School Effectiveness and School Improvement
An International Journal of Research, Policy and Practice

ISSN: 0924-3453 (Print) 1744-5124 (Online) Journal homepage: https://www.tandfonline.com/loi/nses20

Data teams for school improvement

Kim Schildkamp, Cindy L. Poortman & Adam Handelzalts

To cite this article: Kim Schildkamp, Cindy L. Poortman & Adam Handelzalts (2016) Data teams for school improvement, School Effectiveness and School Improvement, 27:2, 228-254, DOI: 10.1080/09243453.2015.1056192

To link to this article: https://doi.org/10.1080/09243453.2015.1056192

Published online: 26 Jun 2015.

School Effectiveness and School Improvement, 2016
Vol. 27, No. 2, 228–254, http://dx.doi.org/10.1080/09243453.2015.1056192

Data teams for school improvement

Kim Schildkamp[a]*, Cindy L. Poortman[a] and Adam Handelzalts[b]

[a]*Faculty of Behavioural, Management and Social Sciences, University of Twente, Enschede, The Netherlands;* [b]*Faculty of Psychology and Education, VU University Amsterdam, Amsterdam, The Netherlands*

(*Received 28 March 2014; final version received 1 May 2015*)

The use of data for educational decision making has never been more prevalent. However, teachers and school leaders need support in data use. Support can be provided by means of professional development in the form of "data teams". This study followed the functioning of 4 data teams over a period of 2 years, applying a qualitative case study design. The findings show that data use is not a linear process, and that teams go through different feedback loops to reach higher levels of depth of inquiry. The data team procedure is a promising way of enhancing data-based decision making in schools.

Keywords: data use; data teams; teacher professional development; qualitative case studies

Introduction

The use of data for educational decision making has never been more prevalent (Datnow, Park, & Kennedy-Lewis, 2013; Marsh & Farrell, 2015; Schildkamp & Kuiper, 2010). Our definition of "data" in the context of education is "*information that is systematically collected and organized to represent some aspect of schools*" (Lai & Schildkamp, 2013, p. 10). This encompasses the multiple types of quantitative as well as qualitative data that teachers and school leaders need for decision making (Lai & Schildkamp, 2013; Wayman, Jimerson, & Cho, 2012). If these data are actually being used by school staff to guide decisions (e.g., with regard to school development or instructional purposes), this is referred to as data-based decision making or data use in short.

Although effective data use can help improve the functioning of schools in terms of increased student achievement (Campbell & Levin, 2009; Carlson, Borman, & Robinson, 2011; McNaughton, Lai, & Hsiao, 2012), teachers often do not use data to its best effect, if at all (Schildkamp & Kuiper, 2010; Schildkamp & Teddlie, 2008). A majority of decisions are based on intuition and limited observations (Ingram, Louis, & Schroeder, 2004). Therefore, valuable time and resources are lost with the implementation of new instructional methods, which, for example, do not coincide with the needs of the students (Earl & Katz, 2006). Professional development (Desimone 2009; Van Veen, Zwart, Meirink, & Verloop, 2010) in data use is therefore urgently needed for improving the quality of schools.

Several studies show that professional development is often more effective when it takes place in professional learning communities (PLCs). These studies show that teacher collaboration can lead to increased teacher and student learning (Borko, 2004;

*Corresponding author. Email: k.schildkamp@utwente.nl

Darling-Hammond, 2010; Stoll, Bolam, McMahon, Wallace, & Thomas, 2006; Vescio, Ross, & Adams, 2008). According to Stoll et al. (2006; see also Lomos, Hofman, & Bosker, 2011), PLCs consist of teachers, and sometimes school leaders, working together. PLCs have several defining characteristics: shared values and visions, collective responsibilities, engagement in reflective professional inquiry, collaboration, promoting both group and individual learning, mutual trust, inclusive membership, and openness. According to Ball and Cohen (1999), the inquiry component of PLCs is crucial as this can improve instruction, which in turn can lead to increased student achievement.

To support teachers in using data effectively, we developed the professional development program “the data team procedure” (Schildkamp, Handelzalts, & Poortman, 2012). Data teams can be seen as PLCs engaging in collaborative inquiry, focused on data use (Ball & Cohen, 1999; Cochran-Smith & Lytle, 2009; Stoll et al., 2006). Data teams consist of teachers and school leaders who analyze and use data collaboratively to improve their educational practice. Professional development in the form of data teams is a promising way of enhancing the effectiveness of data-based decision making (Earl & Katz, 2006; Wayman, Midgley, & Stringfield, 2006). Collaboration helps teachers to learn from each other how to use data, and allows for a fertile exchange of ideas and strategies (Nelson & Slavit, 2007; Wohlstetter, Datnow, & Park, 2008). In addition, data teams incorporate other characteristics of effective professional development (Borko, 2004; Desimone, 2009; Garet, Porter, Desimone, Birman, & Yoon, 2001; Guskey, 2002; Jimerson & Wayman, 2012; Van Veen et al., 2010; Wayman et al., 2012; Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009), because the data team procedure:

- focuses on sharing and discussing opinions with colleagues and taking collective responsibility for the goals and organization of learning;
- is attentive to the context of educators’ work and usability for daily practice, because the focus is on the school’s own educational problems and related data;
- takes place over a longer period of time;
- is coherent, because data teams follow a structured and organized eight-step procedure.

To obtain a more “robust understanding of practice” (Little, 2012), more research into data use is important (see also Datnow et al., 2013). There is a need to open the “black box” of teacher professional development to explore the process of teacher development and interaction between teachers (Butler & Schnellert, 2012; see also Fullan, 2000). Therefore, the main question of this study is: *How do data teams function during the implementation of the data team procedure?*

The data team procedure

Data teams consist of a data expert, four to six teachers and one to two (assistant) school leaders, who collaboratively use data to solve an educational problem within the school, using a structured approach. The data team members are trained in the approach by a university researcher over a period of 2 years. The researcher facilitates the process and visits the data team’s school every month for a data team meeting. At the start of the process, the facilitator explains the steps of the process, but the data team members execute the steps. The facilitator guides the process and redirects when necessary.

School leaders are part of a data team as they often have a different perspective on the educational problem to be solved. Furthermore, their participation will help ensure

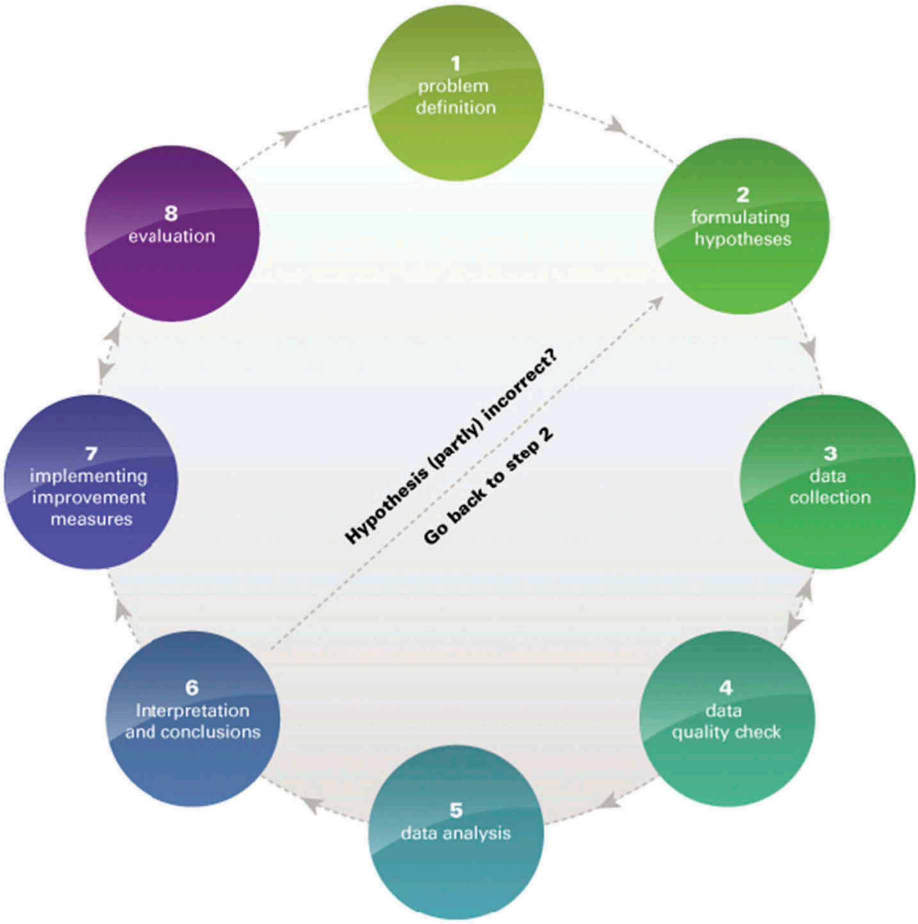

Figure 1. The data team procedure (Schildkamp & Ehren, 2013, p. 56).

