



## Teachers' views on the use of assessment for learning and data-based decision making in classroom practice

Wilma B. Kippers\*, Christel H.D. Wolterinck, Kim Schildkamp, Cindy L. Poortman, Adrie J. Visscher

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

### H I G H L I G H T S

- Identifies top five classroom assessments teachers initiate in the classroom.
- Teachers conduct peer and self-assessment in only 10%–25% of their lessons.
- Teachers use data for instruction in only 25%–50% of their lessons.
- Identifies top five prerequisites teachers consider important for AfL and DBDM.
- Highlights the need for professional development for teachers in AfL and DBDM.

### A R T I C L E I N F O

#### Article history:

Received 23 June 2017

Received in revised form

23 May 2018

Accepted 21 June 2018

#### Keywords:

Formative assessment

Assessment for learning

Data-based decision making

Mixed-methods approach

### A B S T R A C T

This paper focuses on classroom assessments, assessment for learning (AfL), and data-based decision making (DBDM) in Dutch secondary education, as well as on prerequisites for implementing AfL and DBDM. Results show that although teachers use various kinds of classroom assessments, such as paper-and-pencil tests and asking students questions, AfL and DBDM have not yet been integrated into teacher practice. Teachers indicated that they conduct peer and self-assessment in only 10% – 25% of their lessons, and use data for instruction in only 25% – 50% of their lessons. A positive attitude towards AfL and DBDM was considered crucial.

© 2018 Elsevier Ltd. All rights reserved.

### 1. Introduction

Assessment is essential for improving the quality of education and learning (Black & Wiliam, 1998; OECD, 2008). In this study, assessment is defined as the use of instruments (e.g., a test or homework assignment) and processes (e.g., asking questions and classroom conversations) for gathering evidence about student learning (Van der Kleij, Vermeulen, Schildkamp, & Eggen, 2015; Stobart, 2008). If assessment has a formative purpose, it is used to support student learning. Formative assessment has the potential to enhance student achievement (Bennett, 2011; Black & Wiliam, 1998, 2009).

Formative assessment can be seen as a concept that covers

various approaches for using assessment to support student learning (Van der Kleij et al., 2015; Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012). Many studies have emphasized the importance of two approaches: assessment for learning (AfL) and data-based decision making (DBDM) (e.g., Wayman, Jimerson, & Cho, 2012b; Wiliam, 2011). AfL has been defined as “part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning” (Klenowski, 2009, p. 264). DBDM refers to the process of “systematically analyzing data sources within the school, applying outcomes of analyses to innovate teaching, curricula, and school performance, and, implementing (e.g. genuine improvement actions) and evaluating these innovations” (Schildkamp & Kuiper, 2010, p. 482). By data, we mean “information that is systematically collected and organized to represent some aspect of schools” (Lai & Schildkamp, 2013, p. 10).

\* Corresponding author.

E-mail address: [w.b.kippers@utwente.nl](mailto:w.b.kippers@utwente.nl) (W.B. Kippers).

AfL and DBDM share a focus on gathering information to adapt education in order to meet student needs (Van der Kleij et al., 2015; Wiliam, 2011). Through AfL and DBDM, teachers and students utilize various ways to gain insight into student learning. Based on this insight, teachers can change the way they teach and students can change the way they learn, in turn enhancing student achievement (Lai & Schildkamp, 2013; Black & Wiliam, 1998). For example, teachers can adapt instruction, and students can distribute their study effort in several short sessions over a longer period of time (rather than practicing the task in a few long sessions over a short period of time). Examples of studies that showed positive effects of AfL and/or DBDM interventions on student achievement are, for example, the study of Lai, Wilson, McNaughton, and Hsiao (2014) in which secondary school students' reading comprehension improved, and the studies by Andersson and Palm (2017) and Keuning and van Geel (2016) in which student achievement for mathematics in primary education improved.

Generally, little attention is paid to AfL and DBDM, including in teacher training colleges, instead summative assessment is focused on (Birenbaum et al., 2015; Mandinach & Gummer, 2016). It is not clear to what extent and how teachers combine AfL and DBDM in their lessons. It is important that teachers blend AfL and DBDM in the classroom because they can complement each other, for instance, because their feedback loops differ in frequency but can be simultaneously active. With AfL, the quality of the learning process during daily everyday practice can be monitored frequently by using information from mostly qualitative assessments (e.g., asking questions and observations). With DBDM, student learning outcomes can be monitored less frequently but information from high-quality, more objective data are used, such as standardized assessments and student questionnaires, leading to less bias in teachers' interpretations (Van der Kleij et al., 2015; Bennett, 2011; Wayman et al., 2012b).

Many studies either focus on AfL or on DBDM, and are often small-scale, qualitative studies (Heitink, van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016; OECD, 2008). This large-scale study tries to fill this knowledge gap by focusing on the extent to which, and how AfL and DBDM are jointly used by teachers in their lessons. Moreover, as far as we know, nobody has quantitatively analyzed differences in how much AfL and DBDM are used by teachers in different grade levels, subjects, and across genders before. Also, we try to gain further insight into which types of assessment instruments and processes are used in daily classroom practice, as various sources of information are needed to use AfL and DBDM (Black & Wiliam, 1998). Furthermore, over 20 prerequisites that may potentially influence the use of AfL and DBDM in the classroom have been identified by researchers, such as teachers' knowledge and skills and the nature of the feedback provided by assessments

(Heitink et al., 2016; Hoogland et al., 2016). To support secondary schools in benefitting from AfL and DBDM, we also aimed at gaining more in-depth insight into which of these prerequisites matter most for teachers wishing to implement AfL and DBDM with the current study. Thus, this mixed-methods study addresses the following research questions:

- Which assessment instruments and processes are most frequently used in classroom practice according to teachers?
- To what extent and how are AfL and DBDM being used in classroom practice according to teachers?
- Which prerequisites do teachers consider important for implementing AfL and DBDM in classroom practice?

## 2. Theoretical framework

### 2.1. Assessment instruments and processes

Teachers can use assessment instruments and processes to gather information about students' learning needs. This information can be used in a formative way, through AfL or DBDM. In this study, we focus on twelve assessment types teachers can use in Dutch secondary education (see Table 1). This is not an exhaustive list, but an overview of the classroom assessments most frequently mentioned in literature (Van der Kleij et al., 2015; Schildkamp & Kuiper, 2010; Ayala et al., 2008; Black & Wiliam, 1998; Newby & Winterbottom, 2011; Ruiz-Primo, 2011).

Although these twelve classroom assessments are presented in this paper as distinct, they are related to each other and even overlap to some extent. For example, the teacher can ask a student a single question which can lead to a classroom conversation if more questions are being asked and if answers are given by both the teacher and the students. In addition, the work collected in a portfolio can include the products of practical tasks and completed homework assignments. In teachers' daily classroom practice, a continuous interaction between various classroom assessments is likely.

### 2.2. Formative assessment

In Fig. 1, our broader formative assessment conceptual framework is presented. Formative assessment starts with teachers or students eliciting information through (high-quality) assessment and considering this information as a form of feedback towards the quality of their own and each other's performance. To give an example, teachers can consider that poor student results might partially be due to their poor lesson preparation, and students can

**Table 1**  
Various types of classroom assessments.

| Types of assessment instruments |  |
|---------------------------------|--|
| Digital tests                   | The student answers questions and completes tasks on a computer.   |
| Homework assignments            | The student completes tasks outside of the lesson.   |
| Oral tests                      | The student answers questions orally.  |
| Paper-and-pencil tests          | The student answers questions and completes tasks on paper (e.g., in the form of a multiple-choice test).                |
| Portfolios                      | The student collects examples of his/her student work (student-developed artifacts) along with his/her (self)reflection. |
| Practical tasks                 | The student completes a practical assignment (e.g., a comic about a book).   |
| Presentations                   | The student presents tasks he/she worked on.   |
| Questionnaires                  | The student completes a questionnaire.   |
| Types of assessment processes   |  |
| Asking questions                | The teacher asks the student a question (e.g., about solving a problem).   |
| Classroom conversations         | An unplanned dialogue in the classroom between the teacher and students.   |
| Student observations            | The teacher observes a student regarding a specific aspect of behavior.  |
| Reflective lessons              | A planned activity in the classroom to evaluate students' prior knowledge (e.g., using concept maps).                    |

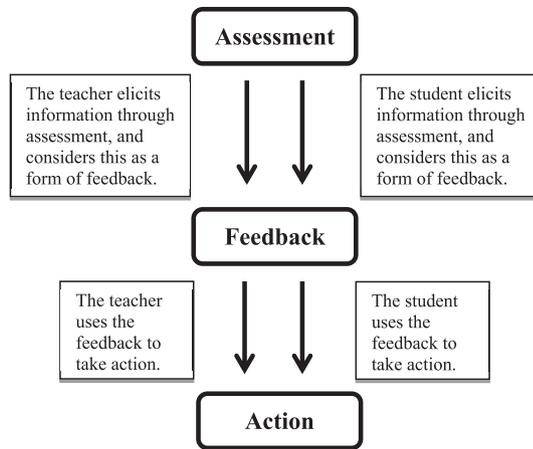


Fig. 1. Conceptual framework of formative assessment (based on Kippers, Schildkamp, & Poortman, 2016).

consider that poor results might partially be due to their lack of motivation to learn. Next, teachers or students can use this feedback to take action. For example, teachers can improve their instruction (e.g., ‘I will provide additional examples when I explain subject matter’), and students can improve their learning strategies (e.g., ‘I will practice in shorter sessions over a longer period of time with these additional exercises to perform calculations with percentages.’) (Van der Kleij et al., 2015; Bennett, 2011; Sadler, 1989; Stobart, 2008). The formative assessment process is presented here as linear and static, but in teachers’ daily classroom practice, formative assessment is an iterative process (i.e., there is continuous interaction between eliciting information and taking action to move learning forward).

### 2.2.1. Afl

One important approach towards formative assessment is Afl. This approach focuses on daily practice in which teachers, students, and peers continuously gather information about student learning processes. They gain insight into where the student is now, where s(he) is going, and how s(he) can progress (Van der Kleij et al., 2015; Klenowski, 2009; Wiliam, 2011). In this study, we focus on teachers’ use of the key Afl-strategies in classroom practice.

