

# A learning theory-based exploratory analysis of teacher professional development interventions for formative assessment

J. A. de Vries  | A. J. Visscher | K. Schildkamp

ELAN Teacher Development, Faculty of Behavioural, Management and Social Sciences, University of Twente, Enschede, The Netherlands

## Correspondence

J. A. de Vries, Universiteit Twente, Faculty of Behavioural, Management and Social Sciences, Ravelijn 5436, Postbus 217, 7500 AE Enschede, The Netherlands.  
Email: [j.a.devries@utwente.nl](mailto:j.a.devries@utwente.nl)

## Abstract

There is substantial evidence in cognitive psychology for the effectiveness of learning theory-based principles, such as activating prior knowledge or retrieval practice, for learning. In this exploratory study, we analysed to what degree such learning theory-based principles are applied in teacher professional development (TPD) interventions for formative assessment and how, and whether differences in the use of these principles between TPD interventions explain differences in the effectiveness of such interventions. We found that the impact of TPD interventions for formative assessment on student achievement was evaluated rigorously in very few studies, with only five in our sample. Our results also showed that most learning theory-based principles were applied only to a very limited degree, and not deliberately. Both findings imply that no patterns relating the application of learning theory-based principles and intervention effectiveness were discernible. However, given the general evidence for the principles, there is probably much to gain by applying the learning theory-based principles when developing TPD interventions for formative assessment.

## KEYWORDS

effectiveness, formative assessment, learning theory, teacher professional development

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Review of Education* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.

## Context and implications

### Rationale for this study

Studies in cognitive psychology have resulted in clear principles to promote learning. It is unknown whether these have been used to improve teacher professional development (for formative assessment).

### Why the new findings matter

The limited use of learning theory-based principles in TPD development is remarkable, as there is strong evidence that these principles are effective.

### Implications for researchers

Only 5 of 44 evaluated studies adhered to high standards for randomised controlled trials set by Cheung and Slavin (2016) and What Works Clearinghouse (2020), meaning that we ultimately have very little reliable data on the effectiveness of TPD interventions for formative assessment. If we want to know what influences improvement in teacher skills when conducting TPD interventions, then better interventions and evaluations are needed.

The learning theory-based principles themselves are the result of thorough research, and it is therefore worthwhile to further investigate how the learning theory-based principles can be used to improve TPD interventions for formative assessment.

## INTRODUCTION

Increasingly, schools around the globe, most prominently in Europe (e.g., Brown et al., 2024; De Vries et al., 2024), North-America (e.g., Lee et al., 2020; Suurtamm & Koch, 2014) and Asia (e.g., Tang et al., 2010; Yan & Brown, 2021), are opting to use formative assessment as an important element of their teacher's teaching and student's learning, as it may significantly improve student learning (e.g., Anders et al., 2022; Sims et al., 2021). *Formative assessment* can be generally defined as 'gathering data about student learning and using this evidence to guide student learning' (Schildkamp et al., 2020). Teacher professional development is required to help teachers implement formative assessment, as it requires a complex set of teacher skills (Heitink et al., 2016). Teacher professional development (TPD), which can be described as a 'structured, facilitated activity for teachers, intended to increase their teaching ability' (Sims & Fletcher-Wood, 2021, p. 7), can take many different forms, ranging, for example, from a one-shot workshop to weekly meetings spread out over one or more school years, in which teachers come together to learn about formative assessment together with being coached in their classrooms. Many initiatives have been undertaken to help teachers develop the skills necessary for formative assessment in the form of teacher professional development; however, only with mixed, weakly substantiated, results in terms of both teaching quality and student outcomes (Randel et al., 2016; Sims & Fletcher-Wood, 2021).

Reviews of what makes TPD in general (not only TPD for formative assessment) effective (e.g., Desimone, 2009; Timperley, 2008) have claimed that some TPD features are decisive for improving teacher and student outcomes. Examples mentioned by Desimone (2009) and Timperley (2008) include focus on subject-matter content, longer duration and active and

collective participation of both teachers and school leaders. The claims for the effectiveness of these TPD features have been criticised for the lack of solid research supporting them as these features are based on research that includes few randomised controlled trials and is theory-poor. Moreover, despite the fact that most TPD interventions adhere to these general guidelines, effectiveness is, especially in the long term, not observed (Sims & Fletcher-Wood, 2021).

A possible way to increase specificity of these guidelines are to study learning-theory-based principles, which have of yet had little attention by TPD developers despite the fact that it is about (teacher) learning. These learning theory-based principles have their basis in cognitive psychology, which, bluntly put, focuses on how people learn through a process of remembering, forgetting and applying (Rosenshine, 2010; Winne & Nesbit, 2010). Even though studies in cognitive psychology have resulted in clear principles that can be used to promote learning in terms of storage in long-term memory and improved skills (Kirschner et al., 2006; Shiffrin & Atkinson, 1969), these learning theory-based principles, to our knowledge, have as yet been used very little to study teacher learning, and more specifically, TPD. Examples of learning theory-based principles include well-known concepts such as retrieval practice and modelling. Retrieval practice is the recall of information or the repetition of procedures after a delay, in order to learn them better (Adesope et al., 2017). Modelling refers to providing learners with examples of effective problem-solving strategies, from which learners, especially novices, learn much more than from trial and error (Creemers & Kyriakides, 2011; Kirschner et al., 2006; Rosenshine, 2010; Soderstrom & Bjork, 2015). It may be evident that the use of these principles is affected by the TPD goal, TPD content to be learned, the TPD method used to learn the TPD content, and the characteristics of the context in which the TPD intervention is implemented. It is likely that, for example, teacher motivation to achieve the TPD goal, which is a context characteristic, may very likely influence the way the TPD method needs to be shaped, such as smaller or bigger changes to be implemented in the classroom (Timperley, 2008). The same is true for the conceptualisation of formative assessment, as a measurement instrument or as a way of teaching (Heritage, 2020), which influences its complexity and the need for a more or less intensive TPD intervention. A more intensive TPD intervention may include more need for the implementation of learning theory-based principles, such as more retrieval practice and more extensive modelling. As such, similar to the idea of 'constructive alignment' (Biggs, 1996), the TPD components, including the learning theory-based principles, need to be in line with each other to produce an effective TPD intervention.

The goal of this exploratory review is to investigate if and how these general learning theory-based principles have been incorporated in the design of TPD interventions for formative assessment. Currently, TPD interventions, which often require much time and effort of teachers, have been as of yet designed with only the guidance of broader guidelines, possibly resulting in mixed results on both teacher and student outcomes. To improve TPD guidelines, we have reviewed current TPD interventions with regard to learning theory-based principles. We have extended the learning-theoretical principles to one coherent model, linking the principles to the TPD goal, TPD content to be learned, the TPD method used to learn the TPD content, and the characteristics of the context in which the TPD intervention is implemented, so as to be able to observe the complexities between the alignment of these components as well.

The following research questions were answered in this study, which aims to address the identified gap:

1. How are various learning theory-based principles incorporated in TPD interventions for formative assessment?
2. How does the application of learning theory-based principles explain differences in the effectiveness of TPD interventions?

# THEORETICAL FRAMEWORK

## Teacher skills for formative assessment

Formative assessment requires the mastery and integration of various teacher skills. The skills hierarchy for formative assessment by Wolterinck et al. (2022) shows the complexity of formative assessment (Figure 1). Ideally the teacher skills form a coherent whole.

Firstly, teachers need to be able to prepare the learning goals for lesson units and single lessons based on students' current learning needs and abilities. To do this, and to execute the teacher skills described below, teachers require strong pedagogical content knowledge (Jones & Moreland, 2005). Learning goals describe for learners 'what the intended outcome

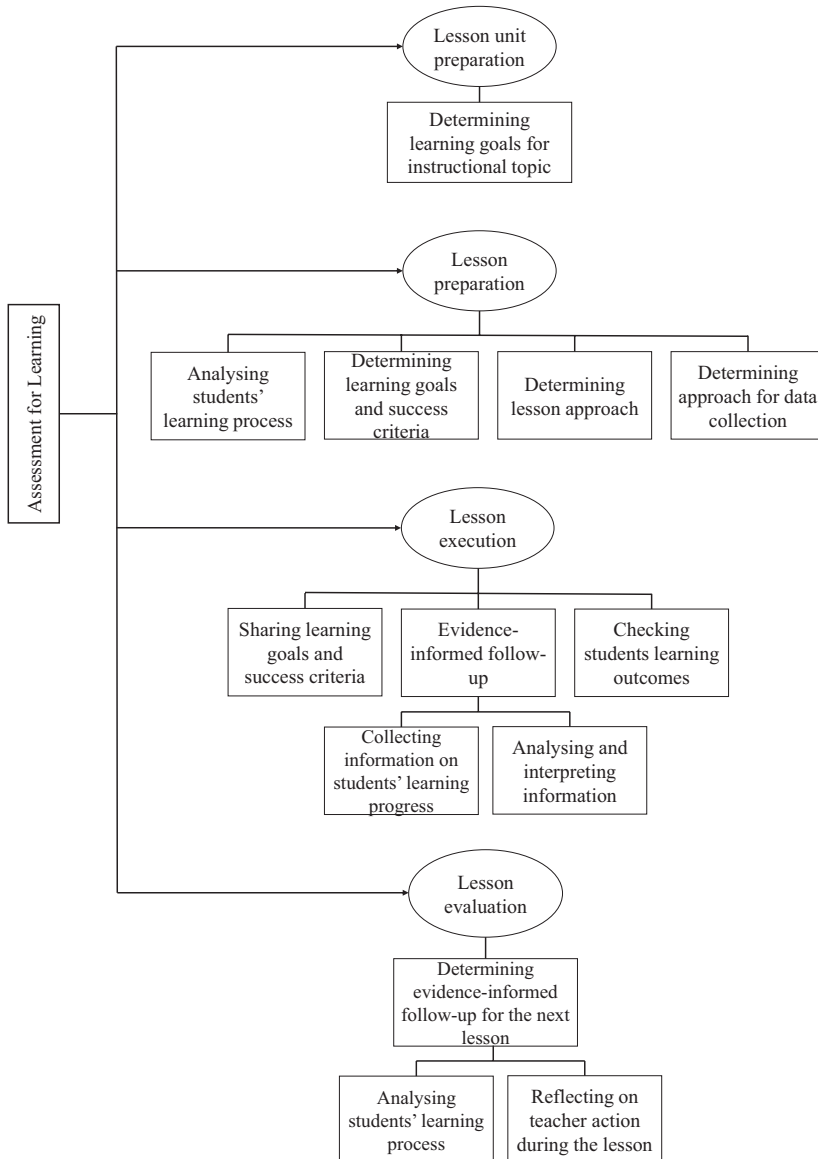


FIGURE 1 The teacher skills needed for formative assessment in hierarchy.

