International Journal of Educational Research 103 (2020) 101602

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

Formative assessment: A systematic review of critical teacher prerequisites for classroom practice

Kim Schildkamp[a,*], Fabienne M. van der Kleij[b,1], Maaike C. Heitink[a], Wilma B. Kippers[a,†], Bernard P. Veldkamp[a]

[a] Faculty of Behavioural, Management and Social Sciences, University of Twente, the Netherlands
[b] Institute for Learning Sciences & Teacher Education, Australian Catholic University, Brisbane, Australia

ARTICLE INFO

Keywords:
Formative assessment
Data-based decision making
Assessment for learning
Teacher prerequisites
Classroom practice
Systematic review

ABSTRACT

Formative assessment has the potential to support teaching and learning in the classroom. This study reviewed the literature on formative assessment to identify prerequisites for effective use of formative assessment by teachers. The review sought to address the following research question: What teacher prerequisites need to be in place for using formative assessment in their classroom practice? The review was conducted using a systematic approach. A total of 54 studies were included in this review. The results show that (1) knowledge and skills (e.g., data literacy), (2), psychological factors (e.g., social pressure), and (3) social factors (e.g., collaboration) influence the use of formative assessment. The prerequisites identified can inform professional development initiatives regarding formative assessment, as well as teacher education programs.

1. Introduction

Using assessment for a formative purpose is intended to guide students' learning processes and improve students' learning outcomes (Van der Kleij, Vermeulen, Schildkamp, & Eggen, 2015; Bennett, 2011; Black & Wiliam, 1998). Based on its promising potential for enhancing student learning (Black & Wiliam, 1998), formative assessment has become a "policy pillar of educational significance" (Van der Kleij, Cumming, & Looney, 2018, p. 620). Although there is still no clear consensus on what the term "formative assessment" encompasses (Van der Kleij et al., 2015; Bennett, 2011; Torrance, 2012; Wiliam, 2011), it is broadly accepted as a good classroom practice for teachers (Torrance, 2012).

1.1. Two formative assessment approaches

Different conceptualizations of formative assessment place different emphases on various aspects of the approach, stemming from different underlying theoretical perspectives (Van der Kleij et al., 2015; Baird, Hopfenbeck, Newton, Stobart, & Steen-Utheim, 2014; Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012). The core unifying characteristic is the focus on gathering evidence about student learning and using this evidence to guide student learning. To this end, feedback is recognised as a crucial aspect of formative assessment (Bennett, 2011; Black & Wiliam, 1998; Sadler, 1989; Stobart, 2008). Hattie and Timperley (2007) defined feedback as

* Corresponding author at: University of Twente, BMS/ELAN, P.O. Box 217, 7500 AE, Enschede, the Netherlands.
E-mail address: k.schildkamp@utwente.nl (K. Schildkamp).
[1] Current affiliation: School of Education, The University of Queensland, St Lucia, Queensland, 4072, Australia.
[†] July 31, 2018.

https://doi.org/10.1016/j.ijer.2020.101602
Received 2 January 2020; Received in revised form 7 April 2020; Accepted 17 May 2020
Available online 16 June 2020
0883-0355/ © 2020 The Authors. Published by Elsevier Ltd.

"information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (p. 81). Evans (2009) added to this definition that feedback may include all "exchanges generated within assessment design, occurring within and beyond the immediate learning context, being overt or covert (actively and/or passively sought and/or received), and importantly, drawing from a range of sources" (p. 71). Teachers can adapt their instruction to the needs of learners based on information derived from assessments as a form of feedback, to modify their teaching and/or provide feedback to students; students can use such feedback to steer their own learning processes directly (Bennett, 2011; Black & Wiliam, 1998; Sadler, 1989).

Two approaches to formative assessment are Data-Based Decision Making (DBDM) and Assessment for Learning (AfL) (Van der Kleij et al., 2015). These approaches can complement each other, and elements of each approach are often used by teachers in their classroom practice (Kippers, Wolterinck, Schildkamp, Poortman, & Visscher, 2018). A brief outline of each approach is provided next (for a more extensive analysis, see Van der Kleij et al., 2015).

1.2. Data-Based Decision Making (DBDM)

DBDM focuses on using data to achieve specific targets in the form of student learning outcomes and achievement (Wayman, Spikes, & Volonnino, 2013). Schildkamp and Kuiper (2010) defined DBDM as "systematically analysing existing 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" (p. 482). DBDM can take place at the level of the school, the classroom and the student. This review focuses on DBDM at the classroom and student levels. Driven by accountability pressures, many teachers internationally are expected to use data to inform decisions in their classroom, for example, with regard to instructional strategies (Ledoux, Blok, Boogaard, & Krüger, 2009; Wayman, Jimerson, & Cho, 2012). When teachers use DBDM effectively, this can lead to improved student learning and achievement (Poortman & Schildkamp, 2016; Carlson, Borman, & Robinson, 2011; Lai, Wilson, McNaughton, & Hsiao, 2014; van Geel, Keuning, Visscher, & Fox, 2016).

The data used in DBDM are collected in a systematic and formal manner and include both qualitative data (e.g., structured classroom observations) and quantitative data (e.g., periodic assessment results) (Wayman, Cho, Jimerson, & Spikes, 2012). DBDM often focuses on (standardized) assessment results as an important source of information for exploring how learning results can be improved. However, in recent years, the focus on data that are available more frequently has increased, as these allow for closer monitoring of student progress. Examples of these types of data include homework assignments, curriculum-embedded assessments, and structured observations from daily classroom practice (Ikemoto & Marsh, 2007).

DBDM is a systematic process and usually starts with a certain purpose, often taking the form of reducing the gap between the current and desired levels of student achievement. Teachers need to be able to identify measurable goals for all students in their classroom. Teachers then need to collect data to determine possible causes of this gap. These data need to be analyzed and interpreted, in order to determine actions that can be taken to reduce the gap, such as making instructional changes. New data need to be collected to evaluate the effectiveness of these instructional changes (Mandinach & Gummer, 2016; Mandinach, Honey, Light, & Brunner, 2008; Marsh, Pane, & Hamilton, 2006; Marsh, 2012).

1.3. Assessment for Learning (AfL)

AfL focuses on the quality of the learning process instead of on its outcomes (Stobart, 2008). It can take place at the level of the classroom and the student. Klenowski (2009) defined AfL 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" (p. 264). The information referred to in this definition is often collected in a less structured and more informal manner and can come from a range of different assessment sources, such as observations, portfolios, practical demonstrations, paper-and-pencil tests, peer-assessment, self-assessment, and dialogues (Gipps, 1994), which are used as a form of continual feedback to steer learning. The focus is on classroom interaction and dialogue in a process of discovering, reflecting, understanding, and reviewing (Hargreaves, 2005). The quality of AfL depends on the teacher's capability to identify usable evidence about student learning, make inferences about student learning, and translate this information into instructional decisions and feedback to students (Bennett, 2011). When used effectively, AfL can lead to increased student learning and achievement (Andersson & Palm, 2017; Fletcher & Shaw, 2012; Pinger, Rakoczy, Besser, & Klieme, 2018; Yin, Tomita, & Shavelson, 2013).

The key element of AfL is the ongoing interaction between learners and the teacher to meet learners' needs. AfL takes place in everyday classroom practice in the form of continual dialogues and feedback loops, in which (immediate) feedback is used to direct further learning (Stobart, 2008). Assessment is thus an integrated element of the learning process. Students play a crucial role in AfL, for example, through self- and peer-assessment, which can stimulate students' understanding of what and why they are learning (Elwood & Klenowski, 2002).

1.4. The role of the teacher in implementing formative assessment

Despite the evidence-based potential of formative assessment, various studies have pointed to its mixed effects in classroom practice (Baird et al., 2014; Furtak et al., 2016). A possible explanation for the lack of positive effects could be that teachers struggle with the use of formative assessment in their classrooms (Bennett, 2011), which is clearly a complex undertaking (Elwood, 2006). Research has suggested that many attempts to implement formative assessment have produced disappointing results because approaches were not used to their full potential, but were rather reduced to mechanistically applying a set of principles (Marshall &

Drummond, 2006; Swaffield, 2011). A key conclusion of several reviews of formative assessment (Black & Wiliam, 1998; Torrance, 2012) is that *how* teachers implement formative assessment is critical for its potential to enhance student learning. Teachers thus play a crucial role in formative assessment. However, the guidance available to teachers is often limited to generic principles (Van der Kleij et al., 2018; Elwood, 2006), or at times even inappropriate (Van der Kleij et al., 2018), resulting in limited use of formative assessment in practice (Torrance, 2012).

One of the problems in implementation of formative assessment is that often only certain 'principles' of formative assessment have been adopted, without much consideration of the broader implications for classroom practice (Elwood, 2006; Torrance, 2012). Formative assessment is not an add-on activity, but rather needs to be an integrated element of instruction, which requires a fundamental change in the role of the teacher in the classroom. It also requires a fundamental shift in the power relations between teachers and students, in which teachers and students become jointly responsible for the quality of teaching and learning in the classroom (Black & Wiliam, 1998).