First, teachers can *share learning intentions and success criteria* with students (Cauley & McMillan, 2010; Heritage, 2007; Lysaght & O’Leary, 2013). Learning intentions are the contents the teacher wants the students to learn, and success criteria are used to check whether student learning activities were successful. If teachers share learning intentions and success criteria with students during the lesson, they both know where the student is progressing towards and they both know how it will be assessed whether the students have learned what the teacher wanted them to learn (Black & Wiliam, 2009; Wiliam & Leahy, 2015).

Second, teachers can *elicit evidence* about student learning processes during everyday practice (Cauley & McMillan, 2010; Heritage, 2007; Lysaght & O’Leary, 2013). They can use various formal and informal assessments as evidence to gather information about students’ learning needs, such as student observations, classroom conversations, and homework assignments. If teachers elicit evidence, they gain insight into students’ prior learning, and where they are in their learning process (Black & Wiliam, 2009; Ruiz-Primo, 2011; Wiliam & Leahy, 2015).

Third, teachers can improve teaching and learning by *using feedback* (Black & Wiliam, 2009; Cauley & McMillan, 2010; Heritage, 2007; Wiliam & Leahy, 2015), defined as information

regarding aspects of student performance or understanding (Hattie & Timperley, 2007). For example, teachers can adapt teaching by using evidence from a classroom conversation with students as a form of feedback. Moreover, teachers can provide feedback to students to explain to them how to progress in their learning (Van der Kleij et al., 2015; Sadler, 1989).

Fourth, teachers can let students *conduct peer and self-assessment* as a part of classroom practice (Cauley & McMillan, 2010; Heritage, 2007; Lysaght & O’Leary, 2013). Peer and self-assessment represent students’ ability to assess peers’ learning or one’s own learning, and relate these outcomes to learning goals to improve learning outcomes. If teachers let students conduct peer and self-assessment, students’ ability to self-regulate their learning can be enhanced and they can develop a sense of ownership of their own learning and that of their peers (Black & Wiliam, 2009; Boekaerts & Corno, 2005; Wiliam & Leahy, 2015; Wiliam, 2011).

Afl concerns the combination of Afl-strategies during daily practice through a collective process. Only then teachers and students can identify the gap between where the student is now and where s(he) is going, and then they can take action to improve student learning (Wiliam, 2011).

### 2.2.2. DBDM

Another important approach towards formative assessment is DBDM. This approach also focuses on gathering information about student learning, but the information is systematically collected and organized (i.e., data sources such as assessment data and interview data). This approach does not directly take place during everyday practice, but the high-quality data can be used formatively to improve student learning outcomes (i.e., attaining results and targets) (Hoogland et al., 2016; Van der Kleij et al., 2015; Schildkamp & Kuiper, 2010). The following steps can be distinguished in the DBDM approach: setting a purpose; collecting, analyzing, and interpreting data; and taking action (Lai & Schildkamp, 2013; Keuning & van Geel, 2016; Mandinach & Gummer, 2016). Combined, these steps form a cyclical process.

In this study, we focused on actions relating to instruction, because we wanted to explore how and the extent to which teachers use data to take instructional actions in the classroom. If teachers *use data for instruction*, data (e.g., final examination results) are used as a form of feedback on the quality of teacher instruction, after which teaching is adapted to students’ learning needs with the goal to improve student achievement (Ebbeler, Poortman, Schildkamp, & Pieters, 2016; Carlson, Borman, & Robinson, 2011).

### 2.3. Prerequisites for implementing Afl and DBDM

Prerequisites that can enable the implementation of Afl and DBDM have been identified by various authors (Heitink et al., 2016; Hoogland et al., 2016; Kerr, Marsh, Ikemoto, Dariley, & Barney, 2006). Most prerequisites apply to both Afl and DBDM, however some prerequisites are unique to one of these two approaches. In this study, we focus on the prerequisites related to three overarching categories: (a) the assessments, (b) the teacher, and (c) the context, see Table 2.

The categories ‘assessments’, ‘teacher’, and ‘context’ have been described as separate categories. However, they are linked. To give an example, the context category (which includes access to technology, such as student monitoring systems) is related to and to some extent overlaps with the teacher category (which includes knowledge and skills to use technology). In addition, the categories also influence each other. For example, influencing the context category (internal and external support) might also affect the teacher category (teachers’ knowledge and skills).

**Table 2**  
Prerequisites for implementing AfL and DBDM.

| The assessments category   |  |
|--|--|
| Alignment between assessments and the curriculum                       | Assessments must be aligned with the curriculum (Heitink et al., 2016; Fuchs, Fuchs, Karns, Hamlett, & Katzaroff, 1999). If assessments are aligned with the curriculum, both teachers and students can use them to evaluate 'how the student is doing' regarding accomplishing curriculum learning goals so that they can take action on 'how the student can proceed'.   |
| Assessments providing detailed feedback about student learning         | Assessments should provide detailed feedback to teachers and students (Heitink et al., 2016; Van der Kleij & Eggen, 2013). Feedback is defined as information regarding aspects of student performance or understanding (Hattie & Timperley, 2007). Teachers can use this detailed data feedback to adapt classroom instruction, and to provide students with detailed information on why mistakes were made and how they can proceed.   |
| Integration of assessments into classroom instruction                  | Especially for AfL, assessments should be integrated into classroom instruction (Heitink et al., 2016; Lee, 2011). Assessments should continuously take place in classroom practice as opposed to being viewed separately from explaining subject matter. If assessments are integrated into classroom instruction, students can improve their learning just-in-time, which fosters teachers' and students' shared responsibility for the student learning processes.  |
| High quality of assessments  | Especially for DBDM, the quality of assessments should be high (i.e., reliability and validity) (Hoogland et al., 2016; Schildkamp & Kuiper, 2010; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006). High quality is essential if teachers want to systematically analyze and interpret assessments, such as standardized tests or structured student observations, to improve teaching and learning. Based on these high-quality assessments, high-stakes decisions can be made to adapt teaching, such as deciding to purchase curricular materials for students in Grade 7. |
| The teacher category   |  |
| Teachers' knowledge and skills to analyze and interpret evidence       | If teachers have the knowledge and skills to analyze and interpret evidence from assessments, they can identify students' strengths and weaknesses (Heitink et al., 2016; Van der Kleij & Eggen, 2013).  |
| Teachers' knowledge and skills for adapting their instruction          | If teachers have the knowledge and skills to adapt instruction, they can try to ensure that their instructional practices meet students' learning needs, which can aid in bridging the gap between current and desired student learning outcomes. For example, knowledge and skills on how to provide remedial instruction and how to select learning activities for a specific group of students is important (Hoogland et al., 2016; Datnow, Park, & Kennedy-Lewis, 2012).   |
| Teachers' knowledge and skills to use technology                       | If teachers have the knowledge and skills to use technology, they can use the digital systems to collect evidence, and subsequently they can monitor student progress through interpreting the evidence from assessments (Heitink et al., 2016; Lee, Feldman, & Beatty, 2012).   |
| Teachers having a positive attitude towards the use of AfL and DBDM    | If teachers have a positive attitude towards AfL and DBDM, this implies that they believe in their usefulness and feel responsible for enhancing student learning. Moreover, they recognize the importance of using assessments as evidence for revising instructional practices, and do not feel forced to do so (e.g., by the Inspectorate of Education) (Heitink et al., 2016; Hoogland et al., 2016; Lee et al., 2012).  |
| The context category   |  |
| Facilitation and support from the school leader                        | If school leaders facilitate teachers' use of AfL and DBDM, for instance by making time available for teachers to participate in training programs on the use of AfL and DBDM, they can foster AfL and DBDM implementation in classroom practice (Heitink et al., 2016; Hoogland et al., 2016; Wayman, Cho, Jimerson, & Spikes, 2012a).  |
| Motivation by the school leader  | If school leaders motivate teachers in using AfL and DBDM, they can, for instance, establish a school-wide AfL and DBDM culture in which goals are formulated and expectations are expressed regarding the use of AfL and DBDM in classroom practice (Heitink et al., 2016; Hoogland et al., 2016; Levin & Datnow, 2012).  |
| School leaders' knowledge and skills to analyze and interpret evidence | Especially for DBDM, if school leaders have the knowledge and skills to analyze and interpret evidence from assessments, they can improve education at the school level and model the use of data, such as by helping teachers to analyze assessment data (Hoogland et al., 2016; Levin & Datnow, 2012; Wayman et al., 2012a).   |
| Teacher collaboration  | Teacher collaboration can be important for using AfL and DBDM. Through teacher collaboration, teachers can collaboratively analyze and interpret evidence, and they can collaboratively discuss student outcomes and develop instructional plans (Heitink et al., 2016; Hoogland et al., 2016; Lee, 2011).   |
| Internal and external support  | Internal and external support (e.g., support from an internal data expert at the school or an external coach) can be essential for using AfL and DBDM successfully (Heitink et al., 2016; Hoogland et al., 2016; Kerr et al., 2006). For example, if teachers want to derive information from assessments, they can be supported by the data expert at the school with analyzing and interpreting evidence from assessments, and with developing instructional plans.  |
| Access to technology   | Access to technology to gather students' responses, such as digital systems, can be important in using AfL and DBDM (Heitink et al., 2016; Hoogland et al., 2016; Lee et al., 2012). Especially for DBDM, student monitoring systems should combine various types of data meaningfully and provide reports on student progress that are easy to interpret (Kerr et al., 2006; Van der Kleij & Eggen, 2013).  |

### 3. Method

#### 3.1. Context

This study was conducted in the context of secondary education in the Netherlands (student ages 12–18). Dutch schools have always been free to choose the religious, ideological and pedagogical principles on which they base their education (Ministry of Education, Culture & Science, 2000). However, the Dutch Inspectorate of Education holds schools accountable for the quality of their education and expects schools to use data to improve their educational quality. In the Dutch context, there are only national standardized tests at the end of secondary education. Important data available within Dutch schools include these standardized final examination tests, student work (e.g., paper-and-pencil tests and homework assignments), and parent and student

questionnaires (Schildkamp & Kuiper, 2010). Generally, Dutch students' performances on PISA tests (Programme for International Student Assessment, student age 15) are higher than average, but just as in other countries, on several aspects student performance is decreasing, such as mathematics (OECD, 2016). Besides, just as in teacher training colleges in other countries, in the Dutch teacher training colleges little attention is paid to AfL and DBDM (e.g., Mandinach & Gummer, 2016).