of the lesson is with regard to their learning' (Crichton & McDaid, 2016, p. 190). These learning goals need to be shared with students to clarify where learners need to be going, which can happen in various ways, for example, by sharing the learning goal on the whiteboard, using exemplars to illustrate learning goals, or asking students to work on a worked example first and relating this to the learning goal. The attainment of the learning goals can be measured by means of success criteria, which are more concrete parameters for the current status of students' learning as compared to the learning goals. In line with the idea of 'constructive alignment' (Biggs, 1996), the learning goals and criteria for success can help determine the approach taken in the lesson and in collecting data on students' learning progress.

Secondly, teachers should be able to determine where students are in comparison with the learning goals by continually collecting information on students' learning progress using assessment techniques (e.g., a curriculum-based assessment, diagnostic questions, homework assignments). When teachers can adequately analyse and interpret this information, it should help them determine what follow-up actions, such as clarifying misconceptions, are necessary.

Finally, teachers can evaluate their lesson by determining what information is needed for follow-up actions in the next lessons that can help close the gap between where the learners are (as shown by assessment results) and where they need to go (the learning goals). One step teachers can take to accomplish this is to provide students with feedback. To make feedback effective, it is important that teachers are able to implement it as interactive sense-making, meaning that it is provided in the form of a dialogue instead of just 'telling' students the feedback (Sadler, 2014). Continual analysis of the students' learning process and reflection on how the teachers' actions have influenced this learning process are part of the lesson evaluation, and may further determine follow-up actions as well as help develop learning goals for the next lessons in the lesson unit.

## A learning theory-based perspective on teacher professional development for formative assessment

It is difficult, if not impossible, for teachers to develop and integrate all required formative assessment skills without any support (Heitink et al., 2016). Teacher professional development interventions for formative assessment have been implemented and evaluated, but without resulting in unambiguous, clear-cut and specified guidelines for the design of such interventions (Sims et al., 2021). A learning theory-based analysis of TPD interventions for formative assessment could provide more insight into what makes TPD for formative assessment effective. We will explain here precisely what such a perspective entails, which is based on extensive evidence in cognitive (educational) psychology (e.g., Coe et al., 2020; Kirschner & Hendrick, 2020) by means of the model in Figure 2. Four core components are important in such a perspective: the TPD goal, TPD content to be learned, the TPD method used to learn the TPD content and the characteristics of the context in which the TPD intervention is implemented. It is essential in the learning theory-based perspective that the four components align with each other: the content and the method have to match with each other and with the TPD goal and the TPD context. In particular, the prior knowledge that participating teachers have about formative assessment, and how homogeneous/heterogeneous they are in terms of their experience and prior knowledge are examples of important context characteristics to align with.

Generally, TPD interventions have a more or less explicitly and clearly specified *goal*, which represents the desired outcomes of the TPD intervention, namely, one or more of the following: teacher knowledge, skills and attitudes (Desimone, 2009). The *TPD content* to be learned by teachers should contribute to the accomplishment of the TPD goal. This is not

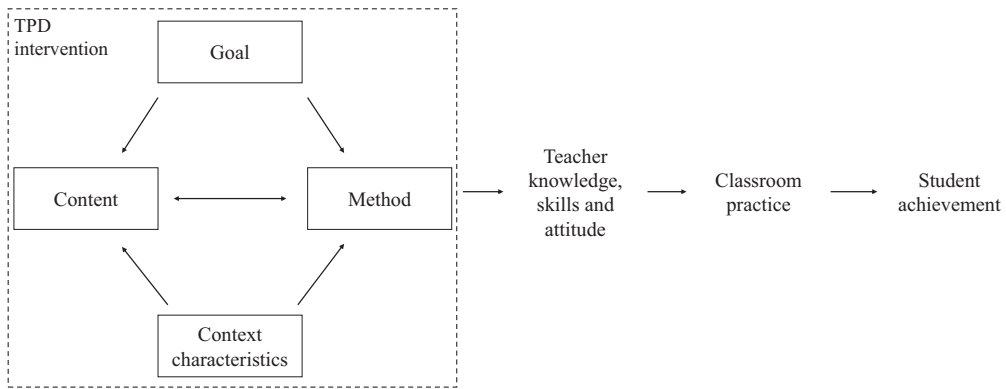


FIGURE 2 A learning theory-based perspective on teacher professional development.

always the case. One-shot workshops in which knowledge is transmitted to teachers so that they learn to carry out complex teacher skills in their classrooms are not rare in TPD interventions (Desimone, 2023). On the other hand, too time-intensive TPD interventions may put too much pressure on teachers as they may disrupt their primary tasks, that is, preparing and executing teaching (Cordingley et al., 2015). The nature of the knowledge, skills and attitudes (the content of a TPD intervention) to be learned by teachers in an intervention can require more or less in terms of the intervention *method* (the learning activities included in the TPD enabling acquisition of the content to be learned; Kennedy, 2014; Visscher, 2021). If teachers only need to acquire new information, then one or more face-to-face meetings can be fine. For learning to enact complex teaching skills in the classroom, classroom coaching will probably also be necessary.

The right part of Figure 2 shows that ideally, newly learned teacher knowledge, skills and attitudes are applied in the classroom and improve the quality of teaching there, which ideally leads to improved student achievement (Desimone, 2009).

## Goal(s) of the TPD intervention

In the case of TPD for formative assessment, the overarching goal is usually equipping teachers with the necessary knowledge, skills and attitudes for implementing some form of formative assessment (Schildkamp et al., 2020); however, there are typically two approaches to formative assessment: a measurement approach and an inquiry approach (Hargreaves, 2005; Heritage, 2020). In the measurement approach, formative assessment encompasses activities such as teachers developing quantitative assessment techniques and recording and using the results of the quantitative assessments to monitor and improve student performance over a longer period of time, such as 1 to 4 weeks (Wiliam, 2006). In this case, instruction and assessment are two separate activities, in which the students are not involved as assessors (van der Kleij et al., 2015). In the inquiry approach, on the other hand, instruction and quantitative and qualitative assessment are integrated activities, which happen continually over the course of periods from 5 seconds to 2 days (van der Kleij et al., 2015; Wiliam, 2006). As opposed to the measurement approach, the involvement of students in assessment and feedback tasks is stimulated in the inquiry approach. It is likely that the inquiry approach trumps the measurement approach regarding the complexity of the competences to be learned, as it includes a far higher frequency of closed feedback loops and stronger involvement of students in the assessment processes. Evidence suggests that

the inquiry approach to formative assessment is likely to be more effective, which, if it is true, may be caused by the more frequent and more rigorous changes in teaching behaviour (William, 2006).

## The content to be learned in the TPD intervention

The emphasis in the TPD content can be on expanding teachers' declarative knowledge (e.g., what is formative assessment), on improving teacher skills (i.e., new procedural knowledge), or a combination of both, possibly also in combination with developing specific teacher attitudes (i.e., developing a positive attitude towards student input in their learning processes). Increased teacher knowledge has been shown to have a positive impact on classroom changes and student outcome (Hill & Chin, 2018). All types of knowledge, also of teachers, is built through experience (or stimuli), rehearsal, encoding and retrieval, explained by the Information Processing Model (or Modal Model), first introduced by Shiffrin and Atkinson (1969)—see Figure 3.

The complexity of the goal, the type of knowledge, influences the TPD content learned in an intervention (Desimone, 2023). Somewhat higher complexity, such as in the inquiry approach to formative assessment, may require more frequent and more varied practice by teachers, as opposed to the measurement approach. The content may simply focus more on declarative knowledge, for example, goodness-of-fit statistics for test instruments, as the content of inquiry. The content needs to address more of formative assessment and also include supporting teachers in how-to's regarding integration of formative assessment in their lessons, each also with their own difficulties (e.g., heterogeneous classes, novice teacher; Feldon, 2007). Through rigorous practice, knowledge on what, for example, learning goals and success criteria are, it may be transformed to procedural knowledge, including knowledge on how to reshape success criteria to better fit with students' individual learning progresses towards these learning goals. As forgetting is a large part of learning (Shiffrin & Atkinson, 1969), it can be necessary to often stimulate retrieval of knowledge. In all types of knowledge, TPD designers need to consider the teacher's individual needs and wishes as well, possibly by letting teachers decide some or all of the content of the TPD.