It is still unknown what the requirements are for teachers to use formative assessment effectively in their classroom practice. Although several theoretical models have provided insights into what factors may be of importance (e.g., Heritage, 2007; Mandinach & Gummer, 2016), these models (1) focus on either AfL or DBDM approaches to formative assessment, and (2) are only partially based on empirical evidence. This study aims to address this gap, by reviewing the available evidence from the literature about prerequisites for teachers' use of formative assessment. This review sought to address the following research question: *What teacher prerequisites need to be in place for using formative assessment in their classroom practice?*

2. Method

2.1. Procedure

This review was part of a larger project, which aimed to identify prerequisites for formative assessment in classroom practice (Schildkamp, Heitink et al., 2014). The review used the methods for conducting systematic literature reviews in the social sciences described by Petticrew and Roberts (2006). After formulating our research question, we defined our search terms, selected literature databases, and started searching for publications. A library expert was consulted during the literature search process. Next, we formulated inclusion criteria, which formed the basis for selecting relevant publications. All relevant publications were read in full and a purposefully developed data extraction form was constructed to enable comparison of the same units of information from each selected publication. Finally, the results from publications judged to be of sufficient quality were synthesized to answer the research question.

2.2. Databases and search terms

Five databases (Education Resources Information Center [ERIC], Web of Science, Scopus, PsychINFO and Picarta) were systematically searched (early 2014) using the same search terms. Initial search terms included 'formative assessment', 'data-based decision making', 'assessment for learning', and related terms as found in a thesaurus and/or terms that were used in other relevant publications. To narrow down the results to publications relevant to formative assessment, the term 'feedback' was added to the search string. The search was further narrowed down by adding the term 'classroom' and related terms to the search string. The retrieved publications were exported to Endnote X6 for systematic selection using the inclusion criteria.

2.3. Inclusion criteria and data extraction

To arrive at a relevant selection of publications, we formulated the following inclusion criteria:

1 The study was published in a scientific peer-reviewed journal or was a PhD thesis. We did not include books, book chapters, reports, and conference proceedings in this review, as it is more difficult to establish the quality of these publications.
2 The study reported on research results. We did not include theoretical papers and reviews.
3 The study was conducted in primary and/or secondary education.
4 The study investigated (aspects of) formative assessment in classroom practice.
5 The study focused (at least in part) on the role of the teacher in implementing formative assessment

The data extraction form (Petticrew & Roberts, 2006) was trialed and modified multiple times to ensure usability and consistency of data extraction across the researcher team. The final data extraction form contained the following sections:

- General information, such as authors, title, and country;
- Research design specifics, such as research questions, instruments, and analysis methods;
- Research sample, such as number of schools, teachers, and students;
- Type(s) of formative assessment approach: DBDM and/or AfL;
- Results, such as the evidence with regard to the role of the teacher in formative assessment (i.e., the prerequisites);
- A set of 11 quality criteria based on Petticrew and Roberts (2006), such as suitable method(s), sampling, and data analysis, to enable selection of high-quality studies.

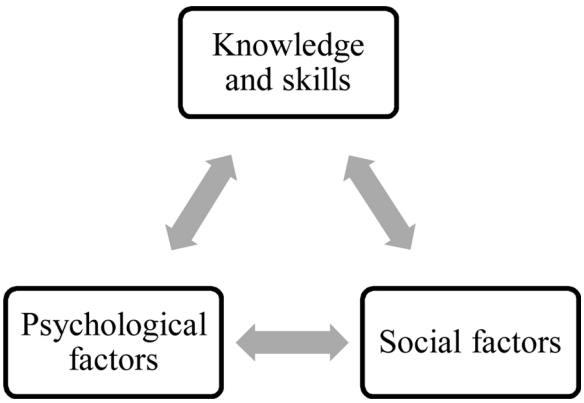


Fig. 1. Three categories of teacher prerequisites for formative assessment.

The reliability of the data extraction process was safeguarded by having two researchers independently code approximately 50 % of the selected publications. Agreement rates of 80 % and a Cohen's Kappa of .620, demonstrated satisfactory inter-coder reliability (Landis & Koch, 1977). Each of the selected publications was scored on 11 quality criteria (see Schildkamp, Heitink et al., 2014). A score of 0, 0.5 or 1 was assigned to each criterion. As we wanted to base this review on high-quality studies only, a publication had to score an average of 7 or higher to be included. Publications with a score between 5 and 7 were discussed between at least two researchers, and publications with a score lower than 5 were excluded. When a study was coded by multiple researchers, the average of the quality score was taken.

2.4. Data analysis

The results were organized around three categories of prerequisites established inductively from the reviewed studies. This process of organizing the data around these categories of prerequisites was discussed with multiple researchers, to overcome the bias that can develop when only one researcher does the process of analysis (Poortman & Schildkamp, 2011; Green, Johnson, & Adams, 2006). We identified the following three categories of teacher prerequisites, which all influence each other: (1) knowledge and skills, (2) psychological factors, and (3) social factors, see Fig. 1.

2.5. Search and selection results

An overview of the search process and results in the form of a PRISMA diagram (Moher, Liberati, Tetzlaff, & Altman, 2009) is provided in Fig. 2.

At the end of the literature identification process, 200 publications were deemed suitable for data extraction. However, screening of the publications led to the realization that we had missed publications in the field of DBDM. For this reason, we conducted an additional DBDM-specific search. This search was conducted without 'feedback' in the search string, as experts in the field had indicated that this specific term is often not used.

A total of 256 publications that were available in full text were screened for relevance using the inclusion criteria. In this screening stage, it was concluded that 125 publications did not meet the inclusion criteria after all (for example, because they focused on the use of data by school leaders or on evaluations at the school level). This left a total of 131 publications suitable for data extraction. Based on the data extraction, 77 publications were either found to be of insufficient quality or lacked too much information from the method section to judge the quality of the study. These publications were removed from the selection. The remaining 54 publications were included in this review (see Table 1).

2.6. Characteristics of selected studies

The selected publications were classified as focusing on either DBDM or AfL. When studies involved a mix of DBDM and AfL, they were classified according to the dominant approach. A total of 29 studies focused on DBDM (e.g., the use of standardized tests, the use of self-evaluation results, and the use of evidence from systematic observations). The DBDM studies were predominantly conducted in primary education (n = 22). Most of these studies were conducted in the US (n = 16) and most were of a qualitative nature (n = 18). The 25 AfL studies focused on different types of assessment occasions, ranging from discussions in the classroom to formal types of assessment such as paper-and-pencil tests. AfL studies were conducted in both primary education (n = 14) and/or secondary education (n = 17). Most AfL studies used either a qualitative design (n = 12) or a mixed method design (n = 9) (see Table 2 for more information).

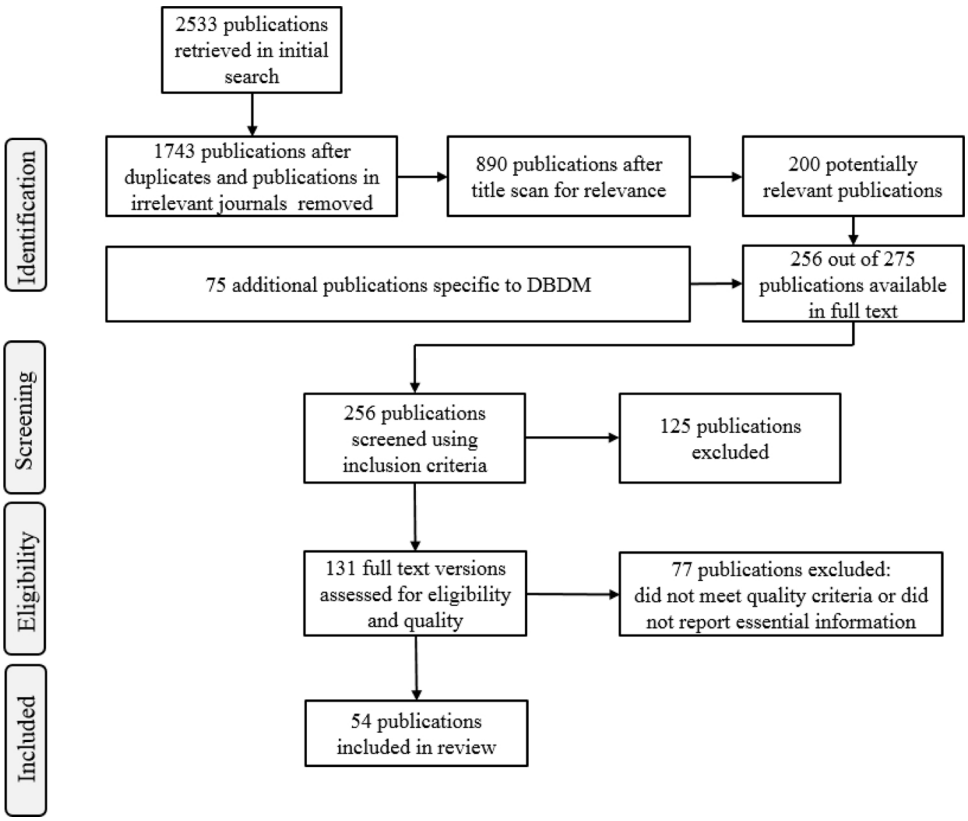


Fig. 2. PRISMA diagram of literature search and selection process.

Table 1
Final selection of publications for review.

	Initial search	DBDM search	Total
Number of publications found for data extraction	200	75	275
Number of publications not found in full text	(15)	(4)	(19)
Number of publications removed based on inclusion criteria after reading the full text	(98)	(27)	(125)
Number of publications removed based on quality criteria	(70)	(7)	(77)
Total	17	37	54

3. Results

Table 3 presents an overview of the studies that investigated specific influential factors related to the role of the teacher (see Table 2). When interpreting these results, it is important to consider that just because certain factors have been studied more than others, that does not imply that these factors are more important.

