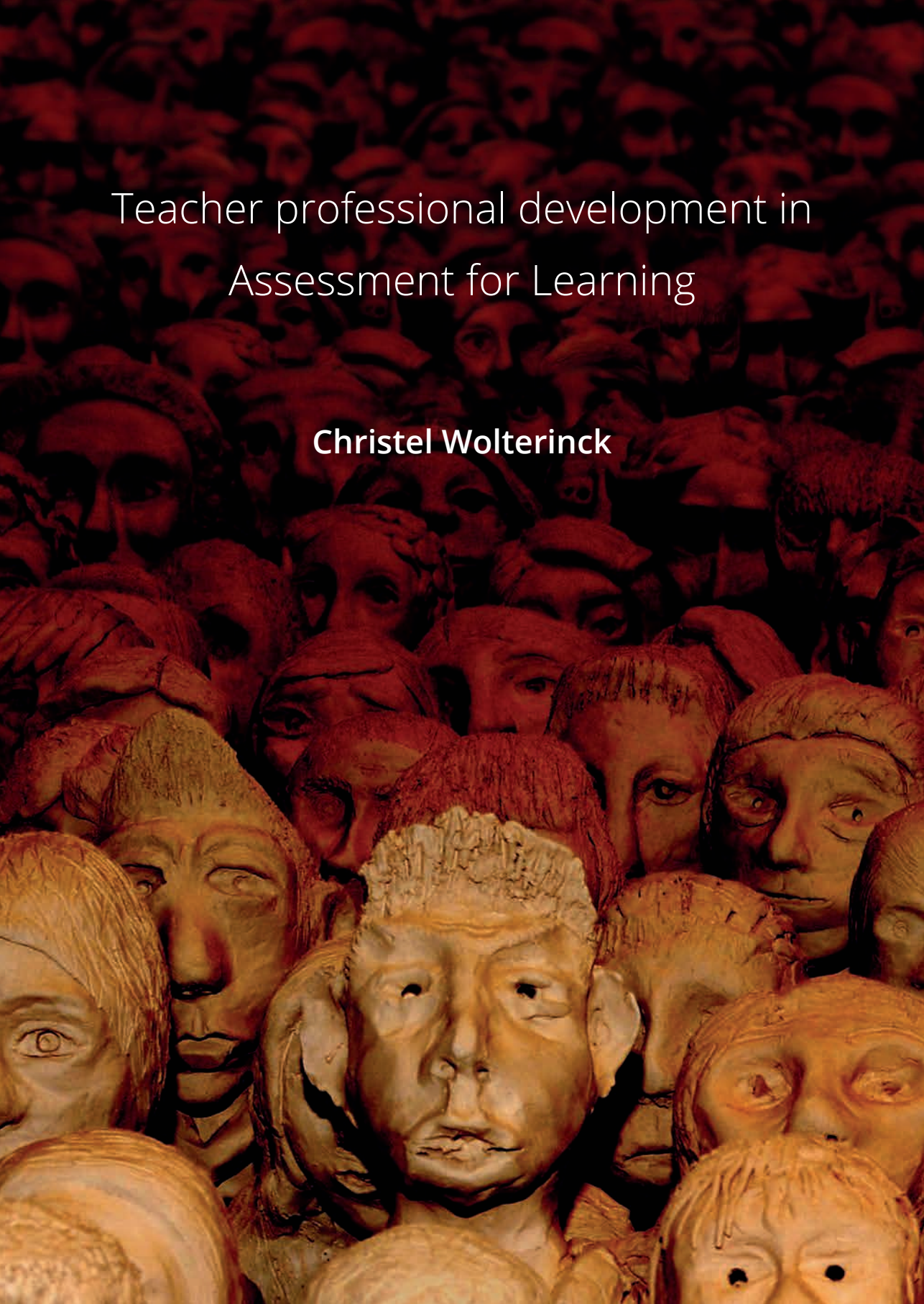
Teacher professional development in Assessment for Learning

Christel Wolterinck



Teacher professional development in Assessment for Learning

Christel Wolterinck

Graduation Committee:

Chair / secretary:	Prof. dr. T. Bondarouk, University of Twente
Supervisors:	Prof. dr. K. Schildkamp, University of Twente Prof. dr. A. J. Visscher, University of Twente
Co-supervisor:	Dr. C. L. Poortman, University of Twente
Committee Members:	Prof. dr. S. E. McKenney, University of Twente Prof. dr. ir. B. P. Veldkamp, University of Twente Prof. dr. D. Joosten - ten Brinke, Open Universiteit Prof. dr. M. Ehren, Vrije Universiteit Amsterdam Dr. L. Baartman, Utrecht University of Applied Sciences

This research was supported by Stichting Carmelcollege and conducted in the context of the research school Interuniversity Centre for Educational research.

Lay-out:	IdeeMedia, Groenlo
Cover photo:	Nicky Konings
Cover design and lay-out:	IdeeMedia, Groenlo
Printed by:	WIKA Druk, Borne
ISBN:	978-90-365-5466-4
DOI:	10.3990/1.9789036554664

Teacher professional development in Assessment for Learning

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. ir. A. Veldkamp,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 2 december 2022 om 12.45 uur

door
Christel Henriëtte Dianne Wolterinck – Broekhuis

geboren op 16 juni 1971
in Enschede

This dissertation has been approved by:

Supervisors
Prof. dr. K. Schildkamp
Prof. dr. A. J. Visscher

Co-supervisor
Dr. C. L. Poortman

CONTENTS

1

Introduction

1.1 PROBLEM STATEMENT

Even when instruction is planned with great care, students often do not learn what was intended (Wiliam, 2011). To improve the quality of students' learning process, it is therefore essential to collect information providing insight into their learning. Information from assessments can be used by both teachers and students to modify teaching and learning activities with the aim of improving students' learning and achievement. This is called formative assessment. An important form of formative assessment is Assessment for Learning (AfL; Klenowski, 2009). With AfL, teachers and students continually gather information about student learning, which is combined with classroom practice to provide teaching that meets learners' needs. Examples are observations in the classroom, the use of portfolios and rubrics, homework assignments and teaching conversations between teacher and students (Sluijsmans et al., 2013). Use of this information that is focused on generating feedback on students' performance in order to improve their learning (Sadler, 1989) is part of AfL. A crucial feature is that teachers and students are jointly responsible for the learning process.

Although AfL can lead to better student learning outcomes (Bennett, 2011), its application in Dutch education still appears to be limited (Kippers et al., 2018; Veugen et al., 2021). AfL is a complex skill and teachers find it difficult to implement in their classroom. As a result, use of AfL does not always lead to improved student achievement (Heitink et al., 2016; Kippers et al., 2018). Teachers need professional development (PD) to support them in developing and implementing AfL in their classrooms (Schildkamp et al., 2020). Research by Heitink and colleagues (2016) showed that various factors influence the implementation of AfL and thus may influence the content of effective PD programs. Teachers' attitudes and beliefs underlying their pedagogical didactic choices can have a significant impact on the implementation of AfL (Hargreaves, 2005). AfL knowledge and skills also play an important role. Teachers need the know-how and ability required to provide feedback that stimulates students' learning, for example.

Implementing AfL in the classroom can be regarded as complex, and therefore the development of teacher PD programs should focus on the learning of that complex skill. In this dissertation, studies are described of our PD intervention that uses a task-centred instructional design model, the four-component instructional design (4C/ID) model (Van Merriënboer & Kirschner, 2017). In this model, the design of authentic learning tasks follows a holistic approach dealing with real-life professional tasks, which allows the learner to practice all of the routine and non-routine aspects of a complex professional

task, in this case AfL, simultaneously (Van Merriënboer, 2007).

This dissertation focuses on teacher professional development for Assessment for Learning and aims to answer the main question: *How can teachers be supported in developing their Assessment for Learning competencies?*

1.2 CONCEPTUAL FRAMEWORK

1.2.1 Assessment for Learning

Assessment is an essential part of education, as assessment is necessary to draw careful conclusions about students' learning and their progress. Assessments not only can help to map students' progress and teaching effectiveness, but also can serve as an instrument for offering tailor-made education. In education, we distinguish between summative and formative use of assessments (Scriven, 1967). Summative assessment is used to assess students' learning achievement in order to decide on the continuation of their (school) career (selection, classification, placement, certification; Sanders, 2011). Formative assessment aims to determine how to optimize students' learning process (Bennett, 2011). Typically, it takes place in daily teaching practice and starts with eliciting information through assessment, which is used as a form of feedback in order to take informed action. In short: the summative use of assessments takes place at the end of a learning trajectory, to determine the results; the formative use of assessment takes place while the learning process is still underway, in order to improve it (Stobart, 2008).

One important approach to formative assessment is AfL. This approach focuses on practices in which teachers and students continually gather information about where students are in their learning process, for the purposes of adapting teaching to the learning needs of students and giving feedback to students about how to go forward with respect to their learning strategies (Klenowski, 2009; Van der Kleij et al., 2015; Wiliam, 2011). Assessment for Learning is a cyclical teaching process consisting of three elements, as shown in Figure 1.1. Ideally, the elements depicted in Figure 1.1 are addressed in interactions between three actors: the teacher, the student and peer students (Wiliam & Thompson, 2008).

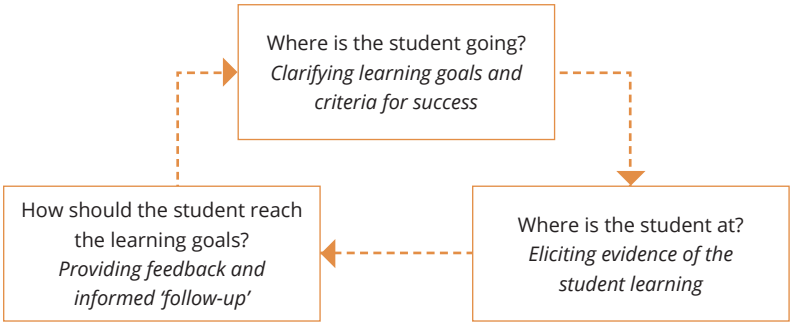


Figure 1.1 Assessment for Learning as a cyclical process

When applying AfL in the classroom, the teacher, and also the students themselves, continually try to answer three formative questions: "Where is the student going?", "Where is the student at?", and "How should the student reach the learning goal?". According to Wiliam (2011), there are several essential strategies that can answer these questions:

Clarifying, sharing, and understanding learning goals and criteria for success

Learning goals are what the teacher wants the student to learn. Criteria for success can be used to assess where students are in their learning process compared with the learning goals (Wiliam & Leahy, 2015). It is important for both teacher and student to know where the student is supposed to be going. Carless (2015) argued that learning goals and criteria for success can best be shared in the form of dialogues in which both the teacher and students play a role. For example, teachers together with their students can compare exemplars, concrete manifestations of quality, in order to formulate the criteria for success.

Eliciting evidence of student learning

For example, teachers can pose hinge questions, which are carefully designed multiple-choice questions that can indicate misconceptions operating in students' learning processes (Wiliam & Leahy, 2015). A wide variety of assessment techniques exist that can give both teacher and students insight into where the students are in their learning process.

Providing feedback and informed follow-up that moves learning forward

Giving feedback based on the information gathered can be an effective way of improving students' understanding and learning (Hattie & Timperley, 2007; Shute, 2008). This may also be the most complex strategy of all, as research (Kluger & DeNisi, 1996) has indicated that 38% of such feedback results in a decrease in student learning. When feedback is given either by the teachers or by the students to each other in a timely and elaborated

manner, then students can be stimulated to improve and self-monitor their own learning process (Nicol & MacFarlane-Dick, 2006). This was confirmed by Clark (2012), who argued that by giving and receiving high-level, formative feedback, students can become aware of their current level of learning progress and can more effectively direct their learning.

Activating students as learning resources for one another and as owners of their own learning
Self- and peer-assessment are mentioned as key strategies that can give students the opportunity to redirect their learning efforts (Nicol & MacFarlane-Dick, 2006). For example, students can assess their own or their peers' task performance by means of a rubric. In this way, students can become more accurate at self-reflection and their improvement efforts can be better directed (Carless & Boud, 2018; Nicol & MacFarlane-Dick, 2006). Sadler (2010) asserted the importance of students' involvement in assessment, stating that student involvement in peer review processes should be a core component of course designs. According to Carless (2015), it is essential to involve students in evaluating their own learning process. He mentioned that without this involvement, students often find the assessment process incomprehensible, which makes it difficult for students to effectively predict, plan, monitor and evaluate their learning. By involving students in evaluating their own learning process, the teacher can give students the opportunity to develop these metacognitive skills (Nicol & MacFarlane-Dick, 2006). Engaging students in the AfL process by means of self- and peer-assessment appears to be an excellent way to make students aware of their own learning progress and to stimulate self-regulated learning. Teachers and students both play a role in using AfL effectively in daily teaching practice. Interweaving AfL and instruction in the classroom and giving students an active role in this can ultimately lead to an improvement of both students' learning process and their learning outcomes.

1.1.2 Professional development

The effectiveness of teacher professional development (PD) has been studied a lot during the last decades. According to Darling-Hammond et al. (2017), effective PD can be defined as structured professional learning that aims to result in changes in teacher practices and the improvement of student learning outcomes. Traditional forms of PD, such as symposia, workshops, conferences, seminars or training sessions, do not seem to have much effect on teacher behaviour (Guskey, 2002). Therefore, other forms of PD are increasingly being sought, such as long-term, job-embedded, inquiry- or learner-centred structures that support teachers as they collaboratively develop the professional knowledge they need to use in their own context (Borko, 2004; Darling-Hammond et al., 2017; van Veen et al., 2012). Although decades of research on the design, implementation and effects of teacher PD has moved the field forward, this has not yet led to a solid knowledge base about

effective teacher PD (Sims & Fletcher-Wood, 2021). Identification of aspects of PD that reliably change teachers' behaviour remains elusive (Desimone & Garet, 2015; Desimone & Stuckey, 2014; Kennedy, 2016). Critical program design features such as program duration, focus on content knowledge, and collective participation may be unreliable predictors of program success (Kennedy, 2016) and do not provide enough guidance. In order to strengthen the design of teacher PD, this study utilizes what is known about learning to complete complex tasks.

A well-known approach to learning to complete complex tasks is learning through authentic tasks in which learners must integrate all elements of a particular skill set (Van Merriënboer, 2007). This approach to teacher PD, which is appropriate for the complexity of AfL, focuses on developing the required competencies for AfL as a coherent set rather than as various skills to be learned separately. From cognitive psychology, we know what is important for learning complex tasks and transfer of learning: whole, meaningful, authentic and varied learning tasks; ordering the learning tasks from simple to complex, in combination with a gradual decrease in learner support; and distinguishing between the routine and non-routine aspects of complex skills (Van Merriënboer et al., 2006). The four-component instructional design (4C/ID) model (Van Merriënboer et al., 2002) incorporates these characteristics, has a strong foundation in research, and its successful application has been described in various contexts (Frerejean et al., 2019). The 4C/ID model includes four interrelated components: a sequence of *learning tasks* based on authentic professional tasks, *supportive information* describing how to approach the tasks and how the domain is organized, *procedural information* describing step-by-step procedures to perform routine aspects of the tasks, and *part-task practice* for repetition of aspects that need to be highly automated (Van Merriënboer & Dolmans, 2015; Van Merriënboer et al., 2003). In this set of intervention studies, the 4C/ID model was used for developing a teacher PD program for AfL.

1.3 STRUCTURE OF THE DISSERTATION

This research aimed to develop a practical basis for supporting teacher professional development for Assessment for Learning. The overarching question that guided this dissertation was: *How can teachers be supported in developing their Assessment for Learning competencies?* To address this question, four studies were conducted, which are presented in chapters 2 to 5. In chapter 2, students' perceptions of the magnitude and nature of

Assessment for Learning in Dutch secondary education were studied. Much research into AfL has focused on teachers, but less research has focused on students. The research question for this study was:

To what extent do students experience AfL strategy use in their classroom?

By means of a survey conducted among 685 students in 12 different Dutch secondary schools, this study aimed to provide insight into the use of AfL strategies in current teaching practice from the perspective of the students. In chapter 3, the skills and knowledge that teachers in secondary education need to apply AfL in the classroom were studied. In addition, factors that reduce or increase the complexity of applying AfL were identified. The research questions for this study were:

What knowledge, skills and attitudes does a teacher need to implement AfL in their lessons?
What factors influence the complexity of applying AfL?

To address these questions, a cognitive task analysis (Clark et al., 2008) was conducted, taking expert teacher behaviour as a starting point and using classroom observations, interviews, and expert meetings for data collection. In chapter 4, the results of the cognitive task analysis were used to develop, implement, and evaluate a teacher professional development program based on the four-component instructional design (4C/ID) model (Van Merriënboer & Kirschner, 2017). The research questions for this study were:

How relevant and usable did teachers find the AfL professional development program that was designed on the basis of the four-component instructional design model?
What knowledge, skills and attitudes needed for AfL did teachers learn as a result of participating in the AfL professional development program?

Ten teachers from four different schools participated in the AfL teacher professional development (AfL-TPD) program and program implementation was studied by means of a questionnaire and a semi-structured interview approach. This study aimed to gain insight into the practicality of a task-centred approach to AfL-TPD as a way to support teachers in developing the competencies for AfL. Chapter 5 reports on the effects of an intensive, 8-month AfL teacher professional development program. This study aimed to measure the effect of the AfL-TPD program on students' self-regulated learning. The research question for this study was:

What is the effect of an Assessment for Learning teacher professional development program on students' self-regulation of learning?

An experimental group of 31 teachers participated in the AfL-TPD program. The effects on students' self-regulation were studied by means of a survey among 286 students in the control group and 301 students in the experimental group. Finally, chapter 6 presents a summary of the findings, the general conclusions and a discussion of the main findings. Moreover, the practical implications of the research presented in this dissertation are discussed. Figure 1.2 provides a schematic overview of the content of this dissertation. Each chapter was written in such a way that it could be read independently from the other chapters. Therefore, some chapters may overlap with respect to the theoretical framework and the description of the teacher professional development program.

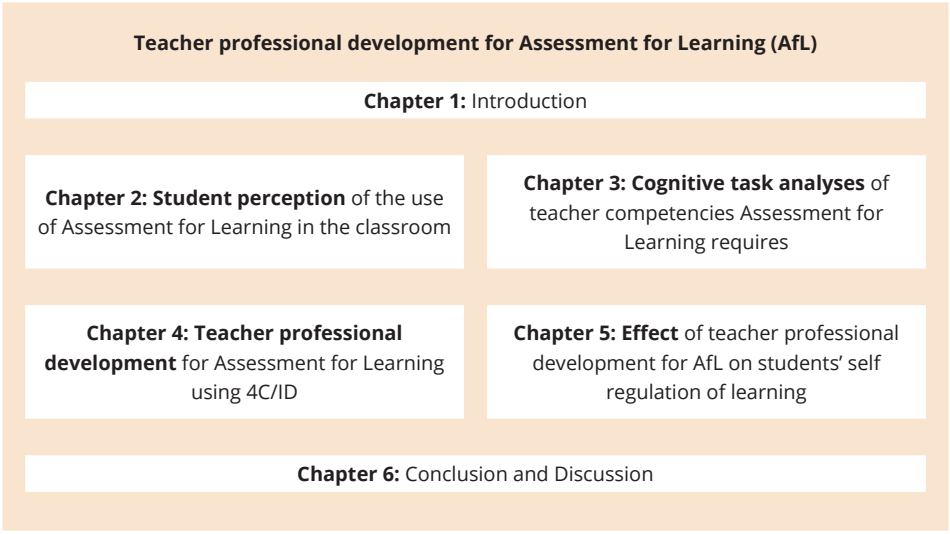
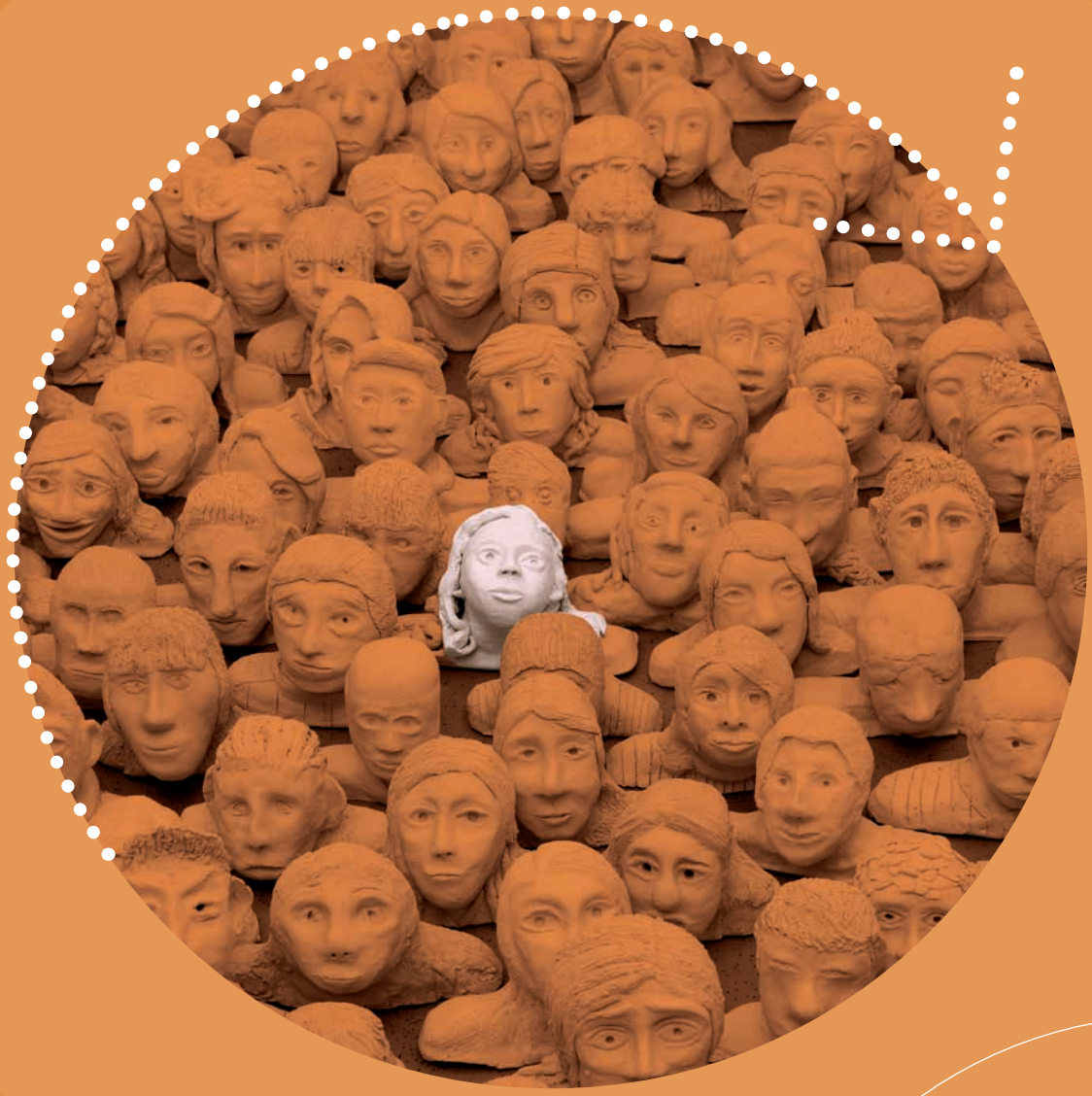


Figure 1.2 Dissertation overview



2

Student Perception of the Use of Assessment for Learning in the Classroom

This chapter is based on

Wolterinck, C. H. D., Poortman, C. L., Schildkamp, K., Visscher, A. J. (2021). *Key Stakeholder Voices: Investigating Student Perceptions of Teachers' Use of Assessment for Learning.* Manuscript submitted for publication

Abstract

Many schools aim to implement Assessment for Learning (AfL) to stimulate students to take more ownership of their learning and develop self-regulatory skills. This survey-based study is among the few in the field showing how students experience the extent of implementation of AfL, here in English language and mathematics classes in 12 Dutch secondary schools. Analysis showed no differences between the subjects. Students (*N* = 685) experienced activities to *clarify the learning intentions and success criteria* and *eliciting evidence on their learning progress* regularly (between 50-74% of the lessons). Students hardly ever experienced activities aimed at *peer- and self-assessment* (less than 25% of the lessons). However, cluster analysis revealed three distinct clusters related to the extent of AfL strategy use students experienced. Overall, we can conclude that AfL is not yet fully integrated into teaching practices. Teachers' skills, knowledge and attitudes required to increase student engagement in AfL practices and strengthen students' self-regulated learning need more attention in future teacher professional development trajectories.

