University of Groningen

Changes in teachers' instructional skills during an intensive data-based decision making intervention

van der Scheer, Emmeline A.; Glas, Cees A. W.; Visscher, A.J.

Published in:
Teaching and Teacher Education

DOI:
[10.1016/j.tate.2017.02.018](https://doi.org/10.1016/j.tate.2017.02.018)

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

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

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

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. <https://doi.org/10.1016/j.tate.2017.02.018>

Copyright

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

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

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

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

Changes in teachers' instructional skills during an intensive data-based decision making intervention

Emmelen A. van der Scheer^{a, b, *}, Cees A.W. Glas^a, Adrie J. Visscher^a

^a University of Twente, Postbus 217, 7500 AE Enschede, The Netherlands

^b University of Groningen, GION education/research, Grote Rozenstraat 3, 9712 TG Groningen, The Netherlands

HIGHLIGHTS

- Changes in teachers' instructional skills after a data use intervention.
- A comparison of skills shown in three lessons prior to and after intervention.
- Teachers showed improved data use related skills after the intervention.
- Teachers with varying initial teaching skills showed similar development patterns.

ARTICLE INFO

Article history:

Received 25 May 2016

Received in revised form

16 February 2017

Accepted 23 February 2017

Available online 7 April 2017

Keywords:

Data-based decision making

Teacher professional development

Item response theory

Evaluation

ABSTRACT

This study evaluates changes in teachers' instructional skills after participating in an intensive data-based decision making (DBDM) intervention for grade 4 teachers. Teachers were recorded three times prior to the intervention, and three times after the intervention, and all recordings were rated by four raters. The data was analyzed by means of advanced item response theory (IRT) techniques, combined with a generalizability model. Teachers significantly improved their DBDM related skills. Teachers' initial basic teaching skills did not seem to matter for the extent to which teachers developed their DBDM related instructional skills. Suggestions for future research are presented.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Improving student performance in schools using achievement data is strongly emphasized in educational policy in several countries (Campbell & Levin, 2009). Based on assessment data, instruction may be tailored to the abilities of all students to maximize the achievement of students (Coburn & Turner, 2012). Although this seems straightforward, current research fails to provide convincing evidence that professional development programs (PDPs), designed to promote data-based decision making (DBDM) in schools, result in the intended improvement of student performance (Carlson, Borman, & Robinson, 2011; Marsh, 2012; Slavin, Cheung, Holmes, Madden, & Chamberlain, 2013). In current PDPs,

the support for teachers regarding the full application of DBDM in practice is limited (Lockwood, Sloan McCombs, & Marsh, 2010). Up to this point PDPs mainly focus on the development of data-analysis skills (Marsh, 2012), while working in a data-based way requires much more from teachers (Kaufman, Graham, Picciano, Popham, & Wiley, 2014). Teachers are supposed to adapt instruction in line with the needs of students (determined on the basis of the analysis of student data), however, teachers rarely make these instructional changes (Datnow & Hubbard, 2015).

Teachers may not be able to implement DBDM in the classroom because they lack the knowledge and skills required for differentiating instruction (Baumann, Hoffman, Duffy-Hester, & Moon Ro, 2000; Holloway, 2000; McGarvey, Marriott, Morgan, & Abbott, 1997; Roy, Guay, & Valois, 2013; Tobin & McInnes, 2008). One way to support teachers in DBDM is to coach them whilst practicing it in the classroom. Coaching is founded on the learning theory that describes that performance can be improved if the instructional material can be practiced, discussed and reflected upon, and if

* Corresponding author. University of Twente, Postbus 217, 7500 AE Enschede, The Netherlands.

E-mail addresses: e.a.vanderscheer@utwente.nl (E.A. van der Scheer), c.a.w.glas@utwente.nl (C.A.W. Glas), a.j.visscher@utwente.nl (A.J. Visscher).

performance feedback is provided by an expert (Lockwood et al., 2010). There is considerable evidence that coaching can improve instructional skills (e.g. Joyce and Showers (2002); Cornett and Knight (2009)). However, very few studies have investigated whether coaching can promote the implementation of DBDM (Lockwood et al., 2010; Marsh, Sloan McCombs, & Martorell, 2010). Marsh et al. (2010) found that support from a reading coach is positively related to the improved teaching of reading, while Lockwood et al. (2010) found a positive effect on student achievement for reading.

Due to the limited knowledge on how working in a data-based way affects teaching practices (Datnow & Hubbard, 2015; Poortman, Schildkamp, & Lai, 2016), in this study it was explored to what extent an intensive, coaching-based DBDM professional development program (PDP) affects instructional practices. It was investigated how the PDP influenced teachers' general teaching skills, but in particular the skills related to DBDM, such as instructional differentiation. To promote the implementation of DBDM in the classroom, this PDP included intensive teacher support consisting of a combination of general meetings for groups of teachers, and coaching sessions during which teachers were coached individually with respect to the implementation of DBDM during their lessons. The following research question is answered here:

To what extent do teachers' instructional skills change after participating in an intensive, coaching-based DBDM professional development program?

2. Theoretical framework

We first will focus on what knowledge and skills are required for teachers to work in a data-based way. Next we will give a description of how we promoted the use of DBDM practices by means of a coaching-based DBDM intervention.

2.1. Knowledge and skills for DBDM

The process of data use is often described as a cyclical and iterative process, including aspects as data analysis and evaluation, transforming data into usable knowledge, and intervening in the classroom to improve student achievement (Keuning, Van Geel, Visscher, Fox, & Moolenaar, 2016; Ikemoto & Marsh, 2007; Mandinach & Gummer, 2016a, 2016b; Poortman et al., 2016). Keuning et al. (2016) described four core components of DBDM in primary education in the Netherlands on the basis of Visscher and Ehren (2011). These components include four teacher activities: 1. Evaluate and analyze student results, 2. set learning and performance goals for each student, 3. determine an instructional strategy to accomplish these goals, and, 4. execute the planned instructional strategies in the classroom. The components are presented in Fig. 1.

As working in a data-based way is rooted in several aspects of teaching, such as planning and delivering instruction, teachers need to master a coherent set of knowledge and skills (Mandinach & Gummer, 2016a). Next to the knowledge and skills required for the separate components in Fig. 1 (as outlined in detail below), Mandinach and Gummer (2016a) argue that foundational knowledge as described by Shulman (1987) is essential during the data use process, to be able to place the data in a meaningful context; curriculum knowledge, (pedagogical) content knowledge, knowledge regarding learners, the educational context and educational ends, purposes and values (Mandinach & Gummer, 2016a, 2016b). As part of the first component, evaluating and analyzing results, teachers have to identify possible data sources (e.g., standardized assessments and the results from student's daily work) and

understand the different purposes of the data (Mandinach & Gummer, 2016a). Furthermore, teachers need to be able to locate the relevant data in a data system (Means, Chen, DeBarger, & Padilla, 2011). Although Dutch primary schools have access to sophisticated student monitoring systems (SMS) that have built-in possibilities to analyze the results of standardized assessments, teachers are generally unaware of the possibilities of their system (Staman, Visscher, & Luyten, 2014). However, research shows that by means of training, teachers learn what these possibilities are, and how to interpret the outcomes of the data analysis activities (Staman et al., 2014). As Mandinach and Gummer (2016a, 2016b) have outlined, many more skills are important for analyzing data (such as manipulating and integrating data). However, our intervention was focused on equipping teachers with the ability to analyze the results of standardized assessments by means of a SMS, and to combine the results with other data sources, because using the extensive built-in possibilities of their SMS was considered to be the most relevant thing to learn for Dutch primary school teachers.

