University of Southern California.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181-199. doi:10.3102/0013189X08331140

Deunk, M., Doolaard, S., Smale-Jacobse, A., & Bosker, R. J. (2015). *Differentiation within and across classrooms: A systematic review of studies into the cognitive effects of differentiation practices*. Groningen, The Netherlands: GION onderwijs/onderzoek.

Dunn, K. E., Airola, D. T., Lo, W.-J., & Garrison, M. (2013). Becoming data driven: The influence of teachers' sense of efficacy on concerns related to data-driven decision making. *The Journal of Experimental Education*, *81*(2), 222-241. doi:10.1080/00220973.2012.699899

Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915-945. doi:10.3102/00028312038004915

Goertz, M. E., Oláh, L. N., & Riggan, M. (2009). *From testing to teaching: The use of interim assessments in classroom instruction*. Philadelphia, PA: Consortium for Policy Research in Education.

Gummer, E. S., & Mandinach, E. B. (2015). Building a Conceptual Framework for Data Literacy. *Teachers College Record*, *117*(4), 1-22.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, England: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112. doi:10.3102/003465430298487

Houtveen, T., & Van de Grift, W. J. C. M. (2012). Improving reading achievements of struggling learners. *School Effectiveness and School Improvement*, *23*(1), 71-93. doi:10.1080/09243453.2011.600534

Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the "data-driven" mantra: Different conceptions of data-driven decision making. In P. A. Moss (Ed.), *Evidence and decision making* (Vol. 106, pp. 105-131). Malden, MA: Blackwell.

Ingram, D., Louis, S. K., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, *106*(6), 1258-1287.

Inspectorate of Education. (2010a). *De staat van het onderwijs. Onderwijsverslag 2008/2009* [The state of education. *Education report 2008/2009]*. Utrecht, The Netherlands: Inspectorate of Education.

Inspectorate of Education. (2010b). *Opbrengstgericht werken in het basisonderwijs* [Data-driven decision making in education]. Utrecht, The Netherlands: Inspectorate of Education.

Inspectorate of Education. (2014). *De staat van het onderwijs. Onderwijsverslag 2012/2013* [The state of education. *Education report 2012/2013]*. Utrecht, The Netherlands: Inspectorate of Education.

Keuning, T., Van Geel, M., Visscher, A. J., Fox, J.-P., & Moolenaar, N. M. (2016). The transformation of schools' social networks during a data-based decision making reform. *Teachers College Record*, *118*(9), 1-33.

Knight, J. (2009). Instructional coaching. In J. Knight (Ed.), *Coaching approaches and perspectives* (pp. 29-55). Thousand Oakes, CA: Corwin Press.

Lai, M. K., & Schildkamp, K. (2013). Data-based decision making: An overview. In M. K. Lai & K. Schildkamp (Eds.), *Data-based decision making in education* (pp. 9-21). Dordrecht, The Netherlands: Springer.

Ledoux, G., Blok, H., & Boogaard, M. (2009). *Opbrengstgericht werken: Over de waarde van meetgestuurd onderwijs* [Data-driven decision making: The value of data-driven education]. Amsterdam, The Netherlands: SCO-Kohnstamm Instituut.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., . . . Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013-3000). Washington, DC: The Institute of Education Sciences.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist*, *57*(9), 705-717.

Lockwood, J. R., Sloan McCombs, J., & Marsh, J. A. (2010). Linking reading coaches and student achievement: Evidence from Florida middle schools. *Educational Evaluation and Policy Analysis*, *32*(3), 372-388. doi:10.3102/0162373710373388

Lou, Y., Abrami, P., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*(4), 423-458.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, *42*(1), 30-37. doi:10.3102/0013189X12459803

Mandinach, E. B., & Gummer, E. S. (2016a). *Data literacy for educators: Making it count in teacher preparation and practice*. New York, NY: Teachers College Press.

Mandinach, E. B., & Gummer, E. S. (2016b). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366-376. doi:10.1016/j.tate.2016.07.011

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, *114*, 30-48.

Marsh, J. A., Sloan McCombs, J., & Martorell, F. (2010). How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools. *Educational Policy*, *24*(6), 872-907. doi:10.1177/0895904809341467

McKenna, M. C., & Walpole, S. (2008). *The literacy coaching challenge: Models and methods for grades K-8*. New York, NY: Guilford Press.

Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011). *Teachers' ability to use data to inform instruction: Challenges and supports*. Washington, DC: U.S. Department of Education Office of Planning, Evaluation and Policy Development.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, *48*, 85-112. doi:10.1016/j.jsp.2009.09.002

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223-231. doi:10.1002/j.2168-9830.2004.tb00809.x

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes and promoting and hindering factors. *Teaching and Teacher Education*, *26*(3), 482-496.

Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2013). Effects of a data-driven district reform model on state assessment outcomes. *American Educational Research Journal*, *50*(2), 371-396. doi:10.3102/0002831212466909

Slavin, R. E., Madden, N. A., Chambers, B., & Haxby, B. (2009). *2 million children: Success for all*. Thousand Oaks, CA: Corwin Press.

Staman, L., Visscher, A. J., & Luyten, H. (2014). The effects of professional development on the attitudes, knowledge and skills for data-driven decision making. *Studies in Educational Evaluation*, *42*, 79-90. doi:10.1016/j.stueduc.2013.11.002

ToetsGids. (2014). *Rekenen-Wiskunde* [Arithmetic-Mathematics]. Retrieved from http://www.toetswijzer.nl/main.asp?Sector=TG

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Conover, L. A., & Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*(2/3), 119-145. doi:10.1177/016235320302700203

Van de Grift, W. J. C. M., Van der Wal, M., & Torenbeek, M. (2011). Ontwikkeling in de pedagogisch didactische vaardigheid van leraren in het basisonderwijs [Development of the pedagogical didactic abilities of teachers in primary education]. *Pedagogische Studiën*, *88*, 416-432.

van der Scheer, E. A. (2016). *Data-based decision making put to the test* (Doctoral dissertation). Enschede, The Netherlands: University of Twente.

Van der Scheer, E. A., Glas, C. A. W., & Visscher, A. J. (2017). Changes in teachers' instructional skills during an intensive data-based decision making intervention. *Teaching and Teacher Education*, *65*, 171-182. doi: 10.1016/j.tate.2017.02.018

van der Scheer, E. A., & Visscher, A. J. (2016). Effects of an intensive data-based decision making intervention on teacher efficacy. *Teaching and Teacher Education*, *60*, 34-43. doi:10.1016/j.tate.2016.07.025

Van der Steeg, M., & Gerritsen, S. (2013). *Evaluatie pilot investeren in kwaliteit leraren*.[Evaluation of pilot investing in teaching quality]. Den Haag, The Netherlands; Centraal Planbureau. Retrieved from: https://www.cpb.nl/sites/default/files/publicaties/download/cpb-achtergronddocument-evaluatie-pilot-investeren-kwaliteit-leraren.pdf

Van Geel, M., Keuning, T., Fox, J.-P., & Visscher, A. J. (2016). Assessing the effects of a school wide data-based decision making intervention on student achievement growth in primary schools. *American Educational Research Journal*, *53*(2), 360-394. doi:10.3102/0002831216637346

Van Kuijk, M. (2014). *Raising the bar for reading comprehension* (Doctoral dissertation). Groningen, The Netherlands: Groningen Institute for Educational Research.

Van Veen, K., Zwart, R., & Meirink, J. (2012). What makes teacher professional development effective? In M. Kooy & K. Van Veen (Eds.), *Teacher learning that matters: International perspectives* (pp. 3-21). London, England: Routledge.

Visscher, A. J., & Ehren, M. C. M. (2011). *De eenvoud en complexiteit van opbrengstgericht werken; Analyse in opdracht van de kenniskamer van het ministerie van onderwijs, cultuur en wetenschap* [The simplicity and complexity of data-driven teaching]. Retrieved from http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2011/07/13/de-eenvoud-en-complexiteit-van-opbrengstgericht-werken.html

Author Biographies

Emmelien A. van der Scheer recently finished her PhD at the University of Twente and currently works at the University of Groningen as researcher and lecturer. Her dissertation focused on the effects of a data-based decision making (DBDM)-intervention on teachers' self-efficacy, their instructional behavior, and on student achievement. Her research interests are in designing interventions to support teacher improvement, and teacher and school effectiveness research.

Adrie J. Visscher is a professor at the University of Twente and the University of Groningen. Over the last 8 years, he, in cooperation with colleagues, has developed and evaluated interventions at school and classroom level, intended to provide support to schools for the utilization of performance feedback from student monitoring systems.