Data literacy: What do educators learn and struggle with during a data use intervention?

Wilma B. Kippers*, Cindy L. Poortman, Kim Schildkamp, Adrie J. Visscher

University of Twente, Faculty of Behavioural, Management and Social Sciences, Department of ELAN, P.O. Box 217, 7500AE Enschede, The Netherlands

ARTICLE INFO

ABSTRACT

Data literacy is a prerequisite for making data-based decisions. This paper focuses on the extent to which educators develop components of data literacy during a 1-year data use intervention, as well as what they learn and struggle with concerning these data literacy components. In the data use intervention, teams of teachers, school leaders and a data expert use data to solve an educational problem at their school. We employed a mixed-methods approach, combining data from a pre- and post-test data literacy test ($N = 27$), interviews ($N = 12$), evaluations of meetings ($N = 33$), and logbooks. Findings show that educators' data literacy increased significantly. Participants and the data coach indicated that educators had learned, for example, to analyze data with Excel, and to refute misconceptions. Still, there is room for further improvement. For example, educators struggled with formulating a data use purpose that is plausible, sufficiently concrete and measurable.

1. Introduction

Every day educators make educational decisions, for example, about the instructional guidance they provide to their students. Educators often use only intuition and experience to make decisions, which can lead to making incorrect decisions, such as adjusting their instructional practice to the wrong group of students. If educators can make high quality, data-based decisions, that will improve the quality of education and learning in the classroom (Ingram, Louis, & Schroeder, 2004; Schildkamp & Lai, 2013).

One way to promote making informed decisions is by implementing data-based decision making (DBDM). In this research, DBDM is defined as making educational decisions, based on a broad range of possible types of data (Ikemoto and Marsh, 2007; Schildkamp & Lai, 2013). By data, we mean 'information that is systematically collected and organized to represent some aspect of schooling' (Schildkamp & Lai, 2013, p. 10). Examples of data that can be used are test results and structured classroom observations (Ikemoto & Marsh, 2007; Schildkamp & Poortman, 2015). DBDM can help educators to use data in a formative way, by identifying students' learning strengths and weaknesses, and, based on that, taking instructional action that is in line with what students need (Hoogland et al., 2016; Van der Kleij, Vermeulen, Schildkamp, & Eggen, 2015). For example, a decision educators may make based on data is to differentiate in the use of curricular materials by providing different materials to specific subgroups of students (Farrell & Marsh, 2016). DBDM can enhance student achievement,

because it can bridge the gap between students' current learning and students' desired learning outcomes (Lai, Wilson, McNaughton, & Hsiao, 2014; Van Geel, Keuning, Visscher, & Fox, 2016). However, educators need to have the ability to use data to make decisions.

Educators' ability to implement DBDM is referred to as 'data literacy', which we define as educators' ability to set a purpose, collect, analyze, and interpret data, and take instructional action (Hamilton et al., 2009; Lai & Schildkamp, 2013; Mandinach & Gummer, 2016a, 2016b; Van Geel et al., 2016). In the research, there is not as yet a consensus on precisely what knowledge underlies this broad and complex concept. For example, content knowledge can be essential for analyzing and interpreting data, and pedagogical content knowledge can be important for determining appropriate instructional actions in the classroom (Mandinach & Gummer, 2016a, 2016b; Van Geel, Keuning, Visscher, & Fox, 2017).

Little attention is devoted to data literacy in teacher training colleges (Bocala & Boudett, 2015; Mandinach, Friedman, & Gummer, 2015), and in-service educators' data literacy could also be improved (Schildkamp, Karbautzki, & Vanhoof, 2014). For instance, they struggle with interpreting data, and with taking instructional steps accordingly (Datnow & Hubbard, 2015; Kippers, Wolterinck, Schildkamp, Poortman, & Visscher, submitted; Marsh, 2012). Several DBDM interventions have been implemented to help educators develop data literacy, such as the Data Wise Project (Bocala & Boudett, 2015) and the Center for Data-Driven Reform in Education intervention (Carlson, Borman, & Robinson, 2011). The findings in terms of effects of data use

---

* Corresponding author.
E-mail address: w.b.kippers@utwente.nl (W.B. Kippers).

interventions are, however, mixed (Marsh, 2012).

Moreover, research on the development of data literacy is scarce. Assessment literacy (i.e., the ability to use assessment data only) has often been studied instead of the broader data literacy concept (i.e. the ability to use different types of data) (Mandinach & Gummer, 2016a; Van Geel et al., 2017). The field has mainly focused on stating that data literacy is a key prerequisite for DBDM, has defined the data literacy concept, and has identified educators' current abilities regarding data use (Hamilton et al., 2009; Kerr, Marsh, Ikemoto, Dariley, & Barney, 2006; Mandinach & Gummer, 2016a; Means, Chen, DeBarger, & Padilla, 2011).

In this study, we aimed to obtain more detailed insight into the extent to which educators develop several *data literacy components* during a data use intervention, by studying their ability to set a purpose, collect, analyze, and interpret data, and take instructional action. In this data use intervention, teams of teachers, school leaders and a data expert learned in 11–13 meetings and 1–2 workshops about data use, by trying to solve an educational problem at their secondary school with support from an external data coach. This study provides an overview of which data literacy components educators develop, or not, and what educators learn and struggle with concerning a number of data literacy components.

2. Theoretical framework

Several data literacy components can be distinguished: set a purpose, collect data, analyze data, interpret data, and take instructional action (Hamilton et al., 2009; Lai & Schildkamp, 2013; Mandinach & Gummer, 2016a, 2016b; Van Geel et al., 2016). Educators use these five data literacy components to implement DBDM. The data use intervention is developed by the University of Twente to support educators with implementing DBDM and consists of eight steps: 1. Problem definition, 2. Formulating hypotheses or questions, 3. Data collection, 4. Data quality check, 5. Data analysis, 6. Interpretation and conclusion, 7. Implementing improvement measures, and 8. Evaluation (Schildkamp & Ehren, 2013). Educators use the five data literacy components several times when following the eight steps of the data use intervention. For example, the ability to collect data is applied by educators in the problem definition (step 1 of the data use intervention), data collection (step 3 of the data use intervention), and evaluation (step 8 of the data

use intervention).

Fig. 1 presents our theory of action. It is a framework linking the components of data literacy (shown in bold text) with the data use intervention (shown in italic text). It is based on studies about data literacy and DBDM (Bocala & Boudett, 2015; Coburn & Turner, 2011; Earl & Katz, 2006; Hamilton et al., 2009; Lai & Schildkamp, 2013; Mandinach & Gummer, 2016a; 2016b Marsh, 2012; Means et al., 2011; Van Geel et al., 2016). In the following section, we describe more precisely how the data use intervention and the concept of data literacy relate to each other.