By using their SMS for data analysis, teachers obtain detailed insight into how their students perform: e.g. in relation to their prior assessment, to a relevant norm group, and whether students master specific mathematical subjects or not. However, data interpretation is essential to give meaning to data within a relevant context (Mandinach & Gummer, 2016a; Means et al., 2011). Combining information from several sources of data, like standardized assessments and results from students' daily work is important in this respect (Schildkamp & Kuiper, 2010).

The second component involves that teachers formulate goals for individual students as well as for the entire class. Locke and Latham (2002) showed that setting SMART (Specific, Measurable, Attainable, Realistic and Time-bound) and *challenging* goals can improve performance considerably (component 2). By formulating and trying to accomplish performance goals, teachers also can improve their own performance (Visscher & Ehren, 2011).

After performance goals have been set, the next step is to determine which instructional strategies are suitable to accomplish these goals (component 3). This requires that teachers *understand the context*, by combining several types of knowledge, such as linking knowledge about learning objectives with teacher knowledge about what a student is capable of (Mandinach & Gummer, 2016a). Furthermore, a teacher has to *determine next instructional steps* (Mandinach & Gummer, 2016a; Means et al., 2011). Teachers therefore should be aware of the different paths of learning among their students, and the need for instructional variation, and also outline these strategies in an instructional plan (Vogt & Rogalla, 2009). Planning extended instruction for weak and strong performing students, in addition to whole classroom instruction, is considered important for effective differentiation in the classroom (Tobin & McInnes, 2008; Tomlinson et al., 2003; Van de Grift, 2007). Other strategies for planning differentiated instruction are working with individualized learning plans, and *flexible* student groups (Means et al., 2011). The last DBDM component involves executing the planned instruction strategies in the classroom. About the extent to which DBDM affects daily classroom practice little is yet known (Mandinach & Gummer, 2016a). However, DBDM supports teachers in making instructional decisions, especially in enhancing that instruction is in line with students' needs (Means et al., 2011). This could, for example, entail that a teacher provides additional support to an individual student. However, such an instructional change is hard to observe in daily lessons as this kind of support is generally not offered on a daily basis. A general teaching skill required to work in a data-based way and which can be observed in lessons, is instructional differentiation.

Roy et al. (2013) define differentiation as "an approach by which

Fig. 1. Four components and levels of DBDM (Keuning et al., 2016).

teaching is varied and adapted to match students' abilities using systematic procedures for academic progress monitoring and data-based decision-making" (p. 1187). Differentiation is complex as it relates to various aspects of instruction (content, process and product) (Williams et al., 2014). As Dutch teachers are required to teach the same topics to all students, *content differentiation* is of less relevance in Dutch primary schools. By contrast, as Dutch classes include mixed ability students, *process differentiation* (adjusting instruction to the varying needs of student groups), and *product differentiation* (adjusting the level of assignments to students' ability levels) are essential during a regular Dutch primary school lesson. Working with small groups of students that vary in ability levels, is also considered an important element of effective differentiation (Tomlinson et al., 2003), and also common practice in Dutch primary schools. This involves the adaptation of instruction in line with students' needs, by providing extended instruction to weak performing students in addition to whole-classroom instruction, and additional instruction aimed at challenging strongly performing students. However, although most Dutch teachers plan to work with small instruction groups in their *lesson plans*, most teachers do not do this in the classroom (Inspectorate of Education, 2014).

Because the use of ability groups (or instruction groups) is considered to be a vital part of DBDM in the Netherlands, it also was an important component of this study. Moreover, to work according to data-based principles, a systematic, goal-oriented approach is required. Within a regular mathematics lesson, this could be promoted by clarifying the *lesson objective* at the start of the lesson (Kyriakides, 2005; Van de Grift, 2014), and by evaluating the degree to which the lesson objective has been accomplished at the end of the lesson.

To conclude, DBDM requires a variety of skills from teachers both prior to lessons (to analyze data, to set goals, and to plan instruction carefully), and during lessons (for example, working with instruction groups, and the presentation and evaluation of lesson objectives). How teachers can be supported to work in a data-based way will be explained in the following.

2.2. Supporting DBDM by means of coaching

To support teachers in the implementation of DBDM, coaching was considered especially important. *Data coaches* are skillful in using data and provide support to data teams (a group of teachers working together to use data) within their own school (Mandinach & Gummer, 2016a). An *instructional coach* can work with a teacher to help him implement research-based instructional practices (Knight, 2009). The coach provides context-embedded feedback and promotes reflection. In several studies coaching proved to be an effective lever for the improvement of instructional skills (Carlisle & Berebitsky, 2011; Cornett & Knight, 2009; Joyce & Showers, 2002; Marsh et al., 2010; McKenna, Walpole, Uribe-

Zarain, & Lamitina, 2010; Sailors & Price, 2015; Teemant, 2014; Teemant, Leland, & Berghoff, 2014; Teemant, Wink, & Tyra, 2011).

Coaching interventions differ in the *position* of the coach (e.g. a fulltime employed coach in a school, or an external coach as part of an intervention), in *what* teachers are coached on (e.g. general pedagogical skills, or subject-related instructional strategies), and in *who* the coach is (a peer teacher, or an expert teacher) (Cornett & Knight, 2009; McKenna & Walpole, 2008). In this study, teachers received coaching that was related to both data use and instructional practices. During central meetings, the trainer supported teachers in interpreting their data and with making instructional plans. In addition to this he also supported teachers in the classroom by coaching them to work in line with the instructional plan that had been based on the data analysis (e.g., coaching them on the use of instruction groups, and the discussion/evaluation of the lesson objective).

For an instructional coach it is important to provide feedback on classroom practices to teachers that is clear, specific and incorporates challenging goals (Coe, Aloisi, Higgins, & Major, 2014). Although the coach in this study aimed for the implementation of differentiation in the classroom, this was deemed an unrealistic goal for every teacher as, for example, quality classroom management is an essential prerequisite for differentiation in the classroom (Kyriakides, Creemers, & Antoniou, 2009; Vogt & Rogalla, 2009), while not all teachers master classroom management skills well. Van de Grift, Van der Wal, and Torenbeek (2011) point to an empirical hierarchy of the instructional activities that teachers employ in their classrooms, meaning that the lowest level of the instructional skill hierarchy (e.g. providing a safe learning climate) is observed most in classrooms whereas the higher levels of the hierarchy (e.g. differentiation) are observed to a lesser degree.