implementation of the action plan at the end of the process. Data teams are a form of professional development with the ultimate goal of school improvement. The teams in this study all worked according to an iterative and cyclic procedure (inspired by Earl & Katz, 2006), consisting of eight steps (see also Figure 1) (Schildkamp & Ehren, 2013; Schildkamp & Handelzalts, 2011; Schildkamp et al., 2012):

(1) Problem definition: The team decides on which educational problem they will focus.
(2) Formulating hypotheses: The team develops hypotheses with regard to what may cause their problem.
(3) Data collection: The team collects data to either confirm or reject the hypotheses. Several types of data can be collected, such as assessment data or inspection reports.
(4) Data quality check: The team assesses the quality of the collected data. If the data are valid and reliable, the team proceeds to Step 5. Otherwise, the team needs to collect additional data.

(5) Data analysis: The team analyzes the data. This can involve qualitative and quantitative data analysis, and simple (descriptive) as well as more sophisticated analyses (e.g., *t* tests, correlations).
(6) Interpretation and conclusions: The team concludes whether their hypothesis about the cause of the problem was correct and either proceeds to Step 7 (correct hypothesis) or returns to Step 2 (incorrect hypothesis) and investigates new hypotheses.
(7) Implementing improvement measures: The team takes action based on the data to solve their educational problem, to improve education.
(8) Evaluation: The team evaluates if their measures were implemented as intended and if these measures were effective.

The data team procedure has several communalities with action research. Carr and Kemmis (1986) define action research as "self-reflective inquiry undertaken by participants in social situations in order to improve the rationality and justice of their own practices, their understandings of these practices, and the situations in which the practices are carried out" (p. 162). Action research is often organized around a certain cycle of inquiry (Johnson, 2003) and can also be conducted collaboratively in a PLC (Capobianco & Feldman, 2010). The focus of data teams, however, is on professional development specifically in data use and school improvement, ultimately measured in student achievement outcomes; rather than on teachers building their own practical theories of teaching (Whitehead & McNiff, 2006).

Theoretical framework

Data-based decision making in data teams

Our data use theory of action (Schildkamp & Poortman, 2015, p. 5) is based on the analyses of several data use models and frameworks (Coburn & Turner, 2011; Ikemoto & Marsh, 2007; Lai & Schildkamp, 2013; Mandinach, Honey, Light, & Brunner, 2008; Marsh, 2012; Schildkamp et al., 2012; Schildkamp & Kuiper, 2010; Schildkamp & Poortman, 2015). In this framework (see Figure 2), the interaction between data and people, in a certain context, results in decisions regarding what action to take.

Data use in data teams starts with a *purpose* in the form of a problem definition and a related goal instead of with data. Next, *data* are collected to investigate possible causes of the problem. The team needs to filter the data (e.g., are the data valid and reliable; if not, additional data need to be collected and a *feedback* loop is created), organize the data to investigate the hypothesis, and analyze and interpret the data. Only then, these data are transferred into *information*. Combined with stakeholder understanding and expertise, this becomes actionable *knowledge*. Data teams can take two possible *actions*: The hypothesis is incorrect and the action is to go back to formulating new hypotheses (a feedback loop is created), or the hypothesis is correct and the data team takes action based on the data. In case of the latter, they also need to evaluate (collect new data) whether their actions have led to the desired *outcomes* and goal; in this way, another feedback loop is created (Marsh, Pane, & Hamilton, 2006; Marsh, 2012). Furthermore, data use does not happen in isolation but is shaped by school organizational characteristics, user characteristics, and data characteristics. The framework acknowledges that data-based decision making is a complex and non-linear process, involving several feedback loops. It involves a number of processes,

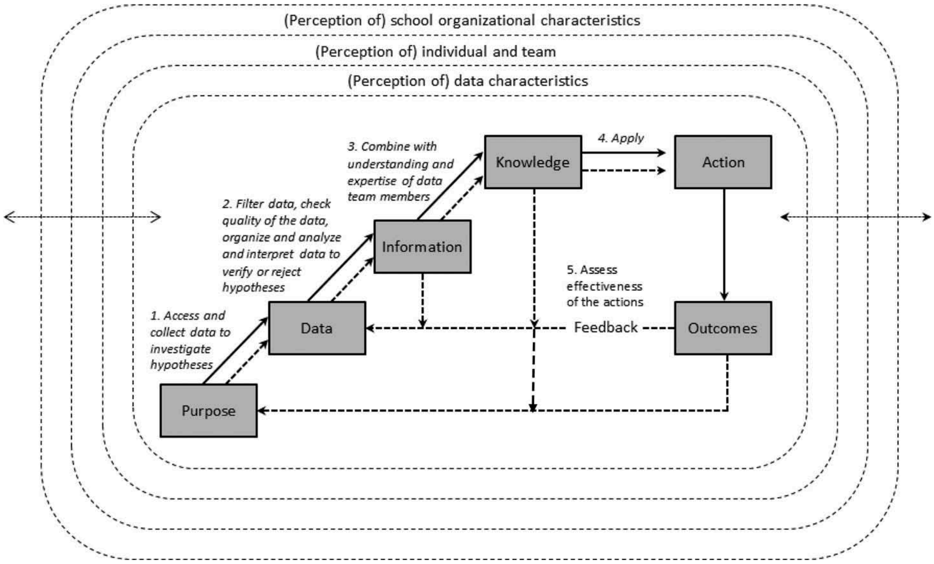

Figure 2. Data use theory of action, and factors influencing data use (Schildkamp & Poortman, 2015, p. 5, based on Marsh, 2012, p. 4; Coburn & Turner, 2011; Ikemoto & Marsh, 2007; Lai & Schildkamp, 2013; Mandinach et al., 2008; Schildkamp et al., 2012; Schildkamp & Kuiper, 2010; Schildkamp & Lai, 2013).

conditions, and contexts that interact in complex ways. For example, data may not always directly lead to actions and outcomes (what Weiss [1998] calls "instrumental use"), but the data might challenge assumptions, undermine myths, and provide the users with new insights and new knowledge. This is what Weiss (1998) refers to as "conceptual use" or "the enlightenment function". Furthermore, there are possible unintended consequences of the use of data, often as a result of increased accountability pressure. For example, teachers and/or school leaders can "misuse" data. A well-known example is teaching to the test and narrowing the curriculum. Moreover, data may be "abused" in that teachers try to improve test scores by focusing only on students who are likely to improve, while ignoring other students (Booher-Jennings, 2005; Wayman, Spikes, & Volonnino, 2013).

The first subquestion regarding the functioning of data teams is: *How do data teams engage in data-based decision making during the implementation of the data team procedure?*

Depth of inquiry and attribution

Research shows that more successful teams (i.e., teams that achieve higher student learning gains) employ more higher level thinking skills (Achinstein, 2002; Stokes, 2001). In these teams, teachers explore underlying assumptions, question their own practice, revise their conceptions of teaching and learning, and adjust their practice accordingly. Consequently, the "depth of inquiry" of data team discussions is important. Henry (2012) defines *depth of inquiry* as the degree to which team conversations express higher level thinking skills, such as analysis, synthesis, goal setting, and reflection. This concerns the extent to which team discussions support the development of new knowledge focused on taking action in their classroom (Henry, 2012, p. 55). Conversations that *lack* depth focus on telling information, retelling, describing, and storytelling.

Teachers have a central role in realizing sustainable and significant improvement of teaching and learning practices. They are the ones who have to change their practice to improve student achievement. Therefore, the extent to which data team conversations are about relationships between teachers, students, and content is a valuable concept in analyzing and understanding teachers' collaborative work in data teams (Henry, 2012; Ikemoto & Marsh, 2007). Conversations about relationships between teachers, students, and content are more prevalent in more successful and effective teams (Henry, 2012). In data teams, this idea is applicable to formulating and researching hypotheses about the cause of their problem: attribution.