#### 3.2. Research design

This research is part of two larger data use projects in which different measures were given to different schools because of different purposes of the two projects. In the first data use project, schools from one of the largest Dutch school boards in secondary education participated, and the main purpose of the project was to

stimulate teachers' use of AfL and DBDM. The teachers from these schools filled out a questionnaire to reflect on the extent to which they already use AfL and DBDM. In the second data use project, schools across the country participated, and the main aim of the project was to stimulate the use of assessment data from computerized adaptive tests for AfL and DBDM purposes. The teachers from these schools participated in interviews and checklists to reflect on which types of assessment instruments and processes are already used, how teachers worked on AfL and DBDM, and which prerequisites matter most for teachers for implementing AfL and DBDM. By employing this mixed methods approach, we use both groups of schools from the two data use projects in one paper because they both provide valuable information to answer the research questions; see Table 3. We have interpreted the data side by side to illustrate what is happening in the field. First, we analyzed the quantitative data (teacher questionnaire). Next, the qualitative data (interview scheme and checklists) were analyzed. To answer the second research question, the qualitative data were used to inform or illustrate the quantitative data (Creswell & Plano Clark, 2011).

Twenty-six of the schools who were invited to fill out the questionnaire (52% of the invited schools) were willing to participate. This convenience sample may not fully represent all invited schools; yet, a mix of denominations, geographical locations, and educational tracks was shown. One other secondary school from another school board also wanted to participate to get insight into the extent to which their teachers use AfL and DBDM. The teachers voluntarily filled out the questionnaire. This school is comparable to the other schools that participated. Thus, a total of 27 schools filled out the questionnaire. In the questionnaire we especially focused on how much teachers used AfL and DBDM in their classroom practice.

Four schools who were invited to participate in interviews and checklists were selected by considering the following criteria: (a) they had to be general secondary or pre-university schools; (b) they were required to have high student achievement results based on their evaluation by the Inspectorate of Education, because we expected that teachers in high-performing schools were using AfL and DBDM in their classrooms ensuring that we could explore *how* they were using this; and (c) they had to focus especially on English language, Dutch language, and mathematics, because these are the three core subjects areas in secondary education. From the nine schools that met these criteria, 4 schools were selected that were distributed across the country. One of the four schools who participated in the interviews and checklists also had filled out the questionnaire. Characteristics of all schools are presented in Table 4 (Ministry of Education, Culture & Science, 2016; <http://www.scholenopdekaart.nl>; <http://www.statline.cbs.nl>).

### 3.3. Respondents

The AfL-DBDM questionnaire was filled out by  $N = 479$  teachers from schools from the first data use project. In the sample, 49.7%

( $N = 238$ ) were female and 49.5% ( $N = 237$ ) were male. In secondary education, on average 52.5% ( $N = 39,703$ ) is female and 47.5% ( $N = 35,980$ ) is male (Ministry of Education, Culture & Science, 2015). We analyzed whether the two groups of teachers, those who participated in the questionnaire and teachers in secondary education in the Netherlands, were similar by calculating chi-square tests of independence regarding gender (Field, 2013); see Table 5. A non-significant association was found between gender and whether or not teachers participated in the questionnaire ( $X^2(1) = 1.049$ ,  $p = 0.306$ ). This suggests that there were no differences in terms of gender. Other teacher characteristics can be found in Table 6.

Based on a convenience sample, twelve teachers from across four schools of the second data use project were selected for an individual interview and for completing checklists: an English teacher, a Dutch teacher and a mathematics teacher, all teaching in the first three years of secondary education. English, Dutch and mathematics are considered core subjects in the Dutch secondary education curriculum. They were all willing to participate in this research. We selected these teachers because we wanted to gain further insight into the current AfL and DBDM situation of schools with high student achievement. Seven of these twelve teachers were female, and five were male. Four teachers had <5 years of teaching experience, two teachers had 5–14 years of teaching experience, five teachers had 15–24 years of teaching experience, and one teacher had  $\geq 25$  years of teaching experience.

**Table 4**  
Characteristics of schools.<sup>a</sup>

|              |  | Schools in this study |         | Schools in the Netherlands |         |
|--------------|--|-----------------------|---------|----------------------------|---------|
|              |  | N                     | (%)     | N                          | (%)     |
| School size  | Small (<500 students)                    | 0                     | (0.0%)  | 158                        | (24.1%) |
|              | Medium (500–1000 students)               | 1                     | (7.7%)  | 95                         | (14.5%) |
|              | Large (>1000 students)                   | 12                    | (92.3%) | 402                        | (61.4%) |
| Denomination | Catholic schools                         | 6                     | (46.1%) | 150                        | (22.9%) |
|              | Interdenominational schools <sup>b</sup> | 3                     | (23.1%) | 66                         | (10.1%) |
|              | Private schools <sup>c</sup>             | 3                     | (23.1%) | 99                         | (15.1%) |
|              | Public school <sup>d</sup>               | 1                     | (7.7%)  | 186                        | (28.4%) |
|              | Other                                    | 0                     | (0.0%)  | 154                        | (23.5%) |

<sup>a</sup> For comparison reasons on the Dutch national level, the schools are compared on school institution level instead of school location level. Therefore, the school characteristics of 13 school institutions are presented instead of 30 school locations.

<sup>b</sup> A Dutch interdenominational school is characterized as a government independent school that is based on a combination of different religions.

<sup>c</sup> A Dutch private school is characterized as a government independent school that is based on a specific educational vision and not on a specific religion.

<sup>d</sup> A Dutch public school is characterized as a government dependent school that is neither based on a specific educational vision nor on a specific religion.

**Table 3**  
Relation between research questions and research design.

| Research questions   | Quantitative research design                | Qualitative research design                        |
|--|---|--|
| Which assessment instruments and processes are most frequently used in classroom practice according to teachers? | –   | Interviews, checklists<br>4 schools<br>12 teachers |
| To what extent and how are AfL and DBDM being used in classroom practice according to teachers?                  | Questionnaire<br>27 schools<br>479 teachers | Interviews<br>4 schools<br>12 teachers             |
| Which prerequisites do teachers consider important for implementing AfL and DBDM in classroom practice?          | –   | Interviews, checklists<br>4 schools<br>12 teachers |

**Table 5**  
Results of the chi-square test of independence for gender.

|                           | Gender            |                   | Total             |
|---------------------------|-------------------|-------------------|-------------------|
|                           | Male              | Female            |                   |
| Teachers in questionnaire | 237.0 (225.9)     | 238.0 (249.1)     | 475.0 (475.0)     |
| Comparison group          | 35980.0 (35991.1) | 39703.0 (39691.9) | 75683.0 (75683.0) |
| Total                     | 36217.0 (36217.0) | 39941.0 (39941.0) | 76158.0 (76158.0) |

### 3.4. Instruments

#### 3.4.1. Questionnaire for teachers

To study the extent to which AfL and DBDM are being used in classroom practice (RQ2), the AfL-DBDM questionnaire was developed. This digital teacher questionnaire was based on (a) existing surveys related to AfL (O'Leary, Lysaght, & Ludlow, 2013; Lysaght & O'Leary, 2013), and (b) an existing survey about DBDM (Schildkamp, Poortman, Luyten, & Ebbeler, 2016). For this specific study, 42 items about the use of AfL and DBDM were relevant. The questionnaire items were set on a five-point Likert scale ranging from 'almost never (it happens in less than approximately 10% of my lessons)' to 'embedded (it happens in more than approximately 90% of my lessons)'. Moreover, 'I don't know' was also a response option. Teachers were asked to fill out the questionnaire by indicating for each item to what extent it applied to their own daily teaching practice. The questionnaire items were in Dutch. Confirmatory factor analysis and reliability analysis were conducted in SPSS (Field, 2013). The factor analysis revealed a 5-factor structure consistent with the theoretical framework: data use for instruction, sharing learning intentions and success criteria, asking questions and classroom discussions, feedback, and peer and self-assessment, see Appendix A. From the 42-items, ten items were deleted because the results of the factor analysis showed that they loaded less than

0.5. Four other items also loaded slightly lower than 0.5, but for theoretical reasons it was decided to keep these items in the scale. Reliability of the scales was sufficient (.70–.80) to good (>.80) (Field, 2013); see Table 7.

#### 3.4.2. Interview scheme

To study the use of assessments in depth, as well as how AfL and DBDM are being used in classroom practice, and which prerequisites are important for AfL and DBDM implementation, an interview scheme with 20 open questions was developed by the first and third author of this paper. Questions were, for example: 'Which assessments do you use in your daily teaching practice?' (RQ1), 'How do you provide feedback to students in the classroom?' (RQ2), and 'Which prerequisites do you consider important for implementing AfL and DBDM in the classroom?' (RQ3). The interview scheme was first tested with a teacher, after which minor adjustments were made. All questions were asked of each teacher. Each interview lasted approximately 45 min and were all held in April 2015 by the first author of this paper. As all interviews were held in Dutch, quotations were translated into English for use in this paper.