Keywords: Assessment for Learning, formative assessment, student voice, co-regulated learning, self-regulated learning.

2.1 INTRODUCTION

Many teachers and schools implement formative assessment with the aim of stimulating students to take more ownership of their learning and develop self-regulatory skills (Panadero et al., 2018). Learning, teaching, and assessment are considered to be interdependent (Black & Wiliam, 1998b). Assessment has two important functions: formative, to provide support for future learning, and summative, to evaluate the achievements or potential of individuals (Bennett, 2011). Formative use of assessment aims to bring about improvements in teaching that will then improve student learning and the outcomes of such learning (Black & Wiliam, 2009). Assessment here refers to various types of assessments providing evidence of students' learning needs, such as diagnostic tests, homework assignments and student observations (Van der Kleij et al., 2015). Distinct approaches to formative assessment have evolved over time, one of which is Assessment for Learning (AfL; Marsh et al., 2006; Wiliam, 2011). AfL is a process in which teachers, together with students, are responsible for goal setting, data collection, sense making, and informed follow-up action in the classroom (Klenowski, 2009). Not only teachers, but also students are crucial stakeholders in the AfL process. Students can use the collected data to actively steer and improve their own learning, by themselves, with their peers, and with their teachers (Schildkamp, 2019). Modern assessment theory emphasizes a critical role for students in assessment, particularly in AfL (Black, 2015; Heritage, 2016; Klenowski, 2009). By participating in this process of AfL students become motivated to be the owners of their own learning and become both more self-regulated and more autonomous in their learning (Stobart, 2008).

AfL can be seen as a skill requiring complex competencies on the part of both teachers and students. For example, students need knowledge and skills to use assessment criteria in self- and peer-assessment, to be able to provide and receive (useful) feedback (Heitink et al., 2016). Teachers, among other things, must be able to interpret assessment information on the spot; they need knowledge and skills to integrate AfL with pedagogical content knowledge (PCK) and need to facilitate classroom discussions engaging students in the assessment process (Heitink et al., 2016). Even though the literature points to the importance of the use of AfL in classrooms for improved student achievement, and despite the fact that formative assessment has been on policy agendas internationally for decades, its implementation has proven to be challenging (Marshall & Drummond, 2006). Numerous studies have found that teachers often lack the skills to implement AfL effectively (e.g., Hubbard et al., 2014; Verhaeghe et al., 2010). Most often, qualitative research has been conducted to study the prerequisites for the use of AfL in the classroom; for example, in a 2016 review by Heitink et al., 48% of the studies were qualitative, 16% quantitative, and

36% mixed method. Moreover, research so far has focused predominantly on the role of the teacher. Detailed understanding is lacking of how students experience classroom practice as far as the use of AfL to move learning forward (Heitink et al., 2016). Therefore, this study focuses on studying students' perceptions of AfL practices in their classrooms.

Moreover, student perspectives can also be relevant for measuring teacher quality: students' perceptions are useful for knowing how the "clients" perceive teaching quality; for example, whether they understood the explanation of the learning goals and success criteria well (Dobbelaer, 2019). Measuring teaching quality is important for a broad array of educational stakeholders (Hill et al., 2012), because such measures can guide the improvement of teaching and support human resource decisions (Haertel, 2013). An easy-to-administer, time- and cost-effective, and non-intrusive method for gathering data on teaching quality in large samples is to evaluate students' perceptions of teaching quality through questionnaires (Kane & Staiger, 2012). Students are important stakeholders, as they are the only ones who observe teachers almost daily and can give feedback not only on 1 or 2 or 3 lessons (as in the case of expensive lesson observations), but on average performance in all lessons (Den Brok et al., 2006; Donahue, 1994). Students know best how teaching is experienced by the client. Gathering student perceptions results in many observations, not only in terms of numbers of lessons, but also in terms of numbers of observers (students), leading to regression to the mean and thus relatively reliable results (Dockterman, 2017). Therefore, we designed and used a questionnaire to study how students experience classroom practice when it comes to AfL. This study addresses the following research question: *To what extent do students experience Assessment for Learning strategy use in their classroom?*

2.2 THEORETICAL FRAMEWORK

2.2.1 Assessment for Learning (AfL)

Teachers and students have shared responsibilities for students' learning processes (Leahy et al., 2005), and therefore need to work together in the AfL process. AfL is an approach to formative assessment that takes place as part of ongoing classroom practice, and focuses on classroom interaction and dialogue between teacher and students and amongst students in a process of discovering, reflecting, understanding and reviewing (Hargreaves, 2005). Through applying AfL, teachers, together with students, find out what students definitely know, what they partially know and what they do not know, so that follow-up activities can advance learning and in turn enhance student achievement (Black & Wiliam, 1998b; Klenowski, 2009). Black and Wiliam (2010) identified five core strategies for AfL practice in the classroom. Each of these core strategies is described below.

Clarifying, sharing and understanding learning goals and criteria for success

This strategy focuses on the students, to have them really understand what their classroom experiences are likely to be and how their success will be measured (Black & Wiliam, 1998b; Crisp, 2012; Heitink et al., 2016; Wiliam & Leahy, 2015). The term *learning goals* indicates what the teacher wants the students to learn, whereas *criteria for success* indicates the criteria used by the teacher to check whether the learning activities in which students were engaged were successful or not. It is important that students are kept engaged and enthusiastic. If teachers share learning goals and success criteria with students and make sure they're clear and that the students understand them, every student knows where to be going and whether the learning goals have been achieved (Black & Wiliam, 2009; Wiliam & Leahy, 2015).

Eliciting evidence of student learning

This strategy involves creating opportunities to gather evidence of student learning through (informal) assessment (Black & Wiliam, 1998b; Crisp, 2012; Gottheiner & Siegel, 2012; Heitink et al., 2016; Wiliam & Leahy, 2015). Finding out what students do and do not know is essential to good teaching. It provides both teachers and students with information that they can use. Observations and classroom interactions, as well as more tangible products such as tests and homework, can be used for gathering evidence about student learning (Stobart, 2008). Teachers become better informed about students' needs, and as a result, instruction can be tailored to the needs of individual students, to maximize their achievement (Coburn & Turner, 2012; Lai & Schildkamp, 2013).

Providing feedback that moves learning forward

A key component of formative assessment is providing feedback, which is beneficial for moving forward with learning. The term feedback signifies the information provided regarding aspects of students' performance or understanding while they are learning, which must contain "where to next/improvement-focused" information in order to stimulate students to act on the feedback they receive (Hattie & Timperley, 2007; Wiliam & Leahy, 2015). Effective feedback supports further learning and stimulates students to think about their learning. For example, feedback can be given in terms of comments that address what a student needs to improve, what he or she needs to do and how (Wiliam & Leahy, 2015).

Activating learners as instructional resources for one another (peer-assessment) and owners of their own learning (self-assessment)

The term *peer-assessment* indicates the type of formative assessment in which learners

act as instructional resources for one another: assessing each other's work, not to judge, but to improve it (Wiliam & Leahy, 2015). To help students to become better learners, they should be given the opportunity to play active roles, talk about their learning, and engage in peer-feedback activities so that they have the opportunity to learn from one another (Black & Wiliam, 1998b; Bryant & Carless, 2010; Crisp, 2012; Harris & Brown, 2013; Heitink et al., 2016; Wiliam & Leahy, 2015). Providing peer-feedback can be beneficial, for example, because it requires students to actively consider the assessment criteria in multiple acts of evaluative judgement, both about the work of peers, and, through a reflective process, about their own work (Nicol et al., 2014). Moreover, providing and utilizing feedback from peers can be considered an important skill for students' future academic careers, and therefore an important learning goal within educational curricula (Huisman et al., 2019). Besides learning from others, it is very important that students are engaged in their own learning. *Self-assessment* focuses on the ability of students to reflect on their learning by assessing their own work (Crisp, 2012; Fletcher & Shaw, 2012; Harris & Brown, 2013; Heitink et al., 2016; Wiliam & Leahy, 2015) Self-assessment entails the students taking ownership of their own learning (Wiliam & Leahy, 2015). According to Harris and Brown (2013), self-assessment is beneficial for students' learning outcomes.

Studies focusing on the implementation of the five AfL strategies in teaching practice emphasise the cyclical character of the AfL process, in which the collected data are analysed (turning data into information) in relation to the learning goals and transformed into decisions, so that teachers can provide feedback for students (Antoniou & James, 2014; Gulikers & Baartman, 2017; Ruiz-Primo & Furtak, 2007; Schildkamp et al., 2020). Schildkamp et al. (2020) stated that the ongoing interaction between learners and the teacher in the form of continual dialogues and short feedback loops is the key element of AfL. Assessment is thus an integrated element of the learning process, and AfL needs to be an integrated element of instruction. One problem in implementation of AfL is that often only certain "principles" of AfL have been adopted, without much consideration of the broader implications for classroom practice (Elwood, 2006; Torrance, 2012). For AfL to lead to improved student learning, it is crucial that teachers, together with their students, actually use all the different AfL strategies coherently (Wiliam & Leahy, 2015).

2.2.2 The role of students in Assessment for Learning practices in the classroom

AfL has the potential to excite co-regulatory activities that foster students' development of self-regulatory skills (Allal, 2020; Andrade & Brookhart, 2020). Positive effects of AfL on students' self-regulated learning can be expected because AfL emphasizes: (1) sharing learning goals and criteria for success in order to help students develop plans to attain goals; (2) assessment in order to monitor where learners are with regard to the set

standards, including peer- and self-assessment; and (3) feedback based on the assessment results, which can be used to adapt learning strategies to move closer to the desired goals. Students improve their self-regulation when they are joint stakeholders in assessment with teachers and peers (Bailey & Heritage, 2018).

Teacher-student interaction during lessons is crucial in the assessment process, and students need to be actively involved in establishing clear learning goals and success criteria, eliciting and interpreting evidence of learning and taking immediate or near-immediate (pedagogical) action based on evidence (Heritage, 2016). This requires purposeful interaction between students and teachers (Carless & Winstone, 2020). Gulikers et al. (2021) found in their study that explaining student-teacher interactions using a student-teacher formative assessment cycle format and explaining student behaviour helps teachers to make their own behaviour in the formative assessment process clearer. These activities make teachers realize the importance of purposefully designing co-regulatory assessment activities that engage and guide students in this regulatory process (Panadero et al., 2019).

2.3 METHOD

2.3.1 Context and participants

To study the use of AfL in classroom practice, we administered an online survey to students (*N* = 685) in secondary education in the Netherlands. The study focused on AfL strategy use in lessons for two core subjects, English and mathematics, essential areas of learning acknowledged as foundational for learning in other areas. Two subjects were chosen to determine if there are any discipline-specific differences in how AfL practices are perceived by students. This study was part of a larger project (Kippers et al., 2018). As described previously (Kippers et al., 2018), we used a convenience sample for this larger study, in which a total of 27 secondary schools participated, 26 of which belonged to one of the largest Dutch school boards in secondary education, which is involved with our university in a research-practice partnership. Although this convenience sample consisted of a mix of participating denominations, geographical locations, and educational tracks (see Table 2.1), the sample was not representative for schools in the Netherlands and caution is advised for generalizing the conclusions. Within this sample of 27 schools, 19 schools offered a senior general secondary education track. The school leaders were asked to inform teachers about this study and to ask them to cooperate by asking the students in their classes to complete the internet survey, and teachers did this in 12 of these schools.

The AfL student questionnaire was offered to students from the fourth (15-16 years old) and fifth (16-17 years old) grade levels of senior general secondary education (the

final two years of high school). We focused on this specific group of students because they were heading towards a final exam to complete their secondary education. We wanted to know to what extent these prospective examinees perceived AfL practices in their lessons that are aimed at gaining insight into one's own learning process and being able to make adjustments on the way to the final examination. The online survey was completed by 685 students (response rate of 21%), from 12 different secondary schools. Parents gave consent for students to participate in the survey. Students were randomly assigned and asked to complete the questionnaire for a single subject, either for English language or mathematics. About half of the students in the sample (51.4%) completed the questionnaire for their lessons in the English language and the remaining 48.6% completed the questionnaire for their mathematics lessons. More than half of the students (58.8%) were female and 40.1% were male. Reports of the overall findings were provided to schools, as a way to inform professional reflection and further discussion.

Table 2.1 Secondary School Characteristics[a]

		Schools in project	**Schools in the Netherlands**
		N (%)	*N* (%)
School size	Small (< 500 students)	6 (22.2)	158 (24.1)
	Medium (500-1000 students)	10 (37.0)	95 (14.5)
	Large (> 1000 students)	11 (40.8)	402 (61.4)
Denomination	Catholic schools	17 (63.0)	150 (22.9)
	Interdenominational schools [b]	4 (14.8)	66 (10.1)
	Generally special schools [c]	5 (18.5)	99 (15.1)
	Public schools [d]	1 (3.7)	186 (28.4)
	Other	0 (0)	154 (23.5)

[a] Ministry of Education, Culture & Science (2016); http://www.scholenopdekaart.nl; http://www.statline.cbs.nl).

[b] A Dutch interdenominational school is characterized as a government-independent school that is based on a combination of different religions.

[c] A Dutch generally special school is characterized as a government-independent school that is based on a specific educational vision and not on a specific religion.

[d] A Dutch public school is characterized as a government-dependent school that is based on neither a specific educational vision nor a specific religion.

2.3.2 Instruments

Questionnaire for students

To study the extent to which various AfL strategies are used in classroom practice in the eyes of students, a student questionnaire was used. The student questionnaire was developed based on an existing reliable teacher self-report instrument related to AfL used in the Dutch context (Kippers et al., 2018). This teacher questionnaire itself was based on an existing reliable instrument to audit teachers' use of AfL: the Assessment for Learning Audit instrument (AfLAi) (Lysaght & O'Leary, 2013; O'Leary et al., 2013). The instrument includes four separate, independent scales based on the key AfL strategies, as outlined in the international literature on formative assessment (Black & Wiliam, 2010): 1) sharing learning goals and success criteria (LISC), 2) eliciting evidence (EE), 3) feedback (FB), 4) peer- and self-assessment (PSA). The 32 items from the teacher questionnaire about the use of AfL in the classroom were reformulated for students. For example, an item from the teacher questionnaire was: "Questions are used to elicit students' prior knowledge on a topic". The parallel item in the student version was: "My teacher asks questions to elicit my prior knowledge on a topic". Students responding to the statements in the questionnaire were asked to report the extent to which the statements reflected current classroom practice using the following rating scale: 5 = *embedded* (it happens in 90% of the lessons), 4 = *established* (it happens in 75% of the lessons), 3 = *emerging* (it happens in 50% of the lessons), 2 = *sporadic* (it happens in 25% of the lessons), 1 = *(almost) never* (it happens in less than 10% of the lessons). For the quantitative analyses, each of the scale points was given a numeric value from 5 to 1 (see Appendix A). "Don't understand" was also omitted. The questionnaire items were in Dutch.

The clarity of the student questionnaire was investigated based on a review by eight students, two teachers and two expert researchers (Dobbelaer, 2019). Each item was checked to determine whether it adequately reflected the construct within the Dutch educational context. In addition, the questionnaire was piloted in two Dutch secondary schools, where it was completed by 63 students. Based on the review and pilot, minor adjustments were made, mostly in terms of formulating the items more clearly and more specifically. The responses to the 32 items about the use of AfL were subjected to confirmatory factor analysis (CFA) and reliability analysis using SPSS version 25 (Field, 2013). In order to obtain conceptually similar and significant categories of items, principal axis factor analysis with Varimax rotation and Kaiser normalization were conducted, as the factors were deemed to be orthogonal. Eigenvalues greater than or equal to 1.00 were extracted, and KMO (0.94) and Bartlett test of sphericity (p = 0.000) values indicated that the data were suitable for factorization. Orthogonal rotation of the variables yielded four factors from the 21 items included, accounting for 12.92%, 11.91%, 10.91%, and 10.61% of

the total variance, respectively, for a total of 46.35% of the total variance explained. The 4-factor structure for the items about the use of AfL was consistent with the theoretical framework: sharing learning goals and success criteria, asking questions and classroom discussions, feedback, and peer- and self-assessment (see Appendix B). Three of the 21 items about the use of AfL were deleted because the results of the factor analysis showed that they loaded insufficiently. Reliability analysis of the scales was acceptable for all four scales (.70 -.80; (Field, 2013); see Appendix B.

2.3.3 Analysis

To answer our research question, we conducted two types of analysis. First, to gain more insight into the data, we carried out descriptive analyses. We analysed the mean and standard error for each scale of the questionnaire. Non-responses to some items of a scale ranged between 37 and 125 students. Moreover, we conducted independent samples t-tests to compare the results for gender (male versus female) and subjects (English language versus mathematics) on the four scales of the student questionnaire (see results section; Field, 2013). Second, we conducted a cluster analysis (Everitt, 1980; Field, 2000; Romesburg, 1990) to further study the differences between students' perceptions of AfL strategy use in their classrooms. We performed a hierarchical cluster analysis in SPSS using Ward's method (Borgen & Barnett, 1987; Romesburg, 1990), applied to a proximity matrix of squared Euclidean distances. As this method is vulnerable to producing clusters influenced by level differences (Borgen & Barnett, 1987), we standardized the data beforehand to enhance validity. We then compared the mean scores for all items for each cluster with the other clusters, to give meaning to the clustering. Missing values were replaced by the mean value for the cluster. Cluster analysis testing revealed that three clusters best fit our data. Furthermore, we wanted to make sure that the clusters were not formed based on variables other than students' experience of the extent of AfL strategy use (Everitt, 1980).

2.4 RESULTS

2.4.1 Students' perceptions of AfL strategy use in teaching practice

The frequencies for the four AfL scales of the student questionnaire are shown in Table 2.2 The highest mean score was for teachers' sharing of learning goals and success criteria, at 3.07 (N = 604; SD = 0.86). Nearly three-quarters of the students (71.0%) indicated that sharing learning goals and success criteria was either emerging (happening in approximately 50% of the lessons) or established (happening in approximately 75% of the lessons) in their classroom. For example, the mean score for "My teacher matches success criteria with

learning goals" was 3.03, and the mean for "My teacher uses child-friendly language to share learning goals with students (e.g., 'We are learning to make a good guess (prediction) about what is likely to happen next in the story')" was 3.13, based on student reports.