The instructional skills that are observed more than differentiation skills in the hierarchy of Van de Grift et al. (2011), relate strongly to the three dimensions of effective teaching in terms of enhancing high levels of student achievement and motivation as identified in several meta-analyses (Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Klieme, Pauli, & Reusser, 2009; Kyriakides, Christoforou, & Charalambous, 2013; Kyriakides et al., 2009; Pianta & Hamre, 2009; Seidel & Shavelson, 2007; Van de Grift, 2007). Although the results of these studies vary to some extent, it is considered essential that a teacher provides a *supportive climate*, with a positive student-teacher relationship that incorporates constructive feedback from teacher to student. Furthermore, *good classroom management* is considered important and entails an orderly and task-oriented classroom, which is a precondition for making students work on their tasks. Lastly, teachers have to ensure that students are cognitively activated, e.g. by ensuring that each student works on challenging tasks, and that prior knowledge is activated. This also involves clear *instruction*, e.g. direct instruction, instruction that matches with what students already know, and instruction that is broken down into small steps

(Van de Grift, 2007). In addition to these three dimensions, teachers also can encourage students to develop learning strategies that support them in performing high-level learning activities (e.g. the teacher stimulates students to plan and check their work when assignments are made) (Van de Grift, 2007, 2014; Van de Grift et al., 2011).

It was assumed here that the basic teaching skills (providing students with a supportive classroom climate; classroom management) are required for mastering higher order teacher skills such as differentiation. Therefore, coaching as part of a DBDM intervention should not focus solely on the implementation of differentiation in the classroom. Teachers who, for example, struggle with the lower level skills, such as classroom organization, may not be able to develop differentiation skills yet. The coach in this study therefore adapted his coaching content and activities to the varying needs of teachers, however always with the aim of implementing DBDM (especially the presentation and evaluation of lesson objectives, and the use of small instruction groups in the classroom) in mind.

In sum, because little is known about to what extent a DBDM impacts teachers' instructional skills, we explored how the PDP affected teachers' instructional skills. Furthermore, we expected a positive change during the PDP on teachers' DBDM skills, but not for teachers who did not yet fully master the basic instructional skills.

3. Method

3.1. Participants

We contacted regular Dutch primary schools with a high percentage of low-SES students by email, to invite them to participate in the project. We decided to approach low-SES schools as the Inspectorate of Education had shown that they underperformed more frequently (Inspectorate of Education, 2010), and hence required more support. Contacted school leaders and teachers were informed about the study design, about what was expected from them, and about the DBDM-PDP content.

The grade 4 teachers of 30 primary schools agreed to participate, resulting in a total of 39 teachers who took part in the PDP during the school year 2013–2014. In this study, 34 teachers (from 26 primary schools) participated as two teachers dropped out due to motivational issues, one teacher refused to be recorded, another could not be recorded for organizational reasons, and one teacher was excluded as he participated in another intensive classroom PDP during the PDP year.

The remaining 34 teachers (71% women, 29% men) had an average teaching experience of 13 years. Eight teacher pairs (two part-time teachers who taught the same classroom) participated together in the PDP. Participating teachers taught a grade 4 class with 9–10 year old students, or a multi-grade classroom with grade 4 students and students from other grades.

3.2. Procedure

This study was based on a short interrupted time series research design as each teacher (also from a teacher pair) was video-recorded during a mathematics lesson three times before the PDP, and three times after the PDP (Shadish, Cook, & Campbell, 2002). Generalizability studies have shown that multiple lessons (below also referred to as time points) and raters are required to achieve a reliable estimate of a teacher's skills (Hill, Charalambous, & Kraft, 2012; Praetorius, Lenske, & Helmke, 2012). The first three recordings were made by the end of the school year 2012–2013, and at the start of the school year 2013–2014 prior to the start of

the coaching sessions. The last coaching session was used to make one of the after-PDP recordings, the other two lessons were recorded during the last month of the school year 2013–2014. It was impossible to record the lessons of three teachers in the last month of the PDP year, due to illness and time restrictions. These recordings were administered at the start of school year 2014–2015. Due to a variety of reasons (e.g. maternity leave) the lessons of five teachers could only be recorded either prior to, or after the PDP. Each of these teachers were part of a teacher pair. Thus, a total of 189 recordings were made, of which 90 were made prior to the PDP, and 99 after the PDP. In most Dutch primary schools, parents provide permission for the recording of any lesson that their child attends at the time of enrollment. In this study, the principal decided what kind of additional action was required to attain parental permission. Only a few parents did not provide permission and their children were not recorded.

Teachers were requested to perform a regular mathematics lesson (in which instruction was included) rather than to adapt their lesson towards what they thought we expected. The lessons were recorded by a rater, by means of the IRIS Connect system. This system consists of two iPods that simultaneously record the lesson: one iPod is focused on the teacher, the other on the students. The lessons were uploaded to a secured online environment, which offered both raters and teachers the possibility to review the lessons whenever they liked to do so.

3.2.1. Raters

The 189 recordings were rated by three raters who had been trained intensively for three days. As part of the training, six observations were rated independently, and rater variation was discussed afterwards, to accomplish consensus about the ratings. As raters could not be appointed as full-time raters, they were allowed to rate the recordings according to their own time schedule. A fourth experienced ICALT-instrument (International Comparative Analysis of Learning and Teaching, the lesson observation instrument used in this study) rater also rated all recorded lessons. The recordings prior to the start of the PDP, and the recordings after the PDP were scored subsequently by each rater. Each rater first received a list of randomly ordered recordings prior to the PDP, and as soon as the post-PDP recordings were available, a new list with randomly ordered recordings was provided. The sequence of watching and scoring the recordings was ordered randomly for each rater, to prevent order-based bias (Shadish et al., 2002).

3.2.2. The professional development program

Teachers, during a school year, followed a DBDM training course, which included seven meetings (attended in five groups of teachers), and four coaching sessions. During these sessions, they were coached in the classroom on how to implement DBDM. Just as in the Data Wise method (Parker Boudett, City, & Murnane, 2007), teachers were taught to go through several phases of DBDM. However in our PDP, in contrast to the Data Wise method, individual teachers (rather than a school team) were trained in DBDM. Moreover, the analysis of data (as mentioned in section 2.1) is less time consuming for Dutch primary school teachers. This resulted in a less comprehensive PDP than the one described by Parker Boudett et al. (2007), however still covering the core DBDM elements and with a stronger emphasis on the implementation of DBDM in the classroom. Moreover, the PDP was designed on the basis of five criteria considered important for effective professional development by Desimone (2009), and Van Veen, Zwart, and Meirink (2012), as is outlined in van der Scheer and Visscher (in press). The PDP was delivered by a trainer who had attained a university Master's degree, had been a primary school teacher, and was experienced in training teachers and schools for DBDM. During the

intervention year, the trainer and research team collaborated closely to discuss the implementation of the intervention. Furthermore, the trainer monitored whether teachers understood the intervention's content, by providing individual feedback on several key aspects of the intervention (e.g., the data-analysis and their instructional plan).