Attribution theory can help to understand what the perceptions are of datateam members with regard to the cause of the problem they are investigating. How individuals attribute causes of a problem has behavioral consequences (Weiner, 2010) in terms of possible improvement actions based on data. Data teams may, for example, attribute problems to causes outside the school (low entry levels of students), causes at the level of the school (school policy regarding failing students), and causes at the level of the students (not motivated) or the teachers' functioning regarding the problem.

We use the concept of "attribution", because this is specifically concerned with ascribing the cause of the problem to something. Investigating causes outside the school contributes to the learning process (learning how to use data), but often these assumptions are incorrect (Ingram et al., 2004). Moreover, it does not lead to solving the problem as often there is nothing that the teachers can do about this external cause. It is important that at some stage teachers take the responsibility for their students' learning and start to look at internal causes (internal attribution).

The second subquestion regarding the functioning of data teams is: *How can we characterize the functioning of data teams in terms of depth of inquiry and attribution during the implementation of the data team procedure?*

Method

We applied a qualitative case study design using a micro-process perspective (see Little, 2012). A qualitative, descriptive approach is most useful to describe the process of the functioning of data teams in its context (Krathwohl, 1998). Because "the corpus of observational research [on data use] remains quite meager" (Little, 2012, p. 159), we studied the functioning of four data teams in depth, using both observations and interviews (Yin, 2003).

Context

This study took place in secondary education in The Netherlands. Dutch schools have a considerable amount of freedom regarding the subject matter taught, textbooks, and the instructional strategies used (Organisation for Economic Co-operation and Development [OECD], 2008, 2010). The only standardized mandatory assessment at this level is a national exam, at the end of secondary education.

However, the Dutch Inspectorate increasingly holds schools accountable for making use of data to improve the quality of their education (Verbeek & Odenthal, 2014). The policy aim of the Ministry of Education is that by 2018, at least 90% of the primary and secondary schools apply data use. In 2010, only 20% of the Dutch secondary schools applied data use according to the Dutch Inspectorate (Inspectie van het Onderwijs, 2011). All schools are visited at least every 4 years by the Inspectorate, to be evaluated on the quality of instruction and student achievement standards. Inspection reports are publically available on the internet. Dutch schools can use several data sources to improve their education, such as data on intake, transfer, and school leavers; school inspection and self-evaluation data; assessment and examination results; and student and parent questionnaire data (Schildkamp & Kuiper, 2010).

Respondents

For this study, four data teams consisting of members from six secondary education schools (one team consisted of teachers from three different school locations; see Table 1) were followed over a period of 2 years. These schools voluntarily approached the University of Twente for participation, because they had heard about the intention to pilot this new form of professional development. The problems these teams tried to solve were:

- high percentage of grade repeaters in the 4th year of secondary education (Team W);
- poor mathematic achievement in the lower grades of secondary education (Teams B and C);
- low(er) performance of pre-vocational education students entering senior general secondary education (Team G).

Data collection and instruments

One researcher developed the initial data team procedure and also facilitated the data team meetings. All 34 data team meetings (1–1.5 hr), were audio-recorded by the same researcher, who also took field notes. Capturing the entire conversations in the meetings provides much more depth and specificity than we would achieve using observation protocols or field notes alone (Little, 2012). In addition, nearly all data team participants were interviewed approximately six months after the data team had started, leading to a total of 24 (1-hr) interviews. The observations focused on the elements of the theoretical framework: data-based decision making, depth of inquiry, and attribution (see Appendix 1). Regarding data-based decision making, we focused on the main activities and outcomes of each step in the data team procedure in terms of the data use theory of action (see Figure 2). Regarding depth of inquiry, we focused on the degree to which team conversations expressed higher level thinking skills. Regarding attribution, we focused on whether teams attributed problems to causes outside the school, to causes at the level of the school, at the level of the students, or at the level of their own functioning.

In the interviews, questions were formulated with regard to the way participants experienced the data teams and used data in these teams. We also focused on data-based decision making, depth of inquiry, and attribution. Example questions are “Please tell me about the activities the data team has undertaken?” and “What data have you collected and how have you used these data?”

Table 1. Team characteristics.

Respondent	Function	Years of experience	Subjects	Data team
Bernhardt	Teacher and mentor	19	History	W
Eve	Teacher and mentor	19	Arts and music	W
Elisha	Teacher and school leader of the HAVO[1] department	15	German	W
Joe	Teacher and HAVO coordinator	27	Dutch	W
Thomas	Teacher	36	English	W
Gerrit	Teacher and mentor	25	English	W
Laura	Data expert[2]	3	NA	W
Boss	Mentor, mathematics teacher, and department head	32	Mathematics	B
Eddy	Mentor and mathematics and science teacher	8	Mathematics and technique	B
James	School leader HAVO and VWO	11	History (but not teaching)	B
Maura	Mathematics, ICT Teacher, ICT coordinator	2	Mathematics and informatics	B
Daisy	Teacher	Not interviewed	Mathematics	B
Laura	Data expert	3	NA	B
Amy	School leader HAVO/ VWO[3]	32	Physical eds. (not teaching)	C
Mary	Teacher and mentor Location 1	25	Mathematics	C
Sue	Teacher and mentor Location 1	4	Mathematics	C
Rita	Teacher and mentor Location 3	32	Mathematics and English	C
Mel	Teacher Location 2	7	Mathematics	C
Jake	Teacher Location 3	30	Mathematics	C
Casper	Data expert	Not interviewed	NA	C
Abel	Teacher	34	Biology	G
Emma	School leader HAVO and VWO	11	Music (not teaching)	G
Gabriel	Teacher	31	Physics and mathematics	G
Mandy	Teacher and teacher trainer	20	Religious studies and social studies	G
Gary	Teacher	25	History and geography	G
Wallis	Data expert	11	NA	G

Notes: [1]Senior general secondary education. [2]The school's data expert has access to data and can help the team to collect, organize, analyze, and interpret the data from their role in the school. [3]Pre-university secondary education.

Data analysis

For objectivity reasons, the researchers who did not facilitate the data team conducted the analyses. Both the interview and observation recordings were transcribed verbatim. The qualitative analysis program Atlas.ti was used to code all the transcripts based on the theoretical framework. Each data team was considered a "unit of analysis", within which we coded the transcripts per data team meeting and per interview. Regarding data-based decision making, we summarized the advancement of the data team per meeting and then per team overall. For example, we summarized how teams accessed and collected data to investigate hypotheses (i.e., from purpose to data, see Figure 2). Furthermore, we coded the level of depth of inquiry throughout the meetings. Based on the literature and an initial round of coding, we identified three levels of depth of inquiry:

- No depth: only storytelling, retelling (known) information and personal anecdotes, not based on systematically collected data.
- Average depth: basic data use and basic understanding and explanations based on data, such as "the percentage of students that pass is too low".
- High depth: data team members developing new knowledge based on data, focused on taking action in their classroom. This refers to analyzing, interpreting, comparing, summarizing, and drawing conclusions based on data, to create new knowledge to solve the data team's problem.

Finally, we coded four types of attribution. Respondents could attribute the problem to factors: outside the school (e.g., problems with primary schools), at the level of school policy (e.g., policy with regard to grade repetition), at the level of students/the classroom (e.g., "not motivated"), at the level of teachers' own functioning (e.g., my instruction is not sufficient). It is important to note that we operationalized the level of teacher's own functioning as the highest level. The cause of a problem might be found at the level of the student, but it is still important that teachers reflect on what this means for their own functioning. So, if the team found the cause at the level of the student, but reflect on what this means for their own functioning, this is referred to as the highest level of attribution.

After initially coding a sample of the data, the coding scheme was refined, based on discussions between the researchers about the applicability of the scheme and the meaning of the data. We first conducted within-case analyses by summarizing the findings per code per case (i.e., a data team). Then we conducted a cross-case analysis by comparing the findings per code per case with the aid of matrix displays as presented in Table 2, and then summarizing the findings for each element of the theoretical framework for all of the cases together. We used (originally Dutch) respondent quotes to illustrate the results and translated these into English.