Students reported experiencing teachers' use of the strategy of eliciting evidence less frequently than sharing learning goals and success criteria, with a mean score of 2.58 (N = 586; SD = 0.79). Three-quarters of the students (75%) indicated that teachers' use of this strategy in their classroom was either sporadic (happening in approximately 25% of the lessons) or emerging (happening approximately in 50% of the lessons). To give an example, "My teacher asks questions to elicit students' prior knowledge on a topic" was scored 2.54 on average, and "My teacher encourages students to share the questioning role with teacher during lessons (e.g., the teacher routinely invites pupils to question their peers' contributions to discussions)" was scored 2.15 on average.

Based on the questionnaire responses, the mean score for teachers' use of the strategy of feedback was 2.33 (N = 544; SD = 0.95). The majority of students (76%) rated the use of feedback as either sporadic (36%; happening in approximately 25% of the lessons) or emerging (40%; happening in approximately 50% of the lessons) in their classroom. For example, the statement "My teacher uses written feedback on pupils' work that goes beyond the use of grades and comments such as "well done" to specify what students have achieved and what they need to do next" received a mean score of 2.39, and the statement "My teacher uses tests diagnostically to tailor the instruction to the needs of the students by taking into account the strengths and needs of students (e.g., extra lessons on adding fractions)" was scored 2.46 on average.

Peer- and self-assessment was the least often used strategy according to students, with a mean score of 1.94 (N = 526; SD = 0.83). Most of the students (80%) rated the use of peer- and self-assessment as happening either (almost) never (32%; happening in approximately less than 10% of the lessons) or sporadically (48%; happening in approximately 25% of the lessons) in the classroom. To give an example, "My teacher stimulates students to assess and comment on each other's work (e.g., they are taught how to use the success criteria for a lesson to judge another pupil's piece of work)" was scored 1.72 on average, and "My teacher encourages students to use a range of assessment techniques to review their own work (e.g., rubric, traffic lights, thumbs up/down, two stairs and a wish)" was scored 2.18 on average.

Table 2.2 AfL Strategy Use Scores

Scale	*N*	Mean (*SD*)	Count (%)				
			(almost) never	sporadic	emerging	established	embedded
LISC	604	3.07 (0.86)	14 (2.3%)	140 (23.2%)	262 (43.4%)	167 (27.6%)	21 (3.5%)
EE	586	2.58 (0.79)	96 (16.3%)	267 (45.6%)	170 (29.0%)	47 (8.0%)	6 (1.0%)
FB	544	2.33 (0.95)	75 (13.8%)	193 (35.5%)	174 (40.0%)	84 (15.4%)	18 (3.3%)
PSA	526	1.94 (0.83)	168 (31.9%)	252 (47.9%)	78 (14.8%)	27 (5.15)	1 (0.2%)

Note: the items in the scales used a 5-point Likert response format: 1 = (almost) never (it happens less than 10% of the time), 2 = sporadic (it happens 25% of the time), 3 = emerging (it happens 50% of the time),
4 = established (it happens 75% of the time), 5 = embedded (it happens 90% of the time).

LISC = sharing learning intentions and success criteria, EE = eliciting evidence, FB = feedback, PSA = peer- and self-assessment.

2.4.2 Independent samples t-tests

We conducted independent samples t-tests to compare the results for gender (male versus female) and subjects (English language versus mathematics) on the four scales of the student questionnaire. The results of the independent sample t-tests (shown in Table 2.3) indicated that the difference between the scores given by male and female students (M = 3.07, SD = .81 and M = 2.89, SD =.81, respectively) for sharing learning goals and success criteria was statistically significant: $t(519) = 2.75$, $p < .01$. The effect size, $d = .22$, represents a small effect (Cohen, 1992; Field, 2013). In addition, the difference between the scores given by male and female students (M = 2.65, SD = .92 and M = 2.35, SD = .94, respectively) for feedback was statistically significant: $t(474) = 3.68$, $p < .01$. The effect size, $d = .32$, represents a moderate effect. The mean differences in reported AfL strategy use in English language lessons compared to mathematics lessons were not statistically significant for any of the scales ($p > .05$).

Table 2.3 Results of Independent Samples t-tests for Gender and Subjects

Scale	Group						
	Male	Female			Mean		
	M (SD) - N	*M (SD) - N*	*t (df)*	*p*	difference	95% CI	*d*
LISC	3.07 (.81) - 243	2.89 (.81) - 358	2.75 (519)	<.01	0.18	.053, .316	0.22
EE	2.29 (.84) - 236	2.16 (.79) - 347	1.99 (487)	.05	0.14	.002, .273	0.16
FB	2.65 (.92) - 218	2.35 (.94) - 325	3.68 (474)	<.01	0.30	.139, .458	0.32
PSA	1.93(.75) - 220	1.82(.75) - 305	1.75 (470)	.08	0.18	-.014, .247	0.15
Scale	Group						
	English	Mathematics			Mean		
	M (SD) - N	*M (SD) - N*	*t (df)*	*p*	difference	95% CI	*d*
LISC	2.91 (.82) - 312	3.02 (.80) - 291	-1.67 (599)	.10	-0.11	-.240, .019	0.14
EE	2.19 (.82) - 303	2.24 (.81) - 281	-0.64 (580)	.53	-0.04	-.175, .090	0.06
FB	2.44 (.93) - 265	2.50 (.95) - 265	-0.75 (538)	.46	-0.06	-.219, .099	0.06
PSA	1.86 (.76) - 267	1.88 (.75) - 258	-0.33 (522)	.74	-0.02	-.151, .108	0.03

Note: the items in the scales used a 5-point Likert response format: 1 = (almost) never (it happens less than 10% of the time), 2 = sporadic (it happens 25% of the time), 3 = emerging (it happens 50% of the time), 4 = established (it happens 75% of the time), 5 = embedded (it happens 90% of the time).

LISC = sharing learning intentions and success criteria, EE = eliciting evidence, FB = feedback, PSA = peer- and self-assessment.

2.4.3 Cluster analysis

The cluster analysis revealed three different clusters based on the extent of AfL strategy use students reported experiencing (Table 2.4):

Cluster 1 (high); students (*N* = 82, 19%) who belong to the first cluster reported that they experienced all four AfL strategies used in 50-74% of their lessons.

Cluster 2 (middle); students belonging to the second cluster (*N* = 183, 41%) reported that they experienced the AfL strategy of sharing learning goals and success criteria most frequently, in 50-74 % of their lessons. The two strategies of eliciting evidence and feedback were used in 25-49% of their lessons. Peer- and self-assessment was experienced less frequently, in less than 25% of their lessons.

Cluster 3 (low); students who belong to cluster three (*N* = 177, 40%) experienced the use of sharing learning goals and success criteria in 25-49% of their lessons. The other three strategies, eliciting evidence, feedback, and peer- and self-assessment, were experienced in less than 25 % of their lessons.

Table 2.4 Clusters Based on Perceived Extent of AfL Strategy Use

AfL strategies	Cluster1: High (*N* = 82)		Cluster 2: Middle (*N* = 183)		Cluster 3: Low (*N* = 177)	
	M	*SD*	*M*	*SD*	*M*	*SD*
LISC	3.59	0.62	3.32	0.61	2.23	0.57
EE	3.35	0.60	2.72	0.47	1.85	0.45
FB	3.41	0.59	2.55	0.66	1.65	0.50
PSA	3.11	0.61	1.71	0.43	1.44	0.40

Note: LISC = sharing learning intentions and success criteria; EE = eliciting evidence; FB = feedback; PSA = peer- and self-assessment.

2.5 DISCUSSION AND CONCLUSIONS

The importance of AfL has been advocated by researchers for many years (e.g., Wiliam, 2011). To support teachers in practicing AfL together with their students, information about the extent to which teachers and students use assessment information on a minute-by-minute, day-by-day basis is helpful (Leahy et al., 2005). In the present study, a questionnaire was developed to investigate students' perceptions of the degree to which AfL takes place in English language lessons and mathematics lessons in Dutch secondary schools.

2.5.1 Students' perceptions of AfL strategy use in teaching practice.

In line with previous studies (Kippers et al., 2018; Veugen et al., 2021), the results of this study suggest that the use of AfL is not yet fully integrated within daily classroom activities, and that there is considerable room for improvement. Students were most positive about the use of sharing learning goals and success criteria and eliciting evidence on students' learning progress. They indicated that these two strategies were often present in their lessons, meaning that both strategies were utilized by their teachers in 50-75% of the lessons. These findings correspond with those from a study of teachers' perceptions of AfL strategy use in Dutch secondary education, where teachers also indicated using these strategies in 50-75% of their lessons (Kippers et al., 2018).

Students indicated that feedback given by their teachers was utilized approximately in 25-50% of the lessons. These findings are also in line with teachers' perceptions of the use of feedback as an AfL strategy in their lessons (Kippers et al., 2018). The results of the independent sample t-tests indicated that the difference between the scores given by male and female students for both strategies, sharing learning goals and success criteria, and feedback, was statistically significant. Male students experience the use of this strategy more frequently than their female colleagues. Feedback has been found to be crucial

in AfL, coming both from teachers and from students to each other. Recent formative assessment research has recognized the need for students to be active participants in feedback processes (Van der Kleij et al., 2019; Winstone et al., 2017). Students need to recognize feedback in order to receive it, and must perceive the feedback as intended by the provider in order for it to be effective (Van der Kleij & Adie, 2020). Carless and Winstone (2020) pointed to the importance of partnership and shared responsibility that underpin the interplay between teacher and student feedback literacy, because feedback processes require investments from both parties. Students may not be the only ones requiring training in providing and utilizing effective feedback; it has also proved to be a complex skill for teachers (Fletcher-Wood, 2018).

Students indicated that conducting peer- and self-assessment was used least frequently, in only 0-25% of the lessons. These findings correspond to those in the study of teachers' perception of AfL strategy use, where teachers also indicated using these strategies in less than 25% of their lessons (Kippers et al., 2018). An explanation for the limited use of peer- and self-assessment may be that teachers tend to stick to teacher-centered teaching; a cultural change in Dutch secondary education may be required to strengthen the use of peer- and self-assessment. Teachers, together with their students, may need to become more proficient in providing and utilizing feedback and in the use of assessment criteria for peer- and self-assessment (Heitink et al., 2016; Veugen et al., 2021).

The results of this study indicate that students do not yet experience AfL as a fully integrated element of their teachers' instructional practices. This may be because students do not recognize AfL in their lessons or because teachers do not actually use this approach, and therefore students do not experience it (Kippers et al., 2018). Making the use of the AfL strategies more explicit for students and therefore more visible and noticeable for them calls for partnership, shared responsibility, and interaction between teachers and students and between students (Carless & Winstone, 2020; Panadero et al., 2018).

2.5.2 Implications for practice

Evaluating existing practices in Dutch secondary education as well as in other countries can identify whether support is needed with regard to Assessment for Learning, and can suggest how to tailor future teacher professional development (TPD) trajectories for teachers to become more proficient users of AfL in their classrooms. The cluster analysis performed in this study revealed three different clusters of students based on the extent of AfL strategy use they experienced: a cluster *high* in which students reported the use of all four AfL strategies as between emerging and established (50-74% of the lessons); a cluster *middle* in which the strategy use experienced was more spread, from (almost) never to established (0-74% of the lessons); and a cluster *low* in which students reported

that they experienced the use of three of the four AfL strategies as between (almost) never and sporadic (0-24% of the lessons), with only the use of sharing learning goals and success criteria experienced as higher, between sporadic and emerging (25-49% of the lessons). Based on the cluster analysis, it can be concluded that the use of the first strategy is largely in order across all three clusters. The clusters are distinguished by differences in the use of the other three strategies. This knowledge can be used when setting up a TPD program. The questionnaire we have developed can be administered to students at the start of a TPD program, for example, and the TPD can then be adapted to the situation of the teachers at that time (assuming that students of the same teacher, in the same classroom, will fall into the same cluster). By working with clusters, differentiation within the TPD program is possible in an efficient way. Furthermore, the questionnaire can be used in schools as a reflection tool to start discussions about AfL in the school and about using student experiences for school improvement (Pekrul & Levin, 2007; Rollett et al., 2021b).

2.5.3 Limitations and implications for further research

It is important to emphasize that in this study, student perception data were collected by means of a questionnaire. Students indicated how often different AfL strategies occurred in their lessons, but we cannot make any statements about the quality of implementation of those strategies. Although the quality of teaching can be assessed efficiently multiple times during a single lesson, i.e. equal to the number of students in a class by means of student perceptions (Kane & Staiger, 2012), some critical concerns need to be taken into account. These concerns are, among others, that students in secondary education might not have fully developed their abstract thinking skills (Roth et al., 2016) and their perceptions might be influenced by both student variables (e.g., ethnicity, student performance; Levy et al., 2003) and teacher variables unrelated to teaching efficacy (e.g., teacher popularity; Fauth et al., 2014). Nevertheless recent studies have illustrated how teachers and teaching can benefit from making use of formative student feedback (Rollett et al., 2021a) and how students' ratings can serve as reliable measurements of teaching quality (Bijlsma et al., 2022). This study shows that students indicated that the use of formative assessment is not fully integrated in their classrooms yet, as was also acknowledged by teachers in a study using the same survey as a self-perception tool (Kippers et al., 2018). Our student self-report data correspond with findings from other studies. For example, Gulikers & Baartman (2017) concluded from a review study including 106 studies from all over the world that the quality of current formative assessment practices is low.

Professional development can explicitly develop teachers' knowledge and skills to integrate different strategies coherently in their classroom practice (Lee, 2011) and can improve how teachers can guide their students to become proficient strategy users as well.

Future research could focus on designing, developing, and implementing professional development interventions in both pre-service and in-service education, to improve teachers' skills and knowledge for student-involved formative assessment practices. The design of professional development trajectories to support teachers is important, as this study demonstrates that students perceive the use of AfL as not fully integrated within their daily classroom activities. Although students play an essential role in AfL, our study shows that involving students in the AfL process can be improved, as peer- and self-assessment were least evident in classroom practice, compared to the other AfL strategies studied. It would therefore be interesting to focus in future research on the role of teachers in putting students more in control of their own learning process, and creating a rich learning environment in which students can practice peer- and self-assessment to the fullest. Research on the skills and knowledge that students need to strengthen their assessment and feedback literacy may also provide an interesting focus, in order to increase student engagement in AfL practices and strengthen their self-regulated learning.

3

Cognitive Task Analyses of Teacher Competencies Assessment for Learning Requires

This chapter is based on

Wolterinck, C. H. D., Van Geel, M., Schildkamp, K., Visscher, A. J. (2021). *A Cognitive Task Analysis of the Skills that Assessment for Learning Requires from Teachers.* Manuscript submitted for publication

Abstract

Assessment for Learning (AfL) is a promising educational innovation. However, a precise description of what teachers should know and be able to do in order to implement AfL is lacking, and therefore a good basis for designing teacher professional development is also lacking. A cognitive task analysis was conducted taking expert teacher behaviour with respect to AfL in secondary education as a starting point. This resulted in an overview of the skills teachers need for AfL in a skills hierarchy, complemented by performance objectives, required teacher knowledge, and the factors that reduce or increase the complexity of AfL. The results also showed that – in line with many other aspects of effective teaching – four closely related phases are important for the task of AfL as a whole: preparing a lesson unit, preparing a lesson, executing the lesson and evaluating the lesson. The insights gained are a valuable contribution to the AfL knowledge base and offer a basis for developing teacher professional development trajectories aimed at equipping teachers with the skills for AfL in the classroom.

Keywords: assessment for learning, formative assessment, cognitive task analysis, secondary education, professional development.

3.1 INTRODUCTION

Assessment for Learning (AfL) takes place as part of ongoing classroom practice, and focuses on the quality of the learning process (Klenowski, 2009). The core idea of AfL is that teachers and students gain insight into where the learner is going, where the learner is now, and how the learner can be moved forward in their learning (Black & Wiliam, 2009). Although AfL can enhance student learning (e.g., Bennett, 2011; Black & Wiliam, 1998; Kingston & Nash, 2011, 2015) and there is wide acceptance of AfL by policymakers in many countries, empirical evidence has indicated that it is difficult to implement consistently and effectively (Torrance, 2012). Research has shown that implementation of AfL is often ineffective because of the lack of a consensus on what the "AfL" approach encompasses precisely, and the resulting wide variety of AfL implementations (Bennett, 2011). Another explanation of the lack of positive effects could be that teachers struggle with implementing AfL in their classrooms. Teachers are increasingly expected to have all of the knowledge, skills and attitudes needed to enact AfL in their classrooms (Heitink et al., 2016), but they may need support in developing these competences (Schildkamp et al., 2020).

In summary, it can be said that AfL, as a complex teacher competence, requires teacher professional development (TPD) to support teachers in integrating AfL strategies coherently within their classroom practice, (Lee, 2011). More research is needed into the characteristics of effective professional development in the use of formative assessment, and into the development, implementation, and evaluation of these professional development interventions (Schildkamp et al., 2020). To enable the development of TPD interventions for AfL, untangling the complexity of AfL is required in order to give a clear picture of the required teacher competences for AfL practices. In this study, a cognitive task analysis (CTA) will be used to examine expert teacher behaviour when implementing AfL in their lessons. By performing a CTA, consisting of detailed observations and in-depth interviews with experts to uncover the skills they perform, the knowledge they possess, how they apply this knowledge, how they approach problems and how they reason, the expert behaviour that underlies observable task performance can be mapped (Clark et al., 2008). The first aim of this study is to provide a coherent description of everything a teacher needs to know and be able to do in order to apply AfL in their lessons. This includes an overview of all constituent skills and the coordination between these skills, and insight into what knowledge and attitudes a teacher needs for AfL. A second aim of this study is to generate an overview of the factors that influence the complexity of applying AfL in the classroom. These complexity-related factors can be used to classify AfL learning tasks into more simple or more complex categories, or "task classes" (Van Merriënboer & Kirschner, 2017). This information can be used to design professional development activities aimed

at developing and/or improving teachers' AfL skills that start with simple learning tasks and continue with AfL learning tasks that gradually become more complex. Consequently, this study addresses the following research questions:
What knowledge, skills and attitudes does a teacher need to implement AfL in their lessons?
What factors influence the complexity of AfL?

3.2 THEORETICAL FRAMEWORK

3.2.1 Assessment for Learning

AfL has become prevalent in educational systems around the world (Hopfenbeck & Stobart, 2015). It is increasingly agreed-upon that assessment practices in the classroom can be a powerful catalyst for learning (Torrance, 2012). According to Klenowski (2009, p. 264), "AfL is part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance on-going learning". The formative use of classroom assessments can enhance teachers' effectiveness and improve students' in-class learning and subsequent performance on (accountability) tests, by fundamentally transforming the way teachers teach (Popham, 2008). The core of AfL is the communication between teacher and student, not only about how learning related to a specific learning task in order to achieve a specific learning goal is progressing, but also to improve the students' experience of learning and to promote student understanding and autonomy in learning (Torrance, 2012).

In this study, assessment is defined as the use of processes (e.g., asking questions and classroom conversations) and instruments (e.g., a test or homework assignment) for gathering evidence about student learning (Stobart, 2008; Van der Kleij et al., 2015). Popham (2011) emphasised that the formative use of assessments needs to be a planned process in which assessment-elicited evidence of students' progress towards the learning goals is used by teachers, to adjust their ongoing instructional procedures, and/or by students to adjust their current learning tactics. In order to shape this process, different studies in which the AfL cycle has been described were used to distinguish key strategies (KS) for AfL practice in the classroom (Gulikers et al., 2021; Ruiz-Primo & Li, 2013; Wiliam, 2011): (KS1) clarifying expectations (i.e., learning goals and success criteria); (KS2) eliciting evidence of student learning (including self- and peer-assessment); (KS3) analysing student responses; (KS4) communicating about results (including feedback); (KS5) taking concrete actions to adjust teaching and/or learning. AfL can be considered a highly complex competence that requires the mobilization and integration of a series of teacher skills related to these key strategies (Lee, 2011). To get a more detailed picture of the teacher skills and knowledge

needed for these AfL key strategies, a more in-depth and detailed approach to AfL as a complex skill in practice may be needed, by means of a cognitive task analysis.

3.2.2 Cognitive task analysis

In order to identify, analyse, and structure the skills and knowledge used by teachers who are considered experts in AfL, a cognitive task analysis (CTA; Clark, 2014) will be used. CTA is "cognitive" in the sense that it attempts to identify the mental processes and decisions that experts use to achieve a goal and/or solve a complex problem (Clark, 2014). Experts are largely unaware of how they decide about and analyse problems in their expertise area. A CTA attempts to help identify more of the specific, operational elements of an expert's cognitive processes. A CTA leads to "an integrative, coherent description of everything needed to perform professional tasks properly" (Van Merriënboer & Kirschner, 2017); in this study, the professional task is AfL.

This study used the framework described by Clark et al. (2008) and refined by Van Geel et al. (2019) for performing a CTA in the context of education. In the initial stage, *collect preliminary knowledge*, literature was used to identify tasks that became the focus of the CTA. In order to perform a CTA, it is important to become generally familiar with the knowledge domain, content, context, and procedures being analysed. Different studies in which the AfL cycle have been described were used to select classroom situations that call for AfL skills (Gulikers et al., 2021; Ruiz-Primo & Li, 2013; Wiliam, 2011). The next stage in this CTA process was *identify knowledge representations*. In line with Van Geel et al. (2019), the representation forms used were: (a) a skills hierarchy in which all constituent skills, including the relationships between those skills are represented; (b) an overview of the required knowledge to perform these skills; and (c) an overview of the factors related to the complexity of performing the task. Classroom observations of lessons by expert teachers, followed by semi-structured stimulated recall interviews with these individual teachers were then used for applying focused *knowledge elicitation methods*. After a first *analysis of the data*, this *information was verified* in a meeting with expert teachers, as well as in a meeting with content experts. These sessions enabled the definition of a final skills hierarchy, the construction of a table with performance standards (specifying the desired performance of the skill), an overview of the required teacher knowledge, and an overview of the factors influencing the complexity of performing AfL in the classroom. In this way the *results were formatted* for the intended application.