3.1. Teacher prerequisites: knowledge and skills

3.1.1. Data literacy and assessment literacy

Many studies found that data literacy is important for DBDM (Blanc et al., 2010; Brown, De Four-Babb, Bristol, & Conrad, 2014; Christoforidou, Kyriakides, Antoniou, & Creemers, 2014; Datnow, Park, & Kennedy-Lewis, 2012; Fuchs, Fuchs, Karns, Hamlett, & Katzeroff, 1999; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Levin & Datnow, 2012; McNaughton, Lai, & Hsiao, 2012; Schildkamp, Karbautzki, & Vanhoof, 2014; Schildkamp & Kuiper, 2010; Schildkamp, Rekers-Mombarg, & Harms, 2012; Schildkamp & Visscher, 2010a, 2010b; Schildkamp, Visscher, & Luyten, 2009; Van der Kleij & Eggen, 2013; Young, 2006). Several studies found assessment literacy to be important for AfL (Birenbaum, Kimron, & Shilton, 2011; Bryant & Carless, 2010; Gottheiner & Siegel, 2012; Lee, 2011; Lee, Feldman, & Beatty, 2012). Assessment literacy encompasses knowledge and skills with regard to the entire assessment process, from collecting information on student learning to making instructional changes based on that information. Data literacy is broader, including assessment literacy as well as the collection, analysis, and use of other types of data, such as student satisfaction surveys, and background information about students (Mandinach & Gummer, 2016).

Table 2
Overview of studies included in this review.

Reference	Educational setting[1]		FA approach	Context				Sample size	
	PE	SE	DBDM or AfL	Country[2]	Subject	Research design[3]	Schools	Teachers	Students
1. Aschbacher and Alonzo (2006)	x		AfL	US	Science	MM	n/a[4]	25	245
2. Birenbaum et al. (2011)	x	x	AfL	IL	Humanities and arts	MM	n/a	128	22 (QL)
3. Blanc et al. (2010)	x		DBDM	US	n/a[4]	QL	5	150	n/a[4]
4. Brown et al. (2014)	x		DBDM	TT	Math, language, social studies, science	MM	20	143	n/a
5. Bryant and Carless (2010)	x		AfL	HK	Language	QL	1	2	34
6. Christoforidou et al. (2014)	x		DBDM	CY	Math	QT	n/a	178	2358
7. Datnow et al. (2012).		x	DBDM	US	n/a	QL	4	90	n/a
8. Datnow et al. (2013)	x		DBDM	US	n/a	QL	6	76	n/a
9. Farley-Ripple and Buttram (2014)	x		DBDM	US	n/a	QL	4	140	n/a
10. Feldman and Capobianco (2008)		x	AfL	US	Science	QT	n/a	8	n/a
11. Fletcher and Shaw (2012)	x		AfL	AU	Writing	MM	1	16	256
12. Fox-Turnbull (2006)	x		AfL	NZ	Humanities	QL	6	n/a	53
13. Fuchs et al. (1999)	x		DBDM	US	Math	QL	4	16	272
14. Furtak and Ruiz-Primo (2008)		x	AfL	US	Science	QL	1	6	n/a
15. Gamlem and Smith (2013)		x	AfL	NO	n/a	QL	4	6	150
16. Gottheiner and Siegel (2012)		x	AfL	SE	Science	QL	n/a	5	n/a
17. Hargreaves (2013)	x		AfL	UK	Numeracy and literacy	QL	1	1	9
18. Harris and Brown (2013)	x	x	AfL	NZ	Language and math	QL	1	3	99
19. Harris et al. (2014)	x	x	AfL	NZ	n/a	MM	11	13	193
20. Hubbard et al. (2014)	x		DBDM	US	n/a	QL	1	8	n/a
21. Havnes et al. (2012)		x	AfL	NO	Language and math	MM	5	192	391
22. Jimerson (2014)	x		DBDM	US	n/a	QL	4	118	n/a
23. Kay and Knaack (2009)		x	AfL	CA	Science	MM	6	7	213
24. Kennedy and Datnow (2011)		x	DBDM	US	n/a	QL	4	n/a	n/a
25. Kerr et al. (2006)		x	DBDM	US	n/a	QL	72	118	n/a
26. Lachat and Smith (2005)		x	DBDM	US	n/a	QL	5	n/a	n/a
27. Lee (2011)		x	AfL	HK	Language	MM	1	4	138
28. Lee et al. (2012)		x	AfL	US	Math and science	QL	2	18	n/a
29. Levin and Datnow (2012)		x	DBDM	US	n/a	QL	1	20	n/a
30. McNaughton et al. (2012)	x		DBDM	NZ	Reading	QT	39	340	671
31. Newby and Winterbottom (2011)		x	AfL	UK	Science	QL	1	n/a	157
32. Ní Chróinín and Cosgrave (2013)	x		AfL	IE	Physical education	QL	n/a	5	n/a
33. O'Loughlin et al. (2013)	x		AfL	IE	Physical education	QL	1	1	22
34. Park and Datnow (2009)	x	x	DBDM	US	n/a	QL	8	70	n/a
35. Penuel et al. (2007)	x	x	AfL	US	Science	QT	n/a	498	n/a
36. Peterson and Irving (2008)		x	AfL	NZ	Math and English	QT	4	n/a	41
37. Rakoczy et al. (2008)		x	AfL	DE, CH	Math	MM	n/a	n/a	1255
38. Ruiz-Primo and Furtak (2006)	x	x	AfL	US	Science	MM	4	4	99
39. Sach (2013).	x	x	AfL	UK	n/a	QL	3	3	n/a
40. Schildkamp and Teddlie (2008)	x		DBDM	NL, US	n/a	QT	85	284	n/a
41. Schildkamp and Visscher (2009)	x		DBDM	NL	n/a	MM	65	n/a	n/a
42. Schildkamp and Visscher (2010b)	x	x	DBDM	US	n/a	QL	5	n/a	n/a
43. Schildkamp et al. (2009)	x		DBDM	NL	Math and spelling	QT	55	n/a	2431
44. Schildkamp and Kuiper (2010)		x	DBDM	NL	Science	QL	6	32	n/a
45. Schildkamp and Visscher (2010a)	x		DBDM	NL	n/a	MM	79	56	n/a
46. Schildkamp et al. (2012)		x	DBDM	NL	Science	QL	6	32	n/a
47. Schildkamp, Karbautzki et al. (2014)	x	x	DBDM	NL, PL, UK, DE, LV	n/a	QL	16	86	n/a
48. Staman et al. (2014)	x		DBDM	NL	n/a	QT	49	168	n/a
49. Sutherland (2004)	x		DBDM	US	n/a	QL	1	58	n/a

(*continued on next page*)

Table 2 (*continued*)

Reference	Educational setting[1]		FA approach	Context				Sample size	
	PE	SE	DBDM or AfL	Country[2]	Subject	Research design[3]	Schools	Teachers	Students
50. Van der Kleij and Eggen (2013)	x		DBDM	NL	n/a	MM	56	97	n/a
51. Vanhoof et al. (2012)	x		DBDM	BE	Math, reading, and spelling	QT	183	2579	n/a
52. Wayman, Cho et al. (2012), Wayman, Jimerson et al. (2012)	x	x	DBDM	US	n/a	MM	19	198	n/a
53. Yin et al. (2013)	x		AfL	US	Science	QT	1	1	52
54. Young (2006)	x		DBDM	US	n/a	QL	4	n/a	n/a

Note. [1]Primary Education (PE), Secondary Education (SE).[2]Country codes according to ISO: United States of America (US), Trinidad and Tobago (TT), Netherlands (NL), Belgium (BE), Cyprus (CY), Poland (PL), Latvia (LV), Israel (IL), Hong Kong (HK), Taiwan (TA), Australia (AU), New Zealand (NZ), Norway (NO), United Kingdom (UK), Canada (CA), Ireland (IE), Germany (DE), Switzerland (CH).[3]Quantitative research design (QT), Qualitative research design (QL), Mixed methods (MM). *4*n/a = Not applicable or information not available.

Table 3
Overview of studies that investigated influential factors.

	DBDM		AfL		
Factor	Studies[a]	*n*	Studies[a]	*n*	Total
Data/assessment literacy	3, 4, 6, 7, 13, 25, 29, 30, 43, 44, 46, 47, 49, 50, 54	15	2, 5, 10, 13, 16, 23, 27, 28, 32, 35, 53	11	**26**
Collaboration	3, 4, 7, 8, 9, 20, 22, 26, 29, 30, 34, 41, 43, 44, 45, 46, 52, 54	18	2, 5, 10, 13, 21, 23, 27, 39	8	**26**
Attitude	7, 22, 25, 40, 41, 42, 44, 45, 46, 47, 50, 51, 52	13	2, 10, 21, 27, 28	5	**18**
PCK	3, 7, 13, 30	4	1, 2, 10, 11, 12, 13, 16, 19, 23, 27, 35, 53	12	**16**
Goal setting	29, 36, 41, 42, 44, 46, 47, 51	8	2, 17, 21, 31, 35, 36	6	**14**
Feedback	–	0	1, 5, 12, 13, 14, 15, 17, 21, 27, 32, 36, 37	12	**12**
Facilitating classroom discussions	–	0	1, 5, 10, 14, 15, 16, 17, 21, 27, 28, 35, 38	12	**12**
Involving students	24	1	5, 11, 15, 18, 21, 23, 28, 31, 33, 37	10	**11**
ICT skills	7, 40, 41, 42, 45, 48	6	1, 10, 28, 35	4	**10**
Ownership	8, 29, 41, 42, 44, 46, 47	7	1, 39	2	**9**
Social pressure	20, 22, 40, 41, 45, 49	6	1, 28	2	**8**
Perceived control	25, 40, 42, 44, 46	5	2, 39	2	**7**

[a] *Note:* Table 2 identifies the specific studies associated with the reference numbers.