2.1. Set a purpose

The ability to set a purpose for using data is a key component within the concept of data literacy (Hamilton et al., 2009; Earl & Katz, 2006; Lai & Schildkamp, 2013; Mandinach & Gummer, 2016a, 2016b; Means et al., 2011; Van Geel et al., 2016). Educators make use of this ability in the first and second step of the data use intervention. In the first step of the data use intervention, educators learn to set a clear purpose for using data, by thinking about the reason for using data, such as gaining insight into the retention rates in the 3rd year of secondary education or the final examination results. In teams of 4–6 teachers, 1–2 school leaders and a data expert from the same school, educators collaboratively formulate a concrete and measurable *problem definition*, and include the shared goals they want to achieve. These goals are always focused on their own school context, and the team is supported by the coach. For example, 'We are dissatisfied with the% of students failing mathematics in grade 9, because we have been coping with an average percentage of 20% failing students for the past three years. We would like to achieve no more than 15% of students failing mathematics next year and fever than 10% within two years.' Next, in the second step of the data use intervention, educators will *formulate hypotheses* (quantitative research) *or questions* (qualitative research) regarding the underlying causes of the problem, to capture the purpose for using data more specifically. For example, related to the problem definition stated above, 'At least 30% more students from elementary schools A, D and F fail in mathematics in the first year of upper secondary school than students from the other elementary schools.'
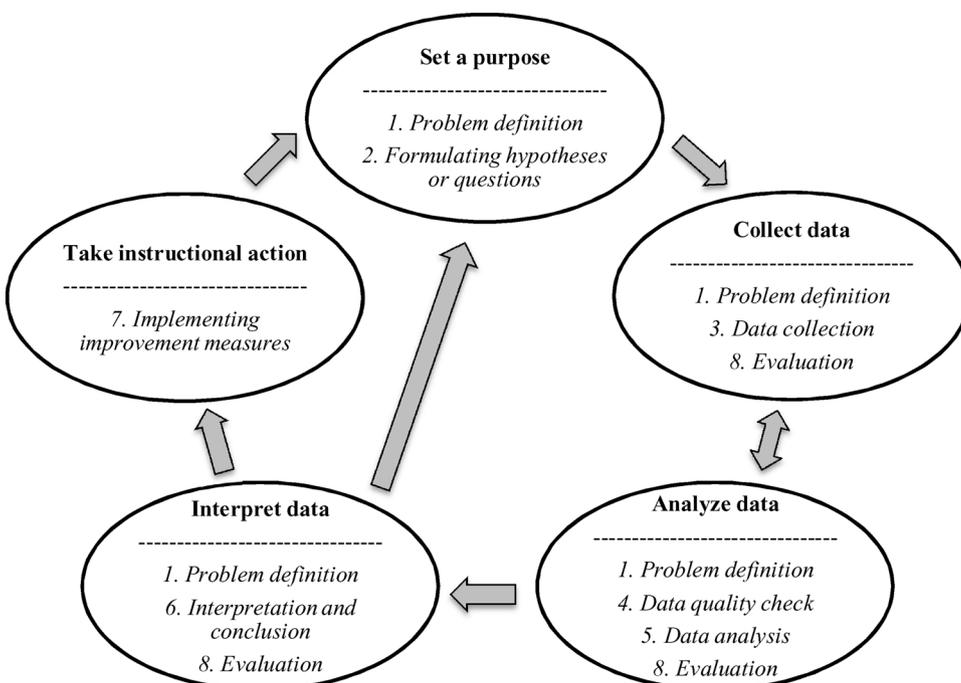

Fig. 1. Links between the concept of data literacy (bold) and the data use intervention (italics).

2.2. Collect data

Being data literate is also reflected in the ability to collect data (Author et al., 2016a, 2016b; Authors, 2013a, 2013b; Bocala & Boudett, 2015; Coburn & Turner, 2011; Earl & Katz, 2006; Hamilton et al., 2009; Mandinach & Gummer, 2016a, 2016b; Marsh, 2012; Means et al., 2011). Educators make use of their ability to collect data when following the first, third and eighth step of the data use intervention. They collaboratively learn to collect and generate multiple types of quantitative and qualitative data, supported by the coach. In the first step, they collect data to ascertain the scope of the problem, and to determine goals. In the third step, they collect data to find out whether the hypothesis regarding the underlying cause of the problem is correct or false, or to find answers to the question posed. In the eighth step, they collect data to evaluate whether the causes of the problem have been eliminated, and whether goals have been accomplished. In these three steps, they can collect student achievement data, or interview data, for example.

2.3. Analyze data

Data literate educators also have the ability to analyze data (Bocala & Boudett, 2015; Coburn & Turner, 2011; Earl & Katz, 2006; Hamilton et al., 2009; Lai & Schildkamp, 2013; Mandinach & Gummer, 2016a, 2016b; Marsh, 2012; Means et al., 2011; Van Geel et al., 2016). Educators can develop this ability to analyze data in the first, fourth, fifth and eighth step of the data use intervention. The coach supports educators in collaboratively learning how to analyze data, such as organizing and prioritizing data, and using statistics (e.g., calculating the mean and the standard deviation). In the data use intervention, educators always first check the quality of data before analyzing the data. In the first step, they check the quality of the data collected regarding the problem (e.g., by checking whether data on several cohorts are collected), and then analyze these data, to formulate evidence-based goals. For example, they organize the collected data in a data table in Excel. In the fourth step, they determine the quality of the data collected about the underlying cause of the problem (hypothesis or question), for example, by determining whether data are missing. If in the fourth step educators determine that the data about the hypothesis or question are of inadequate quality, they decide how to improve the quality, for example, by collecting additional data, and therefore, returning to the third step of the intervention (see Fig. 1). If data about the hypothesis or question are of good enough quality, educators analyze these data in step five of the data use intervention. For example, they create a graph to summarize quantitative data through descriptive analyses in Excel. In the eighth step, educators check the quality of the data collected with regard to the instructional action that was taken (e.g., are the data relevant?), and then analyze these data, to evaluate whether causes of problems have been eliminated, and whether goals have been reached. For example, they create a data table by conducting descriptive analyses in Excel.

2.4. Interpret data

Another important component of the concept of data literacy is the ability to interpret data (Coburn & Turner, 2011; Earl & Katz, 2006; Hamilton et al., 2009; Lai & Schildkamp, 2013; 2016a, 2016b; Marsh, 2012; Means et al., 2011; Van Geel et al., 2016). Educators make use of their ability to interpret data in the first, sixth and eighth step of the data use intervention. They collaboratively learn to transform data into information, by reading and interpreting tables and graphs, and interpreting information from diverse analyzed qualitative data, in order to draw a conclusion. The coach supports educators with interpreting the data. In the first step, they interpret the data regarding the problem, for example, by interpreting the mean of student achievement data, and use this information to formulate their specific data-based goals. In the

sixth step, they interpret data concerning the hypothesis or question, for example, by checking the strength of the correlation between the data, and by concluding whether the hypothesis can be accepted, or should be rejected. They can also check and conclude whether the summary of the interview data has answered their question. If a hypothesis turns out to be false, or no question has been answered, new hypotheses need to be tested, or new questions need to be posed (back to step 2 of the data use intervention, see Fig. 1). If a hypothesis is correct, or a question is answered, educators can take instructional action by proceeding to step 7 of the data use intervention. In the eighth step, educators interpret data regarding the instructional action that was taken, for example, by comparing student outcomes before and after an action was taken, and they use this information to evaluate whether the action had the intended effect, or not. If goals have not been accomplished, they improve the action or think of another instructional action to take (step 7 of the data use intervention, see Fig. 1). This interpretation process continues until the goals are accomplished. In that case, the team can continue with setting a new purpose, and therefore start with a new "step 1".

