AECT | ASSOCIATION FOR EDUCATIONAL COMMUNICATIONS & TECHNOLOGY

Designing Formative Assessment That Improves Teaching and Learning: What Can Be Learned from the Design Stories of Experienced Teachers?

Janneke van der Steen[1,2] · Tamara van Schilt-Mol[2] · Cees van der Vleuten[1] · Desirée Joosten-ten Brinke[3]

Abstract

This article reports on findings of a qualitative study that investigated the difficulties teachers encounter while designing formative assessment plans and the strategies experienced teachers use to avoid those pitfalls. The pitfalls were identified through an analysis of formative assessment plans that searched for potential threats to alignment, decision-driven data collection, and room for adjustment and improvement. The main pitfalls in the design process occurred when teachers did not explicitly and coherently link all elements of their formative assessment plan or when they did not plan to effectively use information about student learning to improve teaching and learning. Through interviews with experienced teachers, we identified seven design strategies they used to design formative assessment plans that were aligned, consisted of decision-driven data collection, and left room for adjustment and improvement. However, these experienced teachers still encountered difficulties in determining how to formulate the right decisions for decision-driven data collection and how to provide students with enough room for improvement. Lessons learned from the design strategies of these experienced teachers are incorporated in design steps for formative assessment plans all teachers can use.

Keywords Assessment for learning · Decision-driven data collection · Constructive alignment · Formative assessment plan · Educational design · Teacher learning

Introduction

Formative assessment can be seen as an ongoing process of monitoring students' learning to decide which teaching and learning actions should be taken to better suit students'

✉ Janneke van der Steen
Janneke.vanderSteen@maastrichtuniversity.nl

Tamara van Schilt-Mol
Tamara.vanSchiltMol@han.nl

Cees van der Vleuten
C.vanderVleuten@maastrichtuniversity.nl

Desirée Joosten-ten Brinke
Desiree.Joosten-tenBrinke@maastrichtuniversity.nl

1 School of Health Professions Education, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands

2 Research Centre Qualities of Teachers, School of Education, HAN University of Applied Sciences, Nijmegen, the Netherlands

3 Department of Online Learning and Instruction, Faculty of Educational Sciences, Open Universiteit, Heerlen, the Netherlands

needs (Allal, 2020; Black & Wiliam, 2009). Activities that are part of effective formative assessment include clarifying expectations, eliciting and analyzing evidence of student learning, communicating the outcomes with students, and performing suitable follow-up activities in teaching and learning (Antoniou & James, 2014; Ruiz-Primo & Furtak, 2007; Veugen et al., 2021). Formative assessment reveals students' learning progress and what is needed to further this learning. Teachers can use this information to make better informed formative decisions about the next steps in teaching (Black & Wiliam, 2009).

Since formative assessment strengthens the connection between teaching and learning, it can be a solid intervention for improving both. However, implementing formative assessment that "works" is challenging for teachers. Research describes many pitfalls teachers can encounter when implementing formative assessment. For example, studies that investigated the implementation of formative assessment in practice conclude that in order to be effective, there needs to be more consideration for the integration, coherency, and alignment of formative assessment in classroom practice (Gulikers et al., 2013; Van Den Berg,

2018; Wylie & Lyon, 2015). Formative assessment should be aligned with learning objectives, lesson activities, and other assessment activities (Biggs, 1996; Gulikers et al., 2013). Moreover, since learning objectives often exceed lessons, this alignment of formative assessment should even be considered for multiple related lessons.

Additionally, Wiliam (2013, 2014) states that formative assessment activities that elicit evidence about student learning should also be designed in alignment with the decisions teachers wish to make based on the outcomes of these activities. Therefore, he recommends decision-driven data collection to ensure teachers and students receive the timely information they need to make well-informed formative decisions about the next steps in teaching and learning (Wiliam, 2013). However, teachers do not always incorporate decision-driven data collection in formative assessment, and this can be a pitfall when, for instance, existing data on student learning does not represent the current situation of learners or comes too late for a meaningful follow-up (Wiliam, 2013).

A recent study conducted by Veugen et al. (2021) examined students' and teachers' perceptions of formative assessment practice and revealed a final example of difficulties teachers seem to encounter when they implement formative assessment. Veugen et al. found that teachers who implement formative assessment use activities that clarify expectations and elicit and analyze student reactions. This results in feedback for the students, but the teachers report few adaptations to teaching and learning based on the outcomes of these activities. Without such follow-up activities, it is unlikely that formative assessment enhances student learning because students do not get the opportunity to use the feedback they were given, and teachers do not get the opportunity to adapt their teaching to students' needs (Black & Wiliam, 2009; Veugen et al., 2021). Formative assessment is not complete without a follow-up where students and teachers have room for adjustment and improvement. This room can be created in lessons that follow the analysis and communication of evidence of student learning. In summary, teachers encounter a range of difficulties when implementing formative assessment. Formative assessment should be aligned, include decision-driven data collection, and leave room for adjustment and improvement but, in practice, these criteria are rarely met.

It seems to be a complex task for teachers to consider these three criteria when conducting formative assessment. As a result, some teachers succeed in enacting formative assessment as recommended in the literature, while others experience more difficulties in reaching this goal (Offerdahl et al., 2018; Veugen et al., 2021). Previous research suggests that pre-planning formative assessment is fundamental for teachers to ensure the effectiveness of these activities by encompassing all essential characteristics (van der Steen et al., 2022). So far, literature focusing on designing formative

assessment has concentrated mainly on designing individual formative assessment activities (Furtak et al., 2018). However, only when teachers design formative assessment in plans that encompass multiple lessons can they tackle difficulties such as alignment and planning follow-up activities in an effective and feasible way (van der Steen et al., 2022; Wiliam, 2013). Taking a broad view of multiple lessons helps teachers consider the alignment between all lessons and activities that contribute to achieving the intended learning objectives. Furthermore, it increases their opportunities to plan for room for adjustment and improvement.

Based on the outcomes of earlier research (van der Steen et al., 2022), 64 teachers from four secondary schools were given time and knowledge to help them design formative assessment plans that met the criteria: alignment, decision-driven data collection, and room for adjustment and improvement. Still, even in this context, differences emerged between teachers when they were designing formative assessment. It seemed that teachers who already had experience with formative assessment in their classroom had an advantage in successfully designing formative assessment plans over teachers who did not yet have this experience.