Reliability and validity

To promote reliability, we used a systematized approach to data collection, consistent with the research questions. All data team meetings and interviews were audio-recorded, allowing for thorough analysis of the data. Also, the instruments were based on our theoretical framework and existing instruments. Finally, two researchers coded approximately 10% of the same observation transcripts. We calculated the inter-rater agreement and found a satisfying Cohen's Kappa of 0.76. Internal validity was enhanced by comparing respondents' experiences and beliefs in a matrix display. We enhanced external validity in terms of analytical generalization by providing case-specific and cross-case thick descriptions (also including citations of respondents), and describing the congruence with the theoretical framework (Poortman & Schildkamp, 2012).

Results: within-case analyses

Here, we give a more extensive description of one of the cases, Team W, as an example of how the subsequent conclusions are based on the data. It is beyond the scope of the paper to discuss all cases in depth: The other cases are therefore summarized more briefly.

Team W

Team W: data-based decision making

Regarding the purpose, in the first meeting Team W decided to focus on the problem of students repeating grades in the middle-level track[1]: on average over 30% per year.

In the next couple of meetings, they were engaged in formulating hypotheses, collecting data and data analysis, and interpreting these data. In general, the different members of the team actively contributed to the discussions, asked each other clarifying questions, and complemented each other.

The team collected data from their own school registration system, which were considered sufficiently reliable and valid. Although some hypotheses about their own functioning and at the level of students were considered, the team mainly focused on hypotheses at the school level in the first meetings. Also, the role of parents was considered: "*The parents make the final decision, and they often want a higher level [for their child] than is advised*".

Their first hypothesis was that department advice might relate to grade repetition. This advice relates to the level of secondary education the students are advised to proceed to.[1] However, the data showed that students who ignored the advice to continue secondary education at a lower level did not have a greater probability to repeat a grade. Other hypotheses concerned strictness of transfer rules and motivation. Data showed that these hypotheses were incorrect. However, the motivation survey results showed that students (both in the failing and passing group) needed help in planning and needed more feedback from teachers on their planning activities. Additionally, it seemed that teachers did not check homework regularly.

The teachers reached two important understandings by the time of Meeting 4: (a) the problem was caused by more than one issue, and (b) they needed to start looking at internal causes, such as problems with their curriculum and their own functioning. The team still discussed the role of parents and school policy; however, they also reflected on their own functioning: "… to offer them (students) more structure …. I think, we do not provide enough direction …. It's the only thing you hear all the time: 'Sir please just give me homework for the next time'."

Furthermore, the hypothesis that student absenteeism was related to students' failing was investigated. The collected data showed that a considerable part of the problem could be explained by failing students' absenteeism (15,6%). The data team turned this new knowledge into action by taking measures regarding absenteeism. They also discussed possible methods for evaluating improvement regarding absenteeism and failing students.

The teachers also decided to take a closer look at curriculum coherence, as recommended by the facilitator. The team administered a curriculum coherence survey constructed by the national curriculum institute (SLO) to all the teachers in the school. The results showed that the connection between the lower and the upper level in the different departments was weak, for example, in terms of content and assessments. The team gave all the subject departments the assignment to work on improving curriculum coherence, by filling out a matrix, regarding the connection upper–lower level. Each subject department had to fill out whether they fulfilled a certain criterion (e.g., "in our team the different types of assessments that are used in lower and upper levels of secondary education are aligned"), why they thought that they fulfilled the criterion, and if not, what action they needed to take to meet the criterion.

The team thus went through the first six steps of the procedure several times, because the first couple of hypotheses that they investigated were rejected, thereby going through several feedback loops, and creating new knowledge (conceptual data use). One teacher stated in this light: "I really thought the school's transition rules caused the problem of grade repetition, but this only contributes very slightly to the problem.... I might be wrong about other things as well. Data are more important than I initially thought."

Team W spent 2 years working on this problem. The team needed a lot of support in data analysis. More in-depth analyses had to be conducted by the researcher. However, the team collectively interpreted the findings.

The action this team took as a result of the knowledge developed in the team (instrumental use), in addition to dealing with absenteeism, concerned implementation of the following measures: discussing the transfer from groups of students from lower to upper level teachers; checking homework more regularly and providing students with more feedback. The subsequent outcome they measured was that the number of students repeating a class turned out to be less than 30%, namely 23%.

Team W: depth of inquiry

Depth of inquiry was nearly all the time confined to the lowest levels in Meetings 1 and 2: The discussion focused mostly on own personal experiences and anecdotes. In Meetings 3 and 4, the mean level increased towards average depth. In Meeting 5, the team compared the results of failing and passing students and tried to relate the answers to different questions; also discussing their own functioning in relation to the students' opinions about this. The team was developing new knowledge based on the collected data (high depth). However, in Meetings 6 and 7 they continued to discuss the results of the student motivation survey at a more basic level (average depth). In Meeting 8, the team started with discussing the absenteeism results. It turned out that this explained a substantial part of the problem, *and* that the school's absenteeism measures should be improved. The team was coming to conclusions based on data, and was looking into how the data could be used to improve practice (high depth). Meetings 9 and 10 concerned the results of the teacher questionnaire regarding curriculum coherence and related measures (high depth) (i.e., using the knowledge gained by the data to improve education).

Team W: attribution

Team W mainly pursued hypotheses at the school level during the first meetings. In Meeting 3, also the roles of parents, students, and sometimes the teachers were discussed. In Meeting 4, the role of parents and school policy were discussed; however, discussions also evolved to the role of students' attitude towards learning and, in some instances, teachers' own functioning in relation to this attitude: "What can we influence most?"

In Meetings 5 and 6, possible causes at the level of students and their own functioning were discussed more. In Meeting 7, the team mainly discussed possible causes at the level of the school (curriculum coherence) and the level of the student (absence). In Meeting 8, the team started with discussing the absenteeism results. In Meetings 9 and 10, teachers addressed what *they* could do to create more curriculum coherence. The data expert stated that it had been important for the team to investigate all these external hypotheses: "… they find out that these did not cause the problem. I think that is really good for awareness …. You can finally let go of these old discussions."

Case summary functioning Team B

Team B: data-based decision making

Team B focused on the problem of low mathematic achievement (3rd year, age 14–15), This was not initially supported by the teachers, as one teacher stated: "Math is a very difficult subject, so I do not see it as a problem that several students are failing."

Setting a purpose was problematic. One of the teachers, for example, consulted a teacher outside the team who felt that 35% of unsatisfactory grades for mathematics was acceptable. Another teacher in the team (also department head), however, reacted that school management would find only 20% acceptable. After extensive discussion, the problem was finally, although not convincingly or explicitly, acknowledged as late as the fifth meeting. When finally moving on to hypotheses and collecting data, many of the proposed data were not available. The limited analyses (using report cards) were mainly descriptive, carried out by their data expert. Team B did not succeed in turning any information into knowledge, action, or subsequent outcomes, therefore. Their process was terminated after Meeting 6 (after 9 months), due to priority issues and school leader changes.

Team B: depth of inquiry

Personal experiences and opinions about solutions were discussed mostly in the first couple of meetings, before the team decided that they needed more and other data to draw conclusions about their hypotheses. This means that depth of inquiry was confined to the lowest levels in the meetings, and typically did not exceed the "average" level of depth.

Team B: attribution

Students lacking arithmetic skills was mentioned as a possible cause for the "problem" of low mathematic achievement in the first meeting. The teachers further mentioned many external possible causes for the problem, such as the difficulty of mathematics or the (starting) level of students. However, they sometimes also brought up their own functioning from the start. The order in which chapters were taught, for example, was also seen as a possible cause.

Case summary functioning Team C

Team C: data-based decision making

The main purpose for Team C was to solve the problem of poor mathematic achievement in the 1st and 2nd year (age 12–14). The team was very eager to go through the eight-step procedure, as stated by one of the teachers: "We have many problems, like all schools do, but we tend to jump to solutions. We can use this (procedure) as an instrument to solve problems instead of constantly implementing short-term and sometimes even not working solutions."

The team struggled with formulating a problem statement with regard to low mathematic achievement. As late as Meeting 5, the exact problem was still not determined. However, the school leader stated: "I think it is great that the team after 2 months concluded that the problem was in fact not the problem. This confirms to me that this (i.e., data team procedure) is working."