2.5. Take instructional action

Being data literate also refers to the ability to take instructional action (Bocala & Boudett, 2015; Coburn & Turner, 2011; Hamilton et al., 2009; Lai & Schildkamp, 2013; Mandinach & Gummer, 2016a, 2016b; Marsh, 2012; Means et al., 2011; Van Geel et al., 2016). Educators can develop this ability in the seventh step of the data use intervention. The coach supports educators in collaboratively learning to build knowledge, by combining the information from data regarding the hypothesis or question with educators' expertise, and by learning to apply this knowledge by taking instructional action to try to reach their goals. They gather ideas for reaching the goals, and formulate concrete actions. For example, they can decide to purchase digital software that students with poor results may use in the classroom to have more practice with specific subject-matter. Moreover, educators create an action plan, in which they agree on who will implement the measures, and when, for example. Furthermore, they communicate the actions to others in the school, because, among other reasons, their colleagues might also be partially responsible for implementing an intended instructional action.

2.6. The data use intervention

The data use intervention is based on several criteria for effective teacher professional development, frequently mentioned in the scientific literature, such as collaboration between colleagues (Desimone, 2009; Marsh, 2012; Timperley, 2008; van Veen, Zwart, & Meirink, 2011) (i.e., by learning about data use in a team of 4–6 teachers, 1–2 school leaders, and a data expert from the same school), active leadership (Marsh, 2012; Timperley, 2008) (i.e., by the school leader's participation in the team), time to learn and change (Desimone, 2009; Marsh, 2012; Timperley, 2008; van Veen et al., 2011) (i.e., a one-year intervention including 11–13 meetings of 90 min each, and 1–2 voluntary data analysis workshops of 4 h each), and support from an expert (Marsh, 2012; Timperley, 2008) (i.e., all teams are supported by the same experienced data coach from the university, who visits them every three weeks for a meeting to work on the steps of the data use intervention by using an 87-page guiding manual, including worksheets). In the guiding manual, each step of the data use intervention is introduced in a separate chapter. After the introduction in which the goal of the step is explained, concrete activities are formulated which have to be followed by educators through filling out worksheets. For example, regarding step 3 of the data use intervention two activities are formulated: "determine required data" and "create a data table". Questions that need to be answered in the worksheets are: "What data are needed to find out whether the hypothesis regarding the underlying cause of the problem is correct or false? Where can we find the data?

Who is going to collect the data? How will we create the data table?". Besides, examples are given of fictional data teams who also followed the activities, and in some chapters a list of possible data sources or possible improvement measures is provided.

In conclusion, in the data use intervention, educators are trained in using data to make decisions to improve the quality of education. We expected to find that educators developed their data literacy to some extent for all components of our data literacy framework during this data use intervention. This leads to the research question: *To what extent do educators show development regarding the data literacy components during the data use intervention, and what do they learn and struggle with concerning these data literacy components?*

3. Method

3.1. Context

We conducted this study in the context of Dutch secondary education. Dutch schools have the freedom and autonomy to choose their own principles (regarding religion, ideology and pedagogy) on which they base the education they provide (Ministry of Education, Culture & Science, 2000). However, the Dutch Inspectorate of Education holds schools accountable for using data to make educational decisions. By 2018, at least 90% of Dutch primary and secondary education schools should be engaged in using data to guide their education (Verbeek & Odenthal, 2014). In the Dutch context, there is no national curriculum and teachers are free to develop assessments, for example, based on the curriculum determined by the school. National standardized assessments are taken only at the end of secondary education. Important data sources available within Dutch secondary schools include the results of those national standardized assessments, curriculum-based assessments, satisfaction questionnaires for students, and inspection data (Schildkamp & Kuiper, 2010). Examples of data that can be used in the data use intervention are final examination results, and data collected by means of student interviews, for example, concerning students' opinions about the quality of teachers' instruction.

3.2. Research design

This research is part of a larger data use project funded by the Dutch Ministry of Education, in which 223 secondary schools in the Netherlands participated. We informed these 223 schools about the opportunity for 6 schools to participate in the data use intervention. Nine out of these 223 schools voluntarily signed up to participate in our research. A criterion sample (Onwuegbuzie & Leech, 2007) of six schools was drawn, based on (1) that they had not already participated in the data use intervention, and (2) that they were distributed throughout the country. Characteristics of these schools are presented in Table 1.

We used a single-group pre-post research design (Field, 2013) and employed a mixed-methods approach to study the development of educators' data literacy during the data use intervention; see Table 2. Prior to and after the intervention, educators took a data literacy test. After the intervention, interviews were conducted with educators about what they had learned and struggled with regarding data literacy. Each meeting, the coach used a logbook to describe what educators said they learned concerning data literacy during that meeting, and she evaluated the meetings with educators halfway through the intervention period with respect to what they had learned so far about data literacy.

3.3. Respondents

From September 2015 till October 2016, two teams focused on the subject of English, two teams on Dutch, and two teams focused on mathematics; see Table 3. Twenty-seven educators participated in a data literacy pre- and post-test. Furthermore, 33 educators participated

in group evaluations of meetings, to provide information on the level of data literacy they had achieved half-way through the data use intervention period. A total of twelve educators were selected based on criterion sampling (Onwuegbuzie & Leech, 2007) for individual interviews at the end of the intervention period. Triangulating different instruments made it possible to gain further and more in-depth insights into what educators learned and struggled with concerning data literacy. For each team, the two educators who attended the most meetings were selected for the interviews, because attending these meetings meant that they at least had a chance to develop data literacy during the data use intervention to some extent.[1]

3.4. Instruments

3.4.1. Data literacy test

The data literacy test was based on (1) an already existing knowledge test related to the eight steps of the data use intervention (Ebbeler, Poortman, Schildkamp, & Pieters, 2017), and (2) recent research into data literacy (e.g., Mandinach & Gummer, 2016a), see Appendix A. The data literacy test was constructed by using an extensive protocol on how to construct open-ended test questions, such as adding answer restrictions to the questions (e.g., 'answer in a maximum *length of five sentences*'), using standard formulations for short-answer questions (e.g., '*identify* two data sources' or '*identify* two quality criteria of data'), and developing questions with others (e.g., the data literacy test was developed and discussed with researchers who also conduct research into DBDM, the data coach, and a teacher with teaching experience in secondary education) (Erkens, 2011). This paper-and-pencil test included twelve open-ended questions, and educators could achieve a maximum score of 25 points. They had to complete the test in 30 min. All data literacy test items were based on our theoretical framework. In Table 4, examples are given of the linkage of each data literacy component with the data use intervention steps and the data literacy test items. The data literacy test items were in Dutch, as the respondents were Dutch. The data literacy test was taken by educators at the start and at the end of the intervention[2] ($N = 27$).

3.4.2. Interview scheme

At the end of the intervention period, one-hour individual interviews were conducted and recorded. Twelve educators were interviewed using an interview scheme based on the theoretical framework. The interviews focused on various aspects of educators' data literacy acquired during the intervention. The main question of the interview was about what educators learned and struggled with concerning DBDM. Examples of these questions are: 'What, if anything, have you learned about data-based decision making during the data use intervention?' and 'What, if anything, have you learned about analyzing data during the data use intervention?'. The interview scheme was first tested with an educator who had already worked with the data use intervention in another project, after which adjustments were made regarding the formulation of the interview questions. The quotes from the respondents in the interviews were translated into English for use in this paper, as all interviews were held in Dutch.