#### 3.4.3. Checklists

For triangulation of the interview data, two checklists for

**Table 6**  
Teacher characteristics of the teachers who filled out the questionnaire.

|                              |                                   | N   | (%)   |
|------------------------------|-----------------------------------|-----|-------|
| Years of teaching experience | <5 years                          | 79  | 16.5% |
|                              | 5–14 years                        | 165 | 34.5% |
|                              | 15–24 years                       | 102 | 21.3% |
|                              | ≥25 years                         | 119 | 24.8% |
|                              | Unknown                           | 14  | 2.9%  |
| Subject area <sup>a</sup>    | Alpha sciences (e.g., English)    | 139 | 29.0% |
|                              | Beta sciences (e.g., mathematics) | 146 | 30.5% |
|                              | Gamma sciences (e.g., history)    | 92  | 19.2% |
|                              | Other (e.g., music)               | 100 | 20.9% |
|                              | Unknown                           | 2   | 0.4%  |
| Grade levels <sup>b</sup>    | Teacher in lower grades           | 197 | 41.1% |
|                              | Teacher in upper grades           | 281 | 58.7% |
|                              | Unknown                           | 1   | 0.2%  |
| Teachers per school          | School A                          | 30  | 6.3%  |
|                              | School B                          | 39  | 8.1%  |
|                              | School C                          | 36  | 7.5%  |
|                              | School D                          | 29  | 6.1%  |
|                              | School E                          | 50  | 10.4% |
|                              | School F                          | 45  | 9.4%  |
|                              | School G                          | 68  | 14.2% |
|                              | School H                          | 86  | 18.0% |
|                              | School I                          | 60  | 12.5% |
|                              | School J                          | 36  | 7.5%  |

<sup>a</sup> Alpha sciences includes language sciences, such as English, Spanish, and Greek. Beta sciences includes exact sciences, such as mathematics, biology, and physics. Gamma sciences includes social sciences, such as history, economy, and philosophy.

<sup>b</sup> The lower grades include Grades 7 and 8 (student ages 12–14) of the educational track that prepares students for vocational education (four years) or Grades 7, 8 and 9 (student ages 12–15) of the track that prepares students for higher education (five years) or university education (six years). The upper grades include Grades 9 and 10 (student ages 14–16) of the educational track that prepares students for vocational education (four years), Grades 10 and 11 (student ages 15–17) of the track that prepares students for higher education (five years), or Grades 10, 11 and 12 (student ages 15–18) of the track that prepares students for university education (six years).

**Table 7**  
Reliability of the scales in the AfL-DBDM questionnaire.

| Scale  | Number of items | Cronbach's alpha | Example items  |
|--|-----------------|------------------|--|
| Data use for instruction                         | 11              | .89              | To what extent do you use data to tailor instruction to individual students' needs?  |
| Sharing learning intentions and success criteria | 4               | .76              | Learning intentions are stated using words that emphasize knowledge, skills, concepts, and/or attitudes, i.e. what the students are learning, NOT what they are doing.   |
| Asking questions and classroom discussions       | 6               | .83              | Questioning goes beyond the 'one right answer style' ( <i>where the focus is often on trying to guess the answer in the teacher's mind</i> ) to the use of more open-ended questions that encourage critical thinking. |
| Feedback   | 5               | .81              | Written feedback on students' work goes beyond the use of grades and comments such as 'well done' to specify what students have achieved and what they need to do next.  |
| Peer and self-assessment                         | 6               | .82              | Students assess and comment on each other's work (e.g., <i>they are taught how to use the success criteria of a lesson to judge another student's piece of work</i> ).   |

teachers were developed. The first checklist described various assessments teachers can use in secondary education (RQ1), such as 'paper-and-pencil tests', 'classroom conversations' and 'homework assignments'. During the interviews (after the teachers had answered questions on the use of assessments), teachers were asked to indicate which of the assessments they use in their classroom, and also whether they use other assessments that were not included in the checklist.

The second checklist listed the prerequisites identified in the literature as important for implementing AfL and DBDM in secondary education (RQ3), such as 'alignment between assessments and the curriculum', 'teachers' knowledge and skills to analyze and interpret evidence', and 'facilitation of teachers' use of AfL and DBDM by the school leader'. During the interviews (after the teachers had answered questions on the importance of prerequisites), teachers were asked to indicate a maximum of five prerequisites that mattered most to them for implementing AfL and DBDM in classroom practice, and also whether additional prerequisites should have been included in the checklist.

### 3.5. Analysis

Descriptive analyses were used to report results. For each scale of the questionnaire, we analyzed the mean and standard error. Non-response to some items of a scale varied between 40 and 83 teachers. Moreover, we conducted independent samples t-tests to compare the results on the questionnaire for gender (male versus female) and grade levels (teacher in lower grades versus teacher in upper grades), and we conducted a one-way between-subjects ANOVA including Tukey's test to compare the results on the questionnaire for the four subject areas (alpha, beta, gamma, other) (see Results section) (Field, 2013). For the checklist, we reported which specific items were ticked by how many teachers. The individual interviews were transcribed verbatim. Based on the theoretical framework, an a priori coding scheme with 32 codes was developed. In line with the theoretical framework, thirteen codes related to the theme "assessments", five codes related to the theme "AfL or DBDM", and fourteen codes related to the theme "prerequisites", see Appendix B. The program Atlas.ti was used for coding the interview data. With regard to which assessments teachers use in secondary education (RQ1), codes were, for example: 'observations', 'paper-and-pencil tests' and 'classroom conversations'. Regarding how teachers use AfL and DBDM in secondary education (RQ2), examples of codes were: 'sharing learning intentions and success criteria', 'feedback', and 'data use for instruction'. Relating to which prerequisites teachers think matter most for using AfL and DBDM in secondary education (RQ3), codes were, for example: 'alignment between assessments and the curriculum', 'teachers' knowledge and skills to analyze and interpret evidence', and 'facilitation of teachers' use of AfL and DBDM by the school leader'. The first two authors of this paper coded the data.

After coding the interview data, we selected a code and summarized what all teachers said during the interviews relating to that code. Then we continued with the next code. Based on the descriptive analyses, we could report detailed answers to the three research questions.

### 3.6. Reliability and validity

A systematic approach to data collection was followed, which was consistent with the research questions. All interviews were audio-taped and transcribed, and the inter-rater reliability between the two coders was calculated across ten percent of the interview data to increase reliability with other researchers (Poortman & Schildkamp, 2012). They found an acceptable Cohen's Kappa of 0.69 (Eggen & Sanders, 1993). The questionnaire was based on existing reliable instruments. Factor analyses and reliability analyses were carried out for the questionnaire data (Field, 2013), and all scales were found to be reliable. The content of the questionnaire, interview scheme, and checklists all link to the theoretical framework, different instruments were used to answer each research question, and all instruments in this study were reviewed by two researchers with teaching experience, to improve construct validity. Detailed descriptions of the results were provided (e.g., by including respondents' quotes). A variety of 30 secondary schools and 491 teachers across the country participated in this study, to promote external validity.

## 4. Results

First, the various assessment instruments and processes used by teachers in secondary education will be presented to answer the first research question. Second, the use of AfL and DBDM will be described to answer the second research question, after which the most important prerequisites for implementing AfL and DBDM effectively will be presented to answer the third research question.

### 4.1. Assessment instruments and processes

Fig. 2 shows the responses of the twelve teachers of the second data use project to the checklist items, in which they had to report which classroom assessments they use in classroom practice. All twelve teachers ticked using *paper-and-pencil tests* and *asking questions* on the checklist, which was in line with the twelve teachers' responses during the interviews.

In their responses to the checklist, eleven teachers ticked using *classroom conversations*. In the interviews, teachers reported that they often spontaneously start a classroom conversation with their students. From the checklist, ten teachers indicated using *homework assignments*, and nine teachers indicated using *student observations* and *reflective lessons*. During the interviews, teachers shared that they do not observe students on a specific aspect, but

rather continuously and often unconsciously observe all students during class. *'If they work on tasks during the lesson, I observe how they work on these tasks by walking around in the classroom.'* In addition, teachers stated that they use a concept map at the start of a lesson or a new chapter to evaluate student understanding and clarify misconceptions. Using *presentations* and *oral tests* were both ticked on the checklist by eight teachers. In the interviews, a teacher stated: *'We expect students to give an 8-minute presentation about a book by using PowerPoint or Prezi.'* Using *portfolios* was ticked by seven teachers, and using *questionnaires* and *practical tasks* by six. For instance, teachers shared that students had to complete questionnaires to evaluate the teacher's teaching quality. Using *digital tests* was indicated on the checklist by three teachers. One teacher reported that digital tests were used for students with learning difficulties, such as dyslexia. Teachers were also asked whether they use other assessments that were not included in the checklists, or not. Two teachers explained that they also use classroom debates in their daily teaching practice, and one teacher pointed to the use of auditory tests.

#### 4.2. The use of AfL and DBDM

In each section of the following paragraphs the results of the questionnaire are first shown, in which 479 teachers from schools from the first data use project filled out *how much* they use AfL and DBDM in their classroom practice. Next, to inform or illustrate the questionnaire data, in each section of the following paragraphs the results of the interviews are shown, in which 12 teachers from schools from the second data use project described *how* they use AfL and DBDM in their classroom practice.

##### 4.2.1. AfL: sharing learning intentions and success criteria

In the questionnaire, teachers scored highest on 'sharing learning intentions and success criteria'. The mean score was 3.30 ( $N = 419$ ;  $SD = 0.79$ ), which can be interpreted as between 'emerging' and 'established' (50% – 75% of the lessons). To give an example, the statement 'Learning intentions are stated using words that emphasize knowledge, skills, concepts and/or attitudes, i.e. what the students are learning, NOT what they are doing,' was scored at 3.03 on average. Fifty teachers answered 'I don't know' on some of the items in this scale. Teachers in the lower grades scored significantly higher ( $M = 3.43$ ;  $SD = 0.80$ ;  $N = 177$ ) than teachers in the upper grades ( $M = 3.21$ ;  $SD = 0.78$ ;  $N = 241$ ) with  $t(416) = 2732$  and  $p = 0.007$ . This was a small effect of  $d = 0.28$  (Field, 2013).