Since there is a lack of literature about how teachers design formative assessment plans (van der Steen et al., 2022), it is unclear how teachers who design and implement formative assessment successfully design their formative assessment plans. Which design strategies do they use, and what can other teachers learn from their experiences? Therefore, the present study focuses on how experienced teachers design formative assessment plans aligned with learning objectives, lessons, and prospective formative decisions while taking follow-up actions into account. Once it becomes clear how teachers design such formative assessment plans, this knowledge can be used to support teachers who struggle with implementing formative assessment as intended. Therefore, the outcomes of this study will result in design steps and strategies for all teachers.

Accordingly, the research question for this study is:

How do experienced teachers design formative assessment plans that are aligned, include decision-driven data collection, and leave room for adjusting and improving teaching and learning?

Sub-questions that help answer this research question are:

1. What are pitfalls that can threaten the design of formative assessment plans that are aligned, include decision-driven data collection, and leave room for adjusting and improving teaching and learning?
2. How do experienced teachers design formative assessment plans that:

   (a) are aligned with learning objectives, lesson activities, and other assessment?

(b)   include decision-driven data collection?

(c)   leave room for adjusting and improving teaching and learning?

Material and Methods

Context

The context of this study is an educational design research project funded by a grant which provided four secondary schools with the opportunity to advance formative assessment in their schools. At these schools, teachers designed formative assessment plans in teacher learning communities. Teacher learning communities are groups of teachers that come together for sustained periods of time — in this case, 15 meetings during a 16-month period — to engage in inquiry and problem solving with the goal of improving student learning (Van Es, 2012). The teacher learning communities in this study focused on improving formative assessment and formative decision-making by designing formative assessment plans. The activities in the teacher learning communities were coordinated by the first author, who provided the teachers with information and support in designing formative assessment plans.

Each school had a teacher learning community that consisted of 11 to 24 teachers — a total of 64 teachers across the four schools — who designed formative assessment plans for their lessons according to five design steps (Fig. 1). These five design steps are based on design principles for formative assessment plans that meet the three quality criteria: alignment (design steps 1 and 5), decision-driven data collection (design steps 2, 3, and 5), and room for adjustment and improvement (design step 4) (van der Steen et al., 2022).

During a previous design cycle, teachers used an earlier version of the design steps for the first time. Thus, most teachers had experience designing a formative assessment plan prior to this study. The design steps were evaluated and adjusted based on group interviews and an analysis
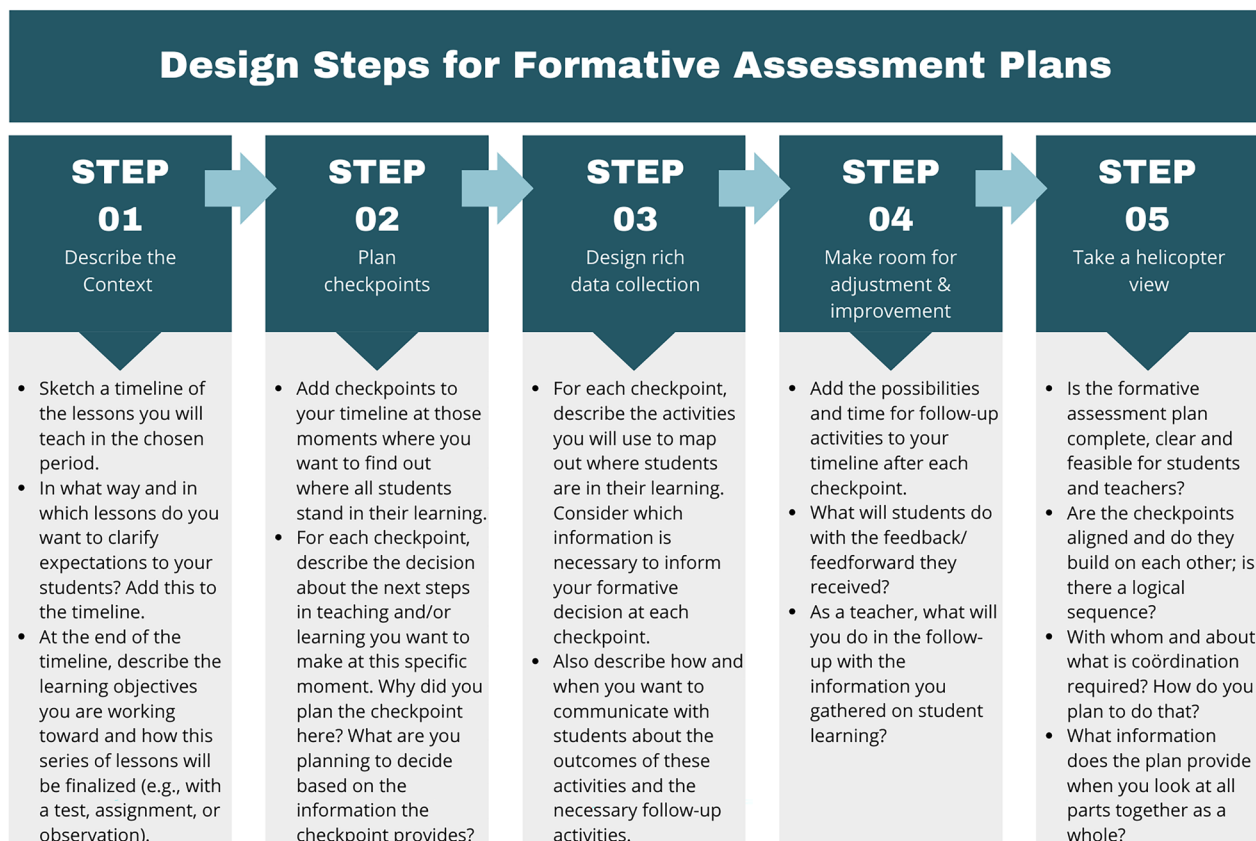