The level of complexity of AfL may differ across situations. Van Geel et al. (2019) found that factors contributing to the complexity of teachers' skill at differentiated instruction were, for example, group composition (diversity, number of grades, students with special needs) and school support. To obtain more insight into the factors contributing to the complexity

of implementing AfL in classroom, the CTA included an analysis of these complexity-related factors. Based on the results of the CTA, new models for teacher professional development (TPD) trajectories could be designed and assessment instruments could be developed, enabling teachers, as well as other stakeholders, to train, assess, and monitor teaching quality with respect to AfL.

3.3 METHOD

3.3.1 Context

This study was conducted in the context of secondary education in the Netherlands (students 12-18 years old). Dutch secondary education is highly selective; it is a tracked system with three school types: pre-vocational secondary education (VMBO – 4-year course), senior general secondary education (HAVO – 5-year course), pre-university education (VWO – 6-year course; Beguin & Ehren, 2011). There is no central curriculum, and there is only one national assessment at the end of secondary education (OECD, 2008). Learning objectives are specified for the different stages and different tracks of the education system, but schools are autonomous in deciding on the teaching and learning methods and the curriculum design, including the subjects to be taught and the course content of these subjects, as long as they ensure the incorporation of these learning objectives (Beguin & Ehren, 2011; OECD, 2008). Teachers can decide how many and what kind of assessments (e.g., practical tests, theoretical tests) they will combine in their classroom assessment, and they have a lot of freedom to make the instructional changes they deem necessary based on assessment data.

3.3.2 Participants

The CTA was aimed at three secondary school subjects: English language, Dutch language and chemistry. Teachers considered to be AfL experts were identified through a network of educational consultants, teacher training institutes, and school boards. A total of 16 teachers (see Table 3.1) participated in this study. Four teachers participated only in the classroom observations followed by interviews. Four teachers participated in both the classroom observations and the expert teacher meeting. Eight others participated only in the expert teacher meeting. Background information about the selected teachers is shown in Table 3.1. In addition, 10 content experts participated in their own expert meeting. These were teacher educators, trainers, and researchers with expertise on AfL and assessment (in secondary education). These experts were recruited through connections of the researchers.

Table 3.1 Characteristics of the Expert Teachers

Teacher[a]	Gender	Age	Years of experience	Characteristics of observed class			
				Subject	School type	Grade[b]	Number of students
Anne[c]	female	50	19	chemistry	A	3	28
Bart[d]	male	49	23	English language	n/a	n/a	n/a
Diede[d]	female	41	15	English language	n/a/	n/a	n/a
Evert[d]	male	42	13	chemistry	n/a	n/a	n/a
Frank[d]	male	33	11	English language	n/a	n/a	n/a
Hermien	female	42	16	chemistry	C	5	19
Karin	female	38	14	Dutch language	A	3	30
Lina[c]	female	34	12	Dutch language	B	3	32
Manon[c]	female	63	32	English language	B	4	30
Marit[d]	female	58	25	chemistry	n/a	n/a	n/a
Mette	female	47	22	English language	C	4	27
Peter[d]	male	42	17	chemistry	n/a	n/a	n/a
Saskia	female	51	23	Dutch language	B	4	28
Thijs[d]	male	39	17	Dutch language	n/a	n/a	n/a
Willem[d]	male	61	10	chemistry	n/a	n/a	n/a
Wilma[c]	female	60	30	Dutch language	C	5	32

Note: A= pre-vocational secondary education, B = senior general secondary education, C = pre-university education.

[a] Names are pseudonyms.

[b] Numbers present year of the x-year program students are in

[c] Observed, but not present at the expert meeting.

[d] Present at the expert meeting, but not observed.

3.3.3 Data collection

Classroom observations and interviews

During the first step of the CTA, two consecutive lessons (in the same class) of eight teachers were observed (see Table 3.1). Thirteen of the 16 observed lessons were videotaped. The other three lessons were only audio-taped, due to technical problems. After each lesson, the teacher participated in a semi-structured interview. The questionnaire used for the interviews consisted of two parts. In the first part, the stimulated-recall method (Vallacher & Wegner, 1987) was used; 22 video-clips from the lesson were reviewed and the researcher asked questions about the teacher's actions and the underlying considerations. The selected video-clips were rich snippets of 4 to 11 minutes of the observed lesson in which the use of key strategies for AfL practice in different phases of the lesson was visible. For example, some video-clips showed the start of a lesson, others the middle part of a lesson or the

wrap-up. When a lesson was not videotaped, the selected lesson situation was described by the researcher. The second part of the interview consisted of more general questions about classroom composition, for example, number of students (Table 3.1), the course of events during the lesson in general, lesson preparation and the wrap-up of the lesson. In addition, the teachers were asked to identify perceived complexity-related factors during the implementation of AfL in the lesson and to categorise them from more complex to less complex. Finally, questions were asked about the teacher's background characteristics (Table 3.1). The interviews were audio-taped and transcribed. Summarized descriptions of the lesson situations in the video clips were made from an observer perspective. In total, 22 video clips, ranging from 4 minutes to 11 minutes, were described.

Meeting with expert teachers

The next step in the CTA was an expert meeting involving 12 expert teachers. The aim of this meeting was to verify and expand on the preliminary results obtained from the classroom observations and interviews in order to obtain a comprehensive image of AfL as a task in teaching practice. In the first part of the meeting, the preliminary results for the skills found were presented in order to verify them. The teachers discussed the results in three different sub-groups, after which the findings were discussed with the whole group. The aim of the second part of this meeting was to identify the complexity-related factors. In two groups, the teachers were asked to describe five (teaching) situations that call upon teachers' AfL skills. Teachers were then asked to supplement the situations described with possible complexity-related factors. The list of complexity-related factors obtained from the interviews was completed with these newly identified factors and a ranking was made from simple to complex. This ranking was then discussed as a group and an overview was made of the factors that make AfL easier or more complex. The aim of the third part of this meeting was to map out the steps, knowledge, actions and decision-making processes of the teachers when applying AfL in the classroom. First, the teachers were divided into four groups and asked to draw up a step-by-step plan for coherently integrating the different AfL key strategies (KS1-KS5) into a lesson. These plans were then discussed as a whole group, and the various step-by-step plans were combined. For this joint step-by-step plan, the knowledge, actions and decision-making processes were then identified using the following questions: What knowledge does a teacher need to carry out the steps? What does success look like when the teacher carries out the step correctly? What decision points does the teacher encounter in the different steps of the process? The discussions in groups as well as the plenary discussions were audio-taped and transcribed.

Meeting with content experts

The findings from the classroom observations, interviews, and the expert meeting with teachers were presented to 10 content experts during a second expert meeting. The first goal was to verify and supplement the list of desired teacher skills and complexity-related factors for implementing AfL in the classroom. The meeting started with a presentation of the preliminary results regarding the skills and complexity-related factors found from the lesson observations, interviews and input from the expert meeting with teachers. The content experts were asked to complete the list with newly identified skills and complexity-related factors. Next, the experts discussed what skills have a temporal relationship, implying that they can be learned and performed successively, simultaneously, or in a random order. Lower-level skills facilitate the learning and performance of the skills higher up in the hierarchy. In this way, a hierarchical classification of skills was created. The second goal of the meeting was to gather input for specifying the standards for acceptable performance by prioritizing AfL skills. The content experts identified those constituent skills that, in their opinion, were crucial for teachers to perform AfL well. Next, the content experts formulated performance standards for those skills. In the last part of the meeting, the activities focused on teachers' AfL approach in the video-recorded lessons, using the video clips from the semi-structured interviews. The following questions were used for discussion: Explain in a step-by-step plan to a beginning teacher what you as expert do? Why? What knowledge do you need to be able to take this step? When are you satisfied? The different discussions in groups as well as the plenary discussions were audio-taped and transcribed.

3.3.4 Data analysis

In line with Van Geel et al. (2019), the data were analysed in an iterative process. That means that after each stage of data collection, the data were analysed and the outcomes were used in the subsequent stage of data collection. In the first stage, the extent to which different key strategies of AfL appeared in the video fragments (used for the structured interviews with teachers) were analysed. The five key strategies for AfL were used to code the teaching situations that were reflected in the descriptions of the video-clips. Because the whole AfL process sometimes takes place within one class session in a short cycle and sometimes over several class periods through a long cycle, the occurrence of the key strategies in each of the 22 descriptions varied. In the second stage, the data from the interview were analysed.

For the different teaching situations extracted from the 22 descriptions, what skills the teacher needs, what knowledge the teacher uses, what complexity-related factors the teacher perceives and what consideration the teacher makes in doing so were noted. In

this way, we linked the teaching situations with the key strategies, required teacher skills, required underlying knowledge and the complexity-related factors for each interview. During the analysis, we also identified overarching teacher skills that are needed in all phases of the AfL process. This category of overarching teacher skills was added to the list of codes. In the third stage, the findings were supplemented and adjusted on the basis of the elaborations from the expert meetings, in order to refine and complement the identified teacher skills, knowledge and complexity-related factors. This generated an overview presented in a skills hierarchy. Data collection and analyses were carried out systematically, and used member checking, that is, verifying the findings by presenting them to the participants in the study (Creswell & Miller, 2000; Stalmeijer et al., 2014). During the expert meetings, whether the interpretation of the data by the researchers was in line with the interpretations by the teachers and/or subject-matter experts was verified.

3.4 RESULTS

3.4.1 Skills

Summarized descriptions were made of the video-taped lesson situations from an observer perspective. The five key strategies for AfL were used to code the teaching situations. An example of such a lesson description linked to the AfL key strategies is presented in Appendix C. In total 22 video clips with variable duration, ranging from 4 minutes to 11 minutes, were described. The analysis of lesson situations (N = 22), in which multiple teaching situations (N = 113) were identified and coded according to the five key strategies in the AfL process, showed that each of the five strategies was well-enough represented for follow-up analyses (Table 3.2). For each of the five key strategies (KS1 to KS5) in the AfL process, an inventory was therefore made of the crucial skills, based on the analysed data from the classroom observations, interview, and expert meetings. During the analysis, we also identified overarching teacher skills that are needed in all phases of the AfL process. For example, efficient lesson organization (including order in the classroom, proper lesson transitions, time management) is one of those overarching teacher skills.

Table 3.2 Key AfL Strategy Use

Key AfL strategy	Occurrences	
	n	%
KS1	17	15
KS2	31	27
KS3	25	22
KS4	21	19
KS5	19	17
Total	113	100

Note: KS1: clarifying, sharing and understanding learning goals and criteria for success; KS2: eliciting evidence of student learning (including self- and peer-assessment); KS3: analysing student responses; KS4: communicating about results (including feedback); KS5: taking concrete actions to adjust teaching and/or learning.

During the iterative process of data analysis, it became clear that classifying teacher skills based on the key AfL strategies KS1 to KS5 did not justify the complexity of AfL. For example, for the key strategy KS1 clarifying, sharing and understanding learning goals and criteria for success, the skills matrix showed that teachers need to prepare a lesson using several other skills, before they can clarify and share the learning goals with the students during the lesson. The teachers need to (a) translate curriculum from the subject area into learning goals and success criteria, (b) adapt learning goals to the target group and identify corresponding success criteria, (c) determine an appropriate lesson approach to share learning goals and success criteria with students. The AfL process during the lesson cannot be separated from the phases of lesson (unit) preparation and evaluation. Therefore, the coherence of the skills based on the teaching phases was examined. For example, what skills does it require for a teacher to prepare, perform, and evaluate an AfL-rich lesson? The following four phases were used to structure the different interrelated teacher skills needed in the AfL process: (1) lesson unit preparation, (2) lesson preparation, (3) lesson execution and (4) lesson evaluation. In this way, a skills hierarchy was created in which the crucial teacher skills necessary for the AfL process were structured according to the four lesson phases (Figure 3.1). Skills that are horizontally adjacent have a temporal relationship, and can be performed consecutively, simultaneously or in any order. Skills that are above or below each other have a vertical relationship, implying that skills lower in the hierarchy are pre-conditions for or supportive of the higher placed skills. For the constituent skills in the skills hierarchy, performance objectives were formulated based on the consultation with the content experts (Table 3.3). Performance objectives are descriptions that clearly reflect the desired performance after training.

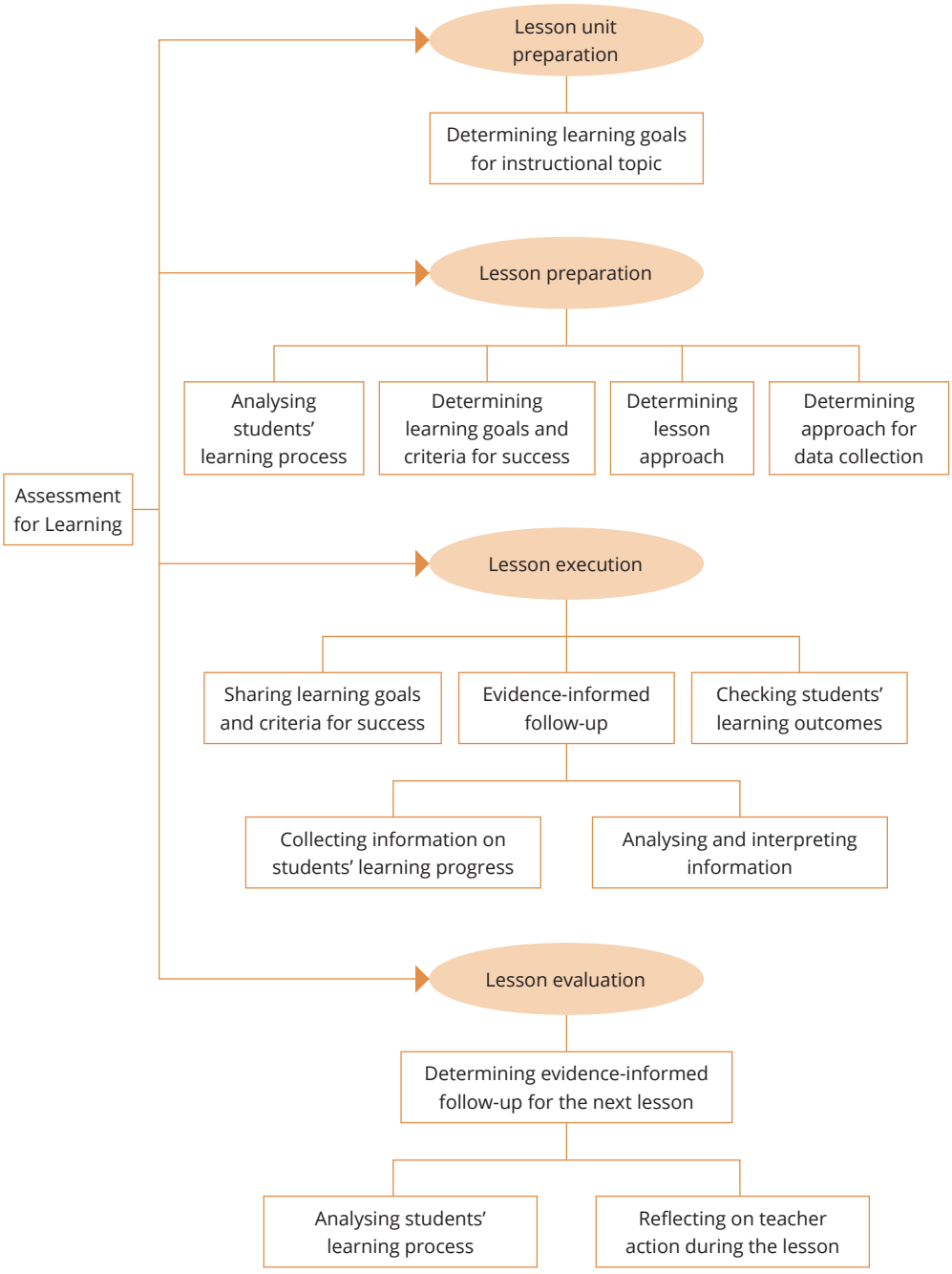


Figure 3.1 Skills Required for Assessment for Learning

Table 3.3 Performance Objectives for Assessment for Learning

Skill	Performance objective
Determining the learning goals for instructional topic	The teacher decides which part of the curriculum should be covered and translates the subject content given in the curriculum for the subject into learning goals. The teacher prepares the lesson series, focusing on determining which learning goals should be mastered at the end of the unit (usually for a period of four to eight weeks). The teacher formulates learning goals based on critical reflection on the instructional method, curriculum and student levels. The learning goals meet the following criteria: 1) specific; 2) ambitious; and 3) suitable for the target group.
Analysing students' learning processes	The teacher analyses the available information about students' learning using the evaluation of the previous lesson and the intended learning goals, and determines the level of differentiation needed to meet the needs of the diverse learners.
Determining learning goals and criteria for success	The teacher determines the learning goals for the lesson and describes the teaching needs related to these goals. The teacher formulates learning goals in understandable language for the entire group of students. Based on the learning goals and analysis, the teacher formulates (differentiated) criteria for success.
Determining lesson approach	The teacher formulates organizational and didactic approaches, following logically from the preceding analysis, and related to the learning goals and criteria for success that were formulated. The teacher describes what materials will be used and how within-lesson transitions will take place. The teacher determines how to share learning goals and criteria for success with students and how students will be engaged in these activities in class.
Determining approach for data collection	The teacher determines how and when information about students' learning will be collected during the lesson, in relation to learning goals and criteria for success.
Sharing learning goals and criteria for success	The teacher shares learning goals and criteria for success with students and makes sure they are clear for the students by calling attention to the goals and criteria at different moments during the lesson and encouraging students to be owners of their learning.

Skill	Performance objective
Evidence-informed follow up	The teacher stimulates students' learning processes by tailoring instruction to the students' learning needs, giving tips, and/or asking questions, to help students correct errors and misconceptions. Collecting information on students' learning progress: during the lesson, the teacher collects information about students' learning related to the learning goals and criteria for success. The teacher shows that mistakes are allowed: this is about insight into learning rather than summative evaluation or grading (providing a safe and stimulating learning climate). Analysing and interpreting information: the teacher (preferably together with students) analyses and interprets the information collected during the lesson, using the intended learning goals and criteria for success, to address mistakes and misconceptions. The conclusions from these analyses are shared with students in order to inform follow-up actions.
Checking students' learning outcomes	The teacher together with the students evaluates the lesson by eliciting information on the extent to which the learning goals of the lesson were achieved.
Determining informed follow-up for the next lesson	The teacher reflects on the lesson in relation to the learning goals and criteria for success – whether the goals were achieved or not – and determines what is needed for the next lesson. Analysing students' learning process: the teacher analyses and interprets the information collected to determine students' learning progress. Reflection on teacher action during the lesson: the teacher reflects on their actions in relation to the lesson preparation and intended learning goals for students.

3.4.2 Knowledge and attitude

The CTA also revealed the basic elements of teacher knowledge and attitudes that are critical for using AfL in the classroom. Teachers need domain-specific knowledge and pedagogical content knowledge to align AfL with instructional goals. To make this alignment, it is important that teachers are knowledgeable about students' previous learning in relation to the curriculum and the learning goals. Knowledge of students' misconceptions is especially essential to be able to look for gaps in students' knowledge regarding the particular curriculum topic. Teachers need to be knowledgeable about a range of assessment strategies. To maximize the opportunities for gathering evidence of learning, it is important to know how to apply different assessment strategies in order to elicit a response from each individual student. As far as teacher attitude is concerned, teachers ideally view AfL as a worthwhile process that yields valuable and actionable information about students' learning, and they preferably view AfL and the teaching process as inseparable. This means

that teachers consider AfL to be a didactic process that is shaped in the classroom, instead of an activity just to gather information. Finally, teachers ideally are aware that AfL requires a joint effort by teachers and students: students' self- and peer-assessments provide important opportunities for establishing students' current learning status.

3.4.3 Complexity-related factors

Five factors related to the perceived complexity of implementing AfL were identified.

1) Student group composition (e.g., the number of students, diversity of student levels of knowledge or ability, and students with special education needs). Greater diversity of student levels, for example, makes using AfL in lessons more complex. Matching the success criteria for a diverse group demands more creativity and experience from the teacher, as well as searching for misconceptions among a group of students at diverse levels.
2) The extent to which the teacher has control during the lesson. Adopting a less teacher-centred approach, for example, may give the teacher less certainty and requires more improvisation. Teachers perceived improvisation during the lesson as more complex.
3) The difficulty level of the content of the lesson (goal and topic). When there is a range of misconceptions, for example, it is even more important to continually check whether students have already (mis)understood it. It is more difficult for the teacher to identify a range of student misconceptions, because there is greater variety. It may therefore be more complex for the teacher to identify them and to take targeted follow-up action. This requires expertise from the teacher.
4) The amount of available support related to the teaching method, such as suggestions for remediation and remediation materials.
5) The amount of school support, for example, collaboration with colleagues and professional development.

3.5 DISCUSSION AND CONCLUSION

The key to the successful implementation of AfL in a lesson is the actual adaptation of teaching to the identified needs of all students, in a process that requires purposeful interaction between students and teachers (Carless & Winstone, 2020). Integrating AfL with daily classroom activities is a complex process, and teachers need to master a set of competencies in order to work on this together with their students (Schildkamp et al., 2020). In this study, we conducted a cognitive task analysis based on classroom observations, interviews, a teacher expert meeting and another expert meeting, to obtain insight into

the skills, knowledge and attitudes teachers need for AfL (RQ1) and into the factors that influence the complexity of implementing AfL (RQ2).