In contrast, during the interviews only two of the twelve teachers explicitly mentioned that they share learning intentions or success criteria with students, such as by writing on the white board that students should be able to identify an object in a sentence at the end of the lesson. Seven other teachers reported that they do not share learning intentions and success criteria but only tell the students the subject topic they will be working on during the lesson, and why. Reasons for not sharing learning intentions given by teachers were that they do not feel the urge to share learning intentions or think sharing learning intentions might intimidate students. *'The learning objectives are given in the teaching method (...) Therefore, I do not feel I have to mention the learning objectives to the students.'*

##### 4.2.2. AfL: asking questions and classroom discussions

In the questionnaire, the mean score for eliciting evidence through 'asking questions and classroom discussions' was 3.07 ( $N = 436$ ;  $SD = 0.80$ ), which can be interpreted as between 'emerging' and 'established' (50% – 75% of the lessons). For example, the statement 'Questions are used to elicit students' prior knowledge on a topic.' received a mean score of 3.39. Thirty-five teachers answered 'I don't know' on some of the items in this scale. There was a significant effect of subject area on 'asking questions and classroom discussions' [ $F(3, 430) = 7.813$ ,  $p = 0.000$ ]. Teachers who teach in Gamma sciences scored significantly higher ( $M = 3.43$ ;  $SD = 0.79$ ;  $N = 83$ ) than teachers who teach in Alpha sciences ( $M = 3.04$ ;  $SD = 0.72$ ;  $N = 132$ ), than teachers who teach in Beta sciences ( $M = 2.96$ ;  $SD = 0.78$ ;  $N = 134$ ), and than teachers who teach in other subject areas ( $M = 2.92$ ;  $SD = 0.85$ ;  $N = 85$ ). These were medium effects of  $d = 0.54$ ,  $d = 0.60$ , and  $d = 0.60$  (Field, 2013).

The interview results also showed that teachers elicit evidence about student learning. All twelve teachers reported that by asking questions or through classroom discussions, they gain insight into such matters as students' prior knowledge and whether students understand the learning content. *'Yes, of course I ask several questions to identify what exactly is the problem. And then I notice, for example, that specific grammar learning content is not understood.'* Moreover, eight teachers talked about eliciting evidence regarding students' strengths and weaknesses and how students learn by observing students in the classroom. Furthermore, two of these teachers reported that they gather information regarding student understanding through homework assignments; one teacher used reflective lessons (i.e., a quiz) to gain insight into students' prior

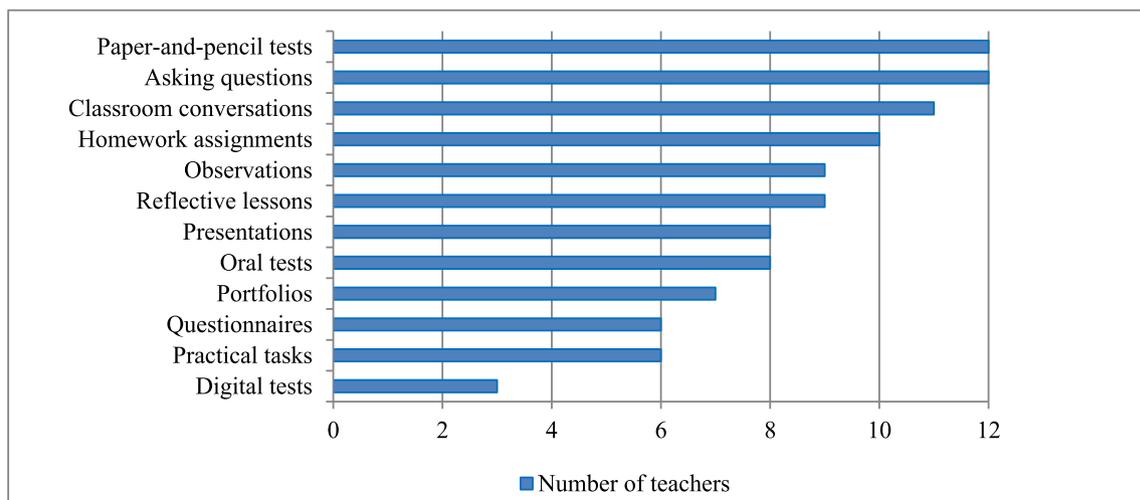


Fig. 2. Teachers' views on the use of various assessments ticked on the checklist.

knowledge.

#### 4.2.3. AfL: feedback

In the questionnaire, the mean score for the application of using 'feedback' was 2.82 ( $N = 424$ ;  $SD = 0.86$ ), which can be interpreted as between 'sporadic' and 'emerging' (25% – 50% of the lessons). To give an example, 'Written feedback on students' work goes beyond the use of grades and comments such as "well done", to specify what students have achieved and what they need to do next.' was scored 2.91 on average. Sixty-six teachers answered 'I don't know' on some of the items in this scale. Female teachers scored significantly higher ( $M = 2.93$ ;  $SD = 0.87$ ;  $N = 211$ ) than male teachers ( $M = 2.72$ ;  $SD = 0.84$ ;  $N = 209$ ) with  $t(418) = -2492$  and  $p = 0.013$ . This was a small effect of  $d = 0.25$  (Field, 2013).

The interview results also showed that teachers made limited use of feedback. Despite the fact that eleven teachers stated that they provide students with oral feedback during the lesson, such as 'Please ensure that if you, for example, are presenting, that you talk loud enough.', only four of these teachers stated that they occasionally alternate between oral and written feedback. Moreover, two teachers expressed that the feedback usually focused on students' current progress and on how to improve rather than on learning intentions. The results of this exploration thus seem to show that teachers can provide different types of feedback to students more frequently during their lessons to improve student learning, as well improve the quality of feedback. During the interviews, only five teachers reported that they used assessments as a form of feedback with regard to their own performance to adapt instruction. They mentioned that they repeat learning content, explain learning content further, provide additional examples on the white board, or work on tasks together with students. 'Yes, then you have to change your lessons. You have to repeat it [the learning content] once more, or you have to explain it in a different way, or you should teach in a different way.' During the interviews, three teachers provided examples in which evidence is only seen as a form of feedback towards students rather than teaching, such as letting students with high grades work on their strengths during projects outside lesson time or letting students with low grades practice more in extra lessons. Based on these results, it seems that some teachers tend to suggest to students what they have to do to improve learning, instead of improving the quality of their own instructional practices.

#### 4.2.4. AfL: peer and self-assessment

In the questionnaire, teachers scored lowest on conducting 'peer and self-assessment'. The mean score was 1.77 ( $N = 396$ ;  $SD = 0.71$ ), which can be interpreted as between '(almost) never' and 'sporadic' (10% – 25% of the lessons). For example, 'Students are encouraged to use a range of assessment techniques to review their own work (e.g., a rubric to provide insight into success criteria, or analyzing a test to identify strengths and weaknesses).' was scored 1.83 on average. One hundred and twelve teachers answered 'I don't know' on some of the items in this scale. There was a significant effect of subject area on 'peer and self-assessment' [ $F(3, 391) = 7.305$ ,  $p = 0.000$ ]. Teachers who teach in the 'other' subject areas group scored significantly higher ( $M = 2.09$ ;  $SD = 0.86$ ;  $N = 72$ ) than teachers who teach in Alpha sciences ( $M = 1.74$ ;  $SD = 0.64$ ;  $N = 122$ ), than teachers who teach in Beta sciences ( $M = 1.62$ ;  $SD = 0.64$ ;  $N = 122$ ), and than teachers who teach in Gamma sciences ( $M = 1.73$ ;  $SD = 0.66$ ;  $N = 79$ ). These were a medium effect of  $d = 0.55$ , a large effect of  $d = 0.73$ , and a medium effect of  $d = 0.55$  (Field, 2013).

In line with this, during the interviews two of the twelve teachers explicitly mentioned that students do not assess themselves, for instance, because the teacher thinks that students lack

self-knowledge to assess themselves. Nine other teachers reported during the interviews that students assess their own learning infrequently. 'They have to assess their own homework assignments, of course. Sometimes, I give them an answer sheet, and then they have to find out what went well and what went wrong.' Two of these teachers explained that students often lack the content knowledge to assess themselves. So, it seems that self-assessment is not conducted often. Moreover, during the interviews eight teachers stated that students only sometimes assess their peer's learning and provide feedback, such as by assessing their peers' assignments or presentations. 'I let them provide feedback. And not just "You did well" but explain to the other why, and what you would do differently so that they gain insight into how the teacher assesses.' Four teachers mentioned during the interviews students' lack of content knowledge or social pressure as reasons for not conducting peer assessment frequently. Moreover, four teachers described during the interviews that students are not aware of learning intentions when assessing one's own learning, or that of their peers.

#### 4.2.5. DBDM: data use for instruction

In the questionnaire, the mean score for the utilization of 'data use for instruction' was 2.85 ( $N = 439$ ;  $SD = 0.75$ ), which can be interpreted as between 'sporadic' and 'emerging' (25% – 50% of the lessons). To give an example, 'I use data to adapt instruction based on the needs of gifted students' had a mean score of 2.77. Seventy-five teachers answered 'I don't know' on some of the items in this scale. Teachers in the lower grades scored significantly higher ( $M = 2.96$ ;  $SD = 0.78$ ;  $N = 174$ ) than teachers in the upper grades ( $M = 2.78$ ;  $SD = 0.73$ ;  $N = 265$ ) with  $t(437) = 2514$  and  $p = 0.012$ . This was a small effect of  $d = 0.25$  (Field, 2013). Besides, there was a significant effect of subject area on 'data use for instruction' [ $F(3, 433) = 2.673$ ,  $p = 0.047$ ], in which teachers who teach Alpha sciences scored significantly higher ( $M = 2.96$ ;  $SD = 0.69$ ;  $N = 132$ ) than teachers who teach Beta sciences ( $M = 2.72$ ;  $SD = 0.75$ ;  $N = 134$ ). This was also a small effect of  $d = 0.32$  (Field, 2013).