Design Steps for Formative Assessment Plans

STEP 01 Describe the Context	STEP 02 Plan checkpoints	STEP 03 Design rich data collection	STEP 04 Make room for adjustment & improvement	STEP 05 Take a helicopter view
• Sketch a timeline of the lessons you will teach in the chosen period. • In what way and in which lessons do you want to clarify expectations to your students? Add this to the timeline. • At the end of the timeline, describe the learning objectives you are working toward and how this series of lessons will be finalized (e.g., with a test, assignment, or observation).	• Add checkpoints to your timeline at those moments where you want to find out where all students stand in their learning. • For each checkpoint, describe the decision about the next steps in teaching and/or learning you want to make at this specific moment. Why did you plan the checkpoint here? What are you planning to decide based on the information the checkpoint provides?	• For each checkpoint, describe the activities you will use to map out where students are in their learning. Consider which information is necessary to inform your formative decision at each checkpoint. • Also describe how and when you want to communicate with students about the outcomes of these activities and the necessary follow-up activities.	• Add the possibilities and time for follow-up activities to your timeline after each checkpoint. • What will students do with the feedback/ feedforward they received? • As a teacher, what will you do in the follow-up with the information you gathered on student learning?	• Is the formative assessment plan complete, clear and feasible for students and teachers? • Are the checkpoints aligned and do they build on each other; is there a logical sequence? • With whom and about what is coördination required? How do you plan to do that? • What information does the plan provide when you look at all parts together as a whole?

Fig. 1 Five design steps for formative assessment plans

of the formative assessment plans designed during that first design cycle. The adjustments mainly concentrated on making the design steps more concise and emphasizing, within the design steps, the importance of communication with students, the link with formative decision-making, and planning room for students and teachers to improve.

In this study, the formative assessment plans the teachers designed and design stories of experienced teachers are used to answer the research questions. Figure 2 shows an example of a formative assessment plan.

This study has a qualitative research design. The first sub-question will be answered by analyzing the collected formative assessment plans for the presence and different appearances of alignment, decision-driven data collection, and room for adjustment and improvement, together with the pitfalls that prevent formative assessment plans from meeting these criteria. The second sub-question about what experienced teachers do to avoid these pitfalls will be answered based on interviews with experienced teachers who participated in this project.

Participants

Thirty-one teachers of 15 subjects from the four participating secondary schools were involved in answering sub-question 1 (see Table 1). All teachers ($n = 64$) that participated in the teacher learning communities at one of the four schools were asked if they could send their formative assessment plans (if they had finalized their plan by that time). To get a representative and information-rich overview of the pitfalls teachers encounter while designing formative assessment, despite experience in teaching and formative assessment or subjects taught, the researchers aimed to gather all finished formative assessment plans. This request resulted in 26 formative assessment plans from 31 teachers (presented in Table 1).

To answer sub-question 2, experienced teachers were recruited from the participating schools. To ensure the interviews provided in-depth information for this study, the teachers had to be actively involved in the design project and have multiple years of experience with formative assessment
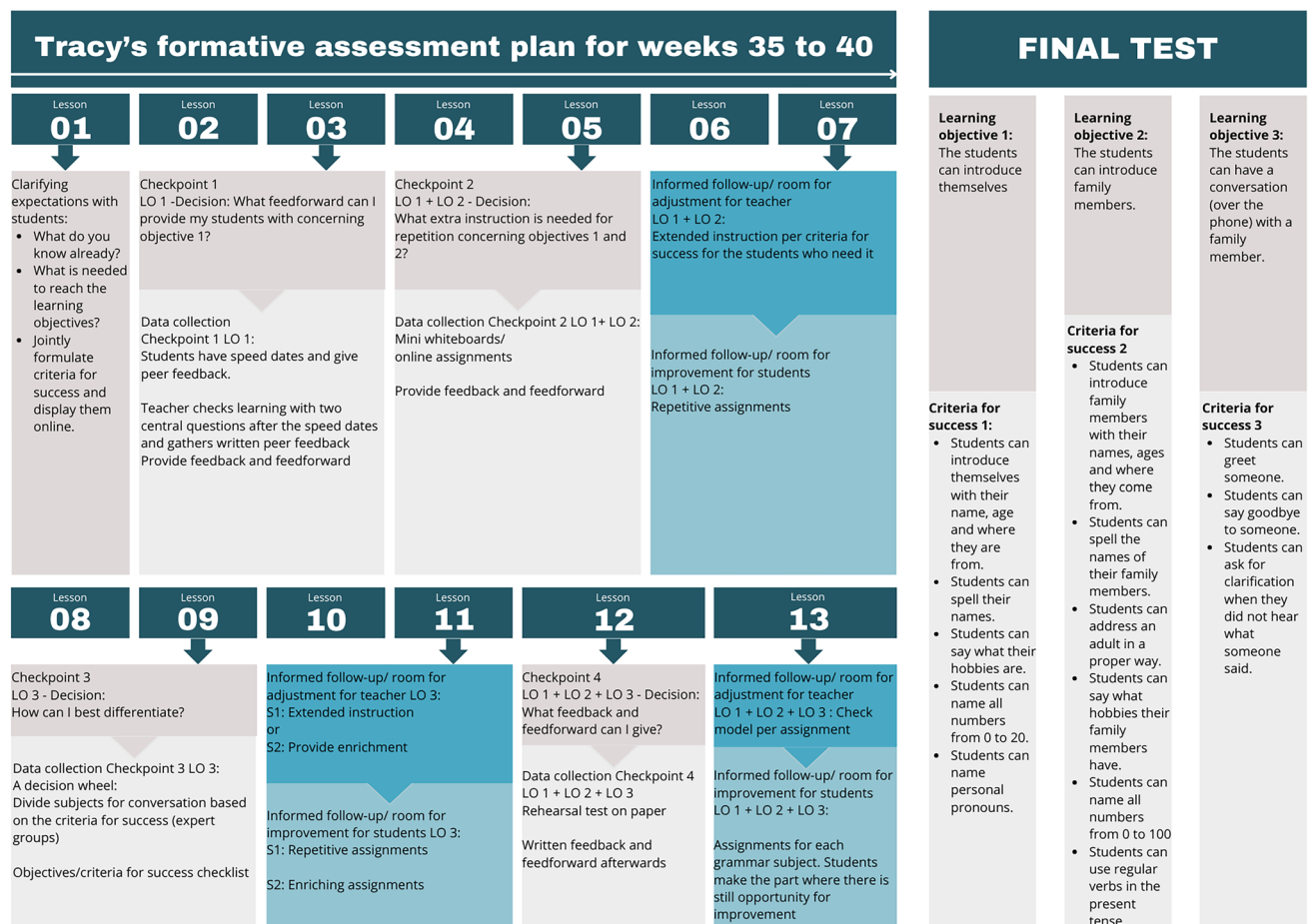