Data literacy and assessment literacy both include knowledge and skills concerning *data collection*. To be able to use formative assessment in the classroom, teachers must be able to collect different types of data. This includes assessment evidence about student achievement, but in the case of DBDM, it also includes data such as information about student characteristics and structured classroom observation data (Kerr et al., 2006). Furthermore, the use of formative assessment requires knowledge and skills related to constructing and using a range of assessment instruments, such as paper-and-pencil tests and homework assignments (Christoforidou et al., 2014; Feldman & Capobianco, 2008; Gottheiner & Siegel, 2012; Kay & Knaack, 2009; Ní Chróinín & Cosgrave, 2013; Yin et al., 2013), and the knowledge and skills to critically evaluate these assessment instruments (Gottheiner & Siegel, 2012).

Data literacy and assessment literacy also include the knowledge and skills to *analyze and interpret* different types of data, such as assessment evidence and classroom observations (Brown et al., 2014; Kerr et al., 2006; Lee et al., 2012; Levin & Datnow, 2012; Schildkamp & Kuiper, 2010; Schildkamp & Visscher, 2010a; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014, 2010b; Schildkamp et al., 2009; Van der Kleij & Eggen, 2013; Young, 2006). Van der Kleij and Eggen (2013), for example, found that some teachers do not possess the basic skills needed to interpret score reports from a pupil monitoring system in primary education. If teachers do not know how to analyze and interpret the collected data, this will lead to a lack of or even inappropriate (e.g., providing students with inaccurate feedback) use of formative assessment.

Further, teachers need to be able to *transform data into information* based on their analysis and interpretation of the data. This step concerns identifying student learning needs (Blanc et al., 2010; Schildkamp & Kuiper, 2010; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014; Young, 2006), and determining appropriate actions to take in the classroom, such as re-teaching certain content, grouping students differently, or differentiating instruction (Blanc et al., 2010; Brown et al., 2014; Datnow et al., 2012; Datnow, Park, & Kennedy-Lewis, 2013; Feldman & Capobianco, 2008; Fuchs et al., 1999; Gottheiner & Siegel, 2012; Kay & Knaack, 2009; Kerr et al., 2006; Lee, 2011; Levin & Datnow, 2012; McNaughton et al., 2012; Penuel, Boscardin, Masyn, & Crawford, 2007; Schildkamp & Kuiper, 2010; Schildkamp & Visscher, 2010a; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014, 2010b; Schildkamp et al., 2009; Young, 2006).

3.1.2. Pedagogical content knowledge

Fifteen studies concluded that teachers need pedagogical content knowledge (PCK) to be able to implement DBDM or AfL (Aschbacher & Alonzo, 2006; Birenbaum et al., 2011; Blanc et al., 2010; Datnow et al., 2012; Feldman & Capobianco, 2008; Fletcher & Shaw, 2012; Fox-Turnbull, 2006; Fuchs et al., 1999; Gottheiner & Siegel, 2012; Harris, Brown, & Harnett, 2014; Kay & Knaack, 2009; Lee, 2011; McNaughton et al., 2012; Penuel et al., 2007; Yin et al., 2013). PCK refers to subject-matter content knowledge, as well as knowledge about how to teach subject-matter knowledge. Assessment evidence can help teachers to identify students' misconceptions. Teachers then need PCK to determine how to provide feedback to students and/or alter their instruction. Gottheiner and Siegel (2012), for example, found that teachers must have knowledge of common misconceptions within the subject. Aschbacher and Alonzo (2006) found that without sufficient content knowledge, teachers are not able to provide students with accurate and complete feedback on their learning and achievement.

3.1.3. Goal setting

Several studies found that goal setting can promote the use of DBDM or AfL (Birenbaum et al., 2011; Hargreaves, 2013; Havnes, Smith, Dysthe, & Ludvigsen, 2012; Levin & Datnow, 2012; Newby & Winterbottom, 2011; Penuel et al., 2007; Peterson & Irving, 2008; Schildkamp & Kuiper, 2010; Schildkamp & Visscher, 2009; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014, 2010b; Vanhoof, Verhaeghe, Van Petegem, & Valcke, 2012). In the case of DBDM, this can refer to goals at the level of the school, the classroom, or the individual student. In the case of AfL, the focus is on student learning goals. When the goals are clear and measurable for both teachers and students, (assessment) data can provide teachers and students with feedback on their progress with regard to these goals (Schildkamp & Kuiper, 2010; Schildkamp & Visscher, 2009; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014, 2010b). Peterson and Irving (2008) found that it is important that teachers and students formulate individual student learning goals together. According to these researchers, this encourages students to reflect and act on feedback.

3.1.4. Feedback

A factor identified in numerous AfL studies, although not found in the DBDM literature, concerns skills related to providing feedback to students (Aschbacher & Alonzo, 2006; Bryant & Carless, 2010; Fox-Turnbull, 2006; Furtak & Ruiz-Primo, 2008; Gamlem & Smith, 2013; Hargreaves, 2013; Havnes et al., 2012; Lee, 2011; Ní Chróinín & Cosgrave, 2013; Peterson & Irving, 2008; Rakoczy, Klieme, Bürgermeister, & Harks, 2008). The following aspects were found to be important for teacher feedback to facilitate student learning:

- Timing of feedback: Think about how and when students can use feedback (Bryant & Carless, 2010), and provide time and opportunity for students to use the feedback (Gamlem & Smith, 2013).
- Type of feedback: Teachers can provide feedback in terms of performance, effort, or achievement (Gamlem & Smith, 2013). With regard to providing students with feedback on their achievement in the form of grades, Peterson and Irving (2008) found that providing students with a grade was essential for their motivation. However, Lee (2011) found that giving grades negatively influenced students' willingness to take risks and learn new things. Birenbaum et al. (2011) also found that a focus on assigning grades and preparing for tests can hinder AfL. Teachers can also provide students with process feedback, focused on strategies to bridge the gap between where students are and where they need to be (Gamlem & Smith, 2013; Havnes et al., 2012; Lee, 2011; Peterson & Irving, 2008). Such strategies may be explicit or more implicit, for example, in the form of cues (Gamlem & Smith, 2013; Hargreaves, 2013; Rakoczy et al., 2008).
- Content: High-quality feedback is needed that is honest, concise and yet sufficiently detailed (Fox-Turnbull, 2006; Hargreaves, 2013; Peterson & Irving, 2008).

3.1.5. Facilitating classroom discussions

Another factor not found in the DBDM literature, but stressed in the AfL literature, is that teachers need to be able to facilitate classroom discussions. This includes teacher questioning to elicit evidence about student learning (Aschbacher & Alonzo, 2006; Feldman & Capobianco, 2008; Fox-Turnbull, 2006; Furtak & Ruiz-Primo, 2008; Gottheiner & Siegel, 2012; Gamlem & Smith, 2013; Hargreaves, 2013; Havnes et al., 2012; Lee et al., 2012; Penuel et al., 2007; Ruiz-Primo & Furtak, 2006). For example, Furtak and Ruiz-Primo (2008) found that it is important that teachers are able to ask the right questions at the right time (planned and unplanned formative assessment). Ruiz-Primo and Furtak (2006) found that it is important that teachers ask 'why/how' questions to get more information about student understanding. They found that discussions with students help to make students' levels of knowledge explicit. Havnes et al. (2012) called these "mutual learning dialogues" (p. 26).

3.1.6. ICT skills

The use of DBDM increasingly requires ICT skills as found in several studies, for example, regarding how to use information management systems and assessment systems (Datnow et al., 2012; Staman, Visscher, & Luyten, 2014), and/or performance feedback tools, such as school self-evaluation systems (Schildkamp & Teddlie, 2008; Schildkamp & Visscher, 2009, 2010b). AfL also requires certain ICT skills as indicated in four studies, for example, with regard to how to use certain digital assessment systems and tools (Aschbacher & Alonzo, 2006; Feldman & Capobianco, 2008; Lee et al., 2012; Penuel et al., 2007).

3.2. Social factors

3.2.1. Collaboration between teachers

Many studies pointed to the importance of collaboration between teachers in regard to DBDM (Blanc et al., 2010; Brown et al., 2014; Datnow et al., 2012, 2013; Farley-Ripple & Buttram, 2014; Hubbard, Datnow, & Pruyn, 2014; Jimerson, 2014; Lachat & Smith, 2005; Levin & Datnow, 2012; McNaughton et al., 2012; Park & Datnow, 2009; Schildkamp & Kuiper, 2010; Schildkamp & Visscher, 2009; Schildkamp et al., 2012, 2010a, 2010b; Wayman, Cho et al., 2012; Young, 2006) and some also found this for AfL (Birenbaum et al., 2011; Bryant & Carless, 2010; Feldman & Capobianco, 2008; Havnes et al., 2012; Kay & Knaack, 2009; Lee, 2011; Sach, 2013). Regarding the nature of collaboration, the following aspects were frequently mentioned, which are closely related to data and assessment literacy:

- Discussing school-, team-, or student-level goals (Farley-Ripple & Buttram, 2014);
- Analyzing and interpreting data (Datnow et al., 2013; Farley-Ripple & Buttram, 2014; Feldman & Capobianco, 2008; McNaughton et al., 2012; Schildkamp & Kuiper, 2010; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014);
- Transforming data into information through discussing and making sense of data/assessment results (Datnow et al., 2012, 2013; Feldman & Capobianco, 2008; Schildkamp & Kuiper, 2010; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014; Young, 2006);
- Discussing and developing plans for action, for example, instructional plans or strategies (Datnow et al., 2012, 2013; Farley-Ripple & Buttram, 2014; Feldman & Capobianco, 2008; Levin & Datnow, 2012; Park & Datnow, 2009; Schildkamp & Kuiper, 2010; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014; Young, 2006).