The team also found it challenging to formulate concrete and measurable hypotheses. They investigated hypotheses with regard to primary schools that did not prepare students well enough, and the advice about the level of secondary education given to students being too high. Both hypotheses turned out to be wrong. However, this led to new knowledge with regard to what did not cause their specific problem (conceptual data use). One teacher stated in this light:

> We are still searching. Sometimes we thought that we found it (i.e., the cause of the problem), but if we looked further it turned out that we were wrong. Some people might perceive this as lost time, but I do not see it that way. Deleting causes is also important.

Next, the team administered a pretest and a posttest to collect data about the hypothesis of whether students "forget" about percentages and fractions between the pre- and posttest. The more advanced analyses, carried out by the external facilitator, showed that the scores dropped considerably. This information could be turned into knowledge and subsequent action (instrumental use): Repetition was considered part of the solution. The team started discussing options to solve the problem, such as regular questions about percentages at the start of lessons: a "5-min pop quiz" in percentages and fractions, and on-line differentiation programs for connecting to different levels of students. By the time of the final meetings, the teachers had structurally implemented the repetition measure in their lessons, both for fractions and percentages, and were using the online programs to differentiate instruction.

The evaluation after a couple of months showed the outcome that students advanced considerably from the pretest to the first posttest, and subsequently student results did not significantly decline from the first to the second posttest.

The team spent 2 years working on this problem. The different members generally contributed actively to the discussion, complemented each other, and asked each other clarifying questions. In some meetings, the discussions were a little chaotic and focused on personal examples, or preconditions, rather than the problem at hand or related hypotheses. One of the challenges was that the data related to the suspected problem were not readily available, and the teachers had to first construct and coordinate the tests to find out more about the problem. The solution regarding repetition, however, was based on data from the final posttest.

Team C: depth of inquiry

In the first two meetings, team members talked about the problem based on their own experiences, and not based on data. For the third meeting, they had collected data: They had administered a test to find out what the exact problem was for students. However, because of the lack of reliable data about the problem, the team kept struggling with formulating a clear problem statement for a couple of meetings. From Meetings 6 to 11, more data were collected and used as a basis for drawing conclusions about the problem, hypotheses, and solutions. In Meeting 12, data resulting from the posttest were discussed, resulting in higher levels of depth of inquiry.

Team C: attribution

The team did not only focus on external causes at the beginning. Even though "primary school" was mentioned as a cause throughout later sessions, causes at the student level, such as students' motivation, were also mentioned. Team members also mentioned causes at the level of their own functioning from the start: "We cannot solve it at the primary school; we have to solve it here."

Starting from Meeting 8, the teachers were concerned with developing a solution at the level of their own functioning: repeating fraction sums and advising students to practice with online material.

Case summary functioning Team G

Team G: data-based decision making

Team G had the purpose to focus on the intake in their senior general secondary education track (middle level), in Year 4, specifically students entering the middle level in the 4th year from pre-vocational education (lower level[1]). The hypotheses the team wanted to investigate concerned, amongst others, the achievement of students entering the 4th year of secondary education at the middle level from different tracks (either from a lower level or from the regular Year 3 middle level) and whether or not the profile[2] they followed in the previous track influenced achievement in the 4th year. The data expert collected the data. With help of the facilitator, it became clear that there were too few cases to draw meaningful conclusions. Where possible, the external facilitator carried out further statistical analyses (by combining the student groups so that more cases were available per question). The team appeared to have difficulty with working on their current hypotheses and kept jumping to new hypotheses. Several times, the facilitator redirected the team to current hypotheses.

According to the analyses by the facilitator, the grades of students entering the 4th year from the vocational lower level track did not differ significantly from the grades of students who already came from the middle level. This came as a surprise to the team: "Our gut feeling … turned out to be wrong. It was even the other way around."

The data team decided to focus further on the general problem of underperformance of students in the fourth grade. Emma, the school leader, underlined the importance of working with hypotheses and checking the hypotheses with data:

> For some people it was quite difficult … not to generate all kinds of solutions… doing, doing, doing …. But it is a shame if you do the wrong things …. That is why we need data, to investigate if our hypotheses are true or not.

The team sometimes focused on a limited part of the data to trigger new hypotheses. In addition, the available data were not always complete and reliable. Furthermore, the (sub)hypotheses were sometimes formulated in a complicated way. This appeared to hinder their progress. From Meeting 3, decisions to study new hypotheses and further analyses and subsequent conclusions were based on data. One of the teachers stated: "Through data you can find out what is happening …. Through facts you can come to entirely different conclusions."

The team spent 1 year working on this problem. They only had six meetings within the period of the pilot study, within which they did not reach the level of turning knowledge into action and outcomes. However, the team did develop new knowledge with regard to their problem (conceptual use).

Team G: depth of inquiry

In the first two meetings, the strong feeling that the lower level intake in the middle-level stream was problematic guided the team in formulating hypotheses. Although these hypotheses were not supported by the data, several team members returned to this feeling in later meetings. Combined with the fact that often too few data were available, it was challenging for this team to develop understanding for taking actions to improve education. In the final meetings, more advanced analyses were carried out leading to more understanding about, for example, whether students performed significantly better in the 3rd year than in the 4th year.

Team G: attribution

Although performance of lower level students was the focus underlying nearly all hypotheses, the team did not only mention causes outside their own functioning during the meetings. As early as Meeting 1, the role of their extra lessons, and in Meeting 2, the role of their own advice regarding sector-profile combinations were mentioned, for example. However, they did not reach the level of implementing measures related to their own functioning.

Results: cross-case analyses

Data-based decision making in the teams

The results of the teams are described in Table 2. With regard to purpose, the teams generally started discussing the problem from their own experience, investigating the more external hypotheses. Although teachers mentioned causes related to their own functioning from the start, they typically did not pursue these until later on in the process, if at all. It generally took some time before the teams could clearly formulate the problem statement and/or the related hypotheses. In addition, the availability of complete and reliable data was sometimes problematic, especially for Teams B and G. The teams needed much support from the facilitator to make sense of the available data. In addition, redirecting the team to the current problem and hypotheses rather than jumping to conclusions, helping with the data collection plan, providing ideas for measures and summarizing progress so far, and providing suggestions for next steps were important supporting activities by the facilitator for all teams. However, all the team members described the decision-making process in the data teams as a joint effort.

Three of the four teams went through several feedback loops and were able to develop new knowledge based on data. The knowledge these teams developed related to: the problem definition not being accurate, a hypotheses being false, and/or a hypothesis being correct. This is called conceptual data use. The data collected by these teams challenged their assumptions, undermined certain myths, and provided the users with new insights and knowledge. In the cases of Team W and Team C, conceptual use also led to instrumental data use. Some of the causes they investigated turned out to be true and could be followed up with effective measures. Moreover, these actions also led to the

Table 2. Data-based decision making in the teams.

W	B	C	G
Purpose	*Purpose*	*Purpose*	*Purpose*
Grade repetition in Senior general secondary education	Low mathematic achievement, only acknowledged hesitantly as late as fifth meeting	Low mathematic achievement	Students entering Year 4 of senior general secondary education from pre-vocational education
Data - Information	*Data-information*	*Data-information*	*Data- information*
School career data of all years; school advice in Year 2 data; data from students over time: final report card Grade 2, grades in Year 3, and grade repetition from 3–4; student survey data; absenteeism data; teacher curriculum coherence survey. Extensive data analysis, by department head and facilitator	Many data not available	Teachers constructed tests; initial reliability issues. Pretest and posttest to collect data on hypothesis of whether students "forget" about learning content. Advanced analyses carried out by facilitator	Intake and profile data combined with achievement data; reliability and completeness issues. Sometimes too few cases to draw meaningful conclusions. Where possible, facilitator carried out further statistical analyses
Knowledge	*Knowledge*	*Knowledge*	*Knowledge*
Rejected hypotheses: student motivation. Accepted hypotheses: feedback, planning, homework checking, absenteeism, curriculum coherence.	Rejected hypotheses: non Accepted hypotheses: non	Rejected hypotheses: primary education, narrative sums, advice data.	Rejected hypotheses: intake and profile. Accepted hypotheses: non
Problem caused by more than one issue and internal causes seem more relevant. A considerable part of problem could be explained by students' absenteeism		Accepted hypotheses: repetition	

(*continued*)

Table 2. (*Continued*).