### *Expert-driven or teacher-driven TPD content*

Whether the content of a TPD intervention is more directed by (external) experts or teachers themselves, can influence teachers' motivation to participate (Kennedy, 2014). TPD interventions that have an expert-driven nature often include well-defined materials, pre-determined goals, and clear teacher strategies that are supposed to be implemented by teachers in ways prescribed by experts (Kennedy, 2014). Usually (Sims et al., 2021), teachers are asked to compare the TPD intervention goals with their own teaching

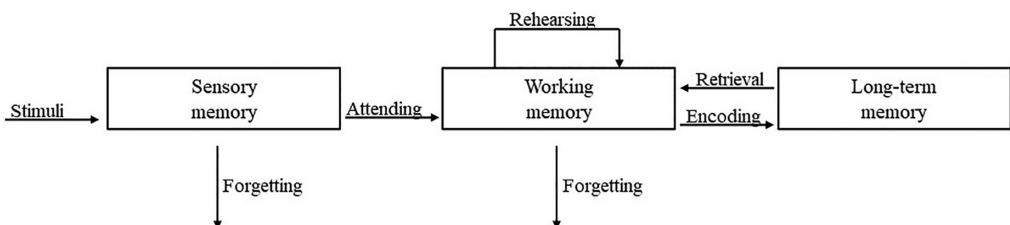


FIGURE 3 Information processing model.

practices and as a result, they become aware of where there is room for expanding their expertise.

More teacher-driven approaches to TPD (e.g., Darling-Hammond et al., 2017; Korthagen, 2017) are characterised by teacher-centred, inquiry approaches in which the expertise of teachers is respected and utilised by not requiring teachers to conform to externally predetermined, fixed teaching strategies (Korthagen, 2017). Gaining teacher buy-in is frequently mentioned as a key factor in the success of professional development (Cordingley et al., 2015). Unfortunately, just like everyone else, teachers may not be able to accurately describe their own shortcomings, and may focus on elements that do not need improvement (Timperley et al., 2007). In this way, teachers may easily confirm each other's (wrong) thinking, as in 'over-assimilation' (i.e., thinking you are already doing well because of some merely superficial changes), preventing actual improvement. For these reasons, a combination of an external expert with some teacher choice might be desirable.

## The method for learning the content to be learned

Some general features of the TPD intervention method include, for example, the duration of the intervention, the frequency and sequence of the TPD meetings, the involvement of a knowledgeable expert or not, and the participation of either one, several or all teachers at schools participating in the TPD (Timperley et al., 2007). These features can matter for TPD effectiveness (Visscher, 2021). For example, if a TPD intervention is designed to have sessions over a broad span of time, teachers have more space to practise the new skills and can receive support (e.g., coaching) for a longer period of time. In this study we focused, as stated, on learning theory-based principles that have been found to make a positive impact on student learning (e.g., Coe et al., 2020; Kirschner & Hendrick, 2020) and that have not been studied much yet in the context of teacher learning. We will now explain these principles.

*Performance standards* represent what the learners are supposed to know and/or be able to do as a result of the TPD. These standards, either more expert- or teacher-driven, can be shared with (co-)teachers to clarify what the desired outcomes of the TPD are. Introducing greater teacher discretion in goal-setting within the TPD could enhance adaptability, enabling instruction more closely fitted to teachers' needs (Kennedy, 2014; Korthagen, 2017). This may also help motivate teachers to change and direct their energy towards this change (Epton et al., 2017). This approach, however, can somewhat compromise the attainment of performance standards as it might cause too much flexibility (Wylie & Lyon, 2015) as too much focus on granular goals may not lead to improvement of formative assessment in the classroom. Formative assessment has been shown to be most effective when all formative assessment strategies are applied as one coherent set ('to the spirit', Brooks et al., 2021), which reduces the room for choice and, thus, teacher-drivenness.

The setting of goals can enable teachers to (1) *evaluate* and (2) *reflect* on and plan their learning process: evaluate where they are compared to the TPD goals, and re-plan their efforts where necessary (the self-regulation principle; Zimmerman, 2002). Self-regulation theory posits that self-evaluation can stimulate learners (in this case, teachers) to reflect on (the effects of) their actions, which in turn can make them more likely to sustainably improve their classroom practice (Sims et al., 2021). As suits a more teacher-driven approach, there is a need in such reflection to deal with the teacher's personal situation of the teacher, what motivates them (Kennedy, 2014; Korthagen, 2017). Accurate self-evaluation can lead to more positive outcomes, but the accurateness of the self-evaluation is often doubtful and needs extra support (Rispoli et al., 2017).



*Monitoring teachers' progress towards the performance standards* by a TPD trainer (or leader) or TPD peers can be done in multiple ways throughout a TPD intervention. For example, teachers can observe (other) TPD participants practising specific teacher skills (e.g., Crawford et al., 2017), or can study, reflect on and discuss each other's lesson plans (e.g., Andersson & Palm, 2017; William & Leahy, 2015). A combination of different (quantitative and qualitative) monitoring methods may be preferable (Timperley, 2008), not only looking at a teacher while teaching, but also hearing the teacher explain why they taught the way they did as a form of a worked example.

Based on the monitoring results, TPD trainers can decide to adapt their instruction through *differentiation*. Differentiation is 'the modification of four curriculum-related elements—content, process, product, and affect—which are based on three categories of student needs and variances—readiness, interest, and learning profile' (Tomlinson & Imbeau, 2010, p. 15). In terms of TPD, differentiation can be done in the explanations of the content, the content itself, the support to teachers, the learning tasks given to teachers and the pace of those learning activities, in line with the differences in progress between teachers. For example, novices to formative assessment may need a stronger focus on the formulation of learning goals, whereas experts can work on more difficult teacher skills such as stimulating students' self-assessment and/or peer feedback (Christoforidou et al., 2014). By adapting the content more proximal to teachers' knowledge, it is likely that they develop (Zone of Proximal Development; Vygotsky et al., 1994).

A TPD session itself may start with the principle of *activating prior knowledge*. This is a good way to start a lesson series (Rosenshine, 2010), as it is a way to activate prior knowledge on the topic to be learned before learning new knowledge, with the aim of assimilating new knowledge into the learner's existing knowledge base (Ausubel, 2012; Wetzels et al., 2011). A TPD trainer, or a critical peer teacher, could do this by questioning or discussing prior knowledge on the topic to be learned. Somewhat similar to activating prior knowledge is *retrieval practice*, which is the recall of information or the repetition of procedures after a delay in order to learn them better (Adesope et al., 2017). It enhances learning by making the retrieval process itself more efficient and improving the accessibility of information in long-term memory (Shiffrin & Atkinson, 1969). Examples of retrieval practice include self-quizzing, flashcards, completing practice problems, or summarising information without looking at the source material. In sum, the difference between these two principles is that activating prior knowledge sets the stage for learning by allowing the connection of new information with existing knowledge, while retrieval practice reinforces learning by actively recalling and reconstructing information from memory (Ausubel, 2012; Nunes & Karpicke, 2015).

It is important that the *learning tasks in a TPD intervention are representative* of desired teacher practice, as this influences the transfer of the teaching skills that are learned to teacher practice (van Merriënboer & Kester, 2014). *Modelling* may be helpful to let teachers gain a better understanding of what 'good' teacher skills in formative assessment look like. Demonstration of the intended instructional behaviour to learners can help them to develop a clear picture of what 'good' looks like: what they are supposed to do and how (van der Linden et al., 2022). Modelling can be done, for example, by trainers who demonstrate (live or with video) the desired teacher behaviour (e.g., classroom management), or who show (and compare) lesson plans. The usefulness of modelling depends partly on the teachers' ability to transfer these modelling examples to their own situation, as no example exactly fits or encapsulates all nuances of the complexity of varying teaching contexts.

Accomplishing a complex TPD goal, such as learning to integrate formative assessment in one's own classroom, is more likely to occur when that complex skill is addressed and practised as a whole. While practising new skills in their own classrooms (or later, if video-recorded), teachers can be given *feedback* by an expert TPD trainer or a peer teacher (van

der Linden et al., 2022). Such feedback can be open-ended or more closed-ended, also depending on teacher's needs and wishes. The first is usually less directive compared to the latter and gives more respect to the teachers' expertise. However, on the other hand, sometimes teachers want and need more directive feedback as this can be considered more helpful, that is, scaffolding (Timperley et al., 2007). Whether such feedback is given once or multiple times depends on whether the learning theory-based principle of *spaced practice* is applied in an intervention. Learning increases if opportunities for practice are distributed over a longer period of time, compared to massed practice (Feldon, 2007; Kang, 2016). It is especially useful if feedback loops are allowed to be closed (i.e., evidence of successful feedback usage; Carless, 2019). This is, of course, only possible when a TPD intervention allows for multiple TPD sessions, and if teachers are facilitated in experimenting with newly learned materials and practices (Heitink et al., 2016).

## Characteristics of the context of the TPD intervention

The TPD content and TPD method are ideally also in line with the *characteristics of the context* in which the TPD intervention takes place (the fourth component of Figure 2). Important context characteristics include the prior knowledge of participating teachers on the topic to be learned, participating teachers' level of similarity with regard to expertise (which may have implications for the need to differentiate between teachers in the TPD intervention), the resources (e.g., time) that can be used for the intervention, and the school staff's attitudes towards the intervention. If a school leader has, for example, established a school-wide formative assessment culture, the TPD does not need to focus on building such a culture (Heitink et al., 2016). Available resources should also have implications for what can be done and achieved with respect to teacher professional development. With less experienced teachers, it may be harder to develop all the complex teacher skills for formative assessment, which might, for example, require leaving out student involvement in the assessment process (Christoforidou et al., 2014).