3.2.2. Involving students

Only one study discussed the importance of involving students in DBDM. Kennedy and Datnow (2011) focused on discussing assessment results with students, and concluded that most schools do not involve students in the use of data. However, students were involved in the analysis of their test scores in some exemplary schools.

Several studies pointed to the importance of involving students in AfL (Fletcher & Shaw, 2012; Havnes et al., 2012; Rakoczy et al., 2008). Harris and Brown (2013) found that good classroom relations between teachers and students are essential, and making mistakes should be viewed as an opportunity to learn by both teachers and students. Fletcher and Shaw (2012) suggested that teachers can involve students in the process of AfL by giving them specific responsibilities, for example, by letting students set their own learning goals and learning paths. According to Havnes et al. (2012), students appreciated such involvement in their own learning.

Several studies focused on peer- and self-assessment as a way of involving students (Bryant & Carless, 2010; Gamlem & Smith, 2013; Harris & Brown, 2013; Kay & Knaack, 2009; Lee, 2011; Newby & Winterbottom, 2011; O'Loughlin, Ní Chróinín, & O'Grady, 2013). Bryant and Carless (2010) emphasized the importance of the teacher knowledge and skills required for using peer- and self-assessment. Harris and Brown (2013) found that it is important that teachers explicitly articulate a rationale for the use of peer- and self-assessments, and communicate this with students. Further, teachers need to provide students with peer- and self-assessment criteria and help students in applying these. The effective use of peer assessment also requires that teachers teach students how to provide useful feedback (Newby & Winterbottom, 2011).

3.3. Psychological factors

3.3.1. Attitude/beliefs

A negative attitude can hinder the use of DBDM or AfL (Birenbaum et al., 2011; Datnow et al., 2012; Feldman & Capobianco, 2008; Havnes et al., 2012; Lee, 2011; Lee et al., 2012; Schildkamp & Kuiper, 2010; Schildkamp & Visscher, 2010a; Schildkamp et al., 2012, 2010b; Wayman, Cho et al., 2012). For example, if teachers are resistant toward data use, do not believe that it can lead to improvement, do not believe that data actually reflect on their teaching, and would rather rely on their experiences and intuition, this can hinder data use in schools (Datnow et al., 2012; Schildkamp & Kuiper, 2010; Schildkamp & Visscher, 2010b; Schildkamp et al., 2012; Wayman, Cho et al., 2012).

In contrast, a positive attitude can enable the use of DBDM or AfL. Birenbaum et al. (2011), Penuel et al. (2007), Rakoczy et al. (2008), and Sach (2013) found that a more constructivist teacher stance (e.g., students should become autonomous and are capable of learning on their own) can enable formative assessment use. Furthermore, buy-in and belief in the use of data are important. Teachers should believe that the use of data can improve the quality of their classroom practice (Jimerson, 2014; Kerr et al., 2006; Schildkamp & Kuiper, 2010; Schildkamp & Visscher, 2009; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014, 2010b; Van der Kleij & Eggen, 2013; Vanhoof et al., 2012; Wayman, Cho et al., 2012). A positive attitude also implies that teachers are not afraid to make changes based on data (Schildkamp & Teddlie, 2008; Schildkamp & Visscher, 2009, 2010b).

3.3.2. Ownership

Ownership over assessment results and student learning can also influence DBDM or AfL implementation. The degree to which teachers feel that they have autonomy to make decisions relates to the ownership they feel over assessment results. Low self-efficacy and lack of ownership (e.g., the quality of my teaching is not reflected in the students' assessment results) can hinder DBDM or AfL (Aschbacher & Alonzo, 2006; Datnow et al., 2013; Levin & Datnow, 2012; Sach, 2013; Schildkamp & Kuiper, 2010; Schildkamp &

Visscher, 2009; Schildkamp et al., 2012; Schildkamp, Karbautzki et al., 2014, 2010b). Teachers need to feel that they are responsible for student learning and achievement in their school, and not just for covering the curriculum (Aschbacher & Alonzo, 2006; Sach, 2013).

3.3.3. Social pressure and perceived control

Social pressure is related to perceived control and autonomy. Several DBDM studies found that a certain amount of social pressure can enable the use of DBDM (Schildkamp & Kuiper, 2010; Schildkamp & Teddlie, 2008; Schildkamp & Visscher, 2010b; Schildkamp et al., 2012). In these cases, social pressure by means of encouragement from the principal enabled the use of DBDM. Sometimes a lot of social pressure, for example, from the accountability system, limits the perceived control that teachers feel they have. Teachers need to feel that they have sufficient autonomy to make changes in instruction and the curriculum based on data, in DBDM (Kerr et al., 2006; Schildkamp & Kuiper, 2010; Schildkamp & Teddlie, 2008; Schildkamp & Visscher, 2010b; Schildkamp et al., 2012) and AfL (Birenbaum et al., 2011; Sach, 2013). Schildkamp and Teddlie (2008), for example, found that the degree of autonomy (e.g., the extent to which teachers feel that they can take measures based on the data) influenced teachers' data use.

In several DBDM studies, teachers indicated they were only using data because they felt forced to do so or felt pressured by the accountability system (e.g., social pressure), and not because they believed data use to be important for improving classroom practice (Hubbard et al., 2014; Jimerson, 2014; Schildkamp & Teddlie, 2008; Schildkamp & Visscher, 2009, 2010b; Sutherland, 2004). In two AfL studies (Aschbacher & Alonzo, 2006; Lee et al., 2012) it was also found that teachers felt pressured by the accountability system. In the DBDM studies, data use was often linked to accountability and high stakes testing, and not to improvement (Hubbard et al., 2014; Jimerson, 2014; Schildkamp & Teddlie, 2008; Schildkamp & Visscher, 2009, 2010b; Sutherland, 2004). For example, Sutherland (2004) concluded that data use is sometimes seen as something that is done *to* the school instead of done *by and for* the school. A study by Hubbard et al. (2014) showed that educators were likely to use data only in subjects in which there are regular benchmark assessments for accountability purposes, such as language and mathematics. Furthermore, too much social pressure sometimes led to overly prescriptive feedback, where the students were required to directly copy information from the blackboard (Aschbacher & Alonzo, 2006). Lee et al. (2012) found that teachers sometimes felt pressured by curricular constraints, which hindered their use of AfL.

4. Discussion

As noted previously, the DBDM and AfL approaches to formative assessment (Van der Kleij et al., 2015) differ with respect to the types of assessment instruments used, the frequency with which they are applied in the classroom, and their relevance at various stages in the learning process. As a result, these two formative assessment approaches require different things from teachers. To address these differences, this review distinguishes between DBDM and AfL when reviewing evidence about critical prerequisites for the teachers' role in formative assessment.

4.1. The role of the teacher in formative assessment

Formative assessment (DBDM and AfL) is not just about the evidence collected; it is mostly about how this evidence is used by teachers and students to influence student learning. As concluded by Black and Wiliam (1998), the teacher plays a fundamental role in formative assessment, and formative assessment can only lead to increased student learning if it is adequately implemented by teachers. In other words, formative assessment is only as effective as the teacher who implements it (Evans, 2009). Therefore, our main research question was: What teacher prerequisites need to be in place for using formative assessment in their classroom practice?

In this review, various prerequisites related to knowledge and skills, psychological factors and social factors were found to influence the use of formative assessment by teachers. Based on the results, Fig. 3 illustrates a conceptual model displaying the various prerequisite categories and their hypothesized relations with each other. For example, teachers with a negative attitude towards formative assessment (psychological factor) are not likely to work on their data and assessment literacy (knowledge and skills). Teachers who collaborate with other teachers and students (social factor) are likely to learn from such interactions (knowledge and skills).

4.2. Data literacy

Regarding teachers' *knowledge and skills*, important factors are adequate levels of data literacy, assessment literacy, pedagogical content knowledge, skills with regard to goal setting, providing feedback, facilitating classroom discussion, and ICT skills. Although discussed as separate skills in this review, several of these skills can be combined in the overarching meta-construct of data literacy (Beck & Nunnaley, 2020; Mandinach & Gummer, 2011, 2013, 2016). Data literacy consists of several subconstructs, as identified by Mandinach and Gummer in their data literacy framework (2011, 2013, 2016). These subconstructs include assessment literacy, pedagogical content knowledge, goal setting, providing feedback, collecting different types of data (including moment-to-moment data, such as information collected based on classroom discussion), as well as the ICT skills needed to store, collect, and analyze data (Beck & Nunnaley, 2020; Mandinach & Gummer, 2011, 2013, 2016).