3.5.1 Skills hierarchy presenting a skillset for providing AfL-rich lessons

The CTA performed in this study provided an overview of all constituent teacher skills for AfL and the coordination between these skills, presented in a skills hierarchy (Figure 3.1). The results of this study showed that – in line with many other aspects of effective teaching – teachers should not only focus on executing the lesson using AfL strategies, but also need to prepare the entire lesson series beforehand, prepare each of the lessons within that series, and evaluate the lessons afterwards in order to make AfL-based teaching effective. That means that the teacher prepares a lesson based on the evaluation of the previous lesson and based also on the preparation of a lesson unit about the instructional topic. This enables the teacher to enact the lessons while tailoring instruction to students' learning needs, eliciting evidence of learning and encouraging students to be owners of their learning processes. Adaptive teaching proves to be important in all phases of instruction, that is, during planning, in the midst of teaching, and when reflecting on instruction (Parsons et al., 2018).

3.5.2 Cognitive task analysis offering the knowledge base for teacher professional development

Because adapting teaching to students' needs is considered to be an important characteristic of effective teaching, both pre-service and in-service teachers could benefit from professional development activities aimed at enhancing teachers' AfL competencies. The idea that TPD can foster improvements in teaching is widely accepted (e.g. Borko, 2004; Darling-Hammond et al., 2017; Desimone, 2009; Guskey, 2002). Despite this widespread agreement about its importance, precisely how TPD can foster teacher learning and improve teaching practice remains elusive (Kennedy, 2016). In response to this challenge, Kennedy (2016) suggested focusing on the underlying theory of learning embraced in the TPD. The four-components instructional design (4C/ID) model can provide this psychology of learning basis for the design of TPD programs that require the integration of teacher skills, teacher knowledge and attitudes (Van Merriënboer & Kirschner, 2017), in this case for the complex teacher skill of AfL.

This study provides an overview of the skills teachers need for AfL, presented in a skills hierarchy (Figure 3.1) that is complemented by formulation of performance objectives (Table 3.3). These indicators specify desired teacher performance after a training intervention, in other words, the desired results of the learning experience of teachers who participate in a TPD intervention for AfL. In addition, insight has been gained into what

knowledge and attitudes a teacher needs for AfL. Teachers' domain-specific knowledge and pedagogical content knowledge proved to be indispensable for implementing AfL in the classroom. To be able to tune AfL to the instructional goals, teachers need to have knowledge of students' previous learning in relation to the curriculum and the learning goals. Teachers also need to be knowledgeable about potential student misconceptions in order to be able determine gaps in students' knowledge for a particular curriculum topic. Moreover, five factors that influence the complexity of applying AfL in the classroom were identified: (1) student group composition, (2) the extent to which the teacher has control during the lesson, (3) the difficulty level of the content of the lesson (goal and topic), (4) the amount of available support provided by the teaching method, such as suggestions for remediation and remediation materials, and (5) the amount of school support, for example, collaboration with colleagues and opportunities for professional development. The insights gained are a valuable contribution to the AfL knowledge base and offer a basis for developing teacher professional development trajectories aimed at equipping teachers with the skills for AfL in the classroom. For example, we showed (Wolterinck et al., 2022) that the four-component instructional design (4C/ID) model for complex learning (Van Merriënboer & Kirschner, 2017) can be well used to design and implement a teacher professional development program for developing teachers' AfL competencies. The complexity-related factors can be used for sequencing learning tasks in the TPD program, from simple to more complex learning tasks.

Performing AfL in the classroom requires not only skills and knowledge from teachers, but certainly also a positive attitude and motivation to enrich your repertoire of actions as a teacher. Teachers ideally view AfL as a worthwhile process that yields valuable and actionable information about students' learning, in which AfL and the other aspects of teaching are inseparable; it is hoped that they are aware that one cannot happen without the other.

3.5.3 Limitations and implications for research and practice

We acknowledge several limitations of our study. The data collection for this study was conducted within one context, namely, Dutch secondary education. We focused on three subjects, Dutch, English language and chemistry. Although this offered the opportunity to thoroughly investigate the processes in one specific context, these exploratory findings cannot be generalized to other settings and populations (Miles et al., 2018; Yin, 2014). The teachers who participated in the observations, interviews and expert teacher meeting were considered to be teachers who apply AfL at a relatively high level. Although the selection of participating teachers was made carefully, the question remains whether all these teachers really are top AfL experts. Nevertheless, when observing enough expert

teachers, it is not a problem that not all of them are top teachers as far as AfL is concerned, because the overall pattern does emerge if a considerable number of teachers is studied and at least some of them are above-average with respect to AfL. Future research could verify whether the results found, the skills hierarchy, teacher knowledge and complexity-related factors, are comparable to what expert teachers report in other countries with similar contexts.

This study focused on teachers, but students also can play a crucial role in AfL, for example, using the key strategies of peer-assessment and self-assessment. Students' ownership of their own learning process can be enhanced if students assess their own learning with the purpose of comparing their current learning status with the learning goals and success criteria, and make judgements about their goal attainment (Hattie & Timperley, 2007; Nicol & MacFarlane-Dick, 2006; Sadler, 1989). These activities can activate students' cognitive and motivational capacities, focus students on their learning goals, and provide feedback and strategies that they can use to help them reach their goals (Panadero et al., 2018). Therefore, it would be interesting to also perform a cognitive task analysis (CTA) focused on students who are experienced in using AfL together with their teachers. The skills, knowledge and attitudes of these above-average (with respect to AfL) students could be systematically analysed and also the complexity-related factors they experience could be mapped. The steps for a CTA described in this study could serve well for such a follow-up study: what skills and knowledge does AfL require from students? This knowledge can be used to enrich TPD programs and make teachers aware of the shared responsibility and interplay between teachers and students and among students in the AfL process in classrooms.

4

Teacher Professional Development for Assessment for Learning Using 4C/ID

This chapter is based on

Wolterinck, C. H. D., Poortman, C. L., Schildkamp, K., Visscher, A. J. (2022). Assessment for Learning: Developing the Required Teacher Competencies. *European Journal of Teacher Education, 1 – 19.* Doi: 10.1080/02619768.2022.2124912

Abstract

Assessment for learning (AfL) is complex requiring teacher competences for collecting, analysing and interpreting evidence of student progress, and for adapting instruction and learning accordingly. In this study, the four-component instructional design (4C/ID) model for complex learning was used to design a teacher professional development (TPD) program for developing teachers' AfL competences. The evaluation of the implementation of the program showed that teachers were positive about the relevance and usability of the program, and about the AfL knowledge, skills and attitudes gained. This study yields important lessons about how to support teachers in learning to apply AfL in their classrooms. The study shows that the 4C/ID model can provide the learning-psychological basis for the design of TPD programs for the acquisition of complex teacher skills, that require the integration of skills, knowledge and attitudes.

Keywords: Assessment for Learning; formative assessment; professional development; teacher education, instructional design.

4.1 INTRODUCTION

Assessment for Learning (AfL) is a conscious attempt to make (classroom) assessment an essential part of teaching and learning (Wiliam, 2011). This implies that teachers and students will ideally use assessment results together to improve ongoing learning as part of everyday practice (Klenowski, 2009). The information derived from different assessment sources can be used as a form of continuous feedback to steer learning (Hattie & Timperley, 2007; Wiliam & Leahy, 2015). Teachers are increasingly expected to have the skills, knowledge and attitudes needed to enact AfL in their classroom (Heitink et al., 2016). Although AfL can enhance student learning (Black and Wiliam, 1998), research has shown that the implementation of AfL often is not effective (Bennett, 2011). No clear consensus on what the approach "AfL" encompasses precisely causes a wide variety of implementations. Another explanation of the lack of positive effects could be that teachers struggle with implementing AfL in their classrooms. Teachers may lack the professional competences needed to implement AfL in their classroom effectively (Schildkamp et al., 2020).

Competencies related to collecting, analysing and interpreting evidence from assessment, and adapting instruction and learning accordingly are crucial for the effective implementation of AfL in classroom practice (Lee, 2011). Research on teachers' use of AfL has indicated room for improvement when it comes to these competencies (Gulikers & Baartman, 2017; Kippers et al., 2018). Therefore, investment in professional development (PD) is crucial (Heitink et al., 2016). To effectively support teachers in using AfL, we developed and implemented the AfL teacher professional development (AfL-TPD) program using the Four-Component Instructional Design (4C/ID) model (Van Merriënboer and Kirschner, 2017). In the design of teacher professional development programs an explicit, validated, learning-psychological basis, that takes into account what it requires to acquire knowledge, skills and attitudes, is not seldom missing. The 4C/ID model can provide such a basis.

The 4C/ID model is a task-centred approach suitable for teaching complex professional tasks that require the integration of knowledge, skills, and attitudes, and the coordination of various skills (Van Merriënboer and Dolmans, 2015). The aim of this study was to evaluate the effects of the AfL-TPD program that was developed based on the following research questions:

How relevant and usable did teachers find the Assessment for Learning - professional development program that was designed on the basis of the four components instructional design model?

What knowledge, skills and attitudes needed for Assessment for Learning did teachers learn as a result of participating in the Assessment for Learning - professional development program?

4.2 THEORETICAL FRAMEWORK

4.2.1 Assessment for Learning

AfL is an approach to formative assessment that takes place as part of ongoing classroom practice and is meant to enhance the quality of students' learning processes (Black and Wiliam 1998; Klenowski 2009; Stobart 2008). AfL can be seen as a process focusing on the interaction between student and teacher and often aims to stimulate learning processes such as self-direction, self-regulation, or learning motivation (Popham 2008; Clark 2012). Wiliam (2011) distinguished five core strategies for implementing AfL in classroom practice: 1. clarifying, sharing and understanding learning goals and criteria for success; 2. engineering effective classroom discussions, activities, and learning tasks that elicit evidence of learning; 3. providing feedback that moves learning forward; 4. encouraging learners to be instructional resources for one another; 5. encouraging learners to be owners of their own learning. Studies focusing on the implementation of the five strategies of AfL in teaching practices emphasise the cyclical character of the AfL process and emphasise that the collected data are analysed (turning data into information) and transformed into decisions, so that teachers can provide feedback for students (Antoniou & James, 2014; Gulikers & Baartman, 2017; Ruiz-Primo & Furtak, 2007; Schildkamp et al., 2020). Bennett (2011) states that the term, 'formative assessment', does not represent a well-defined set of artefacts or practices and existing definitions admit such a wide variety of implementations that effects should be expected to vary widely from one implementation and student population to the next. In conclusion, it can be said that AfL is a complex teacher competence, and TPD should explicitly support teachers in integrating AfL strategies coherently in their classroom practice, in order to maximize its potential impact (Lee, 2011).

4.2.2 Design principles for the AfL-professional development program

From cognitive psychology, we know that the following is important for learning complex tasks and to ensure the transfer of learning: whole meaningful, authentic and varied learning tasks; ordering learning tasks from simple to complex, in combination with the gradual decrease in learner support and distinguishing between the non-routine and routine aspects of complex skills (Van Merriënboer et al., 2006). The AfL-TPD program in this study was designed using a task-centred instructional design model, the Four-Component Instructional Design (4C/ID) model, in which the design of authentic learning tasks follows an holistic approach dealing with real-life professional tasks that allow the learner to practice all the nonroutine and routine aspects of a complex professional task (AfL in this study) simultaneously (Van Merriënboer, 2007). The 4C/ID model includes four

interrelated components: a sequence of learning tasks based on authentic professional tasks, supportive information describing how to approach the tasks and how the domain is organized, procedural information describing step-by-step procedures to perform routine aspects of the tasks, and part-task practice for repetition of aspects that need to be highly automated (Van Merriënboer & Dolmans, 2015; Van Merriënboer et al., 2003). The learning tasks for teachers which need to be conducted in the AfL-TPD program are based on the results of a Cognitive Task Analysis (CTA; Clark et al., 2008) of teachers who are AfL-experts.

The CTA was conducted in a previous study which provided an overview of all constituent skills and the relationships between these skills, visualized in a skills hierarchy depicted in Figure 4.1 (Wolterinck et al. 2021). Four main skills make up AfL and are closely interrelated: the teacher prepares a lesson 1), based on the evaluation of the previous lesson 2) and based on the preparation of the set of lessons for an instructional topic 3). This enables the teacher to enact the lessons while tailoring instruction to students' learning needs, eliciting evidence of learning and encouraging students to be owners of their learning 4). Learning and performing these four main skills is facilitated by a subset of constituent skills (i.e., those positioned next to the main four skills in Figure 4.1). These constituent skills are depicted vertically, implying that they must be performed in order. For example, for the main skill of 'executing a lesson' the teacher starts with sharing learning goals and criteria for success, followed by collecting information on students' learning processes, analysing and interpreting that information and, finally, executing evidence-informed follow up.

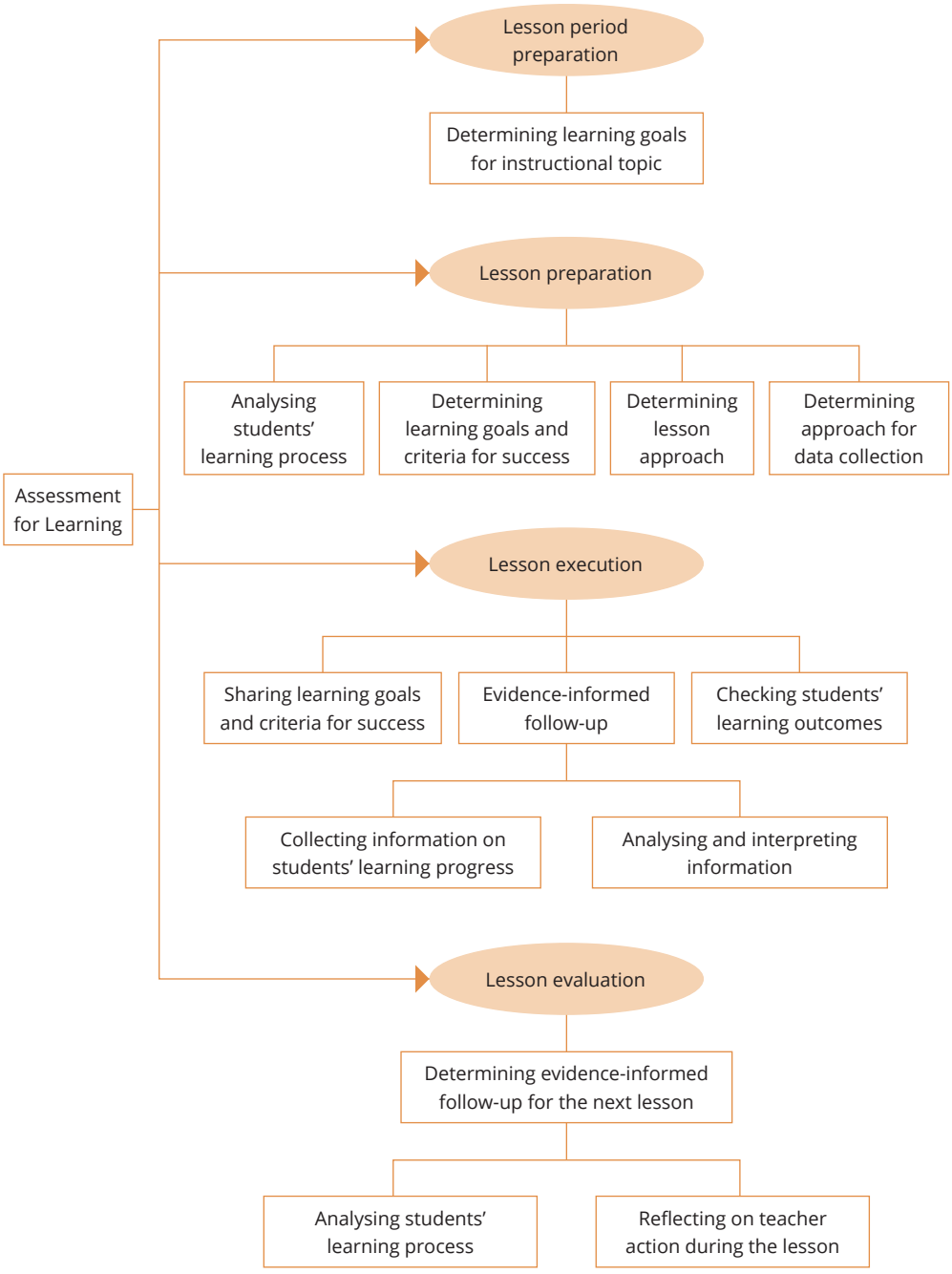


Figure 4.1 Skills Required for Assessment for Learning

Besides the skills hierarchy, the CTA revealed the kinds of teacher knowledge are critical for using AfL successfully in classrooms (Heritage, 2007; Wolterinck et al., 2021): domain knowledge, pedagogical content knowledge, knowledge of students' previous learning, and assessment knowledge. Teachers need to know how to align formative assessment with instructional goals, which requires knowledge of students' previous learning in relation to the curriculum and the learning goals and knowledge of students' misconceptions. Teachers must also know a range of formative assessment strategies, and know how to apply them, to maximize the opportunities for gathering evidence, while understanding that the quality of the assessment is an important concern. Finally, teachers need to know that AfL requires a joint effort by teachers and students: students' self and peer assessments provide important opportunities for establishing their current learning status. With regard to teacher attitude: teachers need to view AfL as a worthwhile process that yields valuable and actionable information about students' learning, and they need to view AfL and the teaching process as inseparable, recognizing that one cannot happen without the other (Heritage, 2007; Wolterinck et al., 2021).

The CTA included an analysis of the factors that influence the complexity of implementing AfL: student group composition (e.g., number of students, variation in student levels), the content of the lesson (goal and topic), curriculum material (suggestions for remediation and materials or not) and support from the school (collaboration and facilities) (Wolterinck et al., 2021). These factors were used as focal points for selecting learning tasks for the AfL-TPD program according to their complexity.

The 4C/ID design guidelines led to several design decisions. First, the main part of the program took place within schools, using the *whole-task approach* with learning tasks that represent the range of possible real-life tasks the teacher may encounter. This means that teachers practiced AfL within the full instructional teaching sequence of preparing a lesson series for an instructional topic, preparing lessons, enacting those lessons and evaluating the lessons, in which all constituent skills are supposed to be present. Second, the program provided many opportunities for practice and experimentation, by applying newly acquired skills during *daily teaching practice*. The program focused on relevant content, highlighting relevant domain knowledge, and teacher learning is activated by connecting what teachers already know or can do with what has to be learned. Third, the program stimulated *active learning using video-recorded lessons*, demonstrating expert strategies and stimulating reflective skills. Fourth, the program stimulated *collaborative teacher learning*, activating teachers as peers in the feedback process. Finally, the program spans approximately 8 months, to provide *ample time* to reach the desired outcomes. Figure 4.2 shows a schematic overview of the essential building blocks of the AfL-TPD program.

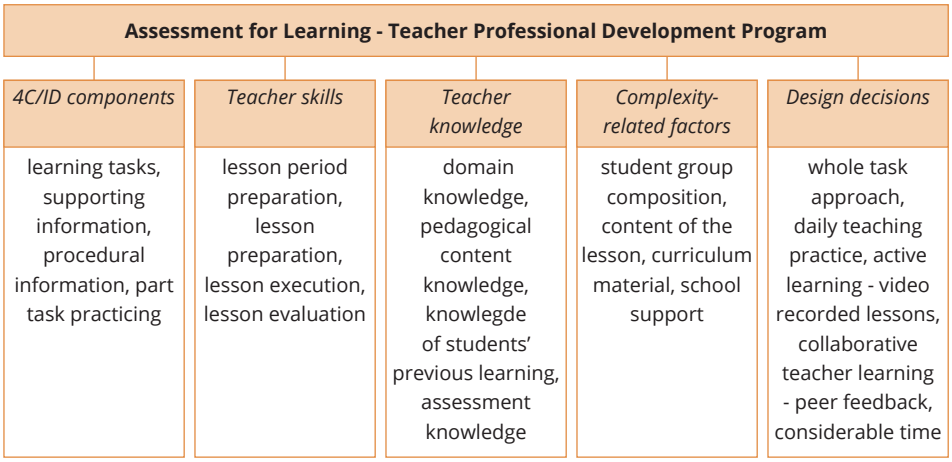


Figure 4.2 Essential Building Blocks of AfL-TPD Program

4.2.3 Evaluation of AfL-Professional Development Program

Merrill's first principles of instruction (Merrill, 2012) provide a system for assessing instruction for complex learning of authentic, real-world, whole tasks. These principles are based on a synthesis of instructional design theories, consistent with the 4C/ID model of instructional design, and therefore are suitable for a formative evaluation of the *relevance and usability* of the AfL-TPD program (Frick et al., 2010; Merrill, 2012). The five criteria that 4C/ID-designed instruction must meet are: 1) instruction is problem- or task-centred, 2) teacher learning is activated by connecting what teachers already know or can do, with what is to be learned, 3) teachers are exposed to demonstrations of what they are to learn, 4) teachers have opportunities to try out what they have learned, in combination with instructor coaching and feedback, and 5) teachers integrate what they have learned into their own classroom practices (Merrill, 2012).