Each of the four DBDM aspects, as outlined in section 2.1, were addressed at least twice by the teachers. During the first, fifth and seventh meeting, participating teachers analyzed students' results on the standardized test, by means of a protocol using their student monitoring system (SMS) in use by their school. These grade level tests were developed to inform instruction and are administered twice a year at the end, and half way a school year. Staman et al. (2014) reported that teachers, who are systematically trained in the use of their SMS, significantly improved their ability to use the SMS. Therefore participating teachers were trained in a way similar to the study by Staman et al. (2014).

Furthermore, teachers had to design, on the basis of the results of their data analysis, an instructional plan twice (during the second and the fifth meeting) which had to include, among others, the instruction groups the teachers composed, the performance goals set for them, and the planned instructional strategies to accomplish the goals. The teacher allocated students to one of the three instruction groups on the basis of students' performance on the standardized test, and additional information about students obtained in the classroom: students who only required brief instruction, students who required basic instruction, and those students who required extended instruction in addition to basic instruction. During the second and third meeting, the trainer discussed goal setting, instructional strategies and several instructional models relevant for grade 4 teachers. This way, both the second and third aspect of DBDM was addressed. The fourth aspect, the implementation of DBDM in the classroom, was strongly emphasized during four individual classroom coaching sessions, and during the fourth and the sixth meeting. During the fourth and sixth meeting teachers watched a video of their own lesson (reduced to approximately 20 min of the original lesson), videos of peers, and they received and provided (to peers) feedback. During four individual coaching sessions each teacher received feedback from the trainer on how (s) he implemented DBDM in his/her classroom, with an emphasis on how the teachers implemented the instructional plan (as this plan was based on the data-analysis). 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 (including differentiating subject matter), independent work by students, and the evaluation of the lesson goal. Teachers first gave their own opinions on their strengths and weaknesses for each lesson phase. Subsequently, the trainer provided his own reflections whilst considering the teacher's zone of proximal development. For example, the PDP was focused on implementing the instructional plan, which also requires that a teacher masters the basic teaching skills such as classroom management. If the trainer identified classroom management problems, the feedback focused on how classroom management could be improved (instead of on differentiating instruction based on student progress data). Finally, the teacher and trainer discussed how the teacher could improve his/her DBDM skills. Each lesson observed by the trainer was recorded, and teachers could watch these lessons again online. However, this was not a mandatory part of the intervention.

An overview of the content of each meeting is presented in Table 1.

3.2.3. The lesson observation instrument

The International Comparative Analysis of Learning and

Teaching (ICALT) instrument was used to measure the instructional skills of teachers, because it includes various DBDM-elements of interest, and because it is a validated lesson observation instrument matching the Dutch context (Van de Grift, 2007). A comparison between ICALT's content and the widely used frameworks by Danielson (2013), and by Pianta and Hamre (2009) can be found in Maulana, Helms-Lorenz, and Van de Grift (2015). The ICALT includes 35 items, each item can be scored on a 4-point Likert scale; predominantly weak (1), 'more' weak than strong (2), more strong than weak (3), and predominantly strong (4).

Based on the constructs in the ICALT instrument, the items were divided over five scales: Basic teaching skills (11 items, including items related to a safe classroom climate, classroom management and student engagement), Clear and activating instruction (13 items, including items related to instruction and an activating lesson), Learning strategies (6 items) and DBDM (5 items). The following five ICALT-items were DBDM-items:

1. The teacher clarifies the lesson objectives at the start of the lesson
2. The teacher evaluates whether the lesson objectives have been achieved at the end of the lesson
3. The teacher adapts instruction to the relevant differences between students
4. The teacher offers weak performing students extra learning- and instruction time
5. The teacher adapts the classroom assignments and processing to the relevant differences between students

It proved to be impossible to observe the fifth DBDM item. In Dutch schools it is common practice to divide students into three instruction groups. Although teachers usually do not explicitly mention differentiation between students in student assignments, the textbook assignments are already organized in line with the three instruction groups (Inspectorate of Education, 2014): relatively easy assignments for students who need additional instruction, relatively hard assignments for students who need little instruction, and 'basic' assignments for those students who receive basic instruction. As students know at which level they have to make assignments, teachers do not explicitly mention this in every lesson. Therefore it was impossible to observe whether teachers adapted their assignments to relevant student differences, this item was therefore excluded from the analyses.

3.3. Analysis

In this section, a global overview of the analyses will be given. For technical details, please refer to Appendix A. The ratings of the observed lessons (referred to as time points) were modeled by means of a combination of an item response theory (IRT) model (Lord, 1980) and a generalizability theory (GT) model (Brennan & Johnson, 1995; Brennan, 1992). All 34 teachers were rated by 4 raters using the same 34 items at all 6 time points. The analyses focused on four subscales (Basic teaching skills, Clear and activating instruction, DBDM, and Learning strategies) and the Total scale. Each scale consisted of a number of Likert items with four response categories. An IRT model was used to transform the discrete item responses for each scale to continuous ratings, that is, to latent variables. Thus, continuous scale ratings were available for each combination of a teacher, scale, rater, and time point. Using an analysis of variance model (more specifically, a GT model), these scale ratings were decomposed into the variance attributable to teachers, time points, raters, and their two-way and three-way interactions. All these components were random effects. Although the combined IRT and GT model was estimated in a

Table 1
Content professional development program.

Month	Duration	Meeting	Content
September	8 h	Meeting 1	<ul style="list-style-type: none"> - The background and meaning of DBDM - Explanation CITO-assessments - Explanation use and interpretation SMS - Analysis and interpretation own data with a SMS by means of a protocol
September	8 h	Meeting 2	<ul style="list-style-type: none"> - Individualized feedback on the SMS-protocol - Drawing up an instructional plan - Formulating SMART goals
October	4 h	Meeting 3	<ul style="list-style-type: none"> - Individualized feedback on instructional plan - Explanation Dutch core goals - The mathematics learning progression - Instructional models - Drawing up personal improvement goals
Coaching session 1 November	4 h	Meeting 4	<ul style="list-style-type: none"> - Observation mastery lessons - Observation recorded lessons in teacher pairs - Peer teachers provide feedback on recorded lessons - Feedback on how to differentiate instruction in the classroom
Coaching session 2 February	4 h	Meeting 5	<ul style="list-style-type: none"> - As in meeting 1
Coaching session 3 April	4 h	Meeting 6	<ul style="list-style-type: none"> - As in meeting 4
Coaching session 4 June	4 h	Meeting 7	<ul style="list-style-type: none"> - As in meeting 1

Bayesian framework (technical details are provided in [Appendix A](#)), traditional terminology (standard errors, significance, confidence intervals) will be used to report the results here. As for only three teacher pairs both the pretest and posttest scores were available, this dependency could not be taken into account.