Pedagogical content knowledge was found to play an important role, for both DBDM and AfL, as it enables teachers to contextualize data within the content domain and its learning stages. Teachers need to understand what the data mean in relation to the goals,

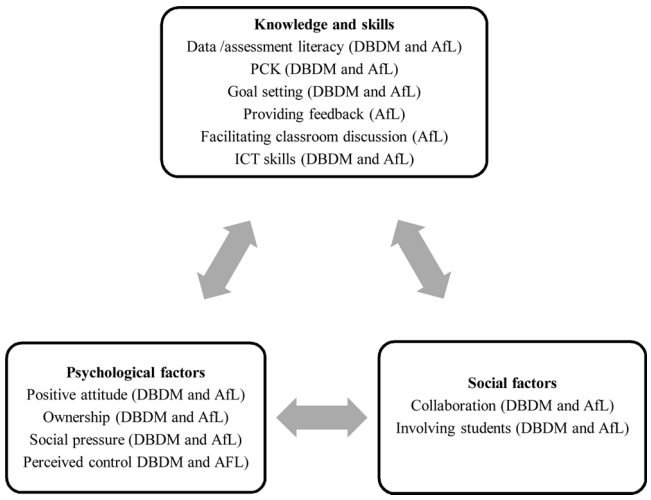


Fig. 3. Critical teacher prerequisites for strengthening formative assessment in classroom practice.

learning objectives and criteria for success of the content domains. Then, they can determine what instructional steps to take or what feedback to provide.

For teachers to be able to determine the next instructional steps, it is important that they have *set clear goals*. This is an important subconstruct in the Mandinach and Gummer (2011, 2013, 2016) data literacy framework, as well as the first aspect of the data literacy continuum developed by Beck and Nunnaley (2020). Teachers should share learning objectives and criteria for success with students during the lesson, so that both students and teachers are aware of the learning goals, and how it will be determined whether students have achieved these goals. Moreover, the AfL literature has emphasized the need to involve students in developing their own goals and criteria for success (Black & Wiliam, 2009; Wiliam & Leahy, 2015). Although it did not emerge as a finding from the empirical literature in this review, several authors have highlighted the importance of involving students in setting their own learning goals in DBDM (e.g., Hamilton et al., 2009).

Data literacy (Beck & Nunnaley, 2020; Mandinach & Gummer, 2011, 2013, 2016) also involves collecting, managing, and organizing a variety of high-quality data. For example, Mandinach and Gummer's literacy framework (2011, 2013, 2016) includes the ability of educators to collect and organize a wide variety of data; not only assessment data, but also data such as behavioral and affective data, in order to more holistically analyze academic growth at the student, classroom, and school levels. As emphasized in the AfL literature, this also includes collecting data by means of *facilitating classroom discussions*. For example, teachers can ask open-ended questions that require students to think critically. Answers to these questions will provide in-depth information on student learning, which teachers can use formatively (Wiliam & Leahy, 2015).

Furthermore, it is essential that the collected data are analyzed (turning data into information) and transformed into decisions (Beck & Nunnaley, 2020; Mandinach & Gummer, 2011, 2013, 2016), so that teachers can *provide feedback to students*. For example, feedback can suggest to students how to move their learning forward (Van der Kleij et al., 2015; Sadler, 1989).

Finally, data literacy includes *ICT* skills (Beck & Nunnaley, 2020; Mandinach & Gummer, 2011, 2013, 2016), such as knowing how to work with digital assessment and data systems. However, we want to stress here that although data availability and access may be facilitated by some sort of data or assessment system, it is crucial that these data are perceived as relevant, reliable, and valid by teachers (Wayman, Jimerson et al., 2012). Moreover, ICT skills are an important condition, but they are not sufficient to ensure actual data use in schools on their own (Cho, Allwarden, & Wayman, 2016; Hamilton et al., 2009; Wayman, Jimerson et al., 2012).

4.3. Social factors

Social factors also play a role in teachers' use of formative assessment. Relationships between teachers, as well as between teachers and their students, are vital. These relationships, or social networks, are important because they facilitate the exchange of resources such as information, knowledge, and advice (Daly, 2010). *Collaboration* with colleagues is an important prerequisite for teachers' use of formative assessment, for example, through engaging in discussions regarding how to improve classroom practices based on assessment results. Relationships between teachers and students also play an important role in AfL. For example, teachers can *involve students* by involving them in the process of formative assessment by using forms of peer- and self-assessment in the classroom. This can lead to increased self-regulation and improved learning outcomes (Black & Wiliam, 2009; Wiliam & Leahy, 2015). Although limited empirical evidence was found for involving students in DBDM, some publications (e.g., Hamilton et al., 2009) did emphasize the importance of involving students in DBDM, for example, by teaching students to examine their own data.

4.4. Psychological factors

Finally, the following *psychological factors* can enable teacher use of formative assessment in the classroom. First, it is important that teachers have a *positive attitude* toward the use of formative assessment, and believe that formative assessment can make a difference to their classroom practice and student learning. Second, the degree to which teachers feel *ownership* over the process and results of formative assessment matters. Furthermore, *social pressure* plays a role. When teachers feel too much (accountability) pressure from their district leaders, for example, this may hinder their use of formative assessment. If there is too much social pressure, the focus is often on summative assessment and meeting certain benchmarks. However, a certain degree of social pressure, for example, pressure from the principal to use data, can actually enable the use of formative assessment. Finally, it is important that teachers *perceive control* over what happens in the classroom. They need to feel that they have sufficient autonomy to make decisions about the curriculum, assessment, and instruction. When supporting teachers in the use of formative assessment, it is crucial to take these psychological characteristics into account. Therefore, "much more attention needs to be paid to the psychological states of teachers and leaders, as what they do most likely is derived from what they think about what they do and who they serve." (Evans, 2009, p. 87).

The factors that influenced DBDM and AfL mostly overlapped, but there were also differences. For DBDM, the most evidence was found for data literacy, collaboration in the use of data, a positive attitude around the use of data, and goal setting. For AfL, feedback strategies, PCK, assessment literacy, and the facilitation of classroom discussions were the factors for which most evidence could be found in the literature. In their classroom practice, teachers are likely to integrate aspects of DBDM and AfL (Kippers et al., 2018), which suggests that all of the prerequisites discussed above matter. Further studies can use the framework developed in this review to examine the relative and joint importance of these prerequisites.

4.5. Limitations

Although this review provides a useful overview of critical teacher prerequisites for formative assessment in classroom practice, we must consider the limitations of this study. First, although this review identified various critical teacher prerequisites, we do not claim that this list of factors is exhaustive. It is possible that there are other critical prerequisites that have not yet been studied empirically. Second, it is possible that despite conducting an extensive literature search, some relevant literature was not retrieved. Further, by focusing only on peer-reviewed high-quality publications, we may have missed important information from other sources, for example, book chapters and conference proceedings. However, we choose this focus to ensure that our review only included publications that had undergone a rigorous peer review process. Moreover, a common problem with systematic reviews is that they often reflect a certain type of bias, such as author bias (e.g., the author decides what publications to include, without clear criteria) and publication bias (e.g., publications with positive effects have a higher chance of being published than publications with no effect) (Green et al., 2006). Some of these biases were avoided by employing detailed, rigorous and explicit methods, focused on a specific research question (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000). Furthermore, we developed clear inclusion criteria (Sackett et al., 2000) to overcome possible author biases in selecting literature. Moreover, we described the methodology used in a detailed manner (Green et al., 2006), and used a scoring system to determine the quality of each publication (Sackett et al., 2000). Finally, in the discussion section we linked our results to several well-known, albeit not always empirical, publications in the field (i.e., Beck & Nunnaley, 2020; Daly, 2010; Hamilton et al., 2009; Heritage, 2007; Mandinach & Gummer, 2011, 2013). Our review highlighted the importance of several factors beyond these existing models, most importantly, psychological factors. Because of the rigorous process we followed (Green et al., 2006), we believe that this review makes a valuable contribution to the field of formative assessment, one on which follow-up research can be based.

4.6. Implications for further research

This study provides an overview of the teacher factors enabling or hindering (which often results from the lack of enablers) the use of formative assessment in the classroom. A lot of evidence was found for some factors (e.g., data literacy, collaboration, attitude). However, this does not imply that these are the most important enablers. Less evidence was found for some factors, simply because these factors have not been investigated in many studies. Moreover, most of the studies had qualitative designs. Although these studies provide valuable insights into how certain factors influence teachers' use of formative assessment, they are not informative regarding the extent of the impact of these factors. Future large-scale quantitative studies can address this identified gap in the literature.

Furthermore, the results of this review show that different factors seem to influence the different approaches to formative assessment. For example, for AfL, we found that the use of feedback strategies and involving students influenced the use of AfL in the classroom. Although involving students and the use of feedback are likely to be important for DBDM, we found no studies that addressed these factors. Moreover, Table 2 shows that the majority of DBDM studies did not involve students. Involving students in DBDM research would be a critical first step to gain insights into how students can effectively be involved in DBDM implementation. Further research is needed on how to involve students in the process of formative assessment, and in DBDM specifically, as well as on the use of feedback by students.

4.7. Implications for practice

This review identified various teacher prerequisites needed for the use of formative assessment in classroom practice. There is some evidence that professional development can address (some of) these prerequisites (Schildkamp & Poortman, 2015; Schildkamp & Teddlie, 2008; Schildkamp & Visscher, 2010b; Staman et al., 2014). However, professional development does not always lead to the desired effects. More research is needed into the characteristics of effective professional development in the use of formative assessment, and into the development, implementation, and evaluation of professional development in the use of formative assessment. The evidence-based framework developed in this review can inform such research.