## The current study

Although the learning theory-based principles we discussed, grounded in four core components from our learning theory-based perspective, have been thoroughly examined and proven to be effective for student learning, we do not know what the relation is between their application in TPD and their effectiveness for teacher learning. In this study, we investigated the application of the learning theory-based principles, including the alignment between the TPD goal, TPD content, TPD method and the TPD contextual features. It was our goal to relate the degree to which and way in which the learning theory-based principles were applied and the effectiveness of the TPD interventions in terms of improved student outcomes.

## METHOD

### Selected teacher professional development interventions

A review study on formative assessment practices by Gulikers and Baartman (2017) was used as the primary source for the studies to be selected in this study, which were a total of 43 studies in which TPD for formative assessment was the subject. We only

used studies which took place in the context of primary education ( $n=15$ ), secondary education ( $n=22$ ) or both ( $n=3$ ), excluding five studies in which the focus was placed on teacher educators ( $n=4$ ) or higher education ( $n=1$ ). Furthermore, unfortunately, we had to exclude many studies because they did not include a control group at all ( $n=32$ ). To complement our number of studies, as we found only a very small number of TPD interventions, we took into account eligible studies on TPD on formative assessment of a review done in more recent years (Sims et al., 2021), leading to the inclusion of two extra studies (Anders et al., 2022; Randel et al., 2016). Moreover, we added studies known via personal connections (De Vries et al., 2022; Wolterinck et al., 2021; Veugen et al., 2021; Andersson & Palm, 2017; Schneider & Meyer, 2012). We evaluated the remaining 14 studies further by comparing them to a set of nine criteria based on criteria formulated by Slavin (2008) and the standards from the What Works Clearinghouse (2020), to help us determine the effects of interventions in terms of student achievement in a methodologically unbiased way as possible, and under conditions that are ecologically valid for educational practice (Cheung & Slavin, 2016)—see Table 1. Our inclusion criteria allowed inclusion only of studies that:

1. Assigned teachers to conditions in a randomised way or were based on matching before the intervention, because such studies provide the best basis for causal claims about the effectiveness of interventions.
2. Investigated the difference between experimental and business-as-usual conditions, to be able to analyse the difference between the two conditions (instead of the difference between intervention variants).

**TABLE 1** Adherence of the identified TPD interventions for formative assessment to inclusion criteria ( $n=17$ ).

	1	2	3	4	5	6	7	8	9
Aschbacher & Alonzo, 2006	x	✓	x	✓	x	✓	?	?	✓
Anders et al., 2022	✓	✓	✓	✓	✓	✓	✓	✓	✓
Andersson & Palm, 2017	✓	✓	✓	✓	x	✓	✓	✓	✓
De Vries et al., 2022	✓	x	✓	x	x	✓	✓	✓	?
Beesley, 2009	x	x	✓	✓	x	✓	x	?	?
Brookhart et al., 2010	x	x	x	✓	✓	✓	?	✓	✓
Fuchs et al., 1999	✓	✓	x	✓	x	✓	?	?	✓
Furtak et al., 2016	x	x	x	✓	✓	✓	?	✓	?
Graney & Shinn, 2005	?	x	x	✓	?	x	?	?	?
Heller et al., 2012	✓	✓	✓	✓	x	✓	✓	✓	✓
Herman et al., 2015	✓	✓	x	✓	x	✓	?	?	x
Phelan et al., 2012	✓	✓	x	✓	x	✓	?	?	?
*Randel et al., 2016	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sato et al., 2008	x	x	✓	✓	x	✓	x	?	?
Schneider & Meyer, 2012	✓	✓	✓	✓	✓	✓	✓	✓	✓
Veugen et al., 2021	✓	x	✓	x	x	✓	✓	✓	?
Wolterinck et al., 2021	✓	✓	✓	x	✓	✓	x	✓	✓
Total	11	10	10	14	6	16	7	10	9

Note: ✓ indicates adherence to criterion; x indicates lack of adherence to criterion; ? indicates that no information about this criterion is reported; \* study selected for analysis.

3. Included at least 20 teachers per condition and at least 30 students per condition, as small-scale studies produce larger effects and are less representative of large-scale implementation effects (Cheung & Slavin, 2016).
4. Evaluated TPD effects on student outcomes. TPD can have various desirable effects at the teacher and school levels; however, the most important criterion for its effectiveness should be its impact on student learning (Guskey, 2003)
5. Used standardised student assessment measures, which were developed neither by the researchers nor by the developers of the TPD interventions. Researcher-made tests are biased (Cheung & Slavin, 2016) and produce much higher effects than independent, standardised tests, as the latter tests measure not only the specific content addressed in an intervention, but also the other content students must learn in a specific grade.
6. Lasted at least 12 weeks. Short interventions (such as one-day workshops) may have strong effects on student learning ('super implementation'; Pellegrini et al., 2021), which cannot be maintained over longer periods and as such are not indicative of impacts in educational practice.
7. Had differential attrition between conditions that was less than 15%. Differential attrition refers to the percentage-point difference between the attrition rates of the experimental and control groups. When this value is above 15%, the experimental group and the control group may not be comparable anymore.
8. Used an intent-to-treat approach to data analysis. The intention should be to include all of the schools, teachers and students who participated in the intervention from the start in the analysis of the intervention effects, and not to leave out specific groups of participants.
9. Students' average pre-test mean differences between conditions should be less than 0.25 of a standard deviation. In the case of larger pre-test differences between the conditions, conditions are not comparable anymore, and the differences cannot be adjusted for statistically (Slavin, 2008).

The effects of TPD interventions were often evaluated by means of researcher-made measures, or had experimental and control groups that were too dissimilar on the pre-test scores. The lack of adherence to these criteria causes bias in conclusions about the effectiveness of these interventions. However, for the purpose of our exploratory review, we required adherence to eight instead of nine criteria, as this allowed us to include five instead of only three TPD interventions, possibly giving us richer data to investigate. This also made it possible to study a somewhat richer variety of TPD interventions, while still having relatively strong evidence for the effects of the studied TPD interventions. Table 2 presents an overview of the features and effects of the five selected TPD interventions. Three of the five selected TPD interventions met all nine criteria: Intervention 1 (Anders et al., 2022), Intervention 4 (Randel et al., 2016) and Intervention 5 (Schneider & Meyer, 2012). Intervention 2 (Andersson & Palm, 2017) and Intervention 3 (Heller et al., 2012) lacked a researcher-independent measurement instrument for the evaluation of the intervention's effect on student achievement.

A detailed summary of each intervention and corresponding study may be found at: <https://www.formatief-toetsen.nl/Review-Learning-Theoretical-Analysis/>. The five interventions and their evaluations varied in terms of their implementation scale (number of students), the grade levels of participating students, the type of student achievement test used, TPD duration in years, and the effects at the student level. The effect sizes of the interventions at the student level may have been influenced by the implementation scale, the measurement type, the grade level and the duration of the TPD. A larger number of teachers involved in an intervention, as in Intervention 1, often means a decrease in the effect sizes, as the intervention is carried out by independent trainers who may be less eager on making the intervention work than the designers of an intervention (Cheung & Slavin, 2016).

TABLE 2 Empirical evidence for the effectiveness of the five TPD interventions.

	Anders et al. (2022)	Andersson and Palm (2017)	Heller et al. (2012)	Randel et al. (2016)	Schneider and Meyer (2012)
Intervention (level of effectiveness)	1 (effective)	2 (effective)	3 (effective)	4 (less effective)	5 (less effective)
Effect size at the student level	0.10	0.66	0.31 (short-term effect); 0.57 (long-term effect)	0.01 (not statistically significant)	0.03 (mathematics, not statistically significant); -0.16 (English)
Implementation scale (No. of students)	25,393	738	2902	9596	4774
Implementation scale (No. of teachers)	11,430	45	145	231	151
Grade level	9	3	4	4/5	6
Student achievement test	State-developed English and mathematics tests	Researcher-developed mathematics test	Researcher-developed science tests	Researcher-developed mathematics tests	State-developed English and mathematics tests
TPD duration in years	2	0.5	0.25	1	1
Effect at the teacher level	Positive teacher perceptions of the impact of TPD on their teaching practice (no control group)	Increase of formative assessment activities (no control group)	More content knowledge: 1.81 (short-term effect); 1.45 (long-term effect)	More assessment knowledge: 0.42	More knowledge of assessment practices: 0.11

Interventions are also more likely to produce higher effect sizes when a measure was used that was specifically designed for the purpose of evaluating the effectiveness of the particular TPD intervention, as in Interventions 2, 3 and 4 (Cheung & Slavin, 2016). Students in lower grades, as in Interventions 2, 3 and 4, seem more likely to improve compared to students in higher grade levels, as in Intervention 1, which may be because there is still more to be learned in the lower than in the higher grades (Kraft, 2020). A longer duration of the TPD interventions, as in Interventions 1, 4 and 5, is likely to influence teachers' ability to durably improve their teaching skills as well. Teachers in Intervention 1 had two school years to practice, whereas teachers in Interventions 2 and 3 received only half or a quarter of a school year, respectively. These latter two interventions could have had strong short-term impact on student learning, but without lasting results. However, delayed post-measurements after Intervention 3 showed that short interventions can in fact have a longer lasting, strong impact on student learning.

## Data analysis of the included TPD interventions

The data for this study included publications related to the five interventions, both peer-reviewed publications and, where possible, evaluation reports. Moreover, to complement and verify the data from the publications (because sometimes information was missing), we conducted interviews with the developers of the five interventions, which averaged an hour long. The interviews were transcribed with the use of *Amberscript*. Based on a coding scheme (to be found at: <https://www.formatief-toetsen.nl/review-learning-theory-based-analysis>), all studies were double-coded by the authors of this study using publications, TPD materials and transcribed interviews. When the resulting codes differed between two of the authors, these differences were discussed until consensus between them was reached.