For assessing teachers' *learning* in the AfL-TPD program, teacher skill improvement was examined in terms of the constituent skills described in the skills hierarchy and the related performance objectives. For each of the constituent skills, standards were set for desired performance after completing the program (Appendix A). To examine changes in teacher *knowledge* and *attitudes*, the following elements of teacher knowledge and attitude, based on the CTA, it was evaluated whether teachers: knew how to align formative assessment with instructional goals; had knowledge about a range of formative assessment strategies; had knowledge of students' previous learning in relation to the curriculum and learning goals; had knowledge of students' misconceptions; viewed AfL as a worthwhile process that yields valuable and actionable information about students' learning; viewed AfL and the teaching process as inseparable and recognized that one cannot happen without the other.

4.3 METHOD

This study was conducted in the context of Dutch secondary education (students 12-18 years old). To study teachers' satisfaction with the AfL-TPD program in regard to the 4C/ID principles and teachers' learning in terms of their knowledge, skills, and attitude necessary for implementing AfL in practice, we adopted a mixed-methods methodology. Teachers completed a questionnaire and also participated in focus group interviews for triangulation purposes. Coaches from the AfL-TPD program participated in a separate focus group interview (Cohen et al., 2013).

4.3.1 Assessment for Learning - professional development program

The AfL-TPD program was conducted on a subject-specific basis (English language, Dutch language and chemistry) and had a study load of approximately 40 hours spread over 8 months. This included 20 hours of contact time, spent in 4 meetings, and 20 hours of practicing time, consisting of carrying out learning tasks. In the AfL-TPD program we focused on workplace-based learning and used participants' lesson series and their daily lessons (for the participants their English language, Dutch language and chemistry lessons) as learning tasks. These real-life whole tasks present the full range of variability, require all the necessary skills, and allow daily opportunities for practice. Figure 4.3 shows a schematic overview of the AfL-TPD program and Appendix D presents the final blueprint for the design of the introductory meeting at the start of the program.

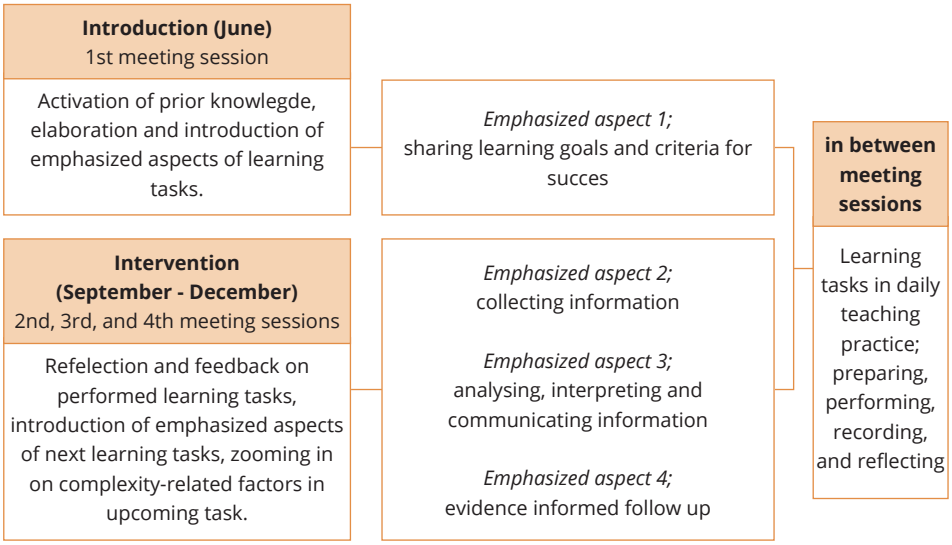


Figure 4.2 Overview AfL-TPD Program

In the 4C/ID model, learning tasks at the beginning of the program are ideally practiced in a safe learning environment in which errors do not have serious consequences and where support and feedback can be easily presented just in time (Van Merriënboer and Kirschner 2017). Because daily lessons were used as learning tasks in the AfL-TPD program, it was not possible to manipulate the complexity-related factors identified in the CTA (e.g., student diversity, group size). In the current design it was for this reason not possible to work with task classes in which each task class includes a set of equivalent learning tasks that are at the same level of complexity. This problem was counteracted by using an emphasis manipulation approach, to sequence learning tasks. In this approach, the learner's attention is actively allocated to an aspect of a certain skill or task (Gopher, 2006), for example, task class 1 focused on sharing learning goals and criteria for success during the lesson (see appendix C for a more extensive description of task class 1).

To compensate for the reduced support during the learning tasks in daily lessons, we increased support before and after the learning tasks (Frerejean et al., 2021). The first face-to-face meeting was planned to prepare teachers for independent practice in their own classrooms focusing on sharing learning goals and criteria for success. The coach guided the teachers in studying exemplars and clarifying expectations regarding what good practices look like (Carless & Chan, 2017). For six weeks, teachers practiced AfL within the full instructional teaching sequence and had to prepare a lesson series for an instructional topic, prepare lessons, enact those lessons and evaluate the lessons. Teachers learned how to apply the whole cycle of AfL, while actively focusing their attention on sharing learning goals and criteria for success during reflection and feedback activities. By the end of the program, all aspects of the task had been addressed. The final task class ended with designing a lesson series; in the fourth and final meeting session, school teams shared their experiences and reflected on their overall improvement. This concluded the 8-month AfL-TPD program.

4.3.2 Participants

Teachers

Four schools voluntarily signed up to participate in the AfL-TPD program, responding to a call communicated via a newsletter sent out by one of the largest school boards in the Netherlands. These schools were each asked to have three teachers participate, preferably one English language teacher, one Dutch language teacher and one chemistry teacher. We started the program with 12 teachers in total, of which, two participants left the program after the second meeting, because the AfL-TPD program was too demanding and too hard for them to combine with their own teaching job. Eight out of ten teachers in the program responded to the questionnaire and participated in the focus groups. For an overview of the key characteristics of the ten teachers see Table 4.1.

Table 4.1 Key Characteristics of Respondents

		N
Subject	English language	2
	Dutch language	4
	Chemistry	4
Gender	Female	9
	Male	1
Teaching experience	< 5 years	1
	5 - 14 years	5
	15 – 24 years	2
	> 25 years	2

Coaches

A design team consisting of the researchers of this paper, and coaches from the Dutch national expertise centre for curriculum design set out to design and implement the AfL-TPD program. The coaches for the AfL-TPD program were subject-matter specialists, one English language, one Dutch language, and one chemistry specialist, and they were experienced in supporting teachers in secondary education. One of the coaches, the English language specialist, had ten years of experience with supporting teachers in using AfL. The other two coaches were less familiar with supporting teachers in using AfL, and were therefore trained by the English language specialist for the AfL-TPD program. The program meetings were carefully prepared by the design team. Before each meeting with teachers, the coaches and researchers participated in two preparatory meetings. The researchers who participated in the design team attended all program meetings for observation purposes, in order to use this input in the preparatory meetings.

4.3.3 Data collection

Questionnaire

In this explorative study a questionnaire was used for evaluating the AfL-TPD program in which participants' satisfaction and whether teachers found the program relevant and usable (utility of the content, timeliness, materials and support, organization as well as satisfaction with the program design) was examined using 27 statements. The statements also covered various other aspects of the program, including the materials and support provided and the organization of the meetings. To optimize its face validity, the instrument was piloted with two expert researchers and four teachers (Cohen et al., 2013). Based on their comments, adjustments were made to the questionnaire. The questionnaire was administered after the final program meeting. Teachers were asked to express their agreement with the statements (e.g., 'In this program, I was satisfied with the course

material that was offered'), using a 5-point Likert scale ranging from 'strongly disagree' to 'strongly agree'. The questionnaire items were in Dutch. In addition, teachers could give a brief explanation of the chosen answer.

Focus group sessions with teachers and coaches

Four weeks after the final program meeting, two focus groups with four teachers each were conducted and one focus group session with the three coaches was held to evaluate from multiple viewpoints (Cohen et al., 2013). Two teachers from each of the four participating schools were invited to attend the focus group sessions for teachers, which resulted in the representation of all three subjects . The goal of these meetings was to evaluate teachers' satisfaction with the program and teachers' learning outcomes (skills, knowledge and attitude). The first goal of the focus group session for coaches was to have the coaches evaluate teachers' satisfaction regarding various aspects of the AfL-TPD program. The second goal was to have the coaches evaluate teachers' learning outcomes (skills, knowledge and attitude regarding AfL).

All three focus group sessions lasted two hours and were audio-recorded. The teachers and coaches were first asked to respond individually to a series of statements and questions based on the theoretical framework and the questionnaire data, followed by a group discussion yielding a collective rather than an individual view (Cohen et al., 2013). The statements (e.g. 'I was satisfied with the approach in which new knowledge was integrated into my own daily teaching practice', 'In order to improve the AfL-TPD program, teachers need more examples from practice') and questions (e.g., 'What did you learn in terms of knowledge of AfL?', 'How satisfied do you think teachers were with the guidance and materials?') used in the focus group session were discussed in advance with a researcher, after which the formulation of the statements and the questions was adjusted.

Analysis

The first research question (teacher satisfaction) was partly answered by analysing the frequency distribution of the questionnaire responses. Additionally, both the first and the second research question (teacher learning) were answered by analysing the recordings of the three focus group sessions. All focus group sessions were audio-taped and transcribed. Based on the theoretical framework, an a priori coding scheme with 20 codes was developed (Table 4.2 presents the coding scheme). The inter-rater reliability between the three coders was calculated across ten percent of the focus group data (Poortman & Schildkamp, 2012). An acceptable Cohen's Kappa of 0.72 was found (Eggen & Sanders, 1993). After coding the focus group data, we summarized what teachers or coaches said during the focus group interviews related to each code. Quotes from the respondents

were translated into English for use in this chapter. Based on the analyses of multiple focus group interviews we could report detailed answers to the two research questions (Cohen et al., 2013).

Table 4.2 Coding Scheme

Concept	Codes
Satisfaction	Design
	Materials and support
	Organization
Learning	Knowledge: general, preparing a lesson period; preparing a lesson; lesson execution; lesson evaluation.
	Skill: general, preparing a lesson period; preparing a lesson; lesson enactment; lesson evaluation.
	Attitude: general, preparing a lesson period; preparing a lesson; lesson execution; lesson evaluation.

4.4 RESULTS

4.4.1 Teacher satisfaction

Quality of the design

The questionnaire showed that participants were satisfied to very satisfied with the content of the AfL-TPD program. For example, all eight participants indicated contentment with the practice-oriented tasks, the structure offered to link new knowledge with prior knowledge, and the examples used for modelling. In the focus groups, teachers reported their satisfaction with the applicability of what was learned in their own teaching practice: "In one of the meetings I prepared a lesson for the next day, the content related directly to what I was doing in my class".

All eight teachers were positive about being actively engaged by the program during the meetings, for example, via discussing examples and giving and receiving feedback. They also appreciated the assignments between meetings, which were linked to their daily practice: "The link with your own teaching practice was very strong, especially because you had to make recordings of your own lesson". In the focus groups, teachers mentioned their appreciation for the exchange of experiences during the meeting sessions and would like to see this happen more frequently between the meetings, when they are applying newly learned insights in their own practice: "For example, sending each other videos between training sessions enables you to give each other more feedback. I could have got more out of the training". Seven out of eight teachers expressed their satisfaction about the structure of the skills hierarchy. It gave them an overview and clarified the cyclical

process of AfL. This also applied to the exemplars used, which stimulated discussion and developing a sense of quality together.

Teachers indicated some room for improvement regarding the program, for example, more explicit use of prior knowledge, and providing success criteria for the assignments. Teachers also mentioned differentiation in the AfL-TPD program as an area for improvement, such that teachers receive instruction and tasks tailored to their needs: "One teacher develops quickly and the other very slowly and there can be more focus on this in the program". Four teachers were only moderately satisfied with the focus on sharing knowledge with colleagues in their school and with how literature was used.

Coaches indicated that the program should focus more on the cyclical process of AfL, because teachers need to understand that AfL is about students' learning processes instead of focusing on students' learning outcomes. Otherwise, teachers tend to use AfL as an instrument for testing without grades: "It is a process, formative assessment is not just a test to see where the student currently is. It's about them getting further and what is needed to do that". Teachers generally came across as satisfied, according to the coaches, as demonstrated by, for example, high turnout at meetings (only one teacher missed a meeting session due to sickness), active participation in discussions, asking questions, and actively working on the assignments in their own teaching practice. One of the teachers indicated to a coach, for example, that after 35 years of teaching, this program had given this teacher's job a new boost. With great enthusiasm, this teacher now reported implementing AfL in her lessons.

Materials and support

The questionnaire showed that all eight teachers were very satisfied with the applicability of materials to their own teaching practice, such as the use of exemplars (video fragments of lessons), discussing examples from participants' teaching practices and sharing experiences and teaching materials. In the focus groups, teachers reported that watching each other's videos of lessons was instructive and gave insight into their own learning processes: "The videos were useful and stimulated the conversation and feedback about own actions". Using peer feedback to look at one's own actions is easier than directly reflecting on one's own actions; "Analysing those videos together was just great fun because colleagues see things you don't see yourself, because you're the one standing there". The book with supportive information was also very practical as a reference work. Teachers added suggestions for improvement in the commentary to the questionnaire, however: "You get a book at the beginning but little attention is paid to it. Coaches may direct how to use the book as a guide". All teachers experienced a lot of guidance and support from the coaches, because each subject had a curriculum specialist present and

the group of teachers was relatively small. Coaches stated that they could almost give one-on-one guidance and many questions could be answered. Teachers also addressed their questions to the coaches between the meetings via e-mail. All eight teachers totally agreed with statements such as: "Questions were adequately answered by the coaches" and "I am satisfied with the course material that was offered".

Teachers indicated some areas for improvement, for example, in the second meeting session the teachers received a syllabus containing an overview of information collection techniques as supportive information for task class 2. Instead of just providing the syllabus, it would be better to share the syllabus as linked with an assignment, for example, asking teachers to choose a technique and develop an application of it in their own lesson. Other areas for improvement reported were, for example, practicing more with formulating and sharing learning goals and success criteria. The assignments should be provided with clear formulated success criteria for the participants: "As participants, we should also set our goals and our success criteria more in these meetings". Teachers suggested that the coaches could demonstrate greater variety in the exemplars (not just language examples), time for reflection to translate the example to one's own teaching practice, and more focus on the content of the book. The meetings could also be less traditional, meaning that coaches could demonstrate the use of AfL more explicitly in their own teaching, or, as one of the teachers stated, "teach as you preach".

Organization and learning management

The questionnaire showed that teachers were very satisfied with the organization of the meetings; for example, seven out of eight teachers indicated their contentment with the spread of the meetings over time, the efficiency of the meetings and the facilities used. All teachers also mentioned their contentment about working in small groups (up to 4) of varying composition, as this allowed discussions and the exchange of experiences. Participating with colleagues from the same school was experienced positively, because it stimulated learning between the meetings in teachers' own teaching practice in their own schools, by giving peer feedback and discussing video recorded lessons.

Both teachers and coaches indicated some room for improvement; for example, they suggested taking more time between the last two meetings to build in peer feedback: "With the feedback obtained, the next lesson can be given and recorded and be compared with the former lesson, in order to make development visible. Learning does not just take place during the meetings, but especially between the meetings when working in your own teaching practice". Both teachers and coaches also suggested setting up a digital environment for sharing the video fragments and subsequently providing each other with feedback, and for conducting intake interviews prior to the AfL-TPD program to get an

idea of teachers' prior knowledge and experience with AfL. Teachers suggested involving school leaders in the program, because school leaders need to support their teachers in broadening AfL at their own school: "Have the school leader join you in one of the first meetings. They should be able to support their teachers to broaden the developments around formative assessment in their own schools".

4.4.2 Teacher learning

Knowledge and skills

Teachers and coaches both indicated that teachers gained more knowledge about AfL and about the context in which AfL can be used in one's own teaching practice. For example, the definition of AfL became clear; teachers were able to use the terminology and see possibilities for applying AfL in their lessons, and teachers indicated that the structure of the skills hierarchy was supportive. The importance of sharing learning goals and criteria for success with the students and the importance of collecting information about student learning based on learning goals are examples of what the focus group teachers said they learned in the program. Another key learning outcome according to the focus group teachers is the awareness of building AfL as a cyclical process into their lessons. Teachers said they learned that AfL is a cyclical process in which teachers must use the strategies coherently together with students, in order to apply it effectively in the classroom: "AfL is not just a tool for gathering information about students' learning processes. It's about how I can help students to develop further and what is required to do so and also to promote students' ownership".

Teachers reported that the AfL-TPD program improved their skills; for example, they prepared the lessons more consciously by thinking about the necessary prior knowledge, checking this with students and building on it further. One teacher indicated that she now differentiates more in her lessons; at the beginning of the lesson, she checks which students have mastered the content and can continue working on it themselves and which students need extended instruction. Teachers also felt more comfortable about the skill of collecting and analysing information on students' learning processes and taking action in their teaching: "I've become much better at collecting information, for example, the use of Socrative [cloud-based student response system] and asking questions, and acting on it. It makes my lessons more effective than before". Both teachers and coaches reported that teachers developed their skills for evaluating the lessons more consciously using the video fragments from their lessons and asking reflective questions such as: "What does this mean for my next lesson and how can I change the daily grind?". The program gave them tools, for example, the use of exit cards to reflect on one's own teaching practice: "As you are constantly collecting information about students, you can also see more about the

effects of your own actions".

The AfL-TPD program helped teachers to start applying and experimenting with AfL in their own practice: "I'm noticing how AfL is changing my lessons and that I approach things differently because I'm more aware of how to achieve more'; '(...) you get to know your students better". Coaches also indicated that teachers showed their AfL skills in several good examples: "One participant developed and taught lessons about organic chemistry in which she determined and shared the learning goals together with her students and then collected information about students' learning based on the learning goals". One teacher indicated that her knowledge and skills had certainly increased; for example, when preparing the lesson, she is now more conscious of students' prior knowledge, what they should know and be able to do, and how she can act on that.

Although teachers indicated that they gained more knowledge and skills related to AfL, both teachers and coaches indicated that more time is needed to practice it properly, for it to become a daily routine: "I have to stimulate myself to let it become a fixed pattern; One good example in the program does not mean that teachers are able to apply AfL in every lesson". According to both teachers and coaches, the strategy of 'formulating learning goals and criteria for success' was still challenging and should be given more attention in the program. At the start, teachers indicated themselves to be skilled at determining learning goals and success criteria, but gradually it turned out to be more complex than teachers thought: "Teachers focus too much on details (little twigs) and not enough on the higher learning goal, the bigger picture (branch). For example, the correct conjugation of the English verbs is not a learning objective, but is a success criterion for the correct writing of a business mail".

Attitude

The AfL-TPD program succeeded in changing teachers' attitude towards the use of AfL in their teaching practice. Teachers said they learned to see AfL as a worthwhile process that yields valuable and actionable information about students' learning, and teachers recognized that AfL and teaching are closely related: "I learned to use AfL more consciously and I became enthusiastic about the concept of AfL". Teachers reported having developed their idea of AfL and related concepts (e.g., learning goals, criteria of success, feedback). They also indicated that the program had taken them out of their routine and made them look more critically at their own teaching practice: "I have become more motivated and positive about AfL, because I started to work with it in my class and it improves the quality of my lessons". Participating in the AfL-TPD program raised awareness that things need to change in secondary education: less use of summative assessments and less whole classroom teaching. Coaches confirmed the change in teachers' attitude towards AfL:

"Teachers think more consciously about AfL and are more enthusiastic about the use of AfL together with students". Coaches reported that teachers were more aware of the importance of setting and sharing the learning goals for the lesson together with the students, the importance of focusing on students' learning processes instead of their learning outcomes, and when to apply AfL strategies in their lessons.

4.5 CONCLUSION AND DISCUSSION

The study yields important lessons about how to support teachers in learning to use AfL in their teaching practice. More in general, the study shows that the 4C/ID model (Van Merriënboer and Kirschner, 2017) can provide the learning-psychological basis for the design of professional development programs for the acquisition of complex teacher skills, that require the integration of skills, knowledge and attitudes.

4.5.1 Contentment with relevance and usability of the AfL-TPD program, and teachers' knowledge, skills, and attitudes related to AfL gained

The evaluation of the implementation of the AfL-TPD program showed that teachers were positive about their active engagement during the meetings, developing a sense of quality together via discussing exemplars and giving and receiving feedback on video recorded lessons. Teachers reported that the content of the AfL-TPD program was highly applicable, using learning tasks in their own teaching practice. The results show that teachers developed their idea of AfL and related concepts and of the contexts in which AfL can be used in their own teaching practice. Teachers reported learning that AfL is a cyclical process, instead of just an instrument for testing without grades, in which teacher must use the strategies coherently together with students, in order to apply it effectively in their own classroom. Teachers became more skilled in analysing students' learning processes and in reflecting on their own actions.

4.5.2 Important design decisions for implementing an AfL-TPD program

First, the program was developed using the 4C/ID model, including a *whole task approach* with real-life learning tasks, decreasing scaffolding, and distinguishing between non-routine and routine aspects of the complex skill (Van Merriënboer et al., 2006). Using a whole-task approach can help teachers transferring newly learned skills to their professional task, their daily lessons in their own classrooms (Van Merriënboer and Kester 2008), and teachers this way experience the complexity of implementing AfL while they practice. Although the design team had little control over task sequence or complexity, because daily lessons

were used as learning tasks, this problem was counteracted by using tasks with a high amount of support and guidance during the face-to-face meetings.