Fig. 2 Example of a formative assessment plan designed by Tracy

Table 1 Overview of collected formative assessment plans ($n = 26$)

School	School A	School B	School C	School D	Total
Number of formative assessment plans	8	5	7	6	26
Number of teachers who designed the formative assessment plans	10 (two English teachers and two mathematics teachers designed in pairs)	6 (two mathematics teachers teamed up)	8 (two physical education teachers teamed up)	7 (two arts and crafts teachers teamed up)	31
Subjects the plans were designed for	d. Mathematics b. Geography c. Arts and crafts d. Physics (2) e. Economics f. Spanish g. English	a. Mathematics b. Chemistry c. Drama d. Biology e. English	a. Mathematics (2) b. Physical education c. German d. Biology e. English f. Human and society	a. Care b. Physical education c. Arts and crafts d. Biology e. Technology f. Human and society	

so they could really understand and explain their choices and considerations in the design process.

All four participating schools were asked to find two teachers for the interviews who (1) agreed to contribute to this research via interviews (2) had finished designing their second formative assessment plan with success and (3) had experience with formative assessment prior to the start of teacher learning communities. For one school at which working with formative assessment was relatively new, no teachers that met these criteria could be found. The other three schools did find two teachers, as presented in Table 2 (all names are pseudonyms).

The table does not show the schools at which these teachers are employed, and a more general description was chosen for the language teachers to ensure their anonymity.

Analysis of Formative Assessment Plans

The 26 formative assessment plans came from 15 subjects and all four schools, varying from five to eight plans per school. That variety makes it likely that this collection of plans can provide a representative sample of formative assessment plans designed based on the five design steps, so there was no need to gather more plans.

Criteria for Analyzing Formative Assessment Plans

Before analyzing the plans to answer sub-question 1, a description was made of the criteria each element must meet to receive a positive evaluation. These were as follows:

- *Alignment*: plans should show proof of coherency between learning objectives, lesson activities, and assessment activities.

- *Decision-driven data collection*: plans should show proof that eliciting data about student learning was linked to a predetermined formative decision about the next steps in teaching and learning.
- *Room for adjustment and improvement*: plans should show proof of space and opportunity for both teachers and learners to adjust and/or improve based on the information about learning that was collected.

Procedure for Analyzing Formative Assessment Plans

The first step in analyzing the formative assessment plans was evaluating the plans on the three criteria: alignment, decision-driven data collection, and room for adjustment and improvement. This analysis was conducted by two researchers individually: the first author and one colleague researcher.

Second, for each criterion, the researchers discussed the differences in appreciation of the quality of the plans before they addressed the differences and similarities between the plans that succeeded in meeting the criterion and the plans that had not. What pitfalls appear in the plans that did not meet a criterion, and what can be learned from the plans that did? Some plans were more elaborately described and explained than others. Therefore, the results in this study are an analysis of the pitfalls in the plans with a clear and elaborate description and the possible pitfalls in the less well described plans.

Interviews with Experienced Teachers

Semi-structured teacher interviews were used to answer sub-question 2 and gain deep insight into the steps

Table 2 All interviewed teachers and the subjects they teach

Patricia	Tracy	Stuart	Bernadette	Lisa		Jenna
Mathematics	Language	Mathematics	Language	Human and society/Geography		Science

experienced teachers take to design their formative assessment plans. What did they do in addition to or differently from the five design steps, and how did this contribute to meeting the three quality criteria for formative assessment plans (sub-question 2)?

Guide for Interviewing Experienced Teachers

In the interviews, teachers were asked about their design process. Each interview started with the question: "How did designing this formative assessment plan come about?" Possible follow-up questions were: (a) "What did you do?," (b) "Which steps did you take in designing this plan?," and (c) "What difficulties did you encounter in designing this plan, and how did you resolve them?"

After the teachers explained their design process in their own words, the conversation turned to comparing the design steps to the design story the teacher had just shared. Where had the teacher followed the five design steps, and where did their process differ? For example, the teachers were asked about their choices in step 2 of the design process about when and why they had planned checkpoints, what they chose to do at each checkpoint to elicit information about student learning, and whether they had linked checkpoints together.

Procedure for Analyzing the Interviews and Writing Narratives of Experienced Teachers

The interviews were transcribed and coded through template analysis (Brooks et al., 2015). The statements and comments about the teachers' decisions and actions during the process of designing their formative assessment plan were coded using the five design steps (Fig. 1) and put into a narrative for each teacher. Based on each narrative, the researchers used the teachers' choices and actions that contributed to alignment, decision-driven data collection, and room for adjustment and improvement to answer the questions about what experienced teachers do in their design process to meet the three criteria this study focuses on (sub-questions 2a, 2b, and 2c).

Results

Pitfalls in Designing Formative Assessment Plans

Sub-question 1 was: "What are pitfalls that can threaten the design of formative assessment plans that are aligned, include decision-driven data collection, and leave room for adjusting and improving teaching and learning?" The results from analyzing the formative assessment plans are presented per criterion.

The main pitfalls related to alignment were:

– Learning objectives, lesson activities, and/or assessment activities were missing from the plan.
– Learning objectives, lesson activities, and/or assessment activities were not clearly described.
– Learning objectives, lesson activities, and/or assessment activities were not explicitly linked.
– There was a mismatch between the learning objectives, lesson activities, and assessment activities. For example, the final test or formative assessment did not match the learning objectives.

The three main pitfalls related to decision-driven data collection were:

– The absence of formative decisions that the data collection was based on. Specific formative decisions were not explicitly linked to each checkpoint and could not be deduced from the planned follow-up. When these decisions are unclear, it is impossible to determine whether the corresponding data collection was accurate for and aligned with the decision at hand.
– Situations in which predetermined formative decisions were present in a plan, but the data collection about student learning was not expected to collect the necessary information to inform the corresponding formative decision. The teacher did not plan to assess what they really needed to know. Four specific examples were found:

  • A mismatch between the data collection method and the necessary data. For example, a teacher who wants to find out about speech through mini whiteboards.
  • Situations in which teachers found a logical method but did not ask the right questions or use the proper assignments to discover where learning on a specific learning objective was and/or discover the next best step.
  • Situations in which teachers only planned to collect information on learning from a few students, but they wanted to use this information to make a formative decision for all students.
  • The planned data collection only shows whether a student has achieved a learning objective; it does not lead to information about what is needed in the follow-up to take learning further.