## RESULTS

### Application of learning theory-based principles

Most of the learning theory-based principles that were mentioned in the theoretical framework were only implemented to a very limited degree in the majority of the five TPD interventions (see Table 3). The developers of these interventions also mentioned that these principles were not taken into account when developing the interventions.

### Performance standards and modelling

The use of performance standards was very limited in all five TPD interventions. For Interventions 3, 4 and 5, the performance standards were only specified for teacher's development of assessment instruments to diagnose students' learning progress (i.e., how to formulate items), and thereby, leaning more to a measurement approach. Also, this shows the difficulty of formulating performance standards for a teacher skill as complex as formative assessment, as almost no developers have ventured to do this. Only Intervention 1 had more concrete performance standards and, in addition, *modelling* was used to concretise the *performance standards* by using eight short videos showing eight separate formative assessment strategies (e.g., no-hands-up technique, random name generator, hinge questions). This helped teachers to direct their energy to change as this not only gives a clear picture of the desired outcome behaviour, but also provides evidence on why it works (Sims

TABLE 3 Degree of application of the learning theory-based principles in the TPD interventions.

	Anders et al. (2022)	Andersson and Palm (2017)	Heller et al. (2012)	Randel et al. (2016)	Schneider and Meyer (2012)
Intervention (level of effectiveness)	1 (effective)	2 (effective)	3 (effective)	4 (less effective)	5 (less effective)
Performance standards	Yes	Yes	Partly	Partly	Partly
Modelling	Teaching examples	No modelling	Narratives	Narratives	Material
Monitoring	Qualitative + quantitative + peer + teacher leader	Qualitative + peer + trainer	Qualitative + trainer	Qualitative + peer	Qualitative + quantitative + trainer
Encouragement of self-regulation	Goal-setting – evaluation – reflection/planning	Reflection/planning	Evaluation – reflection/planning	Evaluation – reflection/planning	Evaluation – reflection/planning
Retrieval practice	Without resources	With resources	Without resources	No activation of prior knowledge	With resources
Activating prior knowledge	Once	Multiple times	Once	Multiple times	Multiple times
Representativeness of learning tasks	Yes + progression	Yes + progression	Partly + progression	Yes + progression	Yes + partly
Spaced practice	High intensity	High intensity	Moderate intensity	High intensity	Moderate intensity
Classroom feedback	Class feedback + peer + benchmarked	No classroom feedback	No classroom feedback	No classroom feedback	No classroom feedback
Differentiation	Content + tasks + feedback	Feedback	Feedback	Feedback	No differentiation

et al., 2023). Other interventions did not apply modelling in the strict sense, but used examples of formative assessment strategies that were regarded by the developers as less costly, yet also successful, types of modelling, such as discussing narratives of lessons and exemplars in the form of materials presenting assessment techniques, for example.

## Monitoring and encouraging self-regulation

In all five TPD interventions, teachers' progress was *monitored by trainers*, often through multiple types of assessment techniques, such as discussions of implementation successes or failures (Interventions 1, 2, 3 and 4) and by reviewing teachers' action plans (Interventions 1 and 2) during every TPD session. Also, the fact that teachers were allowed to formulate their own action plans accounts for at least some teacher-drivenness in these interventions. Teachers in almost all interventions (1, 3, 4 and 5) were asked every session to evaluate their own work or performance and/or to plan new efforts to improve (Interventions 1, 2, 3, 4 and 5), although in most cases not systematically, which can have helped to *stimulate self-regulation*. Only in Intervention 1 were teachers also asked to set their own goals regarding mastery of formative assessment, although not completely teacher-driven, as it only concerned a selection from a set of pre-set goals. In the other interventions, teachers were mainly asked to work on common goals (i.e., implementation of a formative assessment strategy or strategies).

## Activating prior knowledge and retrieval practice

Some evidence pointed in the direction of the application of these two learning theory-based principles, although in none of the cases deliberately. Based on the publications, materials and interviews, we could deduce with some uncertainty that activating prior knowledge happened at least once (Intervention 1 and 3) or multiple times (Interventions 2, 4 and 5) in the interventions. When using qualitative monitoring activities, such as re-discussing a topic, teachers may have needed to activate prior knowledge (i.e., retrieval from long-term memory; Shiffrin & Atkinson, 1969). For example, discussing how teachers tried to implement sharing of learning goals, is likely to lead to teachers also recalling what the learning goals are. In some interventions (1, 2 and 4) teachers were stimulated to think about what their current understanding of formative assessment (strategies) in the classroom was, which could be regarded as a form of *retrieval practice*. However, this was done in combination with support from resources (e.g., the demonstration of a formative assessment model), which is likely to be less effective for long-term learning compared to retrieval practice without support from resources, as it takes less effort to retrieve it from your memory (van Merriënboer & Kester, 2014; Shiffrin & Atkinson, 1969). Still, the combination of retrieval with support from resources may likely be more suitable for more complex skills, as it likely already requires more cognitive resources.

## Learning tasks, spaced practice and classroom feedback

During the TPD sessions in all interventions and between the TPD sessions in some interventions (1, 2 and 4), teachers were assigned *learning tasks* to practise the previously discussed techniques and a combination of formative assessment (strategies) in their own classrooms (Interventions 1 and 2). For example, practising the use of assessment techniques also required knowing what the student learning goals were for their subject matter and the ability to formulate them clearly. *Spaced practice* is also a learning theory-based



principle that was not deliberately included in the TPD interventions. Sometimes, developers (Intervention 2) relied on the fact that the strategies were implemented as a whole, simply because the strategies were considered by them as dependent on each other. It is unlikely, however, that all teachers were able to connect the learning goals with the assessment techniques properly if this task was not situated during the TPD sessions themselves, as this requires strong pedagogical content knowledge (Jones & Moreland, 2005). The learning task of implementing practice was only observed in Intervention 1, which is a prerequisite for creating changes in the classroom, which eventually will turn in to (automatically performed) habits (Feldon, 2007). Teachers were asked to visit and observe each other's lessons between all 18 sessions for 20 min and to give classroom feedback to each other based on an observation sheet. All other interventions made use of feedback, but never on 'real' observed classroom situations. Instead, they discussed classroom situations that were described by the teachers themselves, which is likely not an unbiased representation of the situation.

## Differentiation

In one case *differentiation* took place based on the monitoring of teacher progress to tailor TPD content to teachers' needs (Intervention 1). Trained teacher leaders could adjust the TPD content based on observed teacher needs in their Teacher Learning Community. Most interventions (1, 3 and 4) differentiated through adapting feedback to between-teacher differences in progress, which can be considered differentiation in a very limited sense as we know that feedback is of limited impact when given as a form of one-way communication (Brooks et al., 2021). There was almost no form of differentiation at all in two interventions (2 and 5); the content to be learned, the learning tasks and the feedback were not adjusted to individual teacher needs there. In those interventions, differentiation only emerged as a reduced pace of instruction after observing less teacher progress than expected.

## Alignment of TPD goal, content, method and the context

The features of the (mutual) alignment between the goal, the method (excluding the aforementioned learning theory-based principles) and the content of the five selected TPD interventions for formative assessment are presented in Table 4. Below, we describe our attempt to assess the degree of alignment between all elements of the TPD interventions, including the learning theory-based principles while excluding the context characteristics of the TPD intervention. It appeared that in none of the included interventions was its method or content adapted deliberately to fit with specific context-related (school or teacher) characteristics. This likely indicates that one-size-fits-all interventions were implemented.

Interventions 1, 2, 3 and 5 appeared to have a somewhat higher alignment: the degree of complexity of their goals was reflected in the versatility of the content and the general features of the method. The goal of all interventions was to equip teachers for the use of formative assessment; however, they varied between a more complex inquiry approach (Interventions 1, 2 and 3) and a seemingly less complex measurement approach (Interventions 4 and 5). Only Intervention 5 followed a strict measurement approach, which was likely more aligned with their expert-driven content, along with a lower frequency of meetings over just one school year. This makes it likely to be less problematic that some of the learning theory-based principles, especially classroom feedback, were not applied (deliberately). In keeping with their inquiry approach to formative assessment, the methods used in Interventions 1, 2 and 3 provided more opportunities for the use of the learning theory-based principles. For example, more versatile TPD content and general TPD method features, such as more extensive

TABLE 4 Alignment between goal, method and (source of) content of the TPD interventions.

	Anders et al. (2022)	Andersson and Palm (2017)	Heller et al. (2012)	Randel et al. (2016)	Schneider and Meyer (2012)
Intervention (level of effectiveness)	1 (effective)	2 (effective)	3 (effective)	4 (less effective)	5 (less effective)
Approach to goal	Inquiry approach	Inquiry approach	Inquiry approach	Measurement (and inquiry) approach	Measurement approach
TPD content	Declarative & procedural knowledge	Declarative & procedural knowledge	Declarative knowledge	Declarative & procedural knowledge	Declarative & procedural knowledge
Source of TPD content	Expert- and teacher-driven	Expert- and teacher-driven	Expert and teacher-driven	Expert- and teacher-driven	Expert-driven
TPD hours per teacher	37.5	144	24	60	54
Duration TPD (school years)	2	0.5	0.25	1	1
Number of TPD meetings	18	24	12	10	10
Intervention target group	Whole school	One individual teacher	Several teachers	Whole school	Several teachers
Subject(s) taught	Multiple	Mathematics	Science	Multiple	English and mathematics
Target group experience with formative assessment	Novices and experts	Novices and experts	Novices and experts	Novices and experts	Novices
TPD trainers	Combination	External	External	None	External
Alignment	High	High	High	Medium	High

scaffolding of the implementation of formative assessment (i.e., procedural knowledge) and more frequent TPD meetings (Interventions 1, 2 and 3), sometimes in combination with a longer time frame (Intervention 1), likely gave more opportunities for spaced learning tasks and frequent retrieval practice. In particular, Intervention 1 showed high alignment in this regard: as the focus was on the whole school, Teaching Learning Communities (with 8–14 teachers each) were organised and led by a trained teacher leader. If a higher number of teachers at one school participate (as in Interventions 1 and 4), it is more likely that teachers will have more informal opportunities for (live) modelling and feedback, for example. Content-wise, this may have made it more difficult to align the TPD content with the subject-matter taught by all participating teachers, as all teachers were required to participate. The transfer to their own classroom might have been easier for teachers participating in Interventions 2 and 3, as these specifically focused on one subject or a domain of subjects (i.e., science).