3.4.3. Meeting evaluations

Half-way through the data use intervention period, the coach evaluated the intervention with each team through asking questions.

---

[1] For one team (school F), three educators attended most of the meetings. Therefore, the external data coach supported selecting interviewees from these three by choosing who was most capable of explaining what they had learned and struggled with during the data use intervention.

[2] Two educators joined their team later, in January 2016 and in March 2016. At that time, they participated in the data literacy pre-test. Four educators completed the pre-test outside of the team meeting. Seven educators completed the post-test outside of the team meeting.

Table 1
Characteristics of secondary schools (student ages 12–18).

School	School size	Urbanization	Denomination	Quality label Inspectorate[a]
A	Large (> 1000 students)	Rural	Catholic school	Good
B	Large (> 1000 students)	Urban	Catholic school	Good
C	Medium (500–1000 students)	Urban	Private school[b]	Good
D	Medium (500–1000 students)	Urban	Public school[c]	Weak
E	Large (> 1000 students)	Urban	Catholic school	Good
F	Large (> 1000 students)	Urban	Public school	Weak

[a] We used the judgment by the Dutch Inspectorate of Education, https://zoekscholen.onderwijsinspectie.nl/.
[b] This Dutch private school is characterized as an independent school that is based on neither a specific educational vision nor a specific religion.
[c] A Dutch public school is characterized as a government-supported school that is based on neither a specific educational vision nor a specific religion.

Table 2
Outline of the study.

Prior to intervention	During intervention		After intervention
September 2015	School Year 2015–2016	March 2016	October 2016
Data literacy pre-test	Log entries for meetings	Meeting evaluation	Data literacy post-test Interview

There were also some questions about what educators had learned concerning data literacy, after having participated in the data use intervention for half a year. These questions were based on the theoretical framework, such as: 'What have you learned so far, during the data use intervention, about data-based decision making?'. The questions evaluating the meetings were developed and discussed with the coach, and the evaluation was recorded. The quotes were translated into English for use in this paper.

Table 3
Description of the teams.

School	Subject	# of meetings	# of teachers	# of school leaders	# of internal data experts
A	English	13	3 English teachers 1 special needs teacher	1	–
B	English	11	3 English teachers	2	–
C	Dutch	12	3 Dutch language teachers 1 physics teacher 1 mathematics teacher	1	1
D	Dutch	11	2 Dutch language teachers 1 English teacher 1 physics teacher 1 physical education teacher	1	1
E	Mathematics	11	2 mathematics teachers 1 physics teacher 1 biology teacher	1	1
F	Mathematics	12	2 mathematics teachers 1 Dutch language teacher 1 physics teacher	2	–

Table 4
Link between data literacy, data use intervention, and data literacy test items.

Data literacy	Data use intervention	Data literacy test items
Set a purpose	Problem definition, and formulating hypotheses or questions	Three test items. For example: The first step for a data team is to define a problem for the team to work on. Formulate a problem statement based on Table below, by completing the following sentences. 'We are dissatisfied with…' 'We would like to achieve…'
Collect data	Problem definition, data collection, and evaluation	Three test items. For example: A data team that wants to work on the problem of poor mathematics results in the first year of secondary education thinks this problem is mainly caused by the poor quality of arithmetic lessons in the primary schools their students are from. Identify two data sources this team could use to investigate this cause.
Analyze data	Data quality check and data analysis	Two test items. For example: A data team wants to investigate whether a course for teachers about differentiation has increased student achievement for the subject of Dutch. In Table below you can see the results for four students, before and after the teachers' course. What do their grades tell you? [answer in a maximum length of five sentences].
Interpret data	Interpretation and conclusion, and evaluation	Two test items. For example: The percentages of students with an unacceptable score per subject domain have been comparable for the past three years, see Table above. What conclusions can you draw based on these percentages? [answer in a maximum length of five sentences].
Take instructional action	Implementing improvement measures	Two test items. For example: A data team has finished the first six steps of the intervention. The analysis shows that teachers do not provide feedback and this is one of the causes of disappointing results for the subject of English in the fifth year of pre-university education. The data team wants to continue with step 7: implementing improvement measures. Identify two concrete measures to ensure that teachers provide more feedback to students.

3.4.4. Logbook entries for meetings

All team meetings were described in a logbook by the external data coach. The logbook was based on the theoretical framework, and was developed with the coach. Examples of topics in the logbook were: 'What have educators learned about data-based decision making?', and 'What have educators learned about taking instructional action?'. The coach's descriptions in the logbooks were translated into English for use in this paper.

3.5. Data analysis

The interviews and meeting evaluations were transcribed verbatim. Based on the theoretical framework, a coding scheme with five codes was developed, see Appendix B. For coding the interview data, meeting evaluation data, and logbook data, the program Atlas.ti was used. Codes were, for example: 'analyze data' and 'take instructional action'. We used descriptive analyses to report on each component of data literacy. For each code, we first summarized the evaluations of the meetings. Next, we summarized teachers' answers during the interviews for that code. Subsequently, we summarized the descriptions in the logbooks related to that code. For each code, we combined these three summaries, to report on what educators had learned and struggled with, with respect to each component of data literacy. Then we moved on to the next code.

A scoring sheet was used to score educators' answers on the data literacy test. To analyze whether educators' data literacy developed during the data use intervention, a paired samples t-test was used (Field, 2013). To investigate the extent to which educators had (not) developed certain of the components of data literacy, the percentage of correct answers was calculated for each data literacy component on the pre-test and post-test, and examples were given of the types of mistakes educators had made.

3.6. Reliability and validity

All interviews and evaluations of meetings were audio-taped and transcribed. The interview data, meeting evaluation data, logbook data, and data literacy test data were coded by a researcher. To code the data literacy test, a protocol was used for scoring the educators' answers, to avoid bias (Erkens, 2011). Inter-rater reliability was calculated for the codings (Poortman & Schildkamp, 2012). Approximately 10% of the individual interview data, meeting evaluation data, and logbook data were first also coded by another researcher, and the codings of both researchers agreed substantially, with a Cohen's Kappa of .71. Approximately 20% of the data literacy test data were also coded by this second researcher, and the codings of both researchers agreed substantially, with a Cohen's Kappa of .68 (Eggens & Sanders, 1993). Subsequently, the remainder of the data were coded by one researcher.

The content of the data literacy test, interviews, meeting evaluations and logbooks all closely relate to the theoretical framework presented here, to improve internal validity. All four instruments were developed and discussed with researchers who are experienced with research into DBDM, and the data coach (they commented on the formulation of the items and questions). A protocol regarding construction of open-ended test questions was used to construct the data literacy test, and the interview scheme was tested with an educator. To improve construct validity, all five components of the data literacy concept were measured by both the data literacy test and interviews.