The interview results also showed that teachers made limited use of data for instructional purposes. All twelve teachers used student achievement data to obtain information about matters such as students' strengths and weaknesses. During the interviews, ten teachers explicitly mentioned that they register student grades into the student monitoring system. 'I monitor students' progress and see "well, this group of students is not that good in grammar".' However, three teachers reported during the interviews that they do not gain insight into each student's progress on a regular basis, but only do so sometimes to prepare for parent meetings. In the interviews, eight teachers explicitly described that, based on student achievement data, they provide students with oral or written feedback (e.g., on which answers are wrong and why). However, all teachers stated that they often only use grades as written feedback. Other examples of feedback mentioned by the teachers were not very detailed. 'I write in capital letters on the test "YOU DID NOT LEARN THE IDIOM".' In the interviews, ten teachers reported that they adapt instruction based on data on low-performing students. For instance, in the next chapter they pay attention to previous mistakes, repeat learning content, or explain subject matter to a small group of students. However, two mathematics teachers explained during the interviews that they find it difficult to adjust their instruction based on data, because after a test about specific learning content has been taken other learning content must (also) be taught and assessed. 'I mainly use the information to know what to do next year (...) I do not have the opportunity to use it directly.' This is specifically related to mathematics and less to other subjects. Moreover, eight teachers provided examples during the interviews in which data is only used to criticize students or do them a favour rather than improving the quality of teaching, such as calling

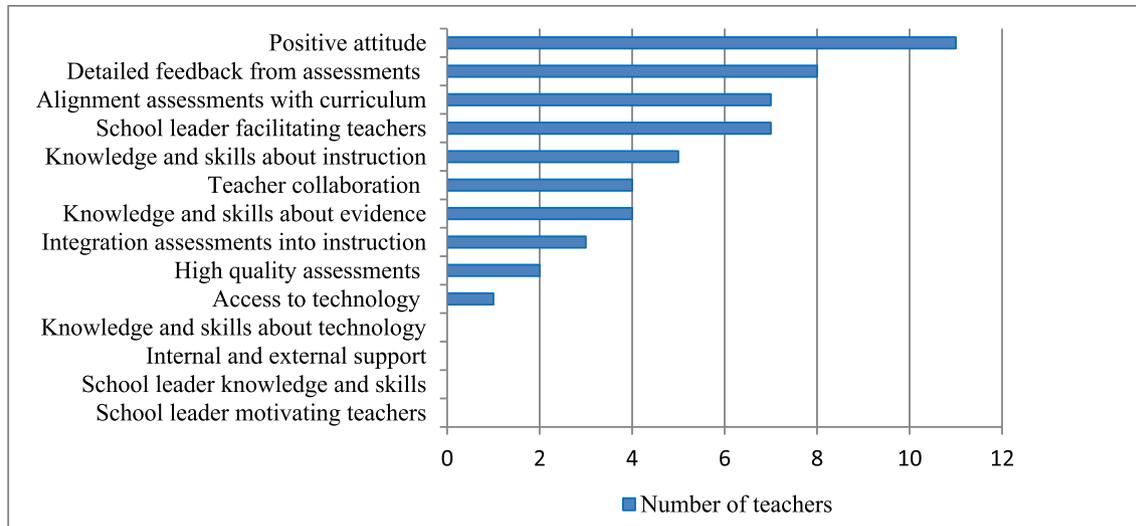


Fig. 3. Teachers' views on the importance of various prerequisites ticked on the checklist.

parents, letting students redo the test with the book, or giving students a higher grade.

#### 4.3. The prerequisites for the implementation of AfL and DBDM

In this section, we will provide an overview of the prerequisites ticked on the checklist and elaborated on during the interviews. We did not distinguish between AfL- and DBDM-prerequisites. Fig. 3 shows the responses of the twelve teachers of the second data use project to the checklist about the prerequisites, where they had to indicate five prerequisites they consider most important for implementing AfL and DBDM in classroom practice. Table 8 elaborates on the responses of the twelve teachers to the checklist and during the interviews.

## 5. Discussion and conclusion

For many years, the importance of AfL and DBDM has been advocated by researchers (e.g., Wayman et al., 2012b; William, 2011). AfL focuses on the use of assessments during daily practice to support student learning and DBDM focuses on the use of high-quality, more objective (assessment) data to support student learning. Therefore, they can complement each other (Van der Kleij et al., 2015; Wayman et al., 2012b). However, previous studies have not investigated the combination of AfL and DBDM, and the extent to which teachers use the key strategies of AfL has also not been studied in detail before in the Netherlands or in other countries.<sup>1</sup> Moreover, most of the completed studies so far are small-scale, qualitative studies (Heitink et al., 2016; OECD, 2008). In the present study, a mixed-methods approach was used to address the current AfL and DBDM situation in Dutch secondary education. We aimed to obtain a better understanding of teachers' use of AfL and DBDM.

### 5.1. Assessments used

The top five classroom assessments that seem to be most

frequently used by our twelve teachers in their daily teaching practice in secondary education are: (a) paper-and-pencil tests, (b) asking questions, (c) classroom conversations, (d) homework assignments, and (e) student observations. With this research, we have explored that these include a variety of assessments: well-known assessments that are traditionally used for a summative purpose (e.g., paper-and-pencil tests) as well as assessments that are more likely to be used formatively (e.g., classroom conversations). It is important to take into account that teachers use these classroom assessments in interaction with each other, and that some of them are related to each other, and overlap to some extent. For example, asking questions (from teacher to student) can lead to a classroom conversation (dialogue between teacher and students). A variety of assessments, such as the five described above, can be used by teachers for a mix of AfL and DBDM purposes. Information from high-quality, more objective assessment data as well as information from assessments collected during everyday practice can inform teachers and students about students' progress from different perspectives (Van der Kleij et al., 2015; Wayman et al., 2012b). Using assessments for AfL and DBDM purposes can improve student learning (even more) (Lai & Schildkamp, 2013; Black & William, 1998).

### 5.2. The use of AfL and DBDM

The results of this study show that teachers' use of AfL and DBDM strategies in classroom practice has considerable room for improvement. Using 'feedback' and 'data use for instruction' was utilized in only approximately 25% – 50% of teachers' lessons, and conducting 'peer and self-assessment' even less, in only approximately 10% – 25% of their lessons. Interviewees also reported the limited use of these strategies. Some teachers provide students with feedback which is not always linked to learning intentions, some teachers attribute underperformance to students (and not as information that can be used to improve teaching), and some teachers feel that students lack the knowledge and skills to conduct peer and self-assessment. In other studies researchers also found limited use of 'data use for instruction', e.g., teachers in these studies reported that they used data for instruction between 'yearly' and 'a couple times per year' (Ebbeler et al., 2016; Schildkamp et al., 2016). Teachers make use of 'sharing learning intentions and success criteria' and 'asking questions and

<sup>1</sup> We have not found studies about the extent to which teachers use the key AfL strategies. Moreover, we e-mailed a well-known researcher, who has conducted research on formative assessment for years, and he explained that the studies he knew were also not about the extent to which teachers use the key AfL strategies.

**Table 8**  
Teachers' responses on the importance of various prerequisites for implementing AfL and DBDM.

| The assessments category  |  |
|---|--|
| Assessments providing detailed feedback about student learning      | Eight out of twelve teachers indicated on the checklist that assessments should provide detailed feedback to teachers and students, see Fig. 3. During the interviews, some teachers also emphasized this. A teacher mentioned that feedback from a test should focus on <i>how</i> students obtain answers instead of only on whether students have answers. Three other teachers explained that when digital tests provide feedback to students, it saves the teacher time and effort, and teachers can get a quick overview of all students.  |
| Alignment between assessments and the curriculum                    | Alignment between assessments and the curriculum was important according to seven teachers. During the interviews, two teachers expressed that criteria for success should be linked to learning intentions in such a way that it is clear to students what has to be learned and why. Another teacher pointed out that close alignment between assessments and the curriculum is a prerequisite, because it is a requirement of the Inspectorate of Education.  |
| Integration of assessments into classroom instruction               | Three teachers referred to the importance of the integration of assessments into classroom instruction. This was not mentioned as important during the interviews.   |
| High quality of assessments   | A high quality of assessments was ticked on the checklist as important by two teachers. In the interview, a teacher described: ' <i>I ask my colleague (...) is the test representative? Is it sufficiently valid?</i> ' Three other teachers mentioned in the interviews that it is important that test items are linked to different levels of student learning to gauge differences in students' knowledge.   |
| The teacher category  |  |
| Teachers having a positive attitude towards the use of AfL and DBDM | A positive attitude towards the use of AfL and DBDM was referred to as important by eleven teachers. One teacher mentioned the importance of this prerequisite during the interview by describing that teachers need to be willing to guide students in their learning.  |
| Teachers' knowledge and skills for adapting their instruction       | Five teachers ticked knowledge and skills for adapting their instruction as important on the checklist. This importance was acknowledged during the interviews by two teachers. A teacher mentioned: ' <i>You must have knowledge of various learning strategies, and know what to do next (...) We have to differentiate.</i> ' Another teacher stated that teachers should adjust instruction to guide students in attaining learning objectives.  |
| Teachers' knowledge and skills to analyze and interpret evidence    | Knowledge and skills for analyzing and interpreting evidence was important in the eyes of four teachers. During the interviews, it was mentioned by four teachers that it is essential that teachers check whether students have learned what they needed to learn (e.g., understanding of learning goals), for example, by using tests or asking questions.   |
| The context category  |  |
| Facilitation and support from the school leader                     | Facilitation and support from the school leader for teachers' use of AfL and DBDM was indicated as important by seven teachers. In the interviews, six teachers stated that more time is needed (e.g., to obtain information about student learning, to differentiate, or provide feedback). In the interviews, a teacher stated: ' <i>The forty-, fifty-minute schedule should be changed to a seventy- or eighty-minute schedule. That would be much better.</i> ' Five teachers explained that having 30 or more students in one classroom prevents teachers from obtaining an overview of each individual student and hinders group work by students. Two teachers stated that having a permanent classroom instead of switching between classrooms is necessary for organizing groups of student tables in order to promote group work when conducting peer assessment. |
| Teacher collaboration   | Four teachers ticked teacher collaboration as important on the checklist. One teacher stated during the interview that teachers should support each other in developing test items, and another teacher expressed that teachers should discuss when to give tests and what to tell to students beforehand about the test items among themselves.   |
| Access to technology  | Access to technology was important according to one teacher. This teacher explained in the interview that a digital system, in which students and teachers can communicate about homework assignments, would enhance teachers' ability to obtain insight into student work and to guide their learning.  |

classroom discussions' in approximately 50% – 75% of their lessons. All interviewees described they elicit evidence to gain insight into student understanding, but seven of the twelve teachers stated that they do not share learning intentions and success criteria with students but only tell them the subject topic they will be working on during the lesson, and why. In conclusion, none of the AfL-strategies and the DBDM-strategy are used in 75% or more of the lessons. Although teachers differ from each other in the extent to which and how they use AfL and DBDM (e.g., the questionnaire results showed standard deviations ranging between 0.71 and 0.86 for the scales, implying that there was much variation between the answers), it seems that on average AfL and DBDM are not yet used in daily classroom practice.