– Situations in which teachers found the right way to find the information on learning they needed, but the outcomes only became visible to the students (e.g., when students only give each other peer feedback). When outcomes only become visible to students, teachers cannot use the information to inform their formative decisions because they do not have it.

The three pitfalls related to room for adjustment and improvement were:

– Including no time for adjustment and improvement in the plan.
– Only including room for teachers to adjust and improve OR only for students to adjust and improve; no room was incorporated and described for both.
– Failing to confirm a follow-up by using another check to determine whether learning on a specific objective had increased because of the planned adjustment and improvement.

The Design of Formative Assessment Plans by Experienced Teachers

We used the narratives of six teachers (as presented in Online Resource 1) to answer sub-question 2: "How do experienced teachers design formative assessment plans that are aligned with learning objectives, lesson activities, and other assessment, include decision-driven data collection, and leave room for adjusting and improving teaching and learning?" The results are presented per criteria.

Alignment with Learning Objectives, Lesson Activities, and Other Assessment

All the teachers started with an existing series of lessons. According to Tracy, this is a coherent foundation from which to start when planning formative assessment if those lessons were designed based on the desired learning objectives (as they were in these cases). Tracy started with a series of lessons that also was aligned externally with an annual schedule that included all learning objectives and criteria for success and was consistent and aligned with the years ahead of or behind the current class. However, Patricia added that these planned series of lessons are not fixed. She stated that aligning formative assessment with existing lessons and lesson activities is an active process of determining whether lesson planning requires something different based on the choices made in designing formative assessment and simultaneously determining how formative assessment can enhance learning.

All the teachers described continuously checking for alignment during the design process, looking at the learning objectives, (formative) assessment, and lesson activities collectively. Stuart even believes checking for coherency and alignment should continue after the design process during and after the execution of the formative assessment plan: "Only then can you really fly over it and notice whether the cohesion is good enough."

All the teachers made sure they understood the learning objectives by taking time to zoom in and formulate criteria for success and/or to zoom out to look for overarching learning objectives to bring everything together. Patricia, Stuart, and Jenna took time to look closely at learning objectives, transcend specifics, discover the coherence between objectives, or find broader objectives that can connect different learning objectives. Another group of teachers (Tracy, Stuart, Bernadette, and Lisa) took the time to formulate criteria for success so objectives would be more specific for teachers and students. Stuart analyzed former lessons, assignments, and previous misconceptions to get a good idea of the criteria for success that should be pursued. Lisa and Tracy formulated these criteria for success together with their students based on examples of work. Patricia, Jenna, and Tracy all mentioned that they think zooming in and out on learning objectives with colleagues is a valuable part of the design process.

After ensuring they comprehended the learning objectives, the teachers planned checkpoints that covered all the learning objectives (five teachers) or a selection of the learning objectives (one teacher). These checkpoints were linked explicitly to the learning objectives. Tracy clearly listed the learning objectives in her formative assessment plan to continuously verify that all activities aligned with what she wanted to accomplish with her class. For Stuart, it was essential to determine which learning objectives and criteria for success were being targeted in each lesson so he could refer to them in his instructions and assignments to the students. This will make it easier for students to reflect on their learning based on these learning objectives and criteria for success because they will be present in each lesson.

Most teachers designed their formative assessment to make it easy to integrate into their existing lesson plans. For example, Tracy said: "I actually really looked at which assignments I use to get them to practice and, based on those assignments, I decided which data collection activity would fit in easily." Patricia planned activities as not only a means to collect data on students' learning but also as an opportunity for students to repeat and rehearse for the selected learning objectives. Stuart and Lisa added that they designed formative assessment to collect data on learning as they would design test items for the final test.

Decision-Driven Data Collection

The teachers wanted to use the information gathered at the checkpoints to make three decisions:

1. Can I go on to the next topic/learning objective/chapter or do I need to spend more time on the current one? (the most common decision)

2. How can I best differentiate in my lessons to support all students in their learning?

3. What do students need to rehearse/repeat/practice/learn more to be prepared for the test?

Sometimes the information gathered at one checkpoint applied to several of these decisions.

When designing decision-driven data collection, the teachers were primarily concerned with how to measure what they needed to know to inform the next formative decision. For example, Lisa planned to use a drawing instead of a question to allow every student to show their learning on the topic unimpeded by their writing skills. Stuart did not want to analyze reflection forms since they only demonstrate students' perceptions of learning, not their progress. Likewise, Jenna did not want to analyze summaries since she does not think they illustrate what students have learned but only whether they can summarize.

Stuart added that he plans decision-driven data collection through exit tickets or mini whiteboards because this gives him more information about all students' learning than he could acquire by walking around while students do their homework in class.

> Because it is quite a pitfall that if you see something go wrong with one person, you zoom in completely on that person. Before you know it, you are working with them for five or six minutes and, supposing they have 15 minutes to work independently, then you have time to see two such students. Then my harvest is two or three students. If I base my data collection on those whiteboards and exit tickets or other formative elements, I get more of an overall picture, and I find that much more valuable.

Patricia emphasized that it is crucial to gather rich information (e.g., about existing misconceptions) instead of a count of how many questions were answered incorrectly. However, she also mentioned that it is not always evident whether a teacher collects rich data in a formative assessment plan: "I believe it is rich information because you certainly address the mistakes you saw and the corresponding underlying misconceptions. Nevertheless, that is so self-evident that it is not mentioned here."

Data used to inform decisions at each checkpoint is collected in different ways. Most of the teachers planned to use combinations of data collection methods to get a complete overview of all students' learning, but they did this in various ways. Some wanted to combine data collection at the checkpoints with planned or on-the-fly daily checks during the lessons before the checkpoint. For example, Stuart plans daily checks of specific lesson goals by using mini whiteboards and exit tickets in addition to formal checkpoints that reveal learning on the overarching learning objectives. Lisa

planned data collection at the checkpoint together with data collection via an online tool that helps her follow how students did on their assignments in the previous lessons. This helps her gather all the information she needs to decide on the next best step after a checkpoint.