Differences between teams were analyzed by conducting a one-way between-subjects ANOVA (Field, 2013), to promote external validity. We used educators' average scores on the data literacy test to compare the effects on educators' data literacy of participating in different teams. There was no significant difference for the scores on the data literacy test between the teams [$F(5, 21) = .530, p = .751$]. The procedures used to arrive at the research results are described as precisely as possible in this paper, and the data literacy test and the coding

scheme are included in the appendix. We strived to make it possible for readers to compare our research methodologies and results with their own research, to promote analytical generalization (Poortman & Schildkamp, 2012). Overall, this study aimed to *explore* the extent to which educators developed a number of data literacy components, and what they had learned and struggled with, which can be expanded in large-scale research.

4. Results

4.1. Data literacy

On the data literacy test, educators scored significantly higher on the post-test ($M = 11.2$; $SD = 3.03$) than on the pre-test ($M = 9.3$; $SD = 2.66$), with $t(26) = -3,113$, $p = 0.004$, and a medium to large effect of $d = 0.71$ (Field, 2013). Still, a score of 11.2 on the post-test is less than half of the maximum score (25) educators could achieve. In the following sections, for each component of data literacy we will discuss the percentages of correct answers on the pre- and post-test, the types of mistakes educators made on the pre- and post-test, and what educators (in their opinion or in the opinion of the data coach) learned and struggled with in terms of the data literacy components.

4.1.1. Set a purpose

The percentage of correct answers for 'setting a purpose' was 30% on the pre-test and 29% on the post-test; see Table 5. By zooming into on the educators' answers on the pre- and post-test, we see that often no plausible purpose had been set. When educators were asked to formulate a hypothesis regarding the cause of a problem, some of them formulated a problem definition or question instead. When they were asked to formulate a qualitative research question, some of them formulated a 'yes, no' question, such as *'Are the students in the correct educational track?'* instead of a 'how, what, why' question, such as 'How do students experience their educational track?'. Although two educators described in an interview and a meeting evaluation that they had learned the difference between a problem definition and a hypothesis, educators more often mixed up the problem definition and hypothesis on the post-test than on the pre-test; this mistake was also reported by the data coach in the logbooks. For example, the data coach felt she had to provide a lot of guidance in formulating hypotheses and questions, because educators found it hard to come up with a hypothesis or question, or because they already wanted to proceed with collecting data. In the meeting evaluations, some educators indicated having learned to start with setting a purpose before collecting the data, because otherwise you get lost in the volume of data, which seems to be in contrast to what the data coach reported educators did. Some educators described in the interviews that they had learned to include in the problem definition what they are dissatisfied about and what they want to achieve with their students. Also, some stated that they had learned to involve colleagues in setting a purpose, to expand possible purposes and to have a shared purpose for using the data.

We also found that often no completely measurable purpose had been set in the responses on the pre- and post-test. On the post-test, an educator reported: *'We are dissatisfied with the English results in grade 3. We would like to achieve an improvement in the transition from grade 2 to*

Table 5
Percentage of correct answers on the data literacy test for each data literacy component.

	Pre-test	Post-test
Set a purpose	30%	29%
Collect data	49%	61%
Analyze data	29%	41%
Interpret data	36%	47%
Take instructional action	64%	76%

grade 3.' In this example, the magnitude of the problem and the time period were not stated (what was the average student grade for the English results in grade 3 for the past three years?), and no goal was reported (what student grades did you want to achieve within what time period?). In line with this, in the logbook entries the data coach often wrote that educators (said they) struggled with formulating a measurable hypothesis, for example, because they found it hard to specify concepts such as 'low' and 'many'; and in her opinion, she had to provide a lot of assistance, or even had to formulate a measurable hypothesis for them. The interviews showed mixed results. On the one hand, many educators reported that they had learned to formulate a measurable problem definition or hypothesis by including numbers or percentages. *'What is disappointing? Is it that 10% of students have unacceptable scores? Is it that 50% of students have unacceptable scores?'* On the other hand, they indeed described that they struggled with stating a measurable problem definition or hypothesis. *'I already mentioned what I now know more about than before [setting a measurable hypothesis], but that still does not mean that I can do it.'*

Setting a purpose that is completely concrete, i.e. including subject area, educational track and grade of students, was also a struggle for educators. On the pre- and post-test, the target group educators were focusing on, the educational track and grade students were in, or the subject area were not always mentioned. For example, as part of a problem definition an educator formulated: *'We are dissatisfied with the results in grade 9 of pre-university education'*, in which the subject area English is missing. This is in line with the data coach's logbook entries, in which she wrote that educators (said they) struggled with formulating a concrete hypothesis, and that in her opinion, she had to provide many examples and guidance on how to state a concrete hypothesis. Educators succeeded better in setting a concrete purpose on the post-test than on the pre-test. This was not the case for setting a plausible or measurable purpose. Educators described in the interviews and in some meeting evaluations that they had learned to specify their purpose by including the target group. *'You have to formulate very clearly which group of students, what grade, and what subject area you are focusing on.'* Some also stated that they had learned the importance of defining concepts (e.g., reading skills) when setting their purpose, as one otherwise does not know what to measure.

4.1.2. Collect data

The percentage of correct answers for 'collecting data' was 49% on the pre-test and 61% on the post-test. In the interviews, educators mentioned that they learned that there are a lot of existing data that can be collected from the student monitoring system, but some also described that they had not collected data since they did not have permission to access the data in the student monitoring system. On the pre- and post-test, we found that the collection of data was often not described concretely enough. The educational track, the grade students were in, or the subject area for they would like to collect data were sometimes missing. For example, an educator formulated: *'By proving this with data.'* in which 'grade 9' and 'pre-university education' are missing.

Moreover, despite the fact that in interviews, logbook entries and meeting evaluation responses educators felt they had learned that you need to collect enough, relevant and recent data to have high quality data, on the pre- and post-test, some educators did not mention data that were relevant for the specific test item, or did not describe collection of data for several cohorts. For example, on the post-test, one educator mentioned results in grade 2 of secondary education, although the test item was about data sources concerning primary education. In some logbook entries, the data coach reported that educators said they struggled with collecting relevant data that were linked to the hypothesis, or that they said that they did not know that data should be collected on several cohorts.

We found mixed results concerning educators' ability to generate data. On the one hand, interview and logbook results showed that

educators indicated having learned how to develop student questionnaires and student interviews. On the other hand, interviews and logbook entries revealed that some educators (said they) struggled with defining concepts in developing questionnaires, that some found it hard to develop interview schemes, and that some had difficulties with probing additional questions during interviews with students.

The interview, meeting evaluation, and logbook results revealed that some educators indicated having learned the importance of storing data in a more accessible way. During the intervention, they struggled with finding the data they wanted, because for several years, data from the same student tests had been named differently in their student monitoring system, or because no data had been stored. Some also stated that they could distinguish between quantitative data and qualitative data, and see the benefits of both. The interviews and logbook entries showed that some educators, in their opinion, now had the skills to export the data from their student monitoring system, and to develop a summary table to display the data accordingly. *'Filter the correct grade, class of students, subject area and tests, and then you can export the data'.* Some also stated that they had learned to make arrangements with colleagues, such as who is collecting the data, and when.