A reason for this underutilization might be that teachers lack the knowledge and skills to apply AfL- and DBDM-strategies (Heitink et al., 2016; Hoogland et al., 2016; Hubbard, Datnow, & Pruyne, 2014). Generally, little or no attention is paid to AfL and DBDM in teacher training colleges, and summative assessment has been the standard in schools for many years (Birenbaum et al., 2015; Mandinach & Gummer, 2016). Thus, benefitting from AfL and DBDM requires teachers to make major changes to their classroom practice. AfL- and DBDM-strategies will be very new for most teachers and changing classroom strategies, such as adapting instruction based on data (instead of providing the same instruction

to all students), can be difficult for teachers. It also presupposes a positive attitude towards AfL and DBDM, and a school culture in which assessment is seen as a way to improve student learning. A reason for the low levels of 'peer and self-assessment' might be the traditional, teacher-led nature of Dutch classroom practice: the teacher is the one who elicits evidence, and uses and provides feedback, and students merely function as recipients (Heitink et al., 2016). A classroom culture that expects and facilitates students to apply peer and self-assessment and views assessment as an opportunity to learn is required to stimulate use of and buy-in for AfL and DBDM.

### 5.3. Prerequisites for formative assessment

The five prerequisites most frequently indicated by our twelve teachers as important for the use of AfL and DBDM in secondary education are: (a) teachers having a positive attitude towards the use of AfL and DBDM, (b) assessments providing detailed information about student learning that can be used by teachers and students as a form of feedback, (c) alignment between assessments and the curriculum, (d) facilitation and support from the school leader, and (e) teacher knowledge and skills for adapting their instruction. These five prerequisites relate to each of the three categories we distinguished between: 'assessments', 'teacher' and

'context'. It is important to take into account of all these prerequisites as a whole and not just focus on each of them separately, because they are related. If teachers develop more knowledge and skills for adapting instruction, their attitude towards AfL and DBDM might also positively change, for example. In other studies on AfL and DBDM researchers identified many prerequisites (Heitink et al., 2016; Hoogland et al., 2016). With the current research, we have explored which of these prerequisites matter most for teachers wishing to implement AfL and DBDM.

#### 5.4. *Limitations of the study*

It is important to emphasize that in this study, teacher self-report data were collected by means of a questionnaire, an interview scheme, and checklists. Even though we asked teachers to specify examples of how they use AfL and DBDM, the data might be biased to some extent and not precisely reflect the actual levels of AfL and DBDM.

Furthermore, some teachers of twelve of the twenty-seven schools who participated in the questionnaire were already familiar with using assessment data as a result of participating in a data use intervention, see Ebbeler et al. (2016). In that data use intervention, 6–8 teachers and school leaders of a school collaboratively try to solve an educational problem in their school by using data. Some teachers of the twelve schools might have a more positive attitude towards (AfL and) DBDM than teachers in other schools, and they might use DBDM better compared to teachers in other schools.

Four schools were selected for the qualitative part of this study, partly based on their evaluation by the Inspectorate of Education (who had rated them as schools with high levels of student achievement), and are therefore not representative for all schools in the Netherlands. It may be that the teachers in these schools use AfL and DBDM more frequently and better compared to teachers in other schools.

When drawing conclusions based on this current study, some significant differences for the results on the questionnaire regarding gender, grade levels, and subject areas should be taken into account. Yet, many effect sizes were small for the differences found (see implications for research and practice).

#### 5.5. *Implications for practice and research*

To effectively support teachers in using AfL and DBDM, we will first need to formulate the criteria characterizing a teacher who uses AfL and DBDM effectively. What is such a teacher precisely doing in the classroom, which actions does (s)he take, and which decisions does (s)he make when, and why? If this is known, we can determine which knowledge and skills AfL and DBDM requires, and next, how teachers can acquire these best. In other words, given the complexity of the advocated AfL- and DBDM-strategies, it is not surprising that teachers do not yet use much of them in our classrooms. Developing and implementing these strategies in school practice presupposes in-depth analysis, instructional design and training work.

As AfL- and DBDM-strategies are not yet frequently used by teachers (keep in mind that some of these teachers worked in schools with high levels of student achievement or had participated in a data use intervention meaning that teachers in other schools might use AfL- and DBDM-strategies even less), future research should focus on designing, developing, and implementing professional development interventions to stimulate teachers' use of AfL and DBDM strategies further. For example, a model that can serve as an approach to identify the AfL- and DBDM-knowledge and skills and to design the professional development trajectories for

teachers is the four-component instructional design model (for more information on the 4C/ID-model see van Merriënboer & Kirschner, 2013).

The few significant differences for the results on the questionnaire regarding gender, grade levels, and subject areas found in this current study should be taken into account when providing professional development interventions or working with teacher training colleges for teachers. Teachers can be grouped based on their use of AfL and DBDM to fulfill their differentiated needs regarding training. To give an example, in our sample, teachers who teach in Gamma sciences elicited evidence through asking questions and classroom discussion to a greater extent than teachers who teach in the 'other' subject areas group. The teachers who teach in the 'other' subject areas group conduct peer and self-assessment to a greater extent than teachers who teach in Beta sciences. It seems that the subject areas might play a role in whether teachers feel they can or should use certain AfL and DBDM strategies. During training, teachers in different grade levels, subjects, and across genders can focus their effort on different (stages within) AfL and DBDM strategies.

The prerequisites that were found to matter most for teachers in our sample, such as a positive attitude of teachers and facilitation and support from the school leader, can also be addressed in professional development trajectories. For example, by focusing the professional development intervention on teachers' own subject area, and by scheduling the (monthly) intervention in line with the school's annual schedule. After AfL- and DBDM interventions have been developed and implemented, the effectiveness of those interventions on different teachers' knowledge and skills, their classroom teaching, and on student achievement should be evaluated (Hubbard et al., 2014; OECD, 2008).

Although students play an essential role in AfL and DBDM, our study shows that peer and self-assessment are conducted least in classroom practice, compared to the other AfL- and DBDM -strategies studied. It would therefore also be interesting to focus research on the role of students in AfL and DBDM (e.g., students' self-regulated learning behaviors in the classroom) and on how teachers and students interact in using AfL and DBDM (Heitink et al., 2016; Hoogland et al., 2016). Questions that arise are, for instance: What student knowledge, skills and attitudes are needed to conduct peer and self-assessment, and what are the effects of AfL and DBDM on student engagement?

### **Compliance with ethical standards**

#### *Funding*

This study was funded by the Ministry of Education, Culture and Science (61500-120437) and Stichting Carmelcollege. They both had no influence on the study design, the collection, analyses, and interpretation of data, and the writing of the report.

#### *Conflicts of interest*

The authors declare that they have no conflict of interest.

### **Acknowledgments**

The authors would like to express their thanks to the Ministry of Education, Culture and Science and Stichting Carmelcollege for funding this research. Furthermore, the authors give special thanks to the schools and teachers that participated in the research. Moreover, the authors would like to thank the educators who provided assistance during the research.

## Appendices

### Appendix A

**Table A1**

Summary of the factor analysis results for the questionnaire.

|   | Data use<br>for<br>instruction | Sharing learning<br>intentions and success<br>criteria | Asking questions and<br>classroom<br>discussions | Feedback Peer and<br>self-<br>assessment |
|---|--------------------------------|--|--|--|
| To what extent do you use data to:  |                                |  |  |  |
| Set learning goals/targets for individual students  | 0.582                          |  |  |  |
| Determine which topics and skills students do and do not possess  | 0.544                          |  |  |  |
| Determine progress of students  | 0.527                          |  |  |  |
| Tailor instruction to individual students' needs  | 0.713                          |  |  |  |
| Set the pace of my lessons  | 0.618                          |  |  |  |
| Gives students feedback on their learning process   | 0.529                          |  |  |  |
| Form small groups of students for targeted instruction  | 0.573                          |  |  |  |
| Identify instructional content to use in class  | 0.605                          |  |  |  |
| Study why students make certain mistakes  | 0.532                          |  |  |  |
| Adapt instruction based on the needs of gifted students   | 0.635                          |  |  |  |
| Adapt instruction based on the needs of struggling students   | 0.744                          |  |  |  |
| Learning intentions are stated using words that emphasize knowledge, skills, concepts and/or attitudes i.e., what the students are learning NOT what they are doing.  |                                | 0.536  |  |  |
| Students are reminded about the links between what they are learning and the big learning picture (e.g., "We are learning to count money so that when we go shopping we can check our changes").  |                                | 0.472  |  |  |
| Success criteria related to learning intentions are matched to pupils.  |                                | 0.721  |  |  |
| Child-friendly language is used to share learning intentions with students (e.g., "We are learning to make a good guess (prediction) about what is likely to happen next in the story")   |                                | 0.722  |  |  |
| Assessment techniques are used to facilitate class discussion (e.g., brainstorming).  |                                |  | 0.591  |  |
| Questions are used to elicit students' prior knowledge on a topic.  |                                |  | 0.630  |  |
| Students are encouraged to share the questioning role with teacher during lessons (e.g., the teacher routinely invites pupils to question their peers' contributions to discussions).   |                                |  | 0.609  |  |
| Questioning goes beyond the one right answer style (where the focus is often on trying to guess the answer in the teacher's mind) to the use of more open-ended questions that encourage critical thinking                                |                                |  | 0.620  |  |
| Students' incorrect responses are used to guide teaching and learning (e.g., a pupil is asked to explain why he/she gave a particular answer).  |                                |  | 0.514  |  |
| Students can explain to others what they are learning (e.g., if a visitor came to the classroom, pupils could articulate what they are learning in terms that identify the knowledge, skills, concepts and/or attitudes being developed). |                                |  | 0.518  |  |
| Written feedback on pupils' work goes beyond the use of grades and comments such as "well done" to specify what students have achieved and what they need to do next.   |                                |  |  | 0.342                                    |
| Teacher-made tests are used diagnostically to identify strengths and needs in the learning of the student (e.g., identifying common mistakes in the addition of fractions).   |                                |  |  | 0.645                                    |
| Tests are used diagnostically to tailor the instruction to the needs of the students by taking into account the strengths and needs of students (e.g., extra lessons on adding fractions)   |                                |  |  | 0.669                                    |
| Tests are used diagnostically to identify strengths and needs in teaching (e.g., teacher makes a key analysis in order to evaluate its own series of lessons)   |                                |  |  | 0.416                                    |
| The teacher uses the information on the progress of students to provide feedback to students.   |                                |  |  | 0.400                                    |
| Students are given an opportunity to indicate how challenging they anticipate the learning will be at the beginning of a lesson or activity (e.g., by using traffic lights).  |                                |  |  | 0.506                                    |
| Students are encouraged to record their progress using, for example, learning logs.   |                                |  |  | 0.581                                    |
| Students assess and comment on each other's work (e.g., they are taught how to use the success criteria of a lesson to judge another pupil's piece of work).  |                                |  |  | 0.612                                    |
| Students are encouraged to use a range of assessment techniques to review their own work (e.g., rubric, traffic lights, thumbs up/down, two stars and a wish)   |                                |  |  | 0.581                                    |
| A visual record of students' progress is maintained to celebrate students' learning and show areas of/for development (e.g., a bulletin board displaying progression in story writing over a term).                                       |                                |  |  | 0.554                                    |
| Time is set aside during lessons to allow for self- and peer-assessment.  |                                |  |  | 0.612                                    |