It is important to stress here that both DBDM and AfL are needed in schools, as these approaches can complement each other (Van der Kleij et al., 2015). The identified factors can support schools in the implementation of DBDM, or as Jimerson, Garry, Poortman, and Schildkamp (2020) stated, slow down data use. This refers to the process of collective in-depth data use, identifying challenging problems, positing hypotheses related to these problems, and collecting and interpreting data to inform changes in instructional practices. The identified factors also need to be taken into account when implementing AfL, which is faster-paced, and more focused on the use of in-the-moment assessment data by teachers and students to inform teaching and learning in everyday practice (Van der Kleij et al., 2015; Heritage, 2007).

In conclusion, this review focused on an underexposed aspect of formative assessment that is essential for its successful use in classroom practice: teacher prerequisites. This review was conducted in a comprehensive and systematic manner, and synthesized evidence from 54 studies. The results confirm the importance of the role of the teacher in the use of formative assessment, and identify a number of crucial influential factors that need to be taken into account. The prerequisites identified can inform professional development initiatives in schools with regard to DBDM and AfL, as well as teacher education programs. Only when proper support is planned for, and the factors that enable the use of formative assessment are in place, can formative assessment lead to the desired effects: improved student learning and achievement.

Acknowledgment

This paper includes parts of the report “Schildkamp,K. M. Heitink. F. M. van der Kleij. I. Hoogland. A.Dijkstra.W. Kippers. and B. Veldkamp. 2014. ‘Voorwaarden voor effectieveformatieve toetsing: een praktische review. [Prerequisites for effectiveformativeassessment: a practical review].’ Enschede:Universiteit Twente.” This project was funded by NRO-PPO, The Netherlands: Grant number 405-14-534. This communication reflects the views of the authors only, and NROcannot be held responsible for any use that may be made of the informationcontained herein.

References[2]

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.

*Aschbacher, P., & Alonzo, A. (2006). Examining the utility of elementary science notebooks for formative assessment purposes. *Educational Assessment, 11*, 179–203. https://doi.org/10.1207/s15326977ea1103&4_3.

Baird, J.-A., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *Assessment and learning: State of the field review.* Retrieved fromLysaker, Norway: Knowledge Centre for Education. https://www.forskningsradet.no/servlet/Satellite?c=Rapport&cid=1253996755700&lang=en&pagename=kunnskapssenter%2FHovedsidemal.

Beck, J., & Nunnaley, D. (2020). A continuum of data literacy for teaching. *Studies in Educational Evaluation* Pre-online publication.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education Principles Policy and Practice, 18*, 5–25. https://doi.org/10.1080/0969594X.2010.513678.

*Birenbaum, M., Kimron, H., & Shilton, H. (2011). Nested contexts that shape assessment "for" learning: School-based professional learning community and classroom culture. *Studies in Educational Evaluation, 37*, 35–48. https://doi.org/10.1016/j.stueduc.2011.04.001.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education Principles Policy and Practice, 5*, 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.

*Blanc, S., Christman, J. B., Liu, R., Mitchell, C., Travers, E., & Bulkley, K. E. (2010). Learning to learn from data: Benchmarks and instructional communities. *Peabody Journal of Education, 85*, 205–225. https://doi.org/10.1080/01619561003685379.

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*(4), 13–17. https://doi.org/10.1111/j.1745-3992.2012.00251.x.

*Brown, L. I., De Four-Babb, J., Bristol, L., & Conrad, D. A. (2014). National tests and diagnostic feedback: What say teachers in Trinidad and Tobago? *The Journal of Educational Research, 107*, 241–251. https://doi.org/10.1080/00220671.2013.788993.

*Bryant, D. A., & Carless, D. R. (2010). Peer assessment in a test-dominated setting: Empowering, boring or facilitating examination preparation? *Educational Research for Policy and Practice, 9*, 3–15. https://doi.org/10.1007/s10671-009-9077-2.

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

Cho, V., Allwarden, A., & Wayman, J. C. (2016). *Technology is not enough: Shifting the focus to people. Principal leadership*28–31.

*Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher’s skills in assessment. *Studies in Educational Evaluation, 40*, 1–11. https://doi.org/10.1016/j.stueduc.2013.11.006.

Daly, A. J. (2010). Mapping the terrain. Social network theory and educational change. In A. J. Daly (Ed.). *Social network theory and educational change* (pp. 1–16). Cambridge, MA: Harvard University Press.

*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*, 247–265.

[2] References marked with an asterisk (*) indicate studies included in the review.

https://doi.org/10.1080/10824669.2012.718944.
*Datnow, A., Park, V., & Kennedy-Lewis, B. (2013). Affordances and constraints in the context of teacher collaboration for the purpose of data use. *Journal of Educational Administration, 51*, 341–362. https://doi.org/10.1108/09578231311311500.
Elwood, J. (2006). Formative assessment: Possibilities, boundaries and limitations. *Assessment in Education Principles Policy and Practice, 13*(2), 215–232. https://doi.org/10.1080/09695940600708653.
Elwood, J., & Klenowski, V. (2002). Creating communities of shared practice: The challenges of assessment use in learning and teaching. *Assessment and Evaluation in Higher Education, 27*, 243–256. https://doi.org/10.1080/0260293022013860.
Evans, A. (2009). No Child Left Behind and the quest for educational equity: The role of teachers' collective sense of efficacy. *Leadership and Policy in Schools, 8*, 64–91. https://doi.org/10.1080/15700760802416081.
*Farley-Ripple, E. N., & Buttram, J. L. (2014). Developing collaborative data use through professional learning communities: Early lessons from Delaware. *Studies in Educational Evaluation, 42*, 41–53. https://doi.org/10.1016/j.stueduc.2013.09.006.
*Feldman, A., & Capobianco, B. M. (2008). Teacher learning of technology enhanced formative assessment. *Journal of Science Education and Technology, 17*, 82–99. https://doi.org/10.1007/s10956-007-9084-0.
*Fletcher, A., & Shaw, G. (2012). How does student-directed assessment affect learning? Using assessment as a learning process. *International Journal of Multiple Research Approaches, 6*, 245–263. https://doi.org/10.5172/mra.2012.6.3.24510.1037/0022-3514.45.2.357.
*Fox-Turnbull, W. (2006). The influences of teacher knowledge and authentic formative assessment on student learning in technology education. *International Journal of Technology and Design Education, 16*, 53–77. https://doi.org/10.1007/s10798-005-2109-1.
*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*, 609–646.
*Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education, 92*, 799–824. https://doi.org/10.1002/sce.20270.
Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D., et al. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science, 44*, 267–291.
*Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education Principles Policy and Practice, 20*, 150–169. https://doi.org/10.1080/0969594X.2012.749212.
Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment.* London: Falmer.
*Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education, 23*, 531–557. https://doi.org/10.1007/s10972-012-9278-z.
Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Journal of Chiropractic Medicine, 5*, 101–117. https://doi.org/10.1016/S0899-3467(07)60142-6.
Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making (NCEE 2009-4067)*. Retrieved fromWashington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications/practiceguides/.
Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education, 35*, 213–224. https://doi.org/10.1080/03057640500146880.
*Hargreaves, E. (2013). Inquiring into children's experiences of teacher feedback: Reconceptualising Assessment for Learning. *Oxford Review of Education, 39*, 229–246. https://doi.org/10.1080/03054985.2013.787922.
*Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*, 101–111. https://doi.org/10.1016/j.tate.2013.07.008.
*Harris, L. R., Brown, G. T. L., & Harnett, J. A. (2014). Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional responses. *Educational Assessment Evaluation and Accountability,* 1–27. https://doi.org/10.1007/s11092-013-9187-5.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112. https://doi.org/10.3102/003465430298487.
*Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation, 38*, 21–27. https://doi.org/10.1007/s10972-012-9278-z.
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.
*Hubbard, L., Datnow, A., & Pruyn, L. (2014). Multiple initiatives, multiple challenges: The promise and pitfalls of implementing data. *Studies in Educational Evaluation, 42*, 54–62. https://doi.org/10.1016/j.stueduc.2013.10.003.
Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the data-driven mantra: Different conceptions of data-driven decision making. In P. A. Moss (Ed.). *Evidence and decision making* (pp. 105–131). Malden, MA: Wiley-Blackwell.
*Jimerson, J. B. (2014). Thinking about data: Exploring the development of mental models for "data use" among teachers and school leaders. *Studies in Educational Evaluation, 42*, 5–14. https://doi.org/10.1016/j.stueduc.2013.10.010.
Jimerson, J. B., Garry, V., Poortman, C. L., & Schildkamp, K. (2020). Implementation of a collaborative data use model in a United States context. *Studies in Educational Evaluation* Online pre-publication.
*Kay, R., & Knaack, L. (2009). Exploring the use of audience response systems in secondary school science classrooms. *Journal of Science Education and Technology, 18*, 382–392. https://doi.org/10.1007/s10956-009-9153-7.
*Kennedy, B. L., & Datnow, A. (2011). Student involvement and data-driven decision making: Developing a new typology. *Youth & Society, 43*, 1246–1271. https://doi.org/10.1177/0044118X10388219.
*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*, 496–520. https://doi.org/10.1086/505057.
Kippers, W. B., Wolterinck, C. H., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education, 75*, 199–213. https://doi.org/10.1016/j.tate.2018.06.015.
Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education Principles Policy and Practice, 16*, 263–268. https://doi.org/10.1080/09695940903319646.
*Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk, 10*, 333–349. https://doi.org/10.1207/s15327671espr1003_7.
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*, 305–334. https://doi.org/10.1002/rrq.73.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. https://doi.org/10.2307/2529310.
Ledoux, G., Blok, H., Boogaard, M., & Krüger, M. (2009). *Opbrengstgericht werken. Over waarde van meetgestuurd onderwijs [Data-driven decision making. About the value of measurement-oriented education]. SCO-Rapport 812.* Retrieved fromAmsterdam: SCO-Kohnstamm Instituut. http://dare.uva.nl/document/170475.
*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*, 179–201. https://doi.org/10.1080/09243453.2011.599394.
Mandinach, E. B., & Gummer, E. S. (2011). *The complexities of integrating data-driven decision making into professional preparation in schools of education: It's harder than you think.* Alexandria, VA, Portland, OR, and Washington, DC: CNA Education, Education Northwest, and West Ed.