However, other teachers only collect data at the planned checkpoints, at which time they aim to gather rich information on all students by using multiple data collection methods simultaneously. Tracy, for instance, combined walking around the classroom while students worked on their assignments with gathering written peer feedback on the same assignment at her first checkpoint. At the second checkpoint, she combined online assignments with answers from mini whiteboards.

The final variation, mentioned by Bernadette and Tracy, was using checkpoints not to gather new information on learning but to bring together all relevant information on learning from all prior lessons until that moment. This can result in rich information and help teachers make formative decisions that have consequences for all students based on information from all students. For example, when they cannot let every student speak during a lesson, they combine evidence from multiple lessons to get a complete overview of students' speech.

Room for Adjustment and Improvement

Experienced teachers recognize how difficult it is to find time to make room for adjustment and improvement in an overfull curriculum. Follow-up is the part of formative assessment that takes the most time and thus the first thing to skip when time is tight, some teachers said. However, Jenna also stated that this is the essential part, and the other teachers seemed to agree since they all have strategies to help them make room for adjustment and improvement.

Three teachers prepared different possible follow-ups in advance to help them act on the checkpoints' outcomes immediately. Patricia, Stuart, and Bernadette prepared slides, instructions, tutorials, and/or assignments for students so they would be ready once they knew what was needed to advance students' learning. In this preparation, they considered different possible outcomes and differences between students (e.g., they prepared assignments for students who had reached particular learning objectives and those who still needed to). Instead of only focusing on the students who did not reach a targeted learning objective, all the teachers planned room for adjustment and improvement for all students. Patricia and Tracy created common assignments for their students that would simultaneously advance the learning of students who had as well as students who had not yet reached a targeted learning objective.

Tracy and Jenna warned that when a curriculum contains too much material, teachers have less flexibility in reacting

to outcomes that are different than expected. Therefore, Jenna, Patricia, and Stuart advised their fellow teachers to be aware of and look for possibilities to leave something out or create more room in the curriculum when planning their lessons. Stuart mentioned many examples of how he makes room for adjusting and improving teaching and learning. For example, he and Jenna plan checkpoints to help them decide whether they can skip (parts of) instruction. Additionally, Stuart creates plans that will help the few students who need it by assigning tutorials and extra assignments to complete at home, instead of taking a whole lesson for extra instruction and practice.

Patricia, Jenna, and Stuart also delay more advanced tasks until formative assessment shows that students are ready to take this next step. That gives them more room for adjustment and improvement on the basics at the start of a lesson plan. Jenna:

> If you do not do that well, do not lay that foundation properly, you can start working on another subject soon. However, the students still have not mastered that, and next comes something that practically goes back to this, so the foundation must be set correctly.

Stuart even finds room for adjustment and improvement after the final test, since he plans room for student improvement in the successive lessons and in assignments for future lessons that take place after the chapters' final test: "Then they understand that it is not just for now and the next test, but learning is continuous." According to Stuart, giving students room for improvement could even mean moving a test to a later date. Bernadette, Stuart, Lisa, and Jenna also make sure students take responsibility for designing their own room for improvement by having students think about the best follow-up. Lisa also wants to give students information based on the checkpoints that they can use to make their own improvement plans.

Conclusions and Discussion

This study focused on the research question: How do experienced teachers design formative assessment plans that are aligned, include decision-driven data collection, and leave room for adjusting and improving teaching and learning? Through the interviews with experienced teachers, seven strategies were found that they used to design formative assessment plans that meet these criteria:

1. Starting from existing and consciously planned series of lessons.
2. Taking time to understand learning objectives.

3. Using existing lesson activities to elicit evidence of student learning instead of designing new ones.
4. Precisely measuring what they need to know to inform their formative decisions.
5. Combining various sources of information about all the students' learning to inform their formative decisions.
6. Preparing different follow-ups to match different possible outcomes from formative assessment.
7. Purposely creating time and space for improvement and adjustment in the curriculum.

To discover the difficulties teachers can encounter in designing formative assessment plans that are aligned, include decision-driven data collection, and leave room for adjusting and improving teaching and learning, sub-question 1 asked which pitfalls can threaten the design of formative assessment plans. One frequently found pitfall was that formative assessment plans were incomplete or unclear. A formative assessment plan needs to clearly describe, explicitly link, and consciously match learning objectives, lesson activities, and assessment in line with what Biggs (1996) defined as constructive alignment. Additionally, a formative assessment plan needs to clearly describe, explicitly link, and consciously match intended formative decisions, data collection and follow-up to meet the criteria of decision-driven data collection and room for adjustment and improvement.

The other pitfalls that were found were:

(a) A mismatch between the learning objectives, lesson activities, and/or assessment activities that resulted in less alignment;
(b) The data planned to be collected for information about student learning was not the information that was needed to inform the corresponding formative decision and therefore was not considered decision-driven data collection;
(c) The information on student learning gathered through the planned data collection did not become available or visible for the teacher to help them inform the corresponding formative decision and therefore is not part of decision-driven data collection; and
(d) The planned room for adjustment and improvement was not followed by another checkpoint to establish whether it had helped to achieve the expected learning objectives and assess whether there had been enough room for adjustment and improvement.

Sub-question 2 was aimed at discovering what experienced teachers do to design formative assessment plans that are aligned, include decision-driven data collection, and leave room for adjustment and improvement. The interviewed teachers mentioned seven design strategies that help

them avoid most of the pitfalls listed above and two other pitfalls that they experienced during designing formative assessment. The design strategies will be presented in more detail per criterion with the implications for improving the design steps. All implications are shown in a revised version of the design steps in Online Resource 2. The pitfalls that these teachers did not yet find a solution for will also be presented together with the implications for future research.