## Appendix B

**Table B1**  
Coding scheme.

| Concept                  | Codes   |
|--------------------------|---|
| Assessments              | Paper-and-pencil tests<br>Digital tests<br>Presentations<br>Practical tasks<br>Portfolios<br>Homework assignments<br>Oral tests (formal)<br>Questionnaires<br>Asking questions (informal)<br>Student observations<br>Classroom conversations<br>Reflective lessons<br>Other assessments |
| AfL                      | Sharing learning intentions and success criteria<br>Asking questions and classroom discussions<br>Feedback<br>Peer and self-assessment  |
| DBDM                     | Data use for instruction  |
| The assessments category | Alignment between assessments and the curriculum<br>Assessments providing detailed feedback about student learning<br>Integration of assessments into classroom instruction<br>High quality of assessments  |
| The teacher category     | Teachers' knowledge and skills to analyze and interpret evidence<br>Teachers' knowledge and skills for adapting their instruction<br>Teachers' knowledge and skills to use technology<br>Teachers having a positive attitude towards the use of AfL and DBDM                            |
| The context category     | Facilitation and support from the school leader<br>Motivation by the school leader<br>School leaders' knowledge and skills to analyze and interpret evidence<br>Teacher collaboration<br>Internal and external support<br>Access to technology  |

## References

- Andersson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and Instruction*, 49, 92–102. <https://doi.org/10.1016/j.learninstruc.2016.12.006>.
- Ayala, C. C., Shavelson, R. J., Ruiz-Primo, M. A., Brandon, P. R., Yin, Y., Furtak, E. M., et al. (2008). From formal embedded assessments to reflective lessons: The development of formative assessment studies. *Applied Measurement in Education*, 21(4), 315–334. <https://doi.org/10.1080/08957340802347787>.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18, 5–25. <https://doi.org/10.1080/0969594X.2010.513678>.
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., et al. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117–140. <https://doi.org/10.1177/1478210314566733>.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>.
- Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: International Review*, 54(2), 199–231. <https://doi.org/10.1111/j.1464-0597.2005.00205.x>.
- Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31, 13–17. <https://doi.org/10.1111/j.1745-3992.2012.00251.x/full>.
- 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. <https://doi.org/10.3102/0162373711412765>.
- Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *Clearing House*, 83(1), 1–6. <https://doi.org/10.1080/00098650903267784>.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Datnow, A., Park, V., & Kennedy-Lewis, B. (2012). High school teachers' use of data to inform instruction. *Journal of Education for Students Placed at Risk*, 17(4), 247–265. <https://doi.org/10.1080/10824669.2012.718944>.
- Ebbeler, J., Poortman, C. L., Schildkamp, K., & Pieters, J. M. (2016). Effects of a data use intervention on educators' use of knowledge and skills. *Studies in Educational Evaluation*, 48, 19–31. <https://doi.org/10.1016/j.stueduc.2015.11.002>.
- Eggen, T. J. H. M., & Sanders, P. F. (1993). *Psychometrie in de praktijk [Psychometrics in practice]*. Arnhem: CITO.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage Publications.
- Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal*, 36(3), 609–646. <https://doi.org/10.3102/00028312036003609>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. <https://doi.org/10.3102/003465430298487>.
- Heitink, M. C., van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50–62. <https://doi.org/10.1016/j.edurev.2015.12.002>.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140–145. <https://doi.org/10.1177/003172170708900210>.
- 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. <https://doi.org/10.1016/j.tate.2016.07.012>.
- Hubbard, L., Datnow, A., & Pruyne, L. (2014). Multiple initiatives, multiple challenges: The promise and pitfalls of implementing data. *Studies in Educational Evaluation*, 42(1), 54–62. <https://doi.org/10.1016/j.stueduc.2013.10.003>.
- 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. <https://doi.org/10.1086/505057>.
- Keuning, T., & van Geel, M. (2016). *Implementation and effects of a schoolwide data-based decision making intervention: A large-scale study* (Doctoral thesis). the Netherlands: University of Twente.

- Kippers, W. B., Schildkamp, K., & Poortman, C. L. (2016, April). *The use of formative assessment by teachers in secondary education in the Netherlands*. Paper presented at the American Educational Research Association conference, Washington D.C., USA.
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16, 263–268. <https://doi.org/10.1080/09695940903319646>.
- 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. <https://doi.org/10.1002/rrq.73>.
- Lee, I. (2011). Bringing innovation to EFL writing through a focus on assessment for learning. *Innovation in Language Learning and Teaching*, 5, 19–33. <https://doi.org/10.1080/17501229.2010.502232>.
- Lee, H., Feldman, A., & Beatty, I. D. (2012). Factors that affect science and mathematics teachers' initial implementation of technology-enhanced formative assessment using a classroom response system. *Journal of Science Education and Technology*, 21, 523–539. <https://doi.org/10.1007/s10956-011-9344-x>.
- Levin, J. A., & Datnow, A. (2012). The principal role in data-driven decision making: Using case-study data to develop multi-mediator models of educational reform. *School Effectiveness and School Improvement*, 23(2), 179–201. <https://doi.org/10.1080/09243453.2011.599394>.
- Lysaght, Z., & O'Leary, M. (2013). An instrument to audit teachers' use of assessment for learning. *Irish Educational Studies*, 32(2), 217–232. <https://doi.org/10.1080/03323315.2013.784636>.
- Mandinach, E. B., & Gummer, E. S. (2016). *Data literacy for educators. Making it count in teacher preparation and practice*. San Francisco, CA: WestEd.
- 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.
- Ministry of Education, Culture & Science (Ministerie van Onderwijs, Cultuur & Wetenschappen). (2000). *Wet op het onderwijstoezicht [Education Supervision Act]*. The Hague: SDU.
- Ministry of Education, Culture & Science (Ministerie van Onderwijs, Cultuur & Wetenschappen). (2015). *Onderwijspersoneel vo in personen 2011-2015 [Educational staff in secondary education from 2011-2015]*. Retrieved November, 17, 2016 from <http://www.duo.nl>.
- Ministry of Education, Culture & Science (Ministerie van Onderwijs, Cultuur & Wetenschappen). (2016). *Adressen van hoofdvestigingen vo [Addresses of school institutions in secondary education]*. Retrieved November, 22, 2016 from <http://www.duo.nl>.
- Newby, L., & Winterbottom, M. (2011). Can research homework provide a vehicle for assessment for learning in science lessons? *Educational Review*, 63(3), 275–290. <https://doi.org/10.1080/00131911.2011.560247>.
- OECD. (2008). Assessment for learning. Formative assessment. In *OECD/CERI International conference "learning in the 21st Century: Research, innovation and policy"*. Retrieved October, 17, 2016 from <http://www.oecd.org.nl>.
- OECD. (2016). PISA 2015. Results in focus. Retrieved November, 22, 2017 from <http://www.oecd.org.nl>.
- O'Leary, M., Lysaght, Z., & Ludlow, L. (2013). A measurement instrument to evaluate teachers' assessment for learning classroom practices. *The International Journal of Educational and Psychological Assessment*, 14(2), 40–60. <https://doi.org/10.13140/2.1.1289.0241>.
- Poortman, C. L., & Schildkamp, K. (2012). Alternative quality standards in qualitative research? *Quality and Quantity*, 46(6), 1727–1751. <https://doi.org/10.1007/s11135-011-9555-5>.
- Ruiz-Primo, M. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37(1), 15–24. <https://doi.org/10.1016/j.stueduc.2011.04.003>.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144. <https://doi.org/10.1007/BF00117714>.
- 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. <https://doi.org/10.1016/j.tate.2009.06.007>.
- Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (2016). *Factors promoting and hindering data-based decision making in schools. School Effectiveness and School Improvement*. <https://doi.org/10.1080/09243453.2016.1256901>.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Abingdon: Routledge.
- Van der Kleij, F. M., & Eggen, T. J. H. M. (2013). Interpretation of the score reports from the computer program LOVS by teachers, internal support teachers and principals. *Studies in Educational Evaluation*, 39, 144–152. <https://doi.org/10.1016/j.stueduc.2013.04.002>.
- 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. <https://doi.org/10.1080/0969594X.2014.999024>.
- Wayman, J. C., Cho, V., Jimerson, J. B., & Spikes, D. (2012a). District-wide effects on data use in the classroom. *Education Policy Analysis Archives*, 20(25), 1–28. Retrieved February, 18, 2015 from <http://epaa.asu.edu/ojs/article/view/979>.
- Wayman, J. C., Jimerson, J. B., & Cho, V. (2012b). Organizational considerations in establishing the data-informed district. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 23, 159–178. <https://doi.org/10.1080/09243453.2011.652124>.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>.
- William, D., & Leahy, S. (2015). *Embedding formative assessment. Practical techniques for K-12 classrooms*. Florida, USA: Learning Sciences International.