Intervention 4 seemed to have somewhat lower alignment between its TPD elements than the other interventions. Since this intervention included elements of both an inquiry and a measurement approach, while at the same time trying to implement this for the whole school and for multiple subjects, the absence of a (trained) TPD trainer possibly made it harder to keep the learning theory-based principles intact. Without a knowledgeable TPD trainer, for example, it will likely be more difficult to adjust the content in a way that will suit teachers' professional needs (i.e., differentiation). Teachers may, for example, decide to ignore formative assessment strategies that they do not want to use, but that are in fact necessary for coherent implementation of formative assessment. Although Intervention 1 was similar to Intervention 4 in its focus on multiple subjects and the whole school, in Intervention 1 they included external support to keep track of and redirect, if necessary, the progress of the schools throughout their participation in the TPD intervention.

## Potential explanations for differences in effectiveness between interventions

Overall, the analysis of the implementation of learning theory-based principles shows that they were implemented only to a limited degree. In line with this finding, the investigation of the use of learning theory-based principles in relation to the effectiveness of the interventions did not show very clear patterns.

### Differentiation

Effective Interventions 2 and 3 and less effective Interventions 4 and 5 incorporated little or no differentiation at all. In Interventions 3 and 4, adapted feedback was used as a form of differentiation, but, in the case of Intervention 4, it was only from (untrained) peers. In Intervention 5, the developers stated that differentiation was not needed, because the entry level of all teachers was considered to be the same. Effective Intervention 1 did include all differentiation aspects, adaptation of content, feedback and tasks.

### Modelling

Only effective Intervention 1 used teaching examples for the purpose of modelling. Both effective Intervention 3 and less effective Intervention 4 made use of narratives to exemplify formative assessment, but not necessarily to demonstrate high-quality formative assessment.

## Classroom feedback

Effective Intervention 1 was the only intervention that made use of classroom feedback. Classroom feedback, especially in combination with observations, ask much of teachers, with an already full schedule as these interventions all took place within one school-year or less (Cordingley et al., 2015). As such, in all other interventions classroom feedback was not considered as a part of the TPD intervention.

## Goal-setting (i.e., part of self-regulation)

All studied interventions used activities to stimulate teacher self-regulation. However, only in effective Intervention 1 were teachers explicitly asked to set their own goals regarding the implementation of formative assessment, possibly improving teacher motivation (Kennedy, 2014), while keeping sight of effective formative assessment integration as a whole (Brooks et al., 2021; Epton et al., 2017). Effective Intervention 2 used teacher reflection and planning, but not the systematic self-evaluation of teachers' progress. The latter is likely difficult to implement as this often needs to be instructed to teachers (Rispoli et al., 2017), including formulating and establishing a system such as in Intervention 1.

One of the effective interventions (Intervention 1) did include several learning theory-based principles: the large-scale Embedding Formative Assessment (EFA) programme, evaluated by Anders et al. (2022). The study evaluating this intervention was the one with the most rigorous research design, a very large-scale randomised controlled trial ( $n=25,393$  students), older students compared to the other interventions included in this study (grade 9), and researcher-independent student achievement measurement instruments. So, although some learning theory-based principles were not used in effective interventions (Interventions 2 and 3), the implementation of these principles in Intervention 1 shows that when implemented, the inclusion of learning theory-based principles in TPD interventions (for formative assessment) may impact student learning positively.

## CONCLUSION

The aim of this study was to investigate the degree of application of learning theory-based principles in TPD interventions for formative assessment. Moreover, we wanted to explore whether there is a connection between the application of these learning theory-based principles, including the alignment between goal, content, method and context of the TPD interventions, and the effectiveness of the TPD interventions (in terms of improved student achievement). A first result of our study is that, unfortunately, the effects of TPD interventions for formative assessment on student outcomes were evaluated in only a small number of studies. We found only five studies that met the criteria for effect studies in education that we used from Slavin (2008) and the What Works Clearinghouse (2020). This resulted in a more exploratory study into the use of learning theory-based principles in TPD for formative assessment than we had originally intended.

The results of our exploratory analysis show that most of the learning theory-based principles were implemented only to a limited degree in both effective and less-effective TPD interventions for formative assessment. The overall lack of the use of learning theory-based principles is remarkable, as there is strong empirical evidence for the effectiveness of these principles for student learning. We found that retrieval practice was likely used at least, but likely at most, once in all TPD interventions—not as a deliberate component, but through including reflection about the application and effects of previously addressed formative

assessment strategies in teachers' own classrooms. This is despite the fact that differentiation, modelling and classroom feedback were understood as important elements of TPD by developers, but, with the exception of Intervention 1, these principles were considered too time-intensive to organise in a large-scale implementation.

A learning theory-based principle that stood out was how alignment between all TPD elements was considered in the TPD interventions. Generally, the complex TPD goals had a relatively good fit with the TPD content elements, such as provision of knowledge on how to apply formative assessment in the classroom, and TPD method elements, such as having a high number of TPD sessions. Despite these positive indications of overall alignment, the lack of focus on teachers' needs and wishes in goal-setting could have diminished the teacher-drivenness of all five interventions. The interventions were not often shaped to fit with teachers' negative or positive attitude regarding formative assessment (e.g., student involvement in assessment and feedback processes), besides giving teachers the opportunity to discuss their views on formative assessment. Another aspect is that the generally high intensity of the TPD interventions may have disrupted teachers' daily activities too much (Cordingley et al., 2015). This makes implementing formative assessment likely a multiple-year endeavour, instead of mostly one-year interventions described in this paper. Some of the teacher drop-out observed in the interventions could be explained by these factors, as resistance can occur if the context and personal needs of teachers are not included in the design of TPD interventions (Korthagen, 2017).

Intervention 1, which did have positive effects on student outcomes, stood out regarding the positive application of learning theory-based principles, such as: teachers' prior knowledge was activated without resources (activate prior knowledge); teachers were activated to self-reflect on their progress, and to set new goals in line with their own needs and wishes (self-regulation and teacher-driven in TPD content); and both novices and experts in formative assessment were stimulated to participate and were stimulated to focus on their own needs (differentiation). Teachers also regularly needed to practise the teacher skills necessary for formative assessment and also systematically assessed each other based on a benchmark (i.e., performance standards in the form of a rubric). Each TPD session was carefully structured beforehand by TPD developers, with an outlined hand-out suggesting an agenda, but TPD developers did give freedom to teachers in regard to the depth of the content to focus on, and assigned the role of 'critical reflector' to one of the teachers. There was external monitoring of teacher progress, but only once in a while, giving teachers room to explore their needs regarding formative assessment in the classroom, and at the same time, motivating teachers to keep making progress.

## DISCUSSION

Based on the findings and conclusions described above, we formulated four suggestions to improve TPD interventions regarding the use of learning-theoretical principles, which we discuss below.

### **Suggestion 1: Include performance standards as a basis for the (self-)evaluation of teacher progress**

Without clearly defined teacher behaviour standards for formative assessment, teachers can have divergent interpretations of what the goal of the TPD is (Visscher, 2021). Only in Intervention 1 was a clear rubric (called 'Learning Walk') provided to evaluate the quality of teachers' formative assessment behaviour in the classroom systematically. Similar to

theory on Assessment for Learning (Wiliam, 2006), in order to learn, it is likely that teachers need to know where they are now, where they need to go and how they can get there. These are also prerequisites for self-regulation (Zimmerman, 2002) and, therefore, durable teaching changes (Butler et al., 2004). To help teachers motivate themselves, it is important to include some form of teacher-drivenness regarding the TPD content, which is difficult in the case of formative assessment which apparently includes many teachers-need-to's (i.e., procedural knowledge) in order for it to be effective (e.g., share learning goals, use multiple assessment techniques frequently, or let students give each other feedback). To overcome this problem, we suggest to take teachers' current practice to help direct focus within formative assessment, without losing sight of the complexity of formative assessment (for example, supported by a 'Learning Walk' rubric).

## **Suggestion 2: Combine classroom practice with modelling and feedback**

Although much of the changes required in the classroom when implementing for formative assessment are about changing teacher behaviour, this teacher behaviour is in most interventions (except Intervention 1) not observed. By asking teachers to practise the whole task of formative assessment, they can be better supported in transferring the learned task from the intervention to their own classroom (van Merriënboer & Kester, 2014). Practising formative assessment regularly between TPD sessions, which means preparing and developing formative assessment strategies in the classroom, gives opportunities to apply the principles of retrieval practice and spaced practice, likely needed for the complex skill that is formative assessment. This may be done, for example, by using the whole rubric, such as the one used in Intervention 1, when modelling performance standards and giving feedback. Modelling, as shown in Intervention 1, can take the form of demonstrating certain teaching skills required for performance standards, and, in addition, explaining their thinking behind certain teaching actions in the classroom (Sims et al., 2023; van der Linden et al., 2022). The latter is in line with the logic 'seeing is believing', which has been found effective for teacher learning (Grierson & Gallagher, 2009). Giving peer feedback on classroom observations, as in Intervention 1, can make it possible to regularly give feedback and close feedback loops, as is often forgotten and leads to diminished long-term learning (Carless, 2019).