The first analyses were meant to calculate the reliability of the estimation of the teachers' ability. Next, the changes in teacher skills during the intervention were estimated. Besides the formal hypothesis testing using the combined IRT and GT model, exploratory analyses were performed to investigate how the instructional skills of distinguished groups of teachers developed during the intervention year. Therefore, the teacher-time interaction was calculated (prior to and after the intervention, to determine individual growth) and teachers were divided over groups based on the teacher-time interaction on the Basic skills scale, prior to the intervention. Teachers thus were divided into three groups. Nine teachers with the lowest *teacher-time* interaction were assigned to the below average performing group, the next eleven teachers to the average performing group, and nine teachers with the highest *teacher-time* interaction formed the above average performing group. The teacher-time interactions (per scale) were included to [Appendix B](#). However, the graphical representations of these teacher-time interactions were presented in section 4. For five teachers either some recordings prior to, or after the PDP were not available. These teachers were not included in this part of the analyses.

By grouping teachers, it became possible to explore in greater depth how different groups of teachers developed their instructional skills during the PDP (as opposed to describing *on average*, non-differentiated teacher development). The researchers were aware that grouping teachers this way represents *relative* below average, relative average, and relative above average performing teachers, as the entire group of teachers participating in this PDP serves as the reference category. It was decided to group teachers based on the teacher-time interaction on the Basic skills scale as estimated prior to the PDP, as it was assumed that the basic skills are a prerequisite to developing more advanced teaching skills. A distinction on the basis of these skills was expected to provide most

insight into whether various groups of teachers developed differently during the PDP, or not.

4. Results

Results are presented for each of the five scales: for the Basic skills, Clear and activating instruction, DBDM, Learning strategies as well as for the Total scale. First, the results of the estimation of the variance components is presented, next the interaction of time points and the kind of teaching as promoted by the PDP are presented for each of the scales. This is followed by a description for the three groups of teachers we distinguish between per scale to explore whether teachers developed differently with respect to the various teaching skills.

The results of the estimation of the variance components and the rater reliability are presented in [Table 2](#).

The 'Var' columns in [Table 2](#) present the estimates of the variance components, whereas the 'Se' columns provide the standard errors. The variance component for teachers was set to 1.00 to identify the rating scale. That is, the rating scale is a latent variable with an arbitrary mean and variance, and these two needed to be fixed to allow for meaningful statistical analyses. The columns labeled '%' provide the percentages of variance each component explains relative to the scale's total variance. From [Table 2](#) it can be concluded that for the Clear and activating instruction scale, the DBDM scale, the Learning strategies scale, and the Total scale, teachers explained most of the variance: between 35 and 38 percent of the variance. For these scales, the time points form the second largest variance component. The largest variance component for the Basic skills scale was formed by the time points (34%), teachers explain a bit less variance here (33%). This was an interesting finding, as [Praetorius, Pauli, Reusser, Rakoczy, and Klieme \(2014\)](#) found that classroom management was more stable across lessons than teacher skills related to cognitive activation. Differences between raters, fortunately, only account for a small percentage of variance (ranging from 6% to 8%), although rater variance was also present in some of the interactions. The relative agreement among teachers was high, meaning that raters predominantly

Table 2

Estimation of variance components, the percentage of explained variance, and the rater reliability for each of the scales.

	Basic skills			Clear and activating instruction			DBDM			Learning strategies			Total		
	Var	Se	%	Var	Se	%	Var	Se	%	Var	Se	%	Var	Se	%
Teachers	1.00	0.00	32.51	1.00	0.00	34.83	1.00	0.00	35.57	1.00	0.00	37.85	1.00	0.00	37.27
Raters	0.21*	0.04	6.86	0.22*	0.05	7.64	0.21*	0.04	7.37	0.25*	0.05	9.48	0.21*	0.05	7.99
Time	1.04	0.83	33.91	0.73	0.63	25.53	0.61	0.39	21.79	0.45	0.32	17.21	0.62	0.47	23.02
Teachers x Raters	0.18*	0.03	5.78	0.19*	0.03	6.52	0.16*	0.03	5.67	0.17*	0.03	6.46	0.17*	0.03	6.36
Teachers x Time	0.21*	0.03	6.85	0.19*	0.03	6.49	0.43*	0.08	15.17	0.20*	0.03	7.72	0.19*	0.03	7.11
Raters x Time	0.17*	0.03	5.52	0.18*	0.03	6.25	0.22*	0.08	7.97	0.19*	0.04	7.32	0.18*	0.03	6.56
Error	0.26*	0.04	8.56	0.37*	0.05	12.74	0.18*	0.03	6.47	0.37*	0.06	13.95	0.31*	0.04	11.86
	ρ	Se		ρ	Se		ρ	Se		ρ	Se		ρ	Se	
Reliability	0.92	0.01		0.91	0.01		0.89	0.01		0.92	0.01		0.92	0.01	

* $p < 0.05$.

Time stands for time point, Var for variance component and Se for standard error.

agreed upon the ranking of teachers.

4.1. Overall change during intervention year

In Table 3 the mean teacher scores on the latent scale over time as well as the overall PDP-related change (the difference between time point 1 until 3, and time point 4 until 6) and the means are presented for each scale, and it is also indicated whether the PDP-related changes could be interpreted as significant or not.

The columns in Table 3 labeled 'Mean' provide the means of the scale scores over time and the mean change between the first three and last three time points (labeled PDP-change). So the PDP-change is the difference between the average scale score before and after the treatment. The columns labeled 'Se' present the standard errors. Table 3 shows a positive PDP-related change for each of the scales. Though, considering its standard error, the PDP-change is only significant for the DBDM scale, which implies that teachers showed the DBDM skills significantly more after the PDP than prior to the PDP. The size of the change for this scale is approximately one, meaning that the difference in the extent to which DBDM skills were shown by teachers is one standard deviation. This can be interpreted as a (very) large change (Cohen, 1988, for details on the computation refer to Appendix A).

When comparing the average values on the time points, both prior to and after the PDP intervention, the positive averages of the time points proved to be strongest for the *basic skills*. This implies that, on average, the teachers studied mastered these skills better than the skills measured by means of the other scales. Furthermore, a clear ranking is visible: the averages on the time points on the Clear and activating instruction scale were the second highest, followed by the DBDM scale, while the averages on the time points

were the lowest (even negative) for the Learning strategies scale. After the PDP this ranking is less apparent as the averages of the time points for the Clear and activating instruction scale were similar to the averages of the time points on the DBDM scale.

4.2. Changes in instructional skills during the intervention year

In the following section, we describe, per scale, to what extent the instructional skills of teachers changed during the intervention year across the groups we distinguished between. As this is an exploration of our data, we describe patterns on the basis of the data. The conclusions are not based on statistical tests.

4.2.1. Basic skills scale

In Fig. 2 the *teacher-time* interactions on the Basic skills scale are presented for each teacher (each teacher represents a line) prior to, and after the PDP. As discussed in section 3.3, teachers were divided into three groups (below average, average, and above average performing teachers).

Fig. 2 shows that teachers placed in the below average group for basic skills, would still be placed in this group after the PDP. The *teacher-time* interactions in the below average group after the PDP are even lower than the lowest interaction in the average group prior to the PDP. This implies that teachers in the below average group did not show growth in their basic skills. In contrast, some of the teachers in the average and above average performance groups would have switched places on the basis of their scores after the intervention.

Fig. 2 also shows that only the poorest performing teacher in the below average group showed better basic skills after the intervention, while in both other groups more teachers developed their

Table 3

The mean scale scores over time, specified per time point, and the PDP-change for each scale.