W	B	C	G
Action	*Action*	*Action*	*Action*
Strict follow-up after unauthorized student absence, more discipline in homework checking, assignment for subject-level team about improving curriculum coherence	Not reached	A “5-min pop quiz” in percentages and fractures, on-line differentiation programs	The team decided to abandon their original problem and continue with low achievement in 4 HAVO, as data showed that HAVO 3 grades are better than HAVO 4 grades
Outcomes	*Outcomes*	*Outcomes*	*Outcomes*
Grade repetition from over 30% to 23%	Not reached	Students advance from pretest to first posttest, and do not significantly decline from first to second posttest	Not reached

desired outcomes: These teams were able to solve their educational problem and increase student achievement

Depth of inquiry in the teams

In the first meetings, the average level of depth was rather low for all teams. The teams focused on causes of the problem and their personal experiences rather than on data. Although Team C was eager to solve the problem, they defined it rather late in their process. Once they had rejected their first hypotheses, they managed to find a cause of their problem. Team B seemed less eager to solve the problem, because they considered mathematics to be "a very difficult subject". Their "slow start" was caused by late *acknowledgement* of the problem. In contrast with Team C, they did not manage to turn information into knowledge or subsequent outcomes.

Teams W and G defined their problem statements much sooner. Team G was focused on formulating hypotheses, collecting related data and drawing conclusions until the sixth meeting, when the pilot study ended. They only had six meetings because they had started later than the other teams.

Team W started with investigating relevant hypotheses early in the process, and already found out in the fourth meeting that their problem was likely to have multiple causes and that they needed to look at internal causes. They managed to draw conclusions and develop and implement related measures, showing increasing depth in their process.

Altogether, in the final meetings more advanced analyses, comparisons, and conclusions allowed Teams W, C, and G to develop new knowledge based on data, and to reach higher levels of depth of inquiry. Team B remained far behind: Most of their meetings were only focused on defining the problem and setting related criteria and personal experiences and opinions about possible solutions. Figure 3 shows the average levels of depth for all meetings and all teams.

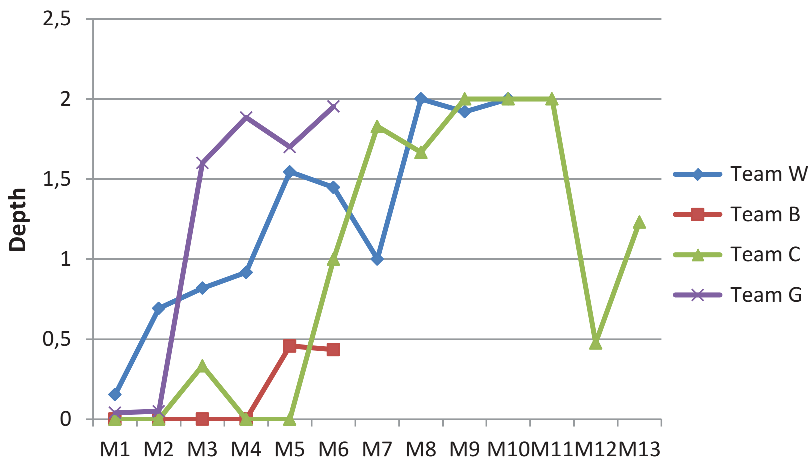

Figure 3. Average levels of depth of inquiry for all meetings for all teams.
Note: We coded the level of depth of inquiry throughout the meetings in three levels: "no depth" (Depth0); "average depth" (Depth1), and "high depth" (Depth2). The average levels of depth of inquiry per meeting were determined by dividing the total "score" of all the depth fragments per meeting by the total amount of depth codes in the particular meeting.

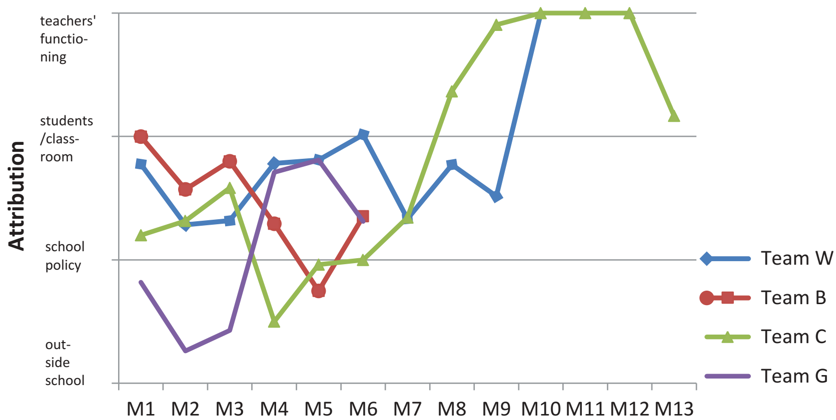

Figure 4. Average levels of attribution for all meetings for all teams.
Note: We coded the level of attribution throughout the meetings. The average levels of attribution were determined by dividing the total "score" of all the attribution fragments in the meeting (0–3, where 0 is attribution at the level outside the school, and 3 is attribution at the level of one's own functioning) by the total amount of attribution codes in the particular meeting.

Attribution in the teams

Although the teams generally pursued more external hypotheses in the first meetings, they did *mention* several causes related to their own function from the start. This is rather remarkable for Team B, for example, because they appeared to start at a rather high level, despite their rather disappointing results regarding decision making and depth. Team G nearly only focused on the level of the students, although they do mention, for example, the role of their own lessons and their own advice regarding sector-profile combinations for students. Only Team C and Team W, however, actually investigated hypotheses related to their own functioning, and consequently took action at this level. Figure 4 shows the average depth of attribution for all meetings for all teams.

Conclusions and discussion

Data-based decision making in the data teams

Although the teams in this study focused on different types of problems, several commonalities can be identified with regard to the teams implementing the data team procedure. Going through the data team procedure was not a linear process. All teams, sometimes frequently, returned to previous steps of the procedure in further meetings. Both Team B and Team C, for example, were still concerned with their problem statement in Meeting 5. In line with our theory of action, which shows that data use is not a linear process, the teams went through several feedback loops (e.g., from knowledge back to purpose and data when a hypothesis turned out to be false).

An important insight from this study is that new knowledge can also be built on hypotheses that turn out to be false. This type of knowledge might be even more important than knowledge on hypotheses that turn out to be true. It might prevent schools from investing time and money in actions that do not solve the problem, because these actions do not address the causes of the problem.

Moreover, the data do not always directly lead to action, but the data have challenged the assumptions of the data team members. The results of the data analysis have undermined certain myths, which Weiss (1998) refers to as the enlightenment function. Finding out that you are wrong (e.g., an external hypothesis) is a powerful way of learning, which may have provided the ultimate argument to turn to more internal (and often more likely) causes of the problem. In the cases of Team W and Team C, such causes turned out to be true and could be followed up with effective measures. This is what Weiss (1998) refers to as instrumental data use.

Corresponding to earlier findings (Butler & Schnellert, 2012; Schnellert, Butler, & Higginson, 2008), the facilitator was crucial in all the teams. This study shows specific ways in which the facilitator helped the team to progress, for example, by pointing out possible data sources (and their quality), carrying out several of the analyses, monitoring progress and drawing intermediate conclusions, and discussing possible solutions. Often, it is assumed that data-based decision making fails, because teachers lack the knowledge and skills to analyze data. The results of our study show that even if teams get beyond the analysis phase (from data to information to knowledge), they still need support in coming up with concrete actions.

Teams W and C managed to successfully complete Steps 1 to 8 within the pilot study period, create new knowledge, and base related decisions on data. After implementing their measures, the number of students repeating a class at Team W's school decreased in 1 year from over 30% to 23%. Team C implemented several measures to improve mathematic achievement, and the outcome was that student results did not significantly decline in percentages and fractions anymore. Although we cannot conclude that these improvements were caused by the work of the data teams alone, it is likely that the work of the data teams contributed to the improvements.