Mandinach, E. B., & Gummer, E. S. (2013). Defining data literacy: A report on a convening of experts. *Journal of Educational Research and Policy Studies, 13*(2), 6–28.
Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education, 60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011.
Mandinach, E., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for data-driven decision-making. In E. B. Mandinach, & M. Honey (Eds.). *Data-driven school improvement: Linking data and learning* (pp. 13–31). New York, NY: Teachers College Press.
Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record, 114*, 1–48. Retrieved from http://www.tcrecord.org/library/Issue.asp?volyear=2012&number=11&volume=114.
Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education. Evidence from recent RAND research.* Retrieved fromSanta Monica, CA: RAND Corporation. https://www.rand.org/content/dam/rand/pubs/occasional_papers/2006/RAND_OP170.pdf.
Marshall, B., & Drummond, M. J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education, 21*, 133–149. https://doi.org/10.1080/02671520600615638.
*McNaughton, S., Lai, M. K., & Hsiao, S. (2012). Testing the effectiveness of an intervention model based on data use: A replication series across clusters of schools. *School Effectiveness and School Improvement, 23*, 203–228. https://doi.org/10.1080/09243453.2011.652126.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*. https://doi.org/10.1371/journal.pmed.1000097.
*Newby, L., & Winterbottom, M. (2011). Can research homework provide a vehicle for assessment for learning in science lessons? *Educational Review, 63*, 275–290. https://doi.org/10.1080/00131911.2011.560247.
*Ní Chróinín, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy, 18*, 219–233. https://doi.org/10.1080/17408989.2012.666787.
*O'Loughlin, J., Ní Chróinín, D., & O'Grady, D. (2013). Digital video: The impact on children's learning experiences in primary physical education. *European Physical Education Review, 19*, 165–182. https://doi.org/10.1177/1356336x13486050.
*Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: District and school connections in data-driven decision-making. *School Leadership and Management, 29*, 477–494. https://doi.org/10.1080/13632430903162541.
*Penuel, W. R., Boscardin, C. K., Masyn, K., & Crawford, V. M. (2007). Teaching with student response systems in elementary and secondary education settings: A survey study. *Educational Technology Research and Development, 55*, 315–346. https://doi.org/10.1007/s11423-006-9023-4.
*Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction, 18*, 238–250. https://doi.org/10.1016/j.learninstruc.2007.05.001.
Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide.* Oxford, UK: Blackwell.
Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018). Implementation of formative assessment–effects of quality of programme delivery on students' mathematics achievement and interest. *Assessment in Education Principles Policy and Practice, 25*, 160–182. https://doi.org/10.1080/0969594X.2016.1170665.
Poortman, C. L., & Schildkamp, K. (2011). Alternative quality standards in qualitative research. *Quality & Quantity, 46*(6), 1727–1751. https://doi.org/10.1007/s11135-011-9555-5.
Poortman, C. L., & Schildkamp, K. (2016). Solving student achievement problems with a data use intervention for teachers. *Teaching and Teacher Education, 60*, 425–433. https://doi.org/10.1016/j.tate.2016.06.010.
*Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction: Grading and feedback in mathematics classrooms. *Zeitschrift für Psychologie/Journal of Psychology, 216*, 111–124. https://doi.org/10.1027/0044-3409.216.2.111.
*Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*, 205–235. https://doi.org/10.1080/10627197.2006.9652991.
*Sach, E. (2013). An exploration of teachers' narratives: What are the facilitators and constraints which promote or inhibit' good' formative assessment practices in schools? *Education 3–13, 43*, 322–335. https://doi.org/10.1080/03004279.2013.813956.
Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to teach and practice.* London: Churchill Livingstone.
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 purposes, and promoting and hindering factors. *Teaching and Teacher Education, 26*, 482–496. https://doi.org/10.1016/j.tate.2009.06.007.
Schildkamp, K., & Poortman, C. L. (2015). Factors influencing the functioning of data teams. *Teachers College Record, 117*, 4. Retrieved from https://www.tcrecord.org/content.asp?contentid=17851.
*Schildkamp, K., & Teddlie, C. (2008). School performance feedback systems in the USA and in the Netherlands: A comparison. *Educational Research and Evaluation, 14*, 255–282. https://doi.org/10.1080/13803610802048874.
*Schildkamp, K., & Visscher, A. (2009). Factors influencing the utilisation of a school self-evaluation instrument. *Studies in Educational Evaluation, 35*, 150–159. https://doi.org/10.1016/j.stueduc.2009.12.001.
*Schildkamp, K., & Visscher, A. J. (2010a). The utilisation of a school self-evaluation instrument. *Educational Studies, 36*, 371–389. https://doi.org/10.1080/03055690903424741.
*Schildkamp, K., & Visscher, A. (2010b). The use of performance feedback in school improvement in Louisiana. *Teaching and Teacher Education, 26*, 1389–1403. https://doi.org/10.1016/j.tate.2010.04.004.
Schildkamp, K., Heitink, M. C., van der Kleij, F., Hoogland, I., Dijkstra, A. M., Kippers, W. B., & Veldkamp, B. P. (2014). *Voorwaarden voor effectieve formatieve toetsing: een praktische review [Prerequisites for effective formative assessment: A practical review].* Enschede: Universiteit Twente.
*Schildkamp, K., Karbautzki, L., & Vanhoof, J. (2014). Exploring data use practices around Europe: Identifying enablers and barriers. *Studies in Educational Evaluation, 42*, 15–24. https://doi.org/10.1016/j.stueduc.2013.10.007.
*Schildkamp, K., Rekers-Mombarg, L., & Harms, T. (2012). Student group differences in examination results and utilization for policy and school development. *School Effectiveness and School Improvement, 23*, 229–255. https://doi.org/10.1080/09243453.2011.652123.
*Schildkamp, K., Visscher, A., & Luyten, H. (2009). The effects of the use of a school self-evaluation instrument. *School Effectiveness and School Improvement, 20*, 69–88. https://doi.org/10.1080/09243450802605506.
*Staman, L., Visscher, A. J., & Luyten, H. (2014). The effects of professional development on the attitudes, knowledge and skills for data-driven decision making. *Studies in Educational Evaluation, 42*, 79–90. https://doi.org/10.1016/j.stueduc.2013.11.002.
Stobart, G. (2008). *Testing times: The uses and abuses of assessment.* Abingdon: Routledge.
*Sutherland, S. (2004). Creating a culture of data use for continuous improvement: A case study of an Edison Project school. *The American Journal of Evaluation, 25*, 277–293. https://doi.org/10.1177/109821400402500302.
Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning. *Assessment in Education Principles Policy and Practice, 18*, 433–449. https://doi.org/10.1080/0969594X.2011.582838.
Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education, 38*, 323–342. https://doi.org/10.1080/03054985.2012.689693.
*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., Cumming, J. J., & Looney, A. (2018). Policy expectations and support for teacher formative assessment in Australian education reform. *Assessment in Education: Principles, Policy & Practice, 25*, 620–637. https://doi.org/10.1080/0969594X.2017.1374924.
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*, 324–343. https://doi.org/10.1080/0969594X.2014.999024.V.
van Geel, M., Keuning, T., Visscher, A. J., & Fox, J. P. (2016). Assessing the effects of a school-wide data-based decision-making intervention on student achievement growth in primary schools. *American Educational Research Journal, 53*, 360–394. https://doi.org/10.3102/0002831216637346.

*Vanhoof, J., Verhaeghe, G., Van Petegem, P., & Valcke, M. (2012). Flemish primary teachers' use of school performance feedback and the relationship with school characteristics. *Educational Research, 54*, 431–449. https://doi.org/10.1080/00131881.2012.734726.

*Wayman, J. C., Spikes, D. D., & Volonnino, M. (2013). Implementation of a data initiative in the NCLB era. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.). *Data-based decision making in education: Challenges and opportunities* (pp. 135–153). Dordrecht: Springer.

Wayman, J. C., Cho, V., Jimerson, J. B., & Spikes, D. D. (2012). District-wide effects on data use in the classroom. *Education Policy Analysis Archives, 20*, 2–31. Retrieved from http://epaa.asu.edu/ojs/article/view/979.

Wayman, J. C., Jimerson, J. B., & Cho, V. (2012). Organizational considerations in establishing the data-informed district. *School Effectiveness and School Improvement, 23*, 159–178. https://doi.org/10.1080/09243453.2011.652124.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*, 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001.

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment. Practical techniques for K-12 classrooms.* West Palm Beach, FL: Learning Sciences International.

*Yin, Y., Tomita, M. K., & Shavelson, R. J. (2013). Using formal embedded formative assessments aligned with a short-term learning progression to promote conceptual change and achievement in science. *International Journal of Science Education, 36*, 531–552. https://doi.org/10.1080/09500693.2013.787556.

*Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*, 521–548. https://doi.org/10.1086/505058.