Second, the program provided many opportunities for practice and experimentation by applying newly acquired skills in teachers' *own teaching practice* and teachers were actively engaged by means of practice-oriented tasks in which learning was stimulated, by connecting what they already knew or could do with what they had learned. Teachers were asked to practice continuously between the face-to-face meetings. This distribution of practice opportunities is considered important to improve long-term learning, which can be explained by the spaced-practice principle: "practice opportunities distributed across time rather than massed within one session" (Petersen-Brown et al. 2019, p. 977).

Third, the program stimulates active learning using *exemplars and video recorded lessons*, demonstrating expert strategies and stimulating reflective skills. An essential component of the program, encouraging teachers to reflect, was the dialogic use of exemplars (Carless & Chan, 2017). The analysis of exemplars was a powerful way of developing teachers' understanding of the nature of quality by producing accounts of strengths, weaknesses and how the task could have been done better.

Fourth, the use of modelling examples and video fragments from their own teaching practice facilitated the discussions between peers (in small groups) about the performance objectives of AfL and facilitated the process of giving and receiving *feedback* (Lynch et al., 2012). The use of these reflective discussions created a rich learning environment for teachers and facilitated *collaborative teacher learning* (Vangrieken et al., 2015). Also, *a team* of three teachers working in the same school attending the AfL-TPD program, which stimulated opportunities for feedback and reflection between the meetings on their own teaching practice by using these colleagues as peers (Smith, 2016).

Finally, the program spanned approximately 8 months, which could provide *ample time* to develop the basic teacher competencies needed for AfL. More time is needed to further develop these competencies in their own teaching practice (Darling-Hammond et al., 2017; Veen and Zwart 2012).

4.5.3 Improvement suggestions for the AfL-TPD program

The experiences of teachers and coaches with the AfL-TPD program were predominantly positive, although suggestions for improvement were also mentioned (Figure 4.4). First, the use of a video-based professional learning platform should support providing peer-to-peer feedback to allow reflective discussions online between the meeting sessions, and to engage teachers as learning resources for one another. After all, there is strong scientific evidence that feedback can enhance learning (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Second, coaches could provide more variety in the modelling examples

used (e.g., not just videos from language lessons), and stimulate effective translation of the examples to other subject domains. Therefore, teachers need curriculum content knowledge and pedagogical content knowledge to make the transfer to their own subject matter (Mandinach & Gummer, 2016). Third, the coaches could demonstrate the use of AfL more explicitly in their teaching and assignments; for example, the program could be more targeted by determining success criteria for the assignments in advance, and by developing an understanding of the features of quality through discussing exemplars (Carless & Chan, 2017; Hamodi et al., 2017). The coaches should not operate on a one-size-fits-all basis, but should differentiate professional development activities deliberately so that teachers receive instruction matching their needs (George, 2005). Finally, teachers suggested involving school leaders in the AfL-TPD program, because school leaders need to support their teachers in broadening AfL at their own school (Heitink et al., 2016; Schildkamp et al., 2020; Smith & Engelsen, 2013).

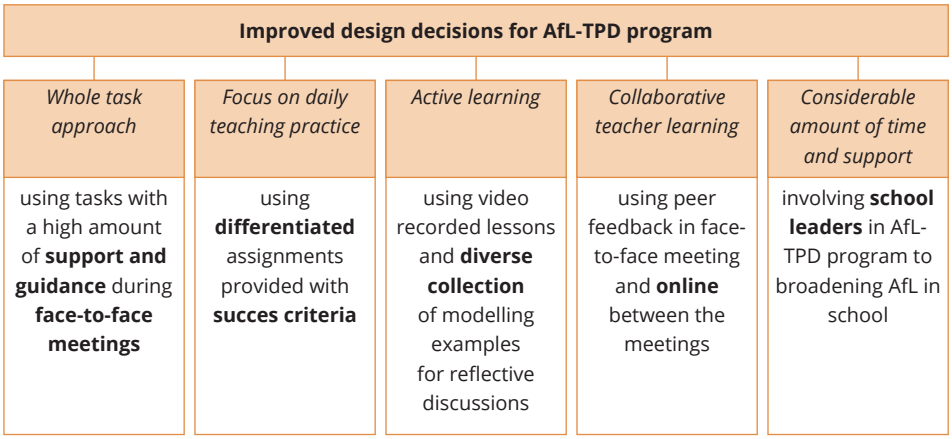


Figure 4.4 Improved Design Decisions

4.5.4 Limitations and implications for practice and research

Several limitations of the study must be mentioned. First, it is important to emphasize that in this study, a small group of 10 motivated teachers participated in the AfL-TPD program and developed their competencies. The context for future professional development programs might be different; for example, scaling up to larger groups might result in less motivated participants, less involved coaches and might result in less successful implementation of AfL in teaching practices. Participating in the AfL-TPD program can be demanding in combination with a fulltime teaching job that might already be overwhelming. In this study, 12 teachers started in the program, two of whom left the program after the second meeting because they felt the AfL-TPD program was too demanding and could

not be combined with their teaching job. Second, teacher self-report data were collected by means of a questionnaire and focus group interviews. We analysed the effects on teacher learning by using teachers' and coaches' perceptions, which could have been biased; for example, the teachers and coaches in the program might have been inclined to respond positively because they had put a lot of effort into it. Although self-reports have disadvantages, in this qualitative study teachers could provide insight into their own goals, thought processes, knowledge and beliefs, and therefore it seems important to involve teachers in the evaluation process (Goe et al., 2008).

In this small-scale study, we were the first to use 4C/ID for the design of an AfL-TPD program. This use of specific design principles for a program provides valuable information and insights for teacher professional development. For example, the dialogic use of exemplars or the use of video-recorded fragments of lessons to stimulate reflective discussions on teaching practice can become a routine within school teams and stimulate a learning culture within the school. Students could also be involved in these reflective discussions of teaching practice, and student feedback might affect teaching quality (Bijlsma et al., 2019). The design principles illustrated in this study can serve as an example for other research studies and programs concerning professional development for complex teacher skills.

Despite the fact that teachers became more knowledgeable about AfL, they indicated that applying AfL was not yet a daily routine. Teachers pointed to the pitfall of easily falling back on their routines, due to the hectic pace of the school day. The AfL-TPD program enabled teachers to start applying AfL in their teaching practice, but learning and coaching should continue in the school, supported by students, colleagues and school leaders (Schildkamp et al., 2020). For example, the teachers who attended the AfL-TPD program can stimulate their own learning processes and transfer to other colleagues within the school using teacher collaboration in professional learning communities, in order to foster effective professional development for a longer period of time, 2 or 3 years (Prenger et al., 2019).

Further research could examine the effects of this AfL-TPD program on teachers' use of AfL in classroom practice and student achievement, using a larger-scale implementation of the program. The new skills can only lead to improved student achievement, the ultimate goal of professional development, if they are applied in practice (Guskey, 2002). An experimental research design could offer more certainty about the causality of the observed effects. Not only should teachers' and coaches' perceptions be considered, but also classroom observations and student performance, to determine to what extent teachers learned the AfL principles so as to apply them successfully in their lessons. It would also be interesting to study the effects of the AfL-TPD program on other aspects of student learning, for example, the effects of AfL on student self-regulation.

5

Effect of Teacher Professional Development for AfL on Students' Self Regulation of Learning

This chapter is based on

Wolterinck-Broekhuis, C. H. D., Poortman, C. L., Schildkamp, K., Visscher, A. J. (2022). *Strengthening Student Self-regulation: Investigating the Effect of Teacher Professional Development for Assessment for Learning.* Manuscript submitted for publication

Abstract

Assessment for Learning (AfL) has the potential of improving self-regulatory skills. However, AfL requires teacher professional development. In this experimental study, we investigated the effect of an intensive, eight-month AfL teacher professional development program on students' ability to self-regulate their learning (26 teachers and 289 students in the control group; 31 teachers and 301 students in the experimental group). The results of our study showed that students in the experimental group did not outperform students in the control group in terms of gains in their ability to self-regulate their learning. Possible explanations include a lack of implementation of all AfL strategies, a lack of attention in the professional development program for stimulating student agency in assessment practices, and teachers needing more time to learn to apply the complex assessment-for-learning skills in their own practice effectively.

Keywords: Assessment for Learning, formative assessment, self-regulated learning, metacognition, professional development, teacher education.

5.1 INTRODUCTION

Assessment for Learning (AfL) is a promising approach to improve student achievement. AfL is intertwined with teaching and learning processes; "...evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers to make decisions about the next steps in instruction, which are likely to be better, or better founded, than the decisions they would have made in the absence of the evidence that was elicited" (Black & Wiliam, 2009, p. 9). The practice of AfL involves establishing clear learning goals and performance criteria (success criteria), eliciting and interpreting evidence of learning, taking immediate or near-immediate pedagogical action based on evidence, and promoting students' active involvement in the assessment process (Wiliam, 2011). Students are stimulated to take an active role in AfL, assessing their own learning. By comparing their current learning status with the goal and assessment criteria they are able to make judgements about their goal attainment (Hattie & Timperley, 2007; Nicol & MacFarlane-Dick, 2006; Sadler, 1989). These processes can activate students' cognitive, and motivational capacities, focus students on their learning goals, and provide feedback and strategies that they can use to help them reach their goals (Panadero et al., 2018).

The active role that students need to adopt in AfL is consistent with the idea of self-regulated learning (SRL), which involves students as metacognitively, motivationally and behaviourally active agents in their own learning (e.g., Boekaerts & Cascallar, 2006; Zimmerman, 1989; Zimmerman & Schunk, 2001). Such active and self-regulated strategy use - including goal setting, defining tasks, planning, and self-reflection among other processes - has been found to be positively related to achievement (Pintrich & De Groot, 1990; Wang et al., 1993; Wolters & Pintrich, 1998; Zimmerman & Martinez-Pons, 1990). For this reason, attention to the relationship between AfL and various aspects of SRL has increased greatly (Andrade & Brookhart, 2016; Heritage, 2018; Panadero et al., 2018).

Unfortunately, the use of AfL in the classroom is not self-evident. One reason is that it requires a complex set of teacher skills (Schildkamp et al., 2020). Too often, teachers find it difficult to put AfL into practice (Kippers et al., 2018; Veugen et al., 2021) as a result of which the potential effects of AfL on the SRL skills of students are not realized. To effectively support teachers in using AfL (Heitink et al., 2016), a Teacher Professional Development (TPD) program was designed (Wolterinck et al., 2022). This program was designed using a task-centred approach to teacher learning that encourages the integration of knowledge, skills, and attitudes, and the coordination of various skills (Van Merriënboer & Dolmans, 2015).

Previous research has shown that such a task-centred approach to an Assessment for Learning teacher professional development program (AfL-TPD program) can result in

improvement in teachers' knowledge, skills and attitudes regarding AfL (Wolterinck et al., 2022). Teachers, for example, became more aware that AfL is a cyclical process, instead of just an instrument for testing without grades, in which teachers must use the strategies coherently together with students, in order to apply it effectively in their own classroom. In this study, the effects of a task-centred approach to AfL-TPD on students' self-regulated learning were investigated. The main question answered in this study is:
What is the effect of an AFL-TPD program on students' self-regulation of learning?

5.2 THEORETICAL FRAMEWORK

5.2.1 Assessment for Learning

AfL, if implemented coherently, can result in higher student achievement (Black & Wiliam, 2003; Kingston & Nash, 2011, 2015). Through AfL, teachers find out what students know, what they partly know and what they do not yet know compared with the learning goals of a lesson. This information can better enable teachers to adapt follow-up activities to the needs of students, which can advance student achievement (Wiliam & Thompson, 2008). Several core strategies for implementing AfL in classroom practice can be distinguished (Black & Wiliam, 2010).

First, teachers have to clarify, share and help students understand the learning goals and criteria for success. This strategy focuses on the students to help them really understand 'where am I going' as a learner, as this will help them in developing plans to attain the expected goals (Black & Wiliam, 1998b; Crisp, 2012; Heitink et al., 2016; Wiliam & Leahy, 2015) The term *learning goals* signifies what the teacher wants the students to learn, whereas *success criteria* signifies the criteria used by the teacher to check whether the students were successful in the learning activities in which they were engaged. Teachers have to share their expectations of students in a meaningful way, helping students to internalize these standards as their own (Brandmo et al., 2020).

Second, teachers have to engineer effective discussions, activities, and learning tasks that *elicit evidence of learning*. This strategy involves generating opportunities to effectively gather evidence of student learning through (predominantly informal) assessment, such as student observations and classroom conversations (Black & Wiliam, 1998b; Crisp, 2012; Gottheiner & Siegel, 2012; Heitink et al., 2016; Wiliam & Leahy, 2015). The teacher together with the students, for example, checks whether the concepts explained in the lesson are understood and uses this information to move learning forward. Students can use assessments to gather information about 'where am I now', as this can help students to continually monitor and review their work against the set standards.

Third, a key component of AfL is providing feedback, based on the gathered evidence and focused on 'how am I getting there', which assists moving forward with learning and stimulates students to think about their learning. The term *feedback* signifies the information provided regarding aspects of students' performance or understanding while they are learning (Hattie & Timperley, 2007; Wiliam & Leahy, 2015). For example, feedback is given in terms of comments that address what the students need to improve, what they need to do and how (Wiliam & Leahy, 2015). Students can use each form of feedback to plan and adapt their efforts to learn, especially feedback at the level of self-regulation (Hattie & Timperley, 2007). This feedback, for example, prompts students to reflect on their performance of the task, or focuses on helping students to make adequate follow-up choices to improve their performance. Self-regulatory feedback is only effective if it convinces students that more attention and effort will lead to the desired result (Hattie & Clarke, 2018).

Fourth, peer- and self-assessment is about the students taking ownership of their own learning (Wiliam & Leahy, 2015). The term *peer-assessment* signifies the type of assessment in which students are activated as instructional resources for one another: assessing each other's work and providing feedback to improve it (Topping, 2017; Wiliam & Leahy, 2015). In order to help students to become better learners, they can be given the opportunity to play an active role such as engage in peer-feedback activity focused on students in classrooms learning from one another (Van der Kleij et al., 2019; Wiliam & Leahy, 2015). The strategy of *self-assessment* focuses on the ability of students to reflect on their own learning by comparing their work and processes against standards, expectations, targets, or goals (Panadero et al., 2016; Wiliam & Leahy, 2015). As students themselves use the data that result from peer- and self- assessments to monitor their learning and generate internal feedback from that comparison, they can become more aware of the development of their own competences and can make better informed decisions about appropriate follow-up choices (Nicol, 2020). This allows students to gain insight into the efficacy of the learning strategies they use; for example, students may decide to start directly with assessment in order to assess what they already know or decide to assess after they are finished with the tasks.

As teachers integrate AfL coherently in their classroom practices, they can allow room for students to regulate and redirect their learning efforts (Heritage, 2018). Students can be active participants in the process of AfL, and teachers can give students the opportunity to play this active role in their own learning process, through activities such as dialogue about the learning goals, sharing success criteria with students, effective questioning and peer-feedback focused on students in classrooms learning from one another (Black & Wiliam, 1998; Bryant & Carless, 2010; Crisp, 2012; Harris & Brown, 2013; Van der Kleij et al., 2019; Wiliam & Leahy, 2015). Teachers who involve students in classroom assessment as jointly responsible for implementing the core strategies of AfL can help their students to

become better self-regulated learners (Clark, 2012; Panadero et al., 2018). When students engage in SRL, they are often more successful learners (Hattie & Timperley, 2007).

5.2.2 Self-regulation

SRL is increasingly considered an important student skill in secondary schools, which implies that students need to learn to set goals, take the initiative and learn to take responsibility for their own learning process (Istance & Dumont, 2010; Pintrich & Zusho, 2002). However, the process of self-regulation does not start by itself and students need to be facilitated. Recent studies on SRL training revealed that both primary and secondary school teachers spent little time explicitly teaching SRL strategies (Dignath & Büttner, 2018; Dignath & Veenman, 2021), although there is evidence that both primary school students and secondary school students benefit substantially from applying these strategies (see e.g., Dignath, Buettner, & Langfeldt, 2008; Hattie, 2008).

Self-regulation can be enhanced as teachers structure the learning environments of their students according to the principles of AfL and provide students with support and guidance in the use of AfL strategies (Nicol & MacFarlane-Dick, 2006). Positive effects of AfL on SRL can be expected because AfL emphasizes: (1) sharing learning goals and criteria for success in order to help students develop plans to attain goals, (2) the emphasis is on assessment in order to monitor where learners are with regard to the set standards, including peer- and self-assessment, and (3) feedback based on the assessment results can be used to adapt learning strategies to move closer to the desired goals. The students improve their self-regulation when they are joint stakeholders in assessment with teachers and peers (Bailey & Heritage, 2018). Self-regulated learners show an active and appropriate use of learning strategies, are metacognitively engaged as they learn, have strong self-efficacy beliefs, and, are motivated to learn (Zimmerman & Schunk, 2001). Pintrich and De Groot (1990) identified five key learning strategies. *Elaboration* and *critical thinking* are strategies that involve active understanding and construction of meaning. Elaboration helps students to build internal connections between items to be learned and prior knowledge; critical thinking includes the evaluation of subject matter while relating it to prior, personalized knowledge and standards of excellence (Garcia & Pintrich, 1996; Pintrich et al., 1993). For simpler learning tasks and the activation of knowledge, *rehearsal* strategies may be effective. *Organisation* is a supporting strategy and helps students to summarise the knowledge to be learned. These four strategies can be categorised as cognitive strategies, applied to learn and perform a task, that refer to information processing. *Metacognition*, comprising strategies such as planning, monitoring and adapting, indicates students' awareness and use of cognitive strategies; in other words, the self-regulated character of their learning. Metacognitive strategies might be more easily

taught directly to older children than to younger children, as older students, who have more advanced self-regulation capacity, can integrate new self-regulation strategies into their learning more easily (Dignath & Büttner, 2018).

Students' active regulation of their motivation has also been identified and described as an aspect of self-regulated learning (Wolters & Benzon, 2013). Self-regulated students are thought to have a set of adaptive beliefs and behaviours that motivate them to participate in and complete learning tasks (Wolters, 2003). These motivational beliefs, attitudes, and values are critical in SRL processes, because they can influence students' goal-setting and choices regarding task engagement, strategy selection, and planning (Wolters, 2011). Self-regulated students attach certain values to learning and school tasks and keep themselves motivated while learning. Several aspects of achievement motivation that students may use to regulate their motivation have been discussed in the literature, such as *goal orientation, task value, self-efficacy and test anxiety* (e.g. Boekaerts, 1997; Pintrich, 1999; Pintrich & De Groot, 1990; Wolters, 2003; Zimmerman & Martinez-Pons, 1990). In addition to learning strategies, the following beliefs and attitudes are most relevant, according to Pintrich and De Groot (1990): *intrinsic and extrinsic goal orientation*; the reason why one undertakes a task, *task value*; the degree to which the task is considered as relevant, important and worthwhile, *self-efficacy*; the students' belief in his or her ability to successfully complete the task, and *test anxiety*; related to students' perception of competence.

5.2.3 Design of the professional development program

AfL is a complex skill requiring multiple teacher competences, and professional development should explicitly show teachers how all five AfL strategies can be coherently integrated in their classroom practice, in order to maximize its impact (Lee, 2011). The AfL-TPD program was designed using the four-component instructional design model (4C/ID) for complex learning, consisting of four interrelated components: *learning tasks* - a sequence of varying and increasingly difficult complex assignments; *supporting information* - the domain knowledge one needs for carrying out tasks; *procedural information* - the information one needs to know about the how-to routine aspects of the task; *practicing part-tasks* – automating routines for task execution (Van Merriënboer & Kirschner, 2017; Wolterinck et al., 2022).

The program was developed using a whole-task approach, ordering authentic learning tasks from simple to complex, in combination with a gradual decrease in learner support at each level of learning task complexity, and distinguishing between non-routine and routine aspects of the complex skill (Van Merriënboer et al., 2006; Wolterinck et al., 2022). The learning tasks are based on the results of a cognitive task analysis (CTA; Clark et al., 2008) of teachers who are AfL-experts, conducted in a previous study which provided an overview of all constituent skills and the relationships between these skills, visualized in

a skills hierarchy (Wolterinck et al., 2021). Several complexity-related factors were used as focal points for organizing learning tasks according to their complexity, for example, classroom student group composition, the content of the lesson, and the availability of remedial curriculum material (Wolterinck et al., 2021).

The program provided many opportunities for practice and experimentation by applying learning tasks in daily teaching practice, enabling the integration of the necessary knowledge, skills and attitudes (Van Merriënboer & Kirschner, 2017). For teachers using AfL in lessons, this meant that they had to prepare a lesson series, prepare lessons, enact lessons and evaluate lessons, in which all constituent skills were supposed to be present (Wolterinck et al., 2022). The necessary teacher knowledge for using AfL successfully in the classroom was also invoked, such as domain knowledge, pedagogical content knowledge, knowledge of students' previous learning, and knowledge of assessment (Heritage, 2007; Wolterinck et al., 2022).

The program stimulated active learning using exemplars and video-recorded lessons in reflective conversations, demonstrating expert strategies and stimulating reflective skills (Carless & Chan, 2017). The use of reflective discussions between peers (teacher participants) facilitated the process of giving and receiving feedback and stimulated collaborative teacher learning. The TPD program provided substantial amount of time, to develop the basic teacher competencies needed for AfL at the teacher and student level, which was expected to result in a positive effect on students' SRL, as explained in the previous section. Figure 5.1 shows a schematic overview of the essential building blocks of the AfL-TPD program.