Depth and attribution in the data teams

The cross-case analysis shows that the average levels of depth (Figure 3) across the meetings are very different for the different teams. The more successful teams W and C proceeded to higher levels of depth after the first meetings (when no data were available yet), meaning that their use of data allowed them to develop new knowledge to become able to improve practice. These teams also moved from external to internal causes of their problem (Figure 4). Hypotheses with regard to their own functioning often turned out to be true. Even if the cause of a problem is external (e.g., students skipping classes), teachers still need to relate this to their own functioning: What can we do about this? Reflecting on their own functioning in the classroom in relation to the data (e.g., do I check homework in the classroom, how much attention do I pay to fractions and percentages) helped them in implementing measures to improve their teaching and related student achievement.

Implications for practice and recommendations for further research

The results of this data team pilot seem to be promising. In three out of four teams, the teachers engaged in using data to improve education and actually developed new knowledge (knowing that a hypothesis is false is also important knowledge), and two of the teams actually succeeded in solving their educational problem. As stated by Marsh and Farrell (2015), data team procedures seem to be promising capacity building interventions.

Furthermore, the study shows that starting with a problem and clear goals which all the data team members agree upon is essential. Data use is more likely to be successful if a team starts with an agreed-upon problem and does not have to go through mounts of data to find a problem (as happened in Team B). Other research also stresses the importance of having a measurable and shared goal for data use (Datnow, Park, & Wohlstetter, 2007; Earl & Katz, 2006; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Wayman & Stringfield, 2006; Wohlstetter et al., 2008; Young, 2006). Starting data-based decision making with a clear purpose is essential (Earl & Timperley, 2008), because it leads to:

- shared understanding of the problem;
- shared cognition, because people are sharing their expertise;
- collective efficacy (i.e., teachers and school leaders feel empowered by data use because they are actually able to solve the problem they are investigating).

This study provided more insight into what happens when teachers engage with data in data teams (opening the "black box" of the collaborative inquiry in PLCs; Butler & Schnellert, 2012). However, we need more knowledge on how to support schools in the use of data. The findings show that our theory of action seems to be applicable in practice, but also that teachers and school leaders go through different feedback loops, and that some are able to reach the action and outcomes phase, whereas others are not.

The difference in the functioning of data teams implementing the data team procedure appears to be influenced by several factors. Some teams were able to engage and use data to make improvements, whereas others were not. Further research is needed into which factors enable the use of data in data teams and in schools. As stated by Horn and Little (2010), contextual factors may account for the differences in the practices of different teams. These may refer to school organizational characteristics, such as the role of the school leader (e.g., facilitation, encouragement) (Knapp, Copland, & Swinnerton, 2007; Knapp, Swinnerton, Copland, & Monpas-Huber, 2006), and the degree to which teams have a measurable and shared goal (Datnow et al., 2007; Earl & Katz, 2006; Kerr et al., 2006; Wayman & Stringfield, 2006; Wohlstetter et al., 2008).

We should therefore take into account the organizational context in which data use is taking place, but also the characteristics of the data and data systems available, as well as the influence of policy on data use at the macrolevel, and personal characteristics of the individual data users at the microlevel, as these all interact (Coburn & Turner, 2011, 2012; Marsh, 2012). According to Spillane (2012), "data do not objectively guide decisions on their own" (p. 114). Teachers and school leaders notice and interpret the available data influenced by their expertise, cognitions, authority, and the broader social setting in which they operate (Spillane & Miele, 2007). Moreover, the development and use of new knowledge is not an automatic outcome, but one that develops over time (Cosner, 2011).

Data use is a complex process. Many schools make little productive use of data, gather almost no systematic data beyond what is given to them, and have had very little training or opportunities to gain practical experience with using data to inform classroom decision making (e.g., Wayman et al., 2013). With the data team procedure, we tried to improve data use practices in schools. Although the (tentative) results described here are based on a small pilot in a specific context, this study indicates the data team procedure seems to be effective in two different manners (a) actually solving certain problems within the school and thereby improving education and (b) educating teachers and school leaders in how to use data to improve education (e.g., a form of professional development) and therefore making this procedure sustainable over time.

Sustainability is important, because studies show that a focus on data use is crucial for long-term school improvement. Data use is, for example, identified as a common core characteristic of high-performing schools (Duke, 2015; Ragland, Clubine, Constable, & Smith, 2002; Schaffer, Reynolds, & Stringfield, 2012; Snipes, Doolittle, & Herlihy, 2002). Duke (2015) found that long-term school improvement requires focus, continuous planning, and inquiry into the causes of problems (e.g., low performance) using data. Therefore, the long- and short-term effects (e.g., in terms of student achievement) of the data teams need to be studied further and on a larger scale.

Acknowledgements

The authors would like to thank the Stichting Carmelcollege for making this project possible.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes

1. In The Netherlands, students enter secondary education at the age of 12. Roughly three tracks can be distinguished. The pre-vocational track (lower level) takes 4 years and prepares for secondary vocational education. Senior general secondary education (middle level) prepares for higher professional education and takes 5 years. The third track (higher level) prepares for college directly and takes 6 years.
2. Students choose a profile in which coherent subjects are offered, for example “culture and society”.

Notes on contributors

Dr. Kim Schildkamp is an associate professor in the Faculty of Behavioural, Management and Social Sciences of the University of Twente. Kim’s research focuses on “data-based decision making for school improvement”. She is a board member of the International Congress for School Effectiveness and Improvement (ICSEI) and chair of the ICSEI data use network. She has published widely on the use of data.

Dr. Cindy Poortman is an assistant professor at the University of Twente Institute for Teacher Training and Professional Development. Her main research interests concern teacher professional development in teams.

Dr. Adam Handelzalts is the managing director of the teacher college at the Faculty of Psychology and Education of VU University Amsterdam. His main areas of interest are collaborative (design) work of teachers and teachers’ professional development.

References

Achinstein, B. (2002). Conflict amid community: The micropolitics of teacher collaboration. *Teachers College Record*, *104*, 421–455.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Towards a practice-based theory of professional development. In L. Darling-Hammond & G. Skyes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco, CA: Jossey-Bass.

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, *42*, 231–268. http://dx.doi.org/10.3102/00028312042002231

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15.

Butler, D. L., & Schnellert, L. M. (2012). Collaborative inquiry in teacher professional development. *Teacher and Teacher Education*, *28*, 1206–1220.

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

Capobianco, B. M., & Feldman, A. (2010). Repositioning teacher action research in science teacher education. *Journal of Science Teacher Education*, *21*, 909–915.

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*, 378–398.

Carr, W., & Kemmis, S. (1986). *Becoming critical: Education, knowledge and action research*. Geelong: Deakin University Press.

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, *9*, 173–206.

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

Cochran-Smith, M., & Lytle, S. L. (2009). *Inquiry as stance. Practitioner research for the next generation*. New York, NY: Teachers College Press.

Cosner, S. (2011). Teacher learning, instructional considerations, and principal communication: Lessons from a longitudinal study of collaborative data use by teachers. *Educational Management Administration & Leadership*, *39*, 568–589.

Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.

Datnow, A., Park, V., & Kennedy-Lewis, B. (2013). Affordances and constraints in the context of teacher collaboration for the purpose of data use. *Journal of Educational Administration*, *51*, 341–362.

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, CA: Center on Educational Governance, Rossier School of Education, University of Southern California.

Desimone, L. (2009). Improving impact studies of teacher's professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*, 181–199.

Duke, D. (2015). *Leadership for low-performing schools: A step-by-step guide to the school turnaround process*. Lanham, MD: Rowman & Littlefield.

Earl, L. M., & Katz, S. (2006). *Leading schools in a data-rich world: Harnessing data for school improvement*. Thousand Oaks, CA: Corwin Press.

Earl, L., & Timperley, H. (2008). *Professional learning conversations: Challenges in using evidence for improvement*. Dordrecht: Springer.

Fullan, M. (2000). The three stories of education reform. *Phi Delta Kappan*, *81*, 581–584.

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

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, *8*, 381–391.

Henry, S. F. (2012). *Instructional conversations: A qualitative exploration of differences in elementary teachers' team discussions* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3535056)

Horn, I. S., & Little, J. W. (2010). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal*, *47*, 181–217.

Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the "data driven" mantra: Different conceptions of data-driven decision making. *Yearbook of the National Society for the Study of Education*, *106*(1), 105–131.

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

Inspectie van het Onderwijs. (2011). *De staat van het onderwijs: Onderwijsverslag 2009/2010* [The state of the art of education: Report about education 2009/2010]. Utrecht: Author.