	Basic skills		Clear and activating instruction		DBDM		Learning strategies		Total	
	Mean	Se	Mean	Se	Mean	Se	Mean	Se	Mean	Se
Time 1	0.98	0.37	0.61	0.38	−0.18	0.37	−0.31	0.30	0.56	0.34
Time 2	0.70	0.36	0.50	0.37	−0.65	0.35	−0.68	0.32	0.36	0.32
Time 3	0.91	0.38	0.63	0.38	−0.25	0.37	−0.44	0.30	0.55	0.33
Time 4	1.02	0.37	0.76	0.38	0.54	0.38	−0.41	0.31	0.65	0.34
Time 5	1.05	0.36	0.84	0.38	0.67	0.39	−0.23	0.30	0.75	0.34
Time 6	1.05	0.38	0.82	0.38	0.90	0.38	−0.32	0.32	0.74	0.33
PDP-change	0.17	0.18	0.22	0.18	1.06**	0.23	0.15	0.18	0.22	0.17
Change size	0.11	0.13	0.20	0.13	0.93**	0.18	0.20	0.15	0.20	0.13

Time stands for time point.

** $p < 0.05$.

Fig. 2. Mean teacher-time interactions on the Basic scale prior to and after participation in the PDP.

basic skills during the intervention year. However, these differences do not seem to follow a consistent pattern, implying that the teachers in the three groups did not develop their basic skills differently during the intervention year.

4.2.2. Clear and activating instruction scale

In Fig. 3 we present the teacher-time interactions for the Clear and activating instruction scale. Teachers stayed in the same group as specified in section 4.2.1.

Fig. 3 shows that if teachers would have been grouped on the basis of their teacher-time interaction on their clear and activating instruction skills (instead of on teacher-time interactions of their basic skills), eight out of nine teachers in the below average group would have remained in the same group as well. This is different from the teachers in the average group and the above average group. In total four teachers would have changed groups. This implies that those teachers, who prior to the PDP mastered the basic skills the least, also mastered the clear and activating instruction skills the least after the intervention. On the other hand, teachers marked as 'average', or 'above average' on the basic skills show a less consistent pattern. Moreover, in Fig. 3 it is shown that, except for a single teacher, teachers in the below average performing group remain stable on their clear and activating instruction skills.

In contrast, some teachers in the average and above average performing groups showed growth in these instruction skills, while a few teachers deteriorated or remained stable.

4.2.3. DBDM scale

From Table 3 we concluded that teachers showed more DBDM skills (the use of instruction groups, and the discussion/presentation of the lesson goals) after the intervention year. In this section, we explore whether teachers who showed below average, average and above average basic skills prior to the PDP, developed differently during the intervention year in terms of their DBDM skills. In Fig. 4 teachers' teacher-time interactions for DBDM prior to and after the PDP are presented.

Fig. 4 shows that, in contrast to the results in Figs. 2 and 3, teachers who performed below average in general are not the teachers who would be placed in the below average group on the basis of their DBDM skills prior to the PDP. Even the two teachers with the lowest teacher-time interactions prior to the intervention from the above average group would be placed in the below average group. However, none of the below average performing teachers would have been placed in the above average group on the basis of their DBDM skills prior to the PDP. This indicates that there was not necessarily a relation between basic, and clear and

Fig. 3. Mean teacher-time interactions on the Clear and activating instruction scale prior to and after participation in the PDP.

Fig. 4. Mean teacher-time interactions on the DBDM scale prior to and after participating in the PDP.

activating instruction skills and whether teachers discussed and evaluated the lesson goal, and adapted to a greater extent their instruction to students' needs.

In contrast to Figs. 2 and 3, Fig. 4 shows that a considerable number of teachers exhibited a high gain in DBDM skills, across all three groups. Moreover, it shows that teachers who were placed in the below average group also show a gain in their DBDM skills. The assumption that teachers who do not fully master the basic skills cannot acquire DBDM skills yet, could therefore not be confirmed. Teachers with the lowest basic levels prior to the intervention also showed a great improvement in their DBDM skills.

4.2.4. Learning strategies scale

The teacher-time interactions for learning strategies skills are presented in Fig. 5. As learning strategies were assumed to be more advanced teaching skills, we expected that above average performing teachers would show the greatest gain for the Learning strategies scale.

In line with Figs. 4 and 5 also shows that the grouping on the basis of the basic teaching skills is not in line with the grouping that would have been made on the basis of the learning strategies skills. Furthermore, Fig. 5 shows that only a few teachers exhibited either a gain or decay in their learning strategies skills. Moreover, there were no clear differences between the three groups. As most teachers remained stable, it was concluded that during the intervention year, teachers did not seem to have improved their learning strategies teaching skills.

4.2.5. Total scale

Finally, we explored the PDP-related changes on the Total ICALT scale which included all teaching skills measured. The teachers' scale scores prior to and after the intervention are graphically presented in Fig. 6.

Fig. 6 shows that, similar to what was found for the clear and activating instruction skills, all teachers in the below average performing group would have been placed in this group on the basis of both their scores prior to and after the intervention. Two teachers in the average and above average performing groups would not have been in the same group. Apparently, a teacher's overall performance is more in line with the Basic and Clear and activating instruction scales, than to the Learning strategies and the DBDM scales. Fig. 6 is comparable to Fig. 3, showing that below average performing teachers do not seem to have developed their teaching

skills as much during the intervention year as some of the average and above average performing teachers do.

5. Discussion

Studies into the effects of DBDM have shown that increased access to high level student achievement data is not sufficient for simply improving student achievement (Kaufman et al., 2014). Additional support for the implementation of DBDM seems to be required, especially at the classroom level. Support is needed so that teachers may become more capable at using student performance data to adapt instruction in line with the needs of students (Marsh, 2012). In addition to this, quality research is required to gain more insight into how DBDM can be promoted successfully (Spillane, 2012). This study evaluated the extent to which teachers developed their general teaching skills after participating in a DBDM intervention.

5.1. Findings

This study contributes to DBDM research by providing detailed insight into the development of teachers' instructional skills during a DBDM intervention.

The main finding of this study is that Dutch primary school teachers, in response to an intensive DBDM professional development program, proved to improve their DBDM related teaching skills significantly. The related variance component of 0.93, is large in terms Cohen's criteria (1988). After the intervention, they used instruction groups more often and presented and evaluated lesson objectives more. This empirical and unprecedented finding is important for the promotion of DBDM. However, both DBDM-aspects, using lesson goals and differentiating instruction, relate more to the *organizational* side of differentiation, rather than to the advanced skills required for differentiating instruction. Deunk, Doolaard, Smale-Jacobse, and Bosker (2015) rightly argue that differentiation is more than placing students into groups, it also requires that the content of the instruction matches students' instructional needs.

Our second core finding is that teachers' initial basic teaching skills were not related to the extent to which teachers were able to develop their DBDM skills. This result was not in line with our expectations, but is promising for the promotion of DBDM. The assumption was that teachers who struggle with the basic teaching

Fig. 5. Mean teacher-time interactions on the Teaching strategies scale prior to and after participating in the PDP.