4.1.3. Analyze data

The percentage of correct answers on 'analyzing data' was 29% on the pre-test and 41% on the post-test. One mistake that was made less often on the post-test than on the pre-test, but that was still often made on both, regarded the lack of clear description of quality criteria for data, such as reliability and validity. For example, an educator formulated *'Data have to provide an answer to the research question. Data have to be collected from participants involved in the problem.'* in which reliability and validity are not mentioned or described as quality criteria (explicitly). This is in line with the logbook entries, in which the data coach described that she often had to help remember the teams to check the quality of the data before analyzing it. However, on the post-test, educators more often succeeded in describing one or both of the quality criteria in detail. The interview results showed that educators, in their opinion, had learned to check the quality of the data they had collected, for example, by checking whether enough relevant data were collected, whether data were missing, and whether data had been properly entered into the summary table. *'You have to check the data: are the data reliable and valid?'*

The interviews and logbooks showed mixed results concerning educators' ability to analyze data. On the one hand, some interviews and logbook entries revealed that educators find it hard to analyze data, and that in some teams, only one or two educators analyzed the data most of the time. On the other hand, many interviews and logbook entries showed that educators felt that they had learned how to use statistical functions in Excel, such as calculating the mean, standard deviation, minimum, maximum, median, and correlation. *'I learned to use Excel, and to create sheets and graphs. I was not capable of doing that in the past.'* On the pre- and post-test, educators were asked to analyze data using paper instead of Excel, and results revealed that educators often struggled with this. For example, the distribution of the data had not always been discussed, or no number of unacceptable grades had been calculated. Sometimes no mean had been calculated for the data. The logbook entries and interview results for one team showed that they had learned to analyze interview data qualitatively, such as coding what respondents said instead of (only) counting how often something was said, but that they sometimes were tended to code what they thought students *meant* to say instead of coding what was *really* said, which led to inaccurate analyses.

4.1.4. Interpret data

The percentage of correct answers for 'interpreting data' was higher on the post-test (47%) than on the pre-test (36%). The interviews and a logbook entry showed that educators had learned to stick to their purpose when interpreting data and drawing conclusions. However,

one mistake that was made more often on the pre-test than on the post-test, but that was still sometimes made on both tests, was lack of interpretation of data concerning an action that had been taken. For example, educators did not describe that data must be interpreted and compared before and after an action has been taken, to evaluate whether the action led to reaching the goal, or not.

On the pre- and post-test, we also found that sometimes no concrete data were interpreted, and therefore there was no link between the interpretation of the data and the subject area for which the action had been taken. An educator answered on the post-test: *'To start with, put the student grades from before the action and after the action next to each other to see what the results are.'* In this example, the subject area is missing (student grades from which subject area are interpreted?).

In the interviews, logbook entries, and in one meeting evaluation, educators described now having the ability to give meaning to data, and to take their time drawing conclusions, without using intuition. *'I did not know before that when the data points in a scatterplot are spread all over, there is no correlation, and that when they approach a straight line, there is a correlation.'* However, on the pre- and post-test, sometimes an analysis was described instead of an interpretation, or incorrect conclusions were drawn that did not follow from the analysis. In addition, in some interviews and logbook entries educators described that they struggled with not using intuition when interpreting data, such as when summarizing interview data.

In the interviews and in some logbook entries and meeting evaluations, educators stated that they had learned not to take assumptions as the truth, and to refute misconceptions. However, on the pre- and post-test, sometimes assumptions were made instead of interpretations, and in the interviews some also described that they find refuting misconceptions hard to do, as assumptions are deeply rooted.

In some interviews and meeting evaluations, educators stated that they had learned the importance of interpreting data with other educators in the team, or other colleagues, to combine different insights and draw a common conclusion.

4.1.5. Take instructional action

Educators scored highest on 'taking instructional action' on the pre- and post-test. Their percentage of correct answers was higher on the post-test (76%) than on the pre-test (64%). In the meeting evaluations and logbook entries, no research results emerged concerning taking instructional action. In some interviews, educators stated that they had learned to take an action that is linked to the purpose of using data, but on the pre- and post-test (but more often on the pre-test than on the post-test), not all educators could describe an action that was related to the purpose of using data. For example, on the post-test an educator stated that teachers should take part in professional development programs without mentioning what such programs should be about.

Moreover, in some interviews educators said that they had learned to implement an action that is concrete, but on the pre- and post-test sometimes they did not describe taking a concrete action (although it was linked to the data analysis and interpretation), such as *'Set priorities. First address writing skills.'*. Also, the "action" described was not always an instructional action at all, such as *'Compare with other schools or national results.'*.

In most interviews, educators described learning that the action does not necessarily have to lead to huge changes in the classroom, and that the action should be feasible. *'The action does not have to be drastic (…) small things can also have effects.'* In addition, many educators indicated having learned to involve and communicate with colleagues throughout the process of data use, to obtain ideas about possible actions to take, to ensure that the action is supported by all educators in the school, and to ensure that the action will actually be implemented (by them) in the classroom.

5. Conclusions and discussion

By making data-based decisions, educators can use data in a formative way to improve education. Data literacy plays a key role in successfully implementing DBDM. Several studies have focused on the implementation of DBDM interventions, have reported the effects of DBDM interventions on educators' data literacy, and focused on the data literacy concept (e.g., Bocala & Boudett, 2015; Carlson et al., 2011; Ebbeler et al., 2017; Hamilton et al., 2009; Means et al., 2011; Van Geel et al., 2017). However, no studies have investigated the extent to which educators developed in terms of each of the five data literacy components (setting a purpose, collecting data, analyzing data, interpreting data, and taking instructional action) during a data use intervention; what they learn and struggle with concerning each data literacy component also has not been studied before.

In the current study, a mixed-methods approach was used to fill this gap. We found that educators' data literacy increased significantly after a data use intervention (i.e., the data literacy test showed a medium to large effect). Examples of what educators had learned, in their opinion or in the opinion of the data coach, are the importance of storing data (correctly) in the student monitoring system, using Excel to analyze data, and refuting misconceptions. Still, there is room for further improvement, because educators' score on the data literacy post-test was less than half of the maximum score educators could achieve. Examples of what educators (said they) still struggled with are setting a plausible, and sufficiently concrete and measurable purpose, reporting the collection of data concretely enough, and describing data quality criteria clearly. The fact that educators did not study for the test, that during the meetings with the coach the guiding manual (which includes explicit data literacy principles) was not always used by all educators, and that the intervention was shorter than necessary for developing data literacy might be explanations for low scores on the data literacy post-test (Timperley, 2008; van Veen et al., 2011). The knowledge, skills, and attitudes of the data coach, the teachers and the school leaders are also important factors that might have influenced the (lack of) development of educators' data literacy (Datnow & Hubbard, 2015; Earl & Katz, 2006; Ikemoto & Marsh, 2007; Marsh, 2012). Although the data use intervention was designed to teach educators to use data and to support them in solving educational problems in their school, the design was not specifically based on learning theories concerning how best to teach educators complex tasks. For example, instead of whole-task practice, which is suggested in the four-component instructional design model (e.g., van Merriënboer & Kirschner, 2013) to support the transfer of learning, some learning tasks during the data use intervention were taught separately, and some were also not representative of tasks educators can encounter during daily teaching practice (Timperley, 2008; van Merriënboer & Kirschner, 2013). These characteristics of the intervention might be further reasons for the low post-test score on the data literacy test.