Jimerson, J. B., & Wayman, J. C. (2012, April). *Branding educational data use through professional learning: Findings from a study in three districts*. Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, BC.

Johnson, A. P. (2003). *What every teacher should know about action research* (3rd ed.). Boston, MA: Pearson Education.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvements: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, *112*, 496–520.

Knapp, M. S., Copland, M. A., & Swinnerton, J. A. (2007). Understanding the promise and dynamics of data-informed leadership. *Yearbook of the National Society for the Study of Education*, *106*(1), 74–104.

Knapp, M. S., Swinnerton, J. A., Copland, M. A., & Monpas-Huber, J. (2006). *Data-informed leadership in education*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Krathwohl, D. R. (1998). *Methods of educational and social science research: An integrated approach*. New York, NY: Longman.

Lai, M. K., & Schildkamp, K. (2013). Data-based decision making: An overview. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: challenges and opportunities* (pp. 9–21). Dordrecht: Springer.

Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, *118*, 143–166.

Lomos, C., Hofman, R. H., & Bosker, R. J. (2011). Professional communities and student achievement – A meta-analysis. *School Effectiveness and School Improvement*, *22*, 121–148.

Mandinach, E. B., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for data-driven decision making. In E. B. Mandinach & M. Honey (Eds.), *Data-driven school improvement: Linking data and learning* (pp. 13–31). New York, NY: Teachers College Press.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, *114*(11), 1–48.

Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. *Educational Management Administration & Leadership*, *43*, 269–289.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education: Evidence from recent RAND research*. Santa Monica, CA: RAND.

McNaughton, S., Lai, M. K., & Hsiao, S. (2012). Testing the effectiveness of an intervention model based on data use: A replication series across clusters of schools. *School Effectiveness and School Improvement*, *23*, 203–228.

Nelson, T. H., & Slavit, D. (2007). Collaborative inquiry among science and mathematics teachers in the USA: Professional learning experiences through cross-grade, cross-discipline dialogue. *Journal of In-Service Education*, *33*, 23–39.

Organisation for Economic Co-operation and Development. (2008). *Education at a glance 2008: OECD indicators*. Paris: Author.

Organisation for Economic Co-operation and Development. (2010). *Education at a glance 2010: OECD indicators*. Paris: Author.

Poortman, C. L., & Schildkamp, K. (2012). Alternative quality standards in qualitative research? *Quality & Quantity*, *46*, 1727–1751.

Ragland, M., Clubine, B., Constable, D., & Smith, P. A. (2002). *Expecting success: A study of five high performing, high poverty schools*. Washington, DC: Council of Chief State School Officers and the Charles A. Dana Center at the University of Texas at Austin.
Schaffer, E., Reynolds, D., & Stringfield, S. (2012). Sustaining turnaround at the school and district levels: The high reliability schools project at Sandfields Secondary School. *Journal of Education for Students Placed at Risk*, *17*, 108–127.
Schildkamp, K., & Ehren, M. C. M. (2013). From “intuition” to data-based decision making in Dutch secondary schools? In K. Schildkamp, M. K. Lai., & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 49–67). Dordrecht: Springer.
Schildkamp, K., & Handelzalts, A. (2011, January). *Collaborative data teams for school improvement – Work and sustainability*. Paper presented at the International Congress for School Effectiveness and Improvement, Malmö.
Schildkamp, K., Handelzalts, A., & Poortman, C. (2012, April). *Data teams for school improvement*. Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, BC.
Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*, 482–496.
Schildkamp, K., & Lai, M. K. (2013). Conclusions and a data use framework. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 177–192). Dordrecht: Springer.
Schildkamp, K., & Poortman, C. L. (2015). Factors influencing the functioning of data teams. *Teachers College Record*, *117*(4). Advance online publication. Retrieved from http://www.tcrecord.org/content.asp?contentid=17851
Schildkamp, K., & Teddlie, C. (2008). School performance feedback systems in the USA and in The Netherlands: A comparison. *Educational Research and Evaluation*, *14*, 255–282.
Schnellert, L. M., Butler, D. L., & Higginson, S. K. (2008). Co-construction of data, co-constructors of meaning: Teacher professional development in an age of accountability. *Teaching and Teacher Education*, *24*, 725–750.
Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. Washington, DC: MDCR and the Council of Great City Schools.
Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, *118*, 113–141.
Spillane, J. P., & Miele, D. B. (2007). Evidence in practice: A framing of the terrain. *Yearbook of the National Society for the Study of Education*, *106*(1), 46–73.
Stokes, L. (2001). Lessons from an inquiring school: Forms of inquiry and conditions for teacher learning. In A. Lieberman & L. Miller (Eds.), *Teachers caught in the action: Professional development that matters* (pp. 141–158). New York, NY: Teachers College Press.
Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change*, *7*, 221–258.
Van Veen, K., Zwart, R., Meirink, J., & Verloop, N. (2010). *Professionele ontwikkeling van leraren. Een reviewstudie naar effectieve kenmerken van professionaliseringsinterventies van leraren* [Teacher professional development. A review of studies on effective characteristics of teacher professionalization interventions]. Leiden: ICLON/Expertisecentrum Leren van Docenten.
Verbeek, C., & Odenthal, L. (2014). Opbrengstgericht werken en onderzoeksmatig leiderschap in po en vo [Data use and research leadership in primary and secondary education]. In M. Krüger (Ed.), *Leidinggeven aan onderzoekende scholen* (pp. 67–78). Bussum: Coutinho.
Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, *24*, 80–91.
Wayman, J. C., Jimerson, J. B., & Cho, V. (2012). Organizational considerations in establishing the Data-Informed District. *School Effectiveness and School Improvement*, *23*, 159–178.
Wayman, J. C., Midgley, S., & Stringfield, S. (2007). Leadership for data-based decision-making: Collaborative data teams. In A. B. Danzig, K. M. Borman, B. A. Jones, & W. F. Wright (Eds.), *Learner centered leadership: Research, policy, and practice* (pp. 189–206). Mahwah, NJ: Lawrence Erlbaum.
Wayman, J. C., Spikes, D. D., & Volonnino, M. R. (2013). Implementation of a data initiative in the NCLB era. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in*

education: Challenges and opportunities (pp. 135–153). Dordrecht: Springer. http://dx.doi.org/10.1007/978-94-007-4816-3_8

Wayman, J. C., & Stringfield, S. (2006). Data use for school improvement: School practices and research perspectives. *American Journal of Education*, *112*, 463–468.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the U.S. and abroad*. Dallas, TX: National Staff Development Council.

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educational Psychologist*, *45*, 28–36.

Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, *19*, 21–33.

Whitehead, J., & McNiff, J. (2006). *Action research: Living theory*. London: Sage.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement*, *19*, 239–259.

Yin, R. K. (2003). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education*, *112*, 521–548.

Appendix 1. Observation focus

Theoretical framework element	Explanation	Related observation fragment/summary
Data-based decision making	Main activities and outcomes of each step in the data team procedure in terms of the data use theory of action	In the first meeting, this team decides to focus on the problem of students repeating grades in the senior general secondary education track: on average over 30% per year. The school leader suggests the hypothesis that department advice might relate to grade repetition. This advice relates to the level of secondary education the student should proceed to (ranging from vocational education to pre-university education). Several students ignored the school's advice to continue secondary education at a lower level, and ran into problems in the next grade. The team also discusses other hypotheses. . .
Depth of inquiry	Degree to which the team's conversations expressed higher level thinking skills	• Lack of depth (stories and experiences related to the problem but not based on data): ◦ *In my experience, students just don't like mathematics* ◦ *I just know that students are not motivated these days* • High depth (data team members express higher level thinking skills, such as analysis and reflection): ◦ The team is comparing the results of failing and passing students and is trying to relate the answers to different questions of the questionnaire they administered; they are also discussing their own functioning in relation to the students' opinions about this.
Attribution	Whether teams attributed problems to causes outside the school, to causes at the level of the school, at the level of the students, or at the level of their own functioning	• Outside the school: *the parents make the final decision and they often want a higher level* [for their child] *than is advised* • Own functioning: *Maybe make up something to offer them* [students] *more structure . . .; I think, we do not provide enough direction; . . . It's the only thing you hear all the time: "Sir, please just give me homework for the next time"*.