Design Strategies for Achieving Alignment

The six interviewed teachers used three design strategies to ensure sufficient alignment in their formative assessment plans:

1. Start with existing and consciously planned series of lessons. When these are aligned with the learning objectives, there is a solid foundation for the formative assessment plan.
2. Take time to understand the learning objectives thoroughly. Aligning formative assessment plans requires teachers to take a good look at learning objectives, zoom in and out, transcend specifics to discover the coherence between objectives, and make them more specific for teachers and students to work with. The learning objectives are the thread that links everything together, and teachers use them during design and in their lessons to ensure and continuously check alignment.
3. Do not regard formative assessment as something new or extra that must be added to lesson plans. Instead, teachers should take the opportunity to re-purpose and evaluate known and existing lesson activities in their lesson plans to investigate the possibilities of using them as activities to elicit evidence about students' learning as part of a formative assessment plan.

These findings have implications for design steps 1, 3, and 5:

- *Step 1: Describe the context.* Based on these outcomes regarding alignment, the first design step should suggest teachers to start their design with existing series of lessons that have already been designed, checked, implemented, and evaluated with alignment in mind. Additionally, it should emphasize that teachers should not only describe the learning objectives but take time to thoroughly understand them by zooming in and out to look for overarching learning objectives to bring everything together.
- *Step 3: Design rich data collection.* To contribute to alignment design step three should include the suggestion to re-purpose learning and assessment tasks that are already part of the existing series of lessons and use them formatively instead of adding on formative assessment.

- *Step 5: Take a helicopter view.* This step is now situated at the end of the design process. However, according to the teachers, this step encompasses the total design process since they continuously check for alignment during the design and execution of the formative assessment plan.

Design Strategies for Decision-Driven Data Collection

The experienced teachers used two design strategies to design decision-driven data collection (Wiliam, 2013, 2014):

1. Ensure to precisely measure what you need to know by matching the form and content of the data collection methods to the formative decision the teacher wants to make.
2. Combine various information on the learning of all students to inform formative decisions solidly. The teachers reported three approaches they designed to help them do this:

  - Teachers combined information from different formative checks — planned or on the fly from prior lessons — with data collection conducted up to and around the checkpoints to inform their formative decisions.
  - Teachers collected data only at the planned checkpoints, at which time they aim to gather rich information on all students by using multiple data collection methods simultaneously.
  - Teachers did not use checkpoints to gather new information on learning but used them to bring together all relevant information on learning from all prior lessons until that moment.

  If data is consciously collected to match and inform the formative decisions, all these variations can be considered decision-driven data collection. The important overarching design strategy is that teachers combine information from multiple data collections to inform their formative decisions.

As for decision-driven data collection, the two strategies experienced teachers mentioned are already incorporated in the design steps. However, the importance of these strategies can be emphasized in design step 3 and 5 based on this study:

- *Step 3: Design rich data collection.* To emphasize the collection of rich data, this step could recommend combining multiple data collection methods positioned at the checkpoint and/or in the lessons before the checkpoint to get an honest, reliable, and complete overview of student learning.
- *Step 5: Take a helicopter view.* Since design step 3 already mentions that teachers should consider if the data collection measures what they need to know, this could be

emphasized by including this as an extra check in design step 5.

Design Strategies for Creating Room for Adjustment and Improvement

Most of the experienced teachers in this study acknowledged the difficulty of creating room for adjustment and improvement (Veugen et al., 2021). As a result, the three design strategies the teachers used to create room for adjustment and improvement focus on increasing the possibility to use the outcomes of the checkpoints:

1. Ensure to prepare various follow-ups to ensure teachers always have assignments, tutorials, and instructions ready to go and suitable for different outcomes and students.
2. Consciously create space and time in the curriculum for adjustment and improvement. The teachers reported three approaches that helped them do this:

   - Teachers planned checkpoints with data collection to help them decide whether they can skip, speed up, or delay the curriculum to create room for adjustment and improvement.
   - Teachers planned to delay more advanced tasks until they were sure students had acquired the basics
   - Teachers planned space for student improvement outside the classroom or after the final test.

   Most teachers thought it was also important that students learned to make their own improvement plans and become co-responsible for the learning process.

As for the design steps, the design strategies for creating room for adjustment and improvement may lead to changes in design steps 4 and 5:

- *Step 4. Make room for adjustment and improvement. T*eachers might find it helpful to prepare follow-ups in advance for multiple possible outcomes and all students at each checkpoint. This design step could incorporate this suggestion
- *Step 5*: *Take a helicopter view.* In this step, teachers could be encouraged to think about risks that could hinder the creation of room for adjustment and improvement after a checkpoint and how to prevent this or prepare an alternate plan.

Other Pitfalls

Apart from the design strategies, the interviews also disclosed that even experienced teachers face difficulties in designing formative assessment. The two pitfalls that were mentioned can make it hard for teachers to design and implement formative assessment plans that are aligned, include decision-driven data collection, and make room for adjustment and improvement despite of the strategies they already use.

Narrow Formative Decisions from the Foundation for Decision-Driven Data Collection

The teachers reported that three formative decisions at the checkpoints were the foundation for their formative data collection. These decisions were limited to: "Can I go on with the next lesson/chapter/learning objective or is something else needed?" However, to ensure decision-driven data collection contributes to creating the best suitable follow-up for students, the decisions should go beyond "Can I go on or not" and include "What is the best way to move forward?" Therefore, the decisions that are the foundation of decision-driven data collection should have a double focus.

The first focus refers to the "yes or no" decision (e.g., "Can I go on to the next chapter?" or "Is it necessary to differentiate?"). There, teachers need to add a "how to continue" decision (e.g., "What is the best way to go forward?" or "How should I differentiate in the next lesson?", respectively). When teachers incorporate both in their formative decisions, it is more likely that decision-driven data collection will provide the information needed to choose the follow-up that best suits students' needs.

The second focus involves collaboration. Teachers can use conversations with their students to discover the best suitable follow-up (Allal, 2020). Formative assessment will achieve its full potential when it is perceived as support for development and learning and discovering how to suit students' needs best rather than solely used for control and accountability and solely focused on the go or no go (Ninomiya & Shuichi, 2016).

Limited Room for Improvement

While one teacher advocated for perceiving learning and formative assessment as a continuous process that transcends specific chapters or series of lessons, the other teachers often wondered "How much room for improvement can I give my students?" or "How can I justify continuous and differentiated learning processes that suit students' needs and still achieve all the learning objectives with all my students within a specific time?"