Fig. 6. Mean teacher-time interactions on the Total scale prior to and after participating in the PDP.

skills, would not yet be ready to learn the more advanced teaching skills (Kyriakides et al., 2009; Van de Grift, 2007) as these would be beyond their zone of proximal development. It could be that the teachers involved in our study did not struggle severely with basic teaching skills, as 'below average performance' was relative to the other teachers in this study. In other words, it did not mean that below average performing teachers did not master the basic teaching skills at all. Moreover, according to the Inspectorate of Education (2014) more than 90 percent of Dutch teachers master the basic teaching skills. Thus, the teachers in this study may have been 'too good' to test this specific assumption. On the other hand, the teachers in this study clearly differ from each other in the extent to which they master the basic teaching skills (see for example Fig. 2). Moreover, the below average performing group of teachers did not seem to show as much progress on the Clear and activating instruction scale, and the Total scale as the other teachers did. Van der Steeg and Gerritsen (2013) also found that weak performing teachers required more coaching than strong performing teachers to achieve instructional growth.

In sum, this study has shown that Dutch primary school teachers are able to work in a more goal-oriented fashion, and with instruction groups within the classroom. Before discussing how, in future research, more focus can be placed on how teachers can be

supported in differentiating the content of their instruction, we discuss the strengths and limitations of the study.

5.2. Strengths and limitations of the study

The use of a short interrupted time series design in combination with the rating of six recorded lessons per teacher, and four raters is rare (Gitomer & Bell, 2013). This design provides a solid basis for the evaluation of the development of teachers' instructional skills during a DBDM intervention. Moreover, in this study advanced Bayesian IRT-analyses were conducted in combination with a generalizability model. The combined model supported the proper separation of all variance components, including the variance attributable to the intervention, the rater variance and its' reliability. Another strong aspect of the intervention is the labor-intensive, tailor-made coaching activities provided to teachers. However, the fact that teachers were coached in line with their needs, might limit the possibilities for replicating this study. In this study, *individual* teachers were trained in data use. However, DBDM is not limited to one teacher within a school, but preferably is part of a schoolwide approach to DBDM (Keuning et al., 2016). A data coach, as described in for example Mandinach and Gummer (2016a), rather than an external trainer could facilitate the

schoolwide implementation of DBDM.

Furthermore, it was impossible to do the labor-intensive measurements in a control group of teachers. Therefore it cannot be concluded with certainty that participation in the PDP caused the changes found on teachers' instructional skills. However, a natural change due to learning is unlikely to have occurred as only significant changes after the intervention were found for DBDM skills, the skills which were the central focus of the PDP. The instructional skills measured by means of the other scales remained relatively stable during the course of the PDP. On top of that, the Dutch Inspectorate of Education (2014) for several years in a row found that although teachers make instructional plans, they do not teach according to those plans in the classroom.

We decided to work with four, not fully independent, raters instead of with one, or two fully independent raters as a number of studies have shown that the addition of extra raters results in more reliable estimates (Hill et al., 2012; Praetorius et al., 2012). Due to time restrictions it could not be avoided that the raters in this study were aware of whether lessons had been recorded prior to, or after participating in the PDP. Two raters were strongly involved in both the design of the PDP, and, although this did not apply to the other two raters, those other raters were aware of the PDP content. This may have biased the raters, but it was impossible to appoint four independent raters due to limited resources. Table 2 shows that the relative agreement among raters is high, and that the variance accounted for by raters can be interpreted as low. This implies that the raters agreed in their rankings of teachers. Moreover, rater variance was included in the model to ensure that the other variance components (e.g., the changes over time) were estimated reliably.

5.3. Future research

Both solid professional development trajectories for teachers and lesson observations are expensive and time consuming (Desimone, 2009; Goe, Bell, & Little, 2008). Reviews of professional development programs however have shown that lengthy, well-designed PDPs are required to have an impact on teachers, and student achievement (Desimone, 2009; Van Veen et al., 2012). It is not strictly necessary to use lesson observations to determine the changes of a PDP intervention on teaching behavior. Teacher student ratings and teacher logs are much less expensive and time consuming (De Jong & Westerhof, 2001; Rowan, Jacob, & Correnti, 2009). However, it is unsure whether instructional differentiation can be fully assessed by means of the alternatives. Self-report data may be susceptible to socially desirable responses (Gitomer & Bell, 2013; Goe et al., 2008), and students might not be able to observe all aspects of instruction validly (e.g. whether the teacher differs his/her instructional strategy adequately). Therefore, to obtain additional insight into whether teachers are able to implement DBDM, including lesson observations in a control group, independent raters and a more in depth analysis of differentiation is recommended. Examples of important elements of differentiation that should be investigated in future studies are the pace of the instruction (Tomlinson et al., 2003), the materials used by teachers (Tomlinson et al., 2003) and the instructional strategies used for different groups of students (Deunk et al., 2015).

Another important topic that could not be addressed in this study, is whether teachers who implemented DBDM also achieve higher student gains. Lockwood et al. (2010) found a positive relationship between coaching, the implementation of DBDM and student achievement. The study of Lockwood et al. (2010) was done in a different setting, focused on a different subject, and the coach provided less guidance in the classroom. In a recent study by van der Scheer and Visscher (in press) positive effects of the PDP

evaluated in this study were found for students in the extended instruction group. In future research we would like to investigate the relationship between DBDM skills growth among teachers, and student achievement.

Acknowledgements

This study was funded by the Dutch Ministry of Education, Culture and Science, the Ministry of Economic Affairs and the Ministry of Finance with 500.000 euro.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.tate.2017.02.018>.

References

- Baumann, J. F., Hoffman, J. V., Duffy-Hester, A. M., & Moon Ro, J. (2000). "The first R" yesterday and today: U.S. elementary reading instruction practices reported by teachers and administrators. *Reading Research Quarterly*, 35(3), 338–377.
- Brennan, R. L. (1992). Generalizability theory. *Educational Measurement: Issues and Practice*, 11(4), 27–34. <http://dx.doi.org/10.1111/j.1745-3992.1992.tb00260.x>.
- Brennan, R. L., & Johnson, E. G. (1995). Generalizability of performance assessments. *Educational Measurement: Issues and Practice*, 14(4), 9–12. <http://dx.doi.org/10.1111/j.1745-3992.1995.tb00882.x>.
- Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability*, 21, 47–65.
- Carlisle, J. F., & Berebitsky, D. (2011). Literacy coaching as a component of professional development. *Reading and Writing*, 24, 773–800.
- 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. <http://dx.doi.org/10.3102/0162373711412765>.
- Coburn, C. E., & Turner, E. O. (2012). The practice of data use: An introduction. *American Journal of Education*, 118, 99–111.
- Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). What makes great teaching? London: The Sutton Trust. Retrieved from: <http://www.suttontrust.com/wp-content/uploads/2014/10/What-makes-great-teaching-FINAL-4.11.14.pdf>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
- Cornett, J., & Knight, J. (2009). Research on coaching. In J. Knight (Ed.), *Coaching approaches and perspectives* (pp. 192–216). Thousand Oaks, CA: Corwin Press.
- Danielson, C. (2013). *The framework for teaching: Evaluation instrument*. Princeton, NJ: Danielson Group.
- Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, 117(4), 1–26.
- De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research*, 4, 51–85. <http://dx.doi.org/10.1023/A:1011402608575>.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181–199. <http://dx.doi.org/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: GION onderwijs/onderzoek.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9.
- Gitomer, D. H., & Bell, C. A. (2013). Evaluating teaching and teachers. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, et al. (Eds.), *Vol. 3. APA handbook of testing and assessment in psychology* (pp. 415–444). Washington, DC: American Psychological Association.
- Goe, L., Bell, C. A., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56–64.
- Holloway, J. H. (2000). Preparing teachers for differentiated instruction. *Educational Leadership*, 58(1), 82–83.
- 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.), *Vol. 106. Evidence and decision making* (pp. 105–131). Malden, MA: Blackwell Publishing Inc.
- Inspectorate of Education. (2010). *Opbrengstgericht werken in het basisonderwijs*. Utrecht: Inspectorate of Education.
- Inspectorate of Education. (2014). *De staat van het onderwijs*. Onderwijsverslag