By zooming in on each component of data literacy, we found that educators scored lowest on the post-test for setting a purpose (29% of answers correct), and that they scored highest on taking instructional action (76% of answers correct). Even though educators were trained through the data use intervention that was linked to all five components of data literacy, the extent to which they showed development varied per data literacy component. For collecting, analyzing, and interpreting data, and taking instructional action, the percentage of correct answers increased by 11% to 12%. However, for setting a purpose, the number of correct answers was slightly lower on the post-test (29% of answers correct) compared to the pre-test (30% of answers correct). A possible reason for educators' low post-test score on setting a purpose might be that throughout the data use intervention, the problem definition in step 1 of the intervention had only been formulated once. Furthermore, not all teams conducted qualitative research activities during the data use intervention, and therefore had not formulated a qualitative research question before while this was required by one item on the data

literacy test. Formulating a concrete and measurable hypothesis also might not be a common practice for educators in secondary education compared to, for example, collecting assessment data (Timperley, 2008; van Veen et al., 2011). In this study, educators scored highest on the ability to take instructional action. At first sight, this is not in line with other literature showing that taking action based on data is the hardest part of the process (Datnow & Hubbard, 2015; Marsh, 2012). The result of our study means, however, that educators could mention or describe ways to take instructional action based on data; yet it does not necessarily mean that educators also actually (have the skills to) make such instructional changes in the classroom (e.g., which pedagogy to choose).

5.1. Limitations of the study

This research is an exploratory, small-scale study, and thus the results cannot be generalized. We used a single-group pre-post research design. Therefore, the data literacy test was only administered to the intervention group and not to a control group. We do not make causal claims, because we cannot say with certainty that the effects we found are due to the data use intervention. While the data literacy test was constructed by using a protocol, the quality of the test could be improved further. For example, more items addressing each data literacy component might improve the test.

5.2. Implications for practice

The results of this study show that educators can develop data literacy during the intensive one-year data use intervention. Professional development interventions, such as the data use intervention, can help educators to make data-based decisions in their schools, to improve the quality of the education they provide (Lai et al., 2014; Van Geel et al., 2016). However, educators' data literacy could be developed to a greater extent during the data use intervention. Adapting the data use intervention through applying more explicitly principles of the four-component instructional design model, such as providing educators with whole-task DBDM experiences (van Merriënboer & Kirschner, 2013), could be a possible improvement to make. We could also try to translate the data use intervention to apply more closely in educators' daily teaching practice, and thus focus on a smaller micro-level (Timperley, 2008). For example, each educator could focus on his or her own educational problems in the classroom (e.g., concerning poor mathematics results). The data coach could provide each educator with guidance in completing the steps of the data use intervention.

Teacher training colleges and professional development programs can play a key role in developing the data literacy of educators. Nowadays, courses about DBDM are rarely given in teacher training colleges, and they are often not integrated with the curriculum. It would be helpful for beginning educators to start out with a certain level of data literacy to improve the quality of the education they provide (Mandinach & Gummer, 2016a; Van Geel et al., 2017). Educators who work at teacher training colleges and researchers from universities can collaborate with each other to re-design courses, such

as integrating topics related to DBDM with the current data courses given in teacher training colleges and professional development programs (Datnow & Hubbard, 2015; Mandinach & Gummer, 2016).

5.3. Implications for future research

With this current study, we obtained more insight into the development of educators' ability to use data to improve education. We defined data as 'information that is systematically collected and organized to represent some aspect of schooling' (Lai & Schildkamp, 2013, p. 10), and mainly studied educators' ability to use assessment data or interview data. However, educators can also use other types of information to improve the education they provide, such as existing research evidence, or informal data (e.g., classroom discussions). Insights into the combination of the use of data and the use of research evidence in an inquiry cycle is limited (Brown, Schildkamp, & Hubers, 2017). Moreover, it is not clear how we can support educators in using informal data to improve the quality of their teaching (Kippers et al., submitted). It would be interesting to focus future research on educators' use of research evidence or informal data to make educational decisions.

Although our study focused on educators' data literacy, it would also be interesting to focus future research on the data literacy of others involved in the student learning process, such as student data literacy or parent data literacy (Datnow & Hubbard, 2015; Marsh, 2012). Students have to learn how to use data to self-regulate their learning. Parents also receive data about their children, and they need data literacy to interpret these data and help improve their children in their learning. Further research is needed to identify which knowledge and skills are linked to the data literacy of students and parents, and to what extent students and parents are currently data literate.

Moreover, many data use interventions are implemented in schools, but changing educational practice permanently seems to be a challenge (e.g., Hubers, Schildkamp, Poortman, & Pieters, 2017). Sustainable interventions are essential for long-term educational improvement. It would be interesting to focus future research on how we can promote the sustainability of data use interventions in schools.

Conflicts of interest

None.

Appendix A

Data literacy test

Question 1. Collect data (Step 1. Problem definition)
"A data team would like to investigate poor results for Dutch in the 9th grade (pre-university education). At first, the team needs to determine the scope of the problem. How will the team be able to 'prove' that the results for Dutch in the 9th grade are really a problem?"

Question 2. Collect data (Step 3. Data collection)
"A data team that wants to work on the problem of poor mathematics results in the first year of secondary education thinks this problem is mainly caused by the poor quality of arithmetic lessons in the primary schools their students are from. Identify two data sources this team could use to investigate this cause."

Question 3. Set a purpose (Step 2. Formulating hypotheses or questions)

"A data team wants to focus on the problem regarding low English student achievement results. The results in the second year of pre-vocational education are low in the school years 2012–2013, 2013–2014, and 2014–2015. Formulate a hypothesis about a possible cause of this problem."

Question 4. Collect data (Step 8. Evaluation)

"A mathematics data team has implemented the instructional action that every lesson in the 8th grade (general secondary education) is started with a short repetition of 'fractions'. The team would like to evaluate how the action was implemented and is experienced by teachers. Specifically, they want to answer the question 'How do colleagues experience this action?'. Mention two instruments that can be used to answer this evaluation question."

Question 5. Analyze data (Step 5. Data analysis)

"A data team wants to investigate whether a course for teachers about differentiation has increased student achievement for the subject of Dutch. In Table below you can see the results for four students, before and after the teachers' course. What do their grades tell you? [answer in a maximum length of five sentences]."

Question 6. Analyze data (Step 4. Data quality check)

"Data have to be of a sufficient quality to take conclusions about the problem. Mention two data quality criteria and explain each of them."

Question 7. Take instructional action (Step 7. Implementing improvement measures)

"A data team has finished the first six steps of the data use intervention. The analysis of the data has shown that teachers do not provide enough feedback about students' learning, and that this is one of the causes of disappointing results for English in the 11th grade (pre-university education). The team wants to start with step 7: implementing improvement measures. Identify two concrete measures to ensure that teachers provide more feedback to students."

Question 8. Set a purpose (Step 1. Problem definition)

"The first step for a data team is to define a problem for the team to work on. Formulate a problem statement based on the consolidation table of collected data below, by completing the following sentences. 'We are dissatisfied with…' 'We would like to achieve…'."