The differences between these teachers and the choices they make can be explained in several ways. The teachers' pedagogical foundation and their knowledge and skills regarding formative assessment can play a role, as can the

amount of space they perceive to define and shape their work within the school context (i.e., their professional agency) (Heitink et al., 2016; Oolbekkink-Marchand et al., 2022). These differences can be used to explain why one experienced teacher creates room for improvement during a new chapter or delays a test, while others wonder how much room for improvement they can give their students before they must move forward to the next subject, lesson, or learning objective. The amount of space and freedom teachers experience to let formative assessment lead their teaching decisions and the role teachers' agency and experience play in this process would be interesting subjects for future research.

Limitations and Implications for Future Research

A first limitation of this study was that the conclusions about sub-question 1 are based on formative assessment plans that were only a reality on paper. The outcomes of the analysis were based on the complete plans and the possible risks perceived in the incomplete plans. Since the formative assessment plans are merely a written plan rather than a reflection of action, they do not always show what a teacher really planned to do. During the interviews, it became clear that when teachers explained their formative assessment plans, a lot of information about them had not been written down but was still essential to comprehend what they planned to do. Since only six teachers were interviewed, it is possible that asking more teachers to explain their formative assessment plans might have led to fewer, more, or different pitfalls in designing formative assessment.

Another limitation of this study was the inclusion of only six experienced teachers, mainly from theoretical subjects. It is unclear whether less experienced teachers or teachers of more practical subjects would use the same or other design strategies to meet the three criteria. Future research could focus on collecting a range of design stories from experienced and new teachers, with or without prior knowledge and skills concerning formative assessment, of both theoretical and practical subjects.

In conclusion, it is interesting to note a paradox about formative assessment. The current study showed that when teachers consciously prepare, plan, and match their formative assessment in advance, this helps them to achieve alignment, decision-driven data collection, and room for adjustment and improvement. However, this contradicts Black and Wiliam's (2009) definition, which emphasizes that formative assessment "is concerned with the creation of, and capitalization upon, 'moments of contingency'" (p. 10). When formative assessment is extensively planned, are teachers still able to pick up on the element of surprise? Decision-driven data collection can present them with different outcomes than expected, so an unforeseen follow-up may be needed to best suit students' needs. How can teachers keep an inquiring and open mind when they collect and analyze evidence of learning after they planned their formative decisions, data collection, and follow-ups in detail in advance?

Thus, it would be interesting for future research to discover how teachers use information about students' learning to inform their decisions. Do they analyze this information with enough openness and curiosity to really discern students' needs and adjust their planned follow-up based on unexpected outcomes?

Data Availability The data that support the findings of this study are available from the corresponding author, [JS], upon reasonable request. The data will not become publicly available before the ending of the research project this study is connected to.

Declarations

Ethical Approvement and Consent Approval was obtained from the ethics committee of HAN University of Applied Sciences. The procedures used in this study adhere to the tenets of the Declaration of Helsinki. Informed consent was obtained from all individual participants in the study.

Conflict of Interest The authors declare no competing interests.

References

Allal, L. (2020). Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice, 27*(4), 332–349. https://doi.org/10.1080/0969594X.2019.1609411

Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. *Educational Assessment Evaluation and Accountability, 26*(2), 153–176. https://doi.org/10.1007/s11092-013-918

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

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Brooks, J., McCluskey, S., Turley, E., & King, N. (2015). The utility of template analysis in qualitative psychology research. *Qualitative Research in Psychology, 12*(2), 202–222. https://doi.org/10.1080/14780887.2014.955224

Furtak, E. M., Circi, R., & Heredia, S. C. (2018). Exploring alignment among learning progressions, teacher-designed formative assessment tasks, and student growth: Results of a four-year study. *Applied Measurement in Education*, *31*(2), 143–156. https://doi.org/10.1080/08957347.2017.1408624

Gulikers, J., Biemans, H., Wesselink, R., & van der Wel, M. (2013). Aligning formative and summative assessments: A collaborative action research challenging teacher conceptions. *Studies in Educational Evaluation, 39*(2), 116–124. https://doi.org/10.1016/j.stueduc.2013.03.001

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

Ninomiya, S. (2016). The possibilities and limitations of assessment for learning: Exploring the theory of formative assessment and the notion of "Closing the learning gap". *Educational Studies in Japan: International Yearbook, 10* 79-91.

Offerdahl, E. G., McConnell, M., & Boyer, J. (2018). Can I have your recipe? Using a fidelity of implementation (FOI) framework to identify the key ingredients of formative assessment for learning. *CBE Life Sciences Education, 17*(16), 1–9. https://doi.org/10.1187/cbe.18-02-0029

Oolbekkink-Marchand, H., van der Want, A., Schaap, H., Louws, M., & Meijer, P. (2022). Achieving professional agency for school development in the context of having a PhD scholarship: An intricate interplay. *Teaching and Teacher Education, 113*, 103684. https://doi.org/10.1016/J.TATE.2022.103684

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84. https://doi.org/10.1002/tea.20163

Van Den Berg, M. (2018). *Classroom formative assessment: A quest for a practice that enhances students' mathematics performance*. [Doctoral dissertation, University of Groningen]. University of Groningen Research Outputs. https://research.rug.nl/en/publications/classroom-formative-assessment-a-quest-for-a-practice-that-enhanc

van der Steen, J., van Schilt-Mol, T., van der Vleuten, C., & Joosten-ten Brinke, D. (2022). Supporting teachers in improving formative decision-making: Design principles for formative assessment plans. *Frontiers in Education*, *7*, 925352. https://doi.org/10.3389/feduc.2022.925352

Van Es, E. A. (2012). Examining the development of a teacher learning community: The case of a video club. *Teaching and Teacher Education, 28*, 182–192. https://doi.org/10.1016/j.tate.2011.09.005

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

Wiliam, D. (2013). Assessment: The bridge between teaching and learning. *Voices from the Middle*, *21*(2), 15–20. https://library.ncte.org/journals/vm/issues/v21-2/24461

Wiliam, D. (2014). *Formative assessment and contingency in the regulation of learning processes.* Paper presented at the annual meeting of the American Educational Research Association, Philadelphia, PA. http://dylanwiliam.org/Dylan_Wiliams_website/Papers_files/Formative%20assessment%20and%20contingency%20in%20the%20regulation%20of%20learning%20processes%20%28AERA%202014%29.docx

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