## **Suggestion 3: Assess teachers' needs and their context, to align interventions with both of them**

In some of the interventions, teachers' needs regarding formative assessment were assessed beforehand, but not with the deliberate aim to use this information for adapting the content of the TPD to teachers' learning needs. The characteristics of the TPD context did not influence the design of the TPD interventions. For example, there was no mention of differentiating between low- or high-performing schools/teachers, more or less experienced teachers, or different hours of TPD for schools (in line with their varying levels of expertise), when designing the TPD intervention. As integrating formative assessment requires a considerable investment of time and effort from schools and teachers (Heitink et al., 2016; van der Kleij et al., 2015), it may be worthwhile for developers to align TPD interventions more with such context-related characteristics.

## Suggestion 4: Complex teacher skills require more time than one school year to improve without putting too much pressure on teachers

The complexity of formative assessment may require more practice than currently is given room for. In most interventions, teachers are asked to make sizeable changes to their teaching, but without much time (more than one school year) allowed for practice, classroom feedback or, indeed, many of the learning theory principles (Cordingley et al., 2015; Feldon, 2007). These learning theory-based principles, which are backed by strong evidence to improve learning, can help make formative assessment implementation, which is quite complex, more durable than currently is the case (Gulikers & Baartman, 2017). On the other hand, the learning theory-based principles may put too much pressure on teachers, disrupting their teaching (Cordingley et al., 2015). This is why it is important to take more than one school year to be able to deliberately implement learning theory-based principles.

## Limitations of the study and recommendations for future research

The great majority of the studies that we evaluated to decide whether they could be included in our review did not adhere to many of our inclusion criteria, which were based on the standards by Slavin (2008) and the What Works Clearinghouse (2020). The purpose of using these rigorous criteria was to find studies that are as methodologically unbiased as possible and also as representative for educational practice (e.g., large-scale studies, at least 3 months of implementation), and thus indicative of likely estimates of intervention effects on student learning. The general lack of adherence to these criteria means that we have only limited insight into what formative assessment is capable of in terms of its effects on student achievement (or other outcomes) in educational practice. If we want to know that, then better interventions and evaluations are needed.

Also, often, the effects of the TPD intervention on student achievement were not evaluated, and the studies typically did not include a control group to compare student outcomes between the experimental and the control condition. Desimone (2023) went as far as to say that a TPD intervention that has not changed student outcomes is a failed intervention. This is a cause for concern, as the primary product of educational improvement should ultimately be the improvement of student outcomes, and we thus have very little solid evidence of the effectiveness of formative assessment (Timperley, 2008). We therefore recommend designing studies that specifically evaluate TPD interventions for formative assessment at the student level. Ultimately, we could only include five studies of TPD interventions for formative assessment. Of these five studies, two did not use independent measures to evaluate the effect on student achievement (we allowed non-adherence to one rigorous criterion), possibly resulting in unrealistically large effect sizes. It is important to take note of how this may have limited the conclusions that were drawn in this study (Slavin, 2008; What Works Clearinghouse, 2020). Because of the limitations above, our study became more of an exploration of whether and how learning theory-based principles are included in (in)effective TPD interventions for formative assessment, which may inspire future studies into this topic.

Lastly, the quality with which the principles were applied in our selected interventions must be interpreted with some caution. The reason for this is that our analysis was sometimes based on limited descriptions of the TPD interventions, and sometimes on the uncertain recall about the inclusion of principles by TPD developers (in our interviews). However, the learning theory-based principles themselves are the result of thorough research, and it is therefore worthwhile to further investigate how the learning theory-based principles can be used to improve TPD interventions for formative assessment.

## AUTHOR CONTRIBUTIONS

**A. J. Visscher:** Investigation; supervision; writing – review and editing; conceptualization. **J. A. de Vries:** Conceptualization; investigation; writing – original draft; methodology; visualization; writing – review and editing. **K. Schildkamp:** Investigation; writing – review and editing; supervision.

## ACKNOWLEDGEMENTS

We would like to thank all developers of the selected teacher professional development interventions for participating in the interviews, which were conducted for the purpose of analysing the learning theory-based principles in their interventions.

## FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Research data are not shared.

## ETHICS STATEMENT

Ethical approval for this study was obtained and given by the Faculty of Behavioural, Management and Social Sciences Ethics Committee of the University of Twente under number 211136.

## ORCID

J. A. de Vries  <https://orcid.org/0000-0003-1095-6180>

## REFERENCES

\* Refers to studies included in our review.

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. <https://doi.org/10.3102/0034654316689306>

\*Anders, J., Foliano, F., Bursnall, M., Dorsett, R., Hudson, N., Runge, J., & Speckesser, S. (2022). The effect of embedding formative assessment on pupil attainment. *Journal of Research on Educational Effectiveness*, 15(4), 748–779. <https://doi.org/10.1080/19345747.2021.2018746>

\*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(3–4), 179–203. <https://doi.org/10.1080/10627197.2006.9652989>

Ausubel, D. P. (2012). *The acquisition and retention of knowledge: A cognitive view*. Springer Science & Business Media.

\*Beesley, A. (2009). Measuring classroom assessment with a work sample. In *Paper presented at the annual meeting of the American Educational Research Association*. <https://files.eric.ed.gov/fulltext/ED508465.pdf>

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364. <https://doi.org/10.1007/bf00138871>

Brookhart, S. M., Moss, C. M., & Long, B. A. (2010). Teacher inquiry into formative assessment practices in remedial reading classrooms. *Assessment in Education: Principles, Policy & Practice*, 17(1), 41–58. <https://doi.org/10.1080/09695940903565545>



- Brooks, C., Burton, R., van der Kleij, F., Ablaza, C., Carroll, A., Hattie, J., & Neill, S. (2021). Teachers activating learners: The effects of a student-centred feedback approach on writing achievement. *Teaching and Teacher Education, 105*, 103387. <https://doi.org/10.1016/j.tate.2021.103387>
- Brown, G. T., Andersson, C., Winberg, M., Palmberg, B., & Palm, T. (2024). Swedish teacher conceptions of assessment: A focus on improving outcomes. *Education Inquiry, 1–21*. <https://doi.org/10.1080/20004508.2024.2375111>
- Butler, D. L., Lauscher, H. N., Jarvis-Selinger, S., & Beckingham, B. (2004). Collaboration and self-regulation in teachers' professional development. *Teaching and Teacher Education, 20*(5), 435–455. <https://doi.org/10.1016/j.tate.2004.04.003>
- Carless, D. (2019). Feedback loops and the longer-term: Towards feedback spirals. *Assessment & Evaluation in Higher Education, 44*(5), 705–714. <https://doi.org/10.1080/02602938.2018.1531108>
- Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283–292. <https://doi.org/10.3102/0013189X16656615>
- Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. (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>
- Coe, R., Rauch, C. J., Kime, S., & Singleton, D. (2020). Great teaching toolkit: Evidence review. <https://www.cambridgeinternational.org/Images/584544-great-teaching-toolkit-evidence-review-introduction.pdf>
- Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., Saunders, L., & Coe, R. (2015). *Developing great teaching: Lessons from the international reviews into effective professional development*. Teacher Development Trust.
- Crawford, A., Zucker, T., Van Horne, B., & Landry, S. (2017). Integrating professional development content and formative assessment with the coaching process: The Texas school ready model. *Theory Into Practice, 56*(1), 56–65. <https://doi.org/10.1080/00405841.2016.1241945>
- Creemers, B. P. M., & Kyriakides, L. (2011). *Improving quality in education: Dynamic approaches to school improvement*. Routledge. <https://doi.org/10.4324/9780203817537>
- Crichton, H., & McDaid, A. (2016). Learning intentions and success criteria: Learners' and teachers' views. *The Curriculum Journal, 27*(2), 190–203. <https://doi.org/10.1080/09585176.2015.1103278>
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.
- De Vries, J. A., Dimosthenous, A., Schildkamp, K., & Visscher, A. J. (2022). The impact on student achievement of an assessment for learning teacher professional development program. *Studies in Educational Evaluation, 74*, 101184. <https://doi.org/10.1016/j.stueduc.2022.101184>
- de Vries, J., Van Gasse, R., van Geel, M., Visscher, A., & Van Petegem, P. (2024). Profiles of teachers' assessment techniques and their students' involvement in assessment. *European Journal of Teacher Education, 1-20*, 369–388. <https://doi.org/10.1080/02619768.2024.2354410>
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199.
- Desimone, L. M. (2023). Rethinking teacher PD: A focus on how to improve student learning. *Professional Development in Education, 49*(1), 1–3. <https://doi.org/10.1080/19415257.2023.2162746>
- Epton, T., Currie, S., & Armitage, C. J. (2017). Unique effects of setting goals on behavior change: Systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology, 85*(12), 1182–1198. <https://doi.org/10.1037/ccp0000260>
- Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist, 42*(3), 123–137. <https://doi.org/10.1080/00461520701416173>
- \*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.
- \*Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science, 44*(3), 267–291. <https://doi.org/10.1007/s11251-016-9371-3>
- \*Graney, S. B., & Shinn, M. R. (2005). Effects of reading curriculum-based measurement (R-CBM) teacher feedback in general education classrooms. *School Psychology Review, 34*(2), 184–201. <https://doi.org/10.1080/02796015.2005.12086282>
- Grierson, A. L., & Gallagher, T. L. (2009). Seeing is believing: Creating a catalyst for teacher change through a demonstration classroom professional development initiative. *Professional Development in Education, 35*(4), 567–584. <https://doi.org/10.1080/19415250902930726>
- Gulikers, J. T. M., & Baartman, L. (2017). Doelgericht professionaliseren. Formatief toetsen met effect! Wat DOET de docent in de klas. [Goal-oriented professionalization. Formative assessment with impact! What DOES the teacher DO in the classroom]? Final report NRO-PPO overview study file number 405-15-722. NRO. <https://library.wur.nl/WebQuery/wurpubs/fulltext/440065>