Question 9. Set a purpose (Step 2. Formulating hypotheses or questions)

"A data team would like to investigate the problem of disappointing results for mathematics in 7th grade (pre-vocational education). All team members have the assignment of formulating one (qualitative) research question to study in relation to this problem. Formulate one concrete research question the team could study."

Question 10. Interpret data (Step 6. Interpretation and conclusion)

"The percentages of students with an unacceptable score per subject domain have been comparable for the past three years, see Table above. What conclusions can you draw based on these percentages? [answer in a maximum length of five sentences]. "

Question 11. Take instructional action (Step 7. Implementing improvement measures)

"Mention a concrete measure to solve low Dutch student achievement results, see above. Answer in a maximum length of five sentences."

Question 12. Interpret data (Step 8. Evaluation)

"Describe how you can evaluate whether the measure written above have solved the problem regarding low Dutch student achievement results."

Appendix B

See Table B1.

Table B1
Coding scheme.

Concept	Codes	Explanation
Data literacy	Set a purpose	Educators' ability to set a purpose (e.g., formulate the purpose for using data).
	Collect data	Educators' ability to collect data (e.g., collect quantitative and qualitative data).
	Analyze data	Educators' ability to analyze data (e.g., use statistics to calculate the mean).
	Interpret data	Educators' ability to interpret data (e.g., read and interpret data, such as tables and interviews, and draw conclusions).
	Take instructional action	Educators' ability to take instructional action (e.g., plan and implement instructional action).

References

Bocala, C., & Boudett, K. P. (2015). Teaching educators habits of mind for using data wisely. *Teachers College Record, 117*(4), 1–20.

Brown, C., Schildkamp, K., & Hubers, M. D. (2017). Combining the best of two worlds: a conceptual proposal for evidence-informed school improvement. *Educational Research, 59*(2), 154–172. http://dx.doi.org/10.1080/00131881.2017.1304327.

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. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives, 9*(4), 173–206. http://dx.doi.org/10.1080/15366367.2011.626729.

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.

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.

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

Ebbeler, J., Poortman, C. L., Schildkamp, K., & Pieters, J. M. (2017). The effects of a data use intervention on educators' satisfaction and data literacy. *Educational Assessment, Evaluation and Accountability, 29*(1), 83–105. http://dx.doi.org/10.1007/s11092-016-9251-z.

Eggens, T. J. H. M., & Sanders, P. F. (1993). *Psychometrics in practice. [Psychometrie in de Praktijk].* Arnhem: CITO.

Erkens, T. (2011). Constructing open questions. [Het construeren van open vragen.]. In P. Sanders (Ed.). *Examination at school [Toetsen op school]* (pp. 111–124). Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.

Farrell, C. C., & Marsh, J. A. (2016). Contributing conditions: A qualitative comparative analysis of teachers' instructional responses to data. *Teaching and Teacher Education, 60*, 398–412. http://dx.doi.org/10.1016/j.tate.2016.07.010.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* London: Sage Publications.

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making (NCEE 2009–4067).* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education [Retrieved December, 2, 2016, from http://ies.ed.gov/ncee/wwc/publications/practiceguides/].

Hoogland, I., Schildkamp, K., Van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education, 60*, 377–386. http://dx.doi.org/10.1016/j.tate.2016.07.012.

Hubers, M. D., Schildkamp, K., Poortman, C. L., & Pieters, J. M. (2017). The quest for sustained data use: Developing organizational routines. *Teaching and Teacher Education, 67*, 509–521. http://dx.doi.org/10.1016/j.tate.2017.07.007.

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

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

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*(4), 496–520. http://dx.doi.org/10.1086/505057.

Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (submitted). *Teachers' views on the use of assessment for learning and data-based decision making in classroom practice*. Manuscript submitted for publication.

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, the Netherlands: Springer.

Lai, M. K., Wilson, A., McNaughton, S., & Hsiao, S. (2014). Improving achievement in secondary schools: Impact of a literacy project on reading comprehension and secondary school qualifications. *Reading Research Quarterly, 49*(3), 305–334. http://dx.doi.org/10.1002/rrq.73.

Mandinach, E. B., & Gummer, E. S. (2016a). 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.

Mandinach, E. B., & Gummer, E. S. (2016b). *Data literacy for educators. making it count in teacher preparation and practice.* WestEd.

Mandinach, E., Friedman, J. M., & Gummer, E. (2015). How can schools of education help to build educators' capacity to use data? A systematic view of the issue. *Teachers College Record, 117*(4), 1–26.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record, 114*(11), 1–47.

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

Ministry of Education, Culture & Science (Ministerie van Onderwijs, Cultuur & Wetenschappen). (2000). Wet op het onderwijstoezicht [Education supervision act.]. The Hague, the Netherlands: SDU.

Onwuegbuzie, A. J., & Leech, N. L. (2007). A call for qualitative power analyses. *Quality & Quantity, 41*, 105–121. http://dx.doi.org/10.1007/s11135-005-1098-1.

Poortman, C. L., & Schildkamp, K. (2012). Alternative quality standards in qualitative research? *Quality and Quantity, 46*(6), 1727–1751. http://dx.doi.org/10.1007/s11135-011-9555-5.

Schildkamp, K., Karbautzki, L., & Vanhoof, J. (2014). Exploring data use practices around Europe: Identifying enablers and barriers. *Studies in Educational Evaluation, 42*(1), 15–24. http://dx.doi.org/10.1016/j.stueduc.2013.10.007.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purpose, and promoting and hindering factors. *Teaching and Teacher Education, 26*(3), 482–496. http://dx.doi.org/10.1016/j.tate.2009.06.007.

Schildkamp, K., & Lai, M. K. (2013). Introduction. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.). *Data-based decision making in education: Challenges and opportunities* (pp. 1–7). Dordrecht, the Netherlands: Springer.

Schildkamp, K., & Poortman, C. L. (2015). Factors influencing the functioning of data teams. *Teachers College Record, 117*(4), 1–30.

Timperley, H. (2008). *Teacher professional learning and development.* Brussels, Belgium: International Academy of Education.

Van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. H. M. (2015). Integrating data-based decision making, assessment for learning, and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy & Practice, 22*(3), 324–343. http://dx.doi.org/10.1080/0969594X.2014.999024.

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

Van Geel, M., Keuning, T., Visscher, A., & Fox, J.-P. (2017). Changes in educators' data literacy during a data-based decision making intervention. *Teaching and Teacher Education, 64*, 187–198. http://dx.doi.org/10.1016/j.tate.2017.02.015.

van Merriënboer, J. J. G., & Kirschner, P. A. (2013). *Ten steps to complex learning. A systematic approach to four-component instructional design.* New York: Routledge.

van Veen, K., Zwart, R., & Meirink, J. (2011). What makes teacher professional development effective? A literature review. In M. Kooy, & K. van Veen (Eds.). *Teacher learning that matters* (pp. 3–21). New York: Routledge.

Verbeek, C., & Odenthal, L. (2014). Opbrengstgericht werken en onderzoeksmatig leiderschap in PO en VO [DBDM and research leadership in primary and secondary education.]. In M. Krüger (Ed.). *Leidinggeven aan onderzoekende scholen [Leading researching schools]* (pp. 67–78). Bussum: Coutinho.