- 2012/2013. Utrecht: Inspectorate of Education.
- Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Danvers: ASCD.
- Kaufman, T. E., Graham, C. R., Picciano, A. G., Popham, A. J., & Wiley, D. (2014). Data-driven decision making in the K-12 classroom. In M. J. Spector, D. M. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 337–346). New York: Springer.
- 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.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik, & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster, Germany: Waxmann.
- Knight, J. (2009). Instructional coaching. In J. Knight (Ed.), *Coaching approaches and perspectives* (pp. 29–55). Thousand Oaks, CA: Corwin Press.
- Kyriakides, L. (2005). Drawing from teacher effectiveness research and research into teacher interpersonal behaviour to establish a teacher evaluation system: A study on the use of student ratings to evaluate teacher behaviour. *Journal of Classroom Interaction*, 40(2), 44–66.
- Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143–152.
- Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training an professional development. *Teaching and Teacher Education*, 25, 12–23.
- 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. <http://dx.doi.org/10.3102/0162373710373388>.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N.J.: Erlbaum.
- 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. <http://dx.doi.org/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. <http://dx.doi.org/10.1177/0895904809341467>.
- Maulana, R., Helms-Lorenz, M., & Van de Grift, W. J. C. M. (2015). Development and evaluation of a questionnaire measuring pre-service teachers' teaching behaviour: A Rasch modelling approach. *School Effectiveness and School Improvement*, 26(2), 169–194. <http://dx.doi.org/10.1080/09243453.2014.939198>.
- McGarvey, B., Marriott, S., Morgan, V., & Abbott, L. (1997). Planning for differentiation: The experience of teachers in Northern Ireland primary schools. *Journal of Curriculum Studies*, 29(3), 351–363.
- McKenna, M. C., & Walpole, S. (2008). *The Literacy Coaching Challenge: Models and Methods for Grades K-8*. New York, NY: The Guilford Press.
- McKenna, M. C., Walpole, S., Uribe-Zarain, X., & Lamitina, D. (2010). The relationships between coaching and instruction in the primary grades. *The Elementary School Journal*, 111(1), 115–140.
- Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011). *Teachers' ability to use data to inform instruction: Challenges and supports*. Washington, D.C.: U.S. Department of Education Office of Planning, Evaluation and Policy Development.
- Parker Boudett, K., City, E. A., & Murnane, R. J. (2007). Introduction. In K. Parker Boudett, E. A. City, & R. J. Murnane (Eds.), *Data wise: A step-by-step guide to using assessment results to improve teaching and learning* (pp. 1–8). Cambridge, MA: Harvard Education Press.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Poortman, C. L., Schildkamp, K., & Lai, M. K. (2016). Professional development in data use: An international perspective on conditions, models and effects. *Teaching and Teacher Education*, 60, 363–365. <http://dx.doi.org/10.1016/j.tate.2016.07.029>.
- Praetorius, A.-K., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction*, 22, 387–400. <http://dx.doi.org/10.1016/j.learninstruc.2012.03.002>.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Rowan, B., Jacob, R., & Correnti, R. (2009). Using instructional logs to identify quality in educational settings. *New Directions for Youth Development*, 121, 13–31.
- Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: development and validation of a differentiated instruction scale. *International Journal of Inclusive Education*, 17(11), 1186–1204. <http://dx.doi.org/10.1080/13603116.2012.743604>.
- Sailors, M., & Price, L. (2015). Support for the improvement of practices through intensive coaching (SIPIC): A model of coaching for improving reading instruction and reading achievement. *Teaching and Teacher Education*, 45, 115–127.
- van der Scheer, E. A., & Visscher, A. J. (in press). Effects of a data-based decision making intervention for teachers on students' mathematical achievement. *Journal of Teacher Education*.
- 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.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont: Wadsworth.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(2), 4–14.
- 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. <http://dx.doi.org/10.3102/0002831212466909>.
- Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, 118(2), 113–141. <http://dx.doi.org/10.1086/663283>.
- 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. <http://dx.doi.org/10.1016/j.stueduc.2013.11.002>.
- Teemant, A. (2014). A mixed-methods investigation of instructional coaching for teachers of diverse learners. *Urban Education*, 49(5), 574–604.
- Teemant, A., Leland, C., & Berghoff, B. (2014). Development and validation of a measure of Critical Stance for instructional coaching. *Teaching and Teacher Education*, 39, 136–147.
- Teemant, A., Wink, J., & Tyra, S. (2011). Effects of coaching on teacher use of sociocultural instructional practices. *Teaching and Teacher Education*, 27, 683–693.
- Tobin, R., & McInnes, A. (2008). Accommodating differences: Variations in differentiated literacy instruction in Grade 2/3 classrooms. *Literacy*, 42(1), 3–9.
- 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. <http://dx.doi.org/10.1177/016235320302700203>.
- 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: Routledge.
- Van de Grift, W. J. C. M. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research*, 49(2), 127–152.
- Van de Grift, W. J. C. M. (2014). Measuring teaching quality in several European countries. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 25(3), 295–311. <http://dx.doi.org/10.1080/09243453.2013.794845>.
- 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. *Pedagogische Studiën*, 88, 416–432.
- Van der Steeg, M., & Gerritsen, S. (2013). *Teacher evaluations and pupil achievement. Evidence from classroom observations*. The Hague: CPB Netherlands Bureau for Economic Policy Analysis.
- 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*. Retrieved from <http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2011/07/13/de-eenvoud-en-complexiteit-van-opbrengstgericht-werken.html>.
- Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education*, 25, 1051–1060.
- Williams, R. T., Swanlund, A., Miller, S., Konstantopoulos, S., Eno, J., Van der Ploeg, A., et al. (2014). Measuring instructional differentiation in a large-scale experiment. *Educational Psychological Measurement*, 74(2), 263–279. <http://dx.doi.org/10.1177/0013164413507724>.