- Guskey, T. R. (2003). Analyzing lists of the characteristics of effective professional development to promote visionary leadership. *NASSP Bulletin*, 87(637), 4–20. <https://doi.org/10.1177/019263650308763702>
- Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, 35(2), 213–224. <https://doi.org/10.1080/03057640500146880>
- 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>
- \*Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, 49(3), 333–362. <https://doi.org/10.1002/tea.21004>
- Heritage, M. (2020). Getting the emphasis right: Formative assessment through professional learning. *Educational Assessment*, 25(4), 355–358. <https://doi.org/10.1080/10627197.2020.1766959>
- \*Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., & Timms, M. (2015). Investigating the dynamics of formative assessment: Relationships between teacher knowledge, assessment practice and learning. *Assessment in Education: Principles, Policy & Practice*, 22(3), 344–367. <https://doi.org/10.1080/0969594X.2015.1006521>
- Hill, H. C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal*, 55(5), 1076–1112. <https://doi.org/10.3102/0002831218769614>
- Jones, A., & Moreland, J. (2005). The importance of pedagogical content knowledge in assessment for learning practices: A case-study of a whole-school approach. *Curriculum Journal*, 16(2), 193–206. <https://doi.org/10.1080/09585170500136044>
- Kang, S. H. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights From the Behavioral and Brain Sciences*, 3(1), 12–19. <https://doi.org/10.1177/2372732215624708>
- Kennedy, A. (2014). Understanding continuing professional development: The need for theory to impact on policy and practice. *Professional Development in Education*, 40(5), 688–697. <https://doi.org/10.1080/19415257.2014.955122>
- Kirschner, P., Sweller, J., & Clark, R. E. (2006). Why unguided learning does not work: An analysis of the failure of discovery learning, problem-based learning, experiential learning and inquiry-based learning. *Educational Psychologist*, 41(2), 75–86. [https://doi.org/10.1207/s15326985ep4102\\_1](https://doi.org/10.1207/s15326985ep4102_1)
- Kirschner, P. A., & Hendrick, C. (2020). *How learning happens: Seminal works in educational psychology and what they mean in practice*. Routledge.
- Korthagen, F. (2017). Inconvenient truths about teacher learning: Towards professional development 3.0. *Teachers and Teaching*, 23(4), 387–405. <https://doi.org/10.1080/13540602.2016.1211523>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, 33(2), 124–140.
- Nunes, L. D., & Karpicke, J. D. (2015). Retrieval-based learning: Research at the interface between cognitive science and education. In R. Scott & S. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences*. John Wiley & Sons.
- Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open*, 7, 2332858420986211. <https://doi.org/10.1177/2332858420986211>
- \*Phelan, J. C., Choi, K., Niemi, D., Vendlinski, T. P., Baker, E. L., & Herman, J. (2012). The effects of POWERSOURCE© assessments on middle-school students' math performance. *Assessment in Education: Principles, Policy & Practice*, 19(2), 211–230. <https://doi.org/10.1080/0969594X.2010.532769>
- \*Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research*, 109(5), 491–502. <https://doi.org/10.1080/00220671.2014.992581>
- Rispoli, M., Zaini, S., Mason, R., Brodhead, M., Burke, M. D., & Gregori, E. (2017). A systematic review of teacher self-monitoring on implementation of behavioral practices. *Teaching and Teacher Education*, 63, 58–72. <https://doi.org/10.1016/j.tate.2016.12.007>
- Rosenshine, B. (2010). *Principles of instruction*. Educational practices Series-21. UNESCO International Bureau of Education.
- Sadler, D. R. (2014). Beyond feedback: Developing student capability in complex appraisal. In S. Hatzipanagos & R. Rochon (Eds.), *Approaches to assessment that enhance learning in higher education* (pp. 55–70). Routledge.
- \*Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of national board certification. *American Educational Research Journal*, 45(3), 669–700. <https://doi.org/10.3102/0002831208316955>

- Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, 101602. <https://doi.org/10.1016/j.ijer.2020.101602>
- \*Schneider, M. C., & Meyer, J. P. (2012). Investigating the efficacy of a professional development program in formative classroom assessment in middle school English language arts and mathematics. *Journal of Multidisciplinary Evaluation*, 8(17), 1–24. <https://eric.ed.gov/?id=EJ971499>
- Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, 76(2), 179–193. <https://doi.org/10.1037/h0027277>
- Sims, S., & Fletcher-Wood, H. (2021). Identifying the characteristics of effective teacher professional development: A critical review. *School Effectiveness and School Improvement*, 32(1), 47–63. <https://doi.org/10.1080/09243453.2020.1772841>
- Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Goodrich, J., Van Herwegen, J., & Anders, J. (2023). Effective teacher professional development: New theory and a meta-analytic test. *Review of Educational Research*. <https://doi.org/10.3102/00346543231217480>
- Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Van Herwegen, J., & Anders, J. (2021). *What are the characteristics of effective teacher professional development? A systematic review and meta-analysis*. Education Endowment Foundation.
- Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5–14. <https://doi.org/10.3102/0013189X08314117>
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176–199. <https://doi.org/10.1177/1745691615569000>
- Suurttamm, C., & Koch, M. J. (2014). Navigating dilemmas in transforming assessment practices: Experiences of mathematics teachers in Ontario, Canada. *Educational Assessment, Evaluation and Accountability*, 26(3), 263–287. <https://doi.org/10.1007/s11092-014-9195-0>
- Tang, S. Y. F., Leung, P. P. W., Chow, A. W. K., & Wong, P. M. (2010). A case study of teacher learning in an assessment for learning project in Hong Kong. *Professional Development in Education*, 36(4), 621–636. <https://doi.org/10.1080/19415250903554087>
- Timperley, H. (2008). *Teacher professional learning and development*. *Educational practices Series-18*. UNESCO International Bureau of Education.
- Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development. Best evidence synthesis iteration (BES)*. New Zealand Ministry of Education. <https://www.educationcounts.govt.nz/publications/series/2515/15341>
- Tomlinson, C. A., & Imbeau, M. B. (2010). *Leading and managing a differentiated classroom*. ASCD.
- Van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. (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>
- van der Linden, S., van der Meij, J., & McKenney, S. (2022). Teacher video coaching, from design features to student impacts: A systematic literature review. *Review of Educational Research*, 92(1), 114–165. <https://doi.org/10.3102/00346543211046984>
- van Merriënboer, J. J., & Kester, L. (2014). The four-component instructional design model: Multimedia principles in environments for complex learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia* (pp. 104–148). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.007>
- \*Veugen, M. J., Gulikers, J. T. M., & den Brok, P. (2021). We agree on what we see: Teacher and student perceptions of formative assessment practice. *Studies in Educational Evaluation*, 70, 101027. <https://doi.org/10.1016/j.stueduc.2021.101027>
- Visscher, A. J. (2021). On the value of data-based decision making in education: The evidence from six intervention studies. *Studies in Educational Evaluation*, 69, 100899. <https://doi.org/10.1016/j.stueduc.2020.100899>
- Vygotsky, L. S., van der Veer, R. E., Valsiner, J. E., & Prout, T. T. (1994). *The Vygotsky reader*. Basil Blackwell.
- Wetzels, S. A., Kester, L., & van Merriënboer, J. J. (2011). Adapting prior knowledge activation: Mobilisation, perspective taking, and learners' prior knowledge. *Computers in Human Behavior*, 27(1), 16–21. <https://doi.org/10.1016/j.chb.2010.05.004>
- What Works Clearinghouse. (2020). *What works clearinghouse standards handbook (version 4.1)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/handbooks>
- Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3–4), 283–289. <https://doi.org/10.1080/10627197.2006.9652993>
- Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. Learning Sciences International.
- Winne, P. H., & Nesbit, J. C. (2010). The psychology of academic achievement. *Annual Review of Psychology*, 61, 653–678. <https://doi.org/10.1146/annurev.psych.093008.100348>

- Wolterinck, C. H. D., de Vries, J., Poortman, C. L., Schildkamp, K., & Visscher, A. (2021). Assessment for learning: The impact of a professional development programme on students' metacognitive strategies. Paper presented at 34th International Congress for School Effectiveness and Improvement, ICSEI 2021, Auckland, New Zealand.
- Wolterinck, C., Poortman, C., Schildkamp, K., & Visscher, A. (2022). Assessment for learning: Developing the required teacher competencies. *European Journal of Teacher Education*, *47*, 2124912.
- Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 140–160. <https://doi.org/10.1080/0969594X.2014.990416>
- Yan, Z., & Brown, G. T. L. (2021). Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. *Studies in Educational Evaluation*, *68*, 100985. <https://doi.org/10.1016/j.stueduc.2021.100985>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*(2), 64–70. [https://doi.org/10.1207/s15430421tip4102\\_2](https://doi.org/10.1207/s15430421tip4102_2)

**How to cite this article:** de Vries, J. A., Visscher, A. J., & Schildkamp, K. (2024). A learning theory-based exploratory analysis of teacher professional development interventions for formative assessment. *Review of Education*, *12*, e70009. <https://doi.org/10.1002/rev3.70009>