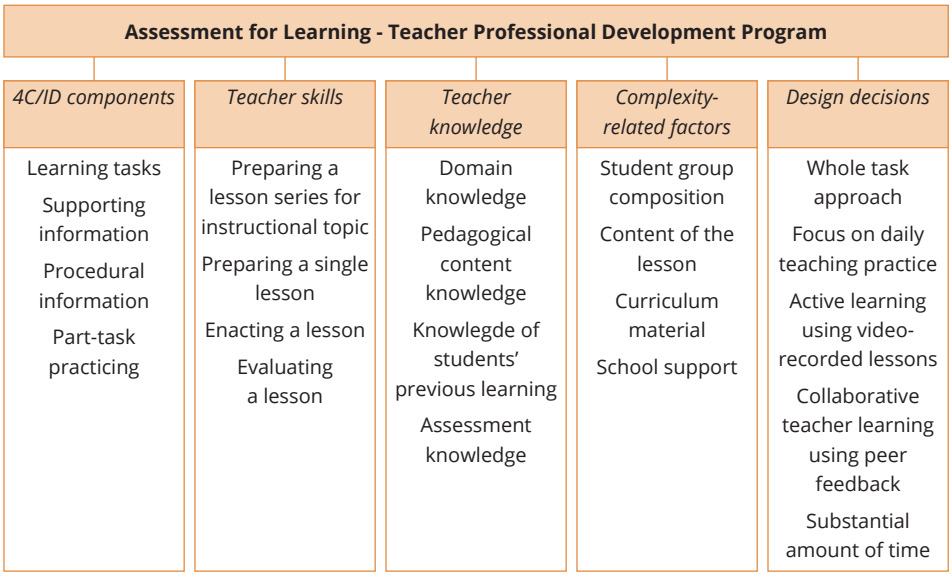


Figure 5.1 Essential Building Blocks of the AfL-TPD Program

5.3 METHODS

This study employed an experimental research design to investigate the effect of the AfL-TPD program on students' ability to self-regulate their learning. Data for this study were gathered from schools for secondary education in the Netherlands from August 2019 to July 2020.

5.3.1 Participants

Schools

A total of 29 schools participated in this study. These schools were recruited through convenience sampling (e.g., schools who had approached the trainers and researchers previously on the topic of AfL). The schools were randomly assigned to either the experimental condition (N = 14) or the control condition (N = 15).

Teachers

Each school in the experimental group participated in the AfL-TPD program with three teachers, one from each of three subjects (English language, Dutch language, and chemistry), with a total of 42 teachers. Participating schools in the control condition similarly designated three teachers each, one from each of those subjects, with a total of 45 teachers. The experimental group followed the TPD program on AfL from September 2019 to June 2020. Due to circumstances such as illness, lack of time, or excessive workload, nine teachers from the control condition and one teacher from the experimental condition dropped out before the pre-test period, in which the students' questionnaire was administered. An additional number of 10 teachers from the control condition and 10 teachers from the experimental condition dropped out before the post-test period, largely due to the extra (work) pressure that coincided with COVID-19, the pandemic that broke out in the middle of our study. The remaining teachers participating in the experimental (N = 31) and control (N = 26) groups were comparable in terms of age and teaching experience.

Students

Each teacher in the experimental condition was asked to focus on AfL in one of their classes for this study, as it would be too difficult to develop AfL in all of their classrooms simultaneously. Each teacher in the control condition was asked to choose one of their classes for participating in this study. Students in the chosen classes totalled 778 in the experimental condition and 587 in the control condition. The high dropout rate for teachers in both the experimental condition and the control condition resulted in the loss of 56.8% of the participating students (experimental condition N = 477, control condition

N = 298). The characteristics of the remaining students in the experimental condition (*N* = 301) and the control condition (*N* = 289) can be found in Table 5.1. Chi-square tests of independence showed that some statistically significant differences existed between the experimental group and the control group with respect to subject area, school track and grade level (Table 5.1), although these differences were small.

Table 5.1 Description of the Sample by Condition

	Experimental *N* (%)	Control *N* (%)	All *N* (%)	Chi-square values
School track:				10.97*
Vocational education	-	10 (3.5)	10 (1.7)	
Higher general education	107 (35.5)	106 (36.7)	213 (36.1)	
Pre-university	194 (64.5)	173 (59.9)	367 (62.2)	
Subject:				15.31*
Chemistry	91 (30.2)	57 (19.7)	148 (25.1)	
Dutch language	135 (44.9)	121 (41.9)	256 (43.4)	
English language	75 (24.9)	111 (38.4)	186 (31.5)	
Grade level:				12.94*
Low [a]	66 (21.9)	102 (35.3)	168 (28.5)	
High [b]	235 (78.1)	187 (64.7)	422 (71.5)	
Student gender:				0.42
Female	167 (55.5)	168 (58.1)	335 (56.8)	
Male	134 (44.5)	121 (41.9)	255 (43.2)	

5.3.2 Instruments

Self-regulated learning: Motivated Strategies for Learning Questionnaire (MSLQ)

The Motivated Strategies for Learning Questionnaire (MSLQ, Pintrich et al., 1993) is a standardized instrument that is used to measure students' self-regulation. It can be used to examine students' cognitive and metacognitive engagement in their own learning and the motivational components that are closely linked with cognitive engagement (Pintrich et al., 1993). The MSLQ has been used to study college students in a wide variety of countries, settings and populations (Credé & Phillips, 2011).

We used the Dutch translation of the MSLQ that was developed and validated for students in Dutch secondary education (Blom & Severiens, 2008). The questionnaire included a learning strategies section with five scales, see Table 5.2. Students were asked to respond to the statements in the 56-item questionnaire on a 7-point Likert scale (1 = not

at all true for me to 7 = very true for me). The intention was that students would fill in the questionnaire before and after the intervention period. Due to COVID-19 and the fact that we had to move the last TPD session to September of the next school year, students filled in the questionnaire after the penultimate TPD session.

With the data from the current study, we were able to re-establish the reliability of most of the scales, which ranged from .61 to .88 (Cronbach's alphas). Three scales had modest reliabilities and should be considered with care: elaboration (.62), critical thinking (.61), and extrinsic goal orientation (.65). One scale, intrinsic goal orientation (0.50), proved to be insufficiently reliable and was left out of the analyses. For the constructs and their characteristics, see Table 5.2.

Table 5.2 Reliability of the Scales in the MSLQ

Scale	Number of items	Cronbach's alpha	Example items
Learning			
Rehearsal	5	.78	When I study for this class, I practice saying the material to myself over and over.
Elaboration	6	.62	When I study for this class, I pull together information from different sources, such as lectures, readings, and discussions.
Organization	6	.80	When I study the readings for this course, I outline the material to help me organize my thoughts.
Critical thinking	5	.61	When a theory, interpretation, or conclusion is presented in class or in the readings, I try to decide if there is good supporting evidence.
Metacognition	6	.67	Before I study new course material thoroughly, I often skim it to see how it is organized.
Motivation			
Intrinsic goal orientation	4	.50	In a class like this, I prefer course material that arouses my curiosity, even if it is difficult to learn.
Extrinsic goal orientation	6	.65	Getting a good grade in this class is the most satisfying thing for me right now.
Task value	4	.74	I am very interested in the content area of this course.
Control of learning beliefs	5	.67	If I try hard enough, then I will understand the course material.
Self-efficacy	7	.88	I'm confident I can do an excellent job on the assignments and tests in this course.
Test anxiety	3	.71	When I take a test, I think about how poorly I am doing compared with other students.

Intervention

The AfL-TPD program was delivered on a subject-specific basis (English language, Dutch language and chemistry) and had a study load of approximately 40 hours, which included 20 hours of contact time in 5 meetings, and 20 hours of practice time spent on carrying out learning tasks in daily teaching practice. The sessions, 4 hours each, were initially planned over 8 months, from October to April. However, due to COVID-19, the final session was rescheduled for September (in the next school year).

The program included one introductory meeting and five consecutive meeting sessions, each emphasizing a different AfL skill. The introductory meeting consisted of activities focused on the activation of prior knowledge (e.g., explanation of the concept "AfL") and theory and examples related to the topic of the first session (learning goals and success criteria). This way, we could activate the cognitive schemas of teachers, which is necessary for linking new knowledge appropriately (Van Merriënboer & Kirschner, 2017). In this meeting session, teachers were also shown the AfL-skills hierarchy and were subsequently asked which five topics they would like to emphasize during the TPD program. Based on this data, we focused on the following topics during the TPD program: 1) sharing learning goals and success criteria (lesson execution); 2) gathering information (preparing a lesson); 3) analysing and interpreting information and stimulating students' learning processes (lesson execution); 4) analysing students' learning processes and reflection (lesson evaluation); 5) determining objectives (preparing a lesson series).

All sessions included some theory and examples of the emphasized AfL skills, followed by extensive reflective discussions in smaller groups about teachers' own recorded lessons. Teachers were asked to give each other feedback on their AfL skills, and also to reflect on their own development in AfL using a "review – reflect – revise" protocol. Between the sessions, teachers were asked to read the theory on a particular AfL skill (Wiliam & Leahy, 2018), and to record part of a lesson in relation to this theory. Teachers needed to upload this recorded lesson to a peer feedback system, in which they were asked to peer review a recorded lesson by another teacher, and to reflect on their own lesson based on peer reviews and improve themselves. By using teachers' own classroom practice as the context for their learning tasks implementing AfL, the TPD program gave them the opportunity to work on the development of their teacher skills every day anew.

5.3.3 Analysis

The data include the results of a student self-report questionnaire. To evaluate the effect of the intervention on students' self-regulation of learning scores a multilevel analyses was performed with SPSS, version 26, and supplemented with the means along with the standard errors and 95% confidence intervals for each variable. Random assignment

of participants to experimental and control conditions was successful; there were no statistically significant differences between the experimental and control group with respect pre-test scores for any of the MSLQ scales. The measures were nested within students, who were nested within classes, and those classes were nested within schools. Therefore, the student responses were analysed with a multilevel linear mixed model. The following fixed factors were included in the model consistently: pre-test scores as covariates and post-test scores as dependent variables. Preliminary assumption-checking revealed that the data were normally distributed.

5.4 RESULTS

Table 5.3 presents the means along with the standard errors and 95% confidence intervals for each variable (except for intrinsic goal orientation, which was omitted due to low scale reliability). The results of the multilevel analysis showed no significant effects of the intervention on students' self-regulation as measured by the five variables in the learning strategies section: rehearsal [$F(1, 586) = .96$, $p = .328$], elaboration [$F(1, 563) = .05$, $p = .827$], organization [$F(1, 552) = .24$, $p = .624$], critical thinking [$F(1, 570) = 3.24$, $p = .072$], metacognition [$F(1, 552) = 1.70$, $p = .193$].

No significant effects of the intervention on students' self-regulated learning as measured by the five analysed variables in the motivational strategies section were found: extrinsic goal orientation [$F(1, 542) = .45$, $p = .502$], task value [$F(1, 552) = 1.15$, $p = .285$], control of learning beliefs [$F(1, 548) = 1.40$, $p = .270$], self-efficacy [$F(1, 548) = .64$, $p = .423$], test anxiety [$F(1, 548) = .23$, $p = .634$]. Intrinsic goal orientation was not analysed, due to low scale reliability.

Table 5.3 MSLQ Pre-test and Post-test Scores, by Condition

Scales		Condition	Pre-test Scores			Post-test Scores		
			M (*SD*)	Min	Max	*M* (*SD*)	Min	Max
1.	Rehearsal	Exp (*N* = 301)	4.93 (1.18)	1.40	7.00	4.90 (1.09)	1.60	7.00
		Control (*N* = 289)	4.98 (1.18)	1.40	7.00	5.00 (1.14)	1.00	7.00
2.	Elaboration	Exp (*N* = 300)	4.21 (0.93)	1.50	6.50	4.52 (0.85)	2.00	6.83
		Control (*N* = 285)	4.34 (0.99)	1.33	6.83	4.58 (0.96)	1.50	6.67
3.	Organization	Exp (*N* = 296)	4.37 (1.14)	1.43	7.00	4.51 (1.09)	1.00	6.71
		Control (*N* = 280)	4.28 (1.15)	1.14	6.86	4.45 (1.14)	1.14	7.00
4.	Critical Thinking	Exp (*N* = 301)	3.56 (1.00)	1.00	6.80	3.64 (1.02)	1.00	6.80
		Control (*N* = 286)	3.55 (1.08)	1.00	6.40	3.77 (1.04)	1.00	6.20
5.	Metacognition	Exp (*N* = 296)	4.10 (1.05)	1.00	6.67	4.17 (0.87)	1.50	6.67
		Control (*N* = 280)	4.19 (1.04)	1.67	7.00	4.31 (0.91)	1.00	6.67
6.	Extrinsic goal orientation	Exp (*N* = 296)	5.36 (1.05)	1.75	7.00	5.79 (1.33)	1.78	7.00
		Control (*N* = 280)	5.49 (1.03)	2.25	7.00	5.93 (1.35)	1.25	7.00
7.	Task value	Exp (*N* = 296)	3.95 (1.23)	1.00	7.00	4.03 (1.11)	1.33	7.00
		Control (*N* = 280)	3.93 (1.28)	1.00	7.00	4.11 (1.21)	1.00	7.00
8.	Control of learning beliefs	Exp (*N* = 296)	5.16 (0.97)	1.75	7.00	5.14 (1.06)	1.75	7.00
		Control (*N* = 280)	5.21 (1.06)	1.25	7.00	5.21 (1.10)	2.00	7.00
9.	Self-efficacy	Exp (*N* = 296)	5.04 (0.99)	2.00	7.00	5.06 (1.05)	1.14	7.00
		Control (*N* = 280)	5.03 (1.04)	2.43	7.00	5.08 (1.07)	1.57	7.00
10.	Test anxiety	Exp (*N* = 296)	4.40 (1.45)	1.00	7.00	4.29 (1.43)	1.00	7.00
		Control (*N* = 280)	4.29 (1.48)	1.00	7.00	4.27 (1.48)	1.00	7.00

5.5 CONCLUSION AND DISCUSSION

In this study, we investigated the effect of an AfL-TPD program on students' ability to self-regulate their learning. Although we did not find an effect, it is important to learn from the results of this study. Therefore, in this section we offer a number of potential explanations for why the expected effects of a TPD program for AfL on students' self-regulated learning were not observed in this study.

5.5.1 COVID-19 school closures and distance education.

Schools were closed for multiple months as a result of COVID-19, meaning that students received only distance education via online communication, and that the last session of

the TPD program could not take place as scheduled. Teachers faced a new challenge giving online lessons, and using AfL in online lessons might have been an even bigger challenge for them. Consequently, teachers probably had less opportunity to stimulate students to engage in AfL activities during the COVID-19 crisis, and students might have had fewer opportunities to develop their SRL skills.

5.5.2 More practice needed to develop AfL skills and SRL skills.

Another explanation might be that it takes more practice for teachers to develop their AfL skills, and consequently more practice for students to develop SRL skills and automate strategy use (Dignath & Büttner, 2008). Teachers might not yet have had enough opportunity in this relatively short period of time (one school year) to develop their own AfL skills and thereby to develop students' use of learning strategies and motivational strategies in order to self-regulate their learning, especially in times without regular lessons.

5.5.3 Insufficient involvement of students in AfL practices.

Teachers starting to implement AfL may choose not to involve students in the assessment process (Panadero et al., 2016) and may make insufficient use of peer- and self-assessment in their lessons (Kippers et al., 2018). Teachers may find it more difficult to use these strategies in their mainly teacher-centred teaching practices and focus primarily on other AfL practices, such as clarifying and sharing assessment criteria, designing learning situations, and providing feedback (Jonsson et al., 2015). For AfL processes to be effective, teachers need to involve students actively in generating, processing and responding to feedback information (Carless, 2020), in order to give students the opportunity to develop the capacity to self-regulate their learning (Panadero, et al., 2016). Self-assessment and task selection are important self-regulated learning skills for secondary school students, but it is very difficult to teach students these complex skills (Nugteren et al., 2018).

Another explanation might be that the AfL-TPD program itself did not focus enough on the active involvement of students in their own learning process, and teachers can be asked more explicitly to include these skills as part of the learning tasks. Only when teachers recognize the importance of student agency, AfL processes can reach their full potential so that students can become self-regulated learners.

5.5.4 Need for multiple approaches to assessing SRL.

Of the SRL instruments currently available, self-report instruments are most frequently used (Roth et al., 2016). Likewise, in this study we used an existing self-report instrument (MSLQ) to assess the use of SRL strategies. Self-report instruments are often criticized regarding the extent to which they reflect actual behaviour (Artelt, 2000). For example,

the validity of questionnaires is influenced by students' ability to relate questions about strategies to their own learning experiences and to analyse the strategic aspects of their learning in a conscious manner (Spörer & Brunstein, 2006). Students in secondary education might not have fully developed such abstract thinking skills (Roth et al., 2016).

Other approaches to assessing SRL and its underlying processes, for example, the development and use of learning diaries, might be preferred in order to resolve persistent problems regarding the validity of SRL measurement (Roth et al., 2016; Veenman, 2011). A new trend in the measurement of SRL is the use of a combination of measurement and intervention, in which the tool used for measurement is also part of an intervention to promote the regulation of learning (Panadero et al., 2016). The use of these multimodal data may offer unobtrusive and scalable measurement solutions, supporting SRL with personalized visualisations and showing how learners regulate their learning over time (Molenaar et al., 2021).

5.5.5 Limitations and implications for research and practice

We acknowledge a number of limitations of our study. We faced a high dropout rate for the teachers (37.9%) and consequently also for the students (56.8%), most probably due to the consequences of COVID-19 in schools. Some statistically significant differences existed between the experimental group and the control group with respect to subject area, school track and grade level, although these differences were small. Moreover, as a result of COVID-19 we were not able to complete the entire professional development program. Furthermore, we examined the effects of an AfL-TPD program on students' ability to self-regulate their learning, using students' perceptions. These students' perceptions were planned to triangulate with classroom observations, to determine to what extent teachers learned the AfL principles and applied them successfully in their lessons. Due to COVID-19 the classroom observations were not feasible.

Although the expected effects of a TPD program for AfL on students' self-regulated learning were not observed in this study, it is important to avoid publication bias (Green et al., 2006), and it is important to learn from less successful interventions (Stoll & Myers, 1998; Stringfield, 1995). Our study shows how difficult it is to improve student agency in assessment practices, even when teachers are supported in this by TPD. Not only students' actual strategy use is important, but also students' knowledge about the usefulness of SRL strategies (Why is it beneficial to do this?). Therefore, teachers' knowledge and beliefs about promoting SRL may be contingent for AfL processes reaching their full potential and students becoming self-regulated learners (Brandmo et al., 2020). The AfL-TPD program might need to be focused more on student involvement in AfL practices and on the development of students' self-regulatory skills to find the expected effects (Heritage,

2018). For example, the reflection and feedback sessions in the AfL-TPD program could be focused more explicitly on what teachers do to promote SRL. The AfL-TPD program should provide teachers with guided and sustained support as they apply this new knowledge to practice. In this way, teachers can experience SRL and AfL just as, ideally, their students will experience them due to these innovations in classroom practice.

6
Conclusion and Discussion

6.1 INTRODUCTION

Assessment for Learning refers to forms of assessment focused on practices in which teachers and students continually gather information about where students are in their learning process. This information is used for the purposes of adapting teaching to students' learning needs and giving feedback to students about how to go forward with respect to their learning strategies (Klenowski, 2009; Van der Kleij et al., 2015; Wiliam, 2011). AfL is a complex skill and teachers find it difficult to implement AfL in their classrooms; as a result, AfL does not always lead to improved student achievement (e.g., Bennett, 2011; Heitink et al., 2016). There is considerable room for teachers to improve their knowledge, skills and attitude with regard to AfL (e.g., Kippers et al., 2018; Schildkamp et al., 2020; Veugen et al., 2021). Through professional development (PD), teachers can improve these competencies and implement AfL effectively in their classrooms.

This brings us to the overarching question that guided this dissertation: *How can teachers be supported in developing their Assessment for Learning competencies?* This final chapter addresses this question, starting with a summary of the findings for each study independently (6.1), followed by an overall conclusion and discussion of the findings (6.2). Final considerations address implications for practice and future research based on the research presented in this dissertation (6.3).

6.2 SUMMARY OF THE FINDINGS

6.2.1 Student perceptions of the use of Assessment for Learning in the classroom

The first research question (Chapter 2) that was addressed in this dissertation was: *To what extent do students experience Assessment for Learning strategy use in their classroom?* To answer this question, a student questionnaire was administered in 12 secondary schools. In total, 685 students from the fourth and fifth grade levels of senior general secondary education filled in the questionnaire. The questionnaire included four separate, independent scales based on the key AfL strategies, as outlined in the international literature on formative assessment (Black & Wiliam, 2010): (1) sharing learning goals and success criteria, (2) eliciting evidence, (3) feedback, (4) peer- and self-assessment. Students were asked to complete the questionnaire for a single subject area, either English language or mathematics, in order to examine the perceived extent of AfL practices in these lessons.

The findings showed that there was no difference between the subjects. Students experienced sharing learning goals and criteria for success and eliciting evidence of students' learning progress as the most frequently used strategies in their lessons. They indicated that feedback and peer- and self-assessment were the least often used strategies, although these strategies make AfL more explicit for students, and therefore more visible and noticeable. This calls for partnership, shared responsibility and interplay between teachers and students and between students (Carless & Winstone, 2020; Panadero et al., 2018).

From the results of this study, we can conclude that there is considerable room for improvement concerning the integration of AfL in daily classroom activities. The cluster analysis performed in this study revealed three different clusters showing how students experience the extent of implementation of AfL in their English language and mathematics lessons: a cluster *'high'* in which students reported the use of all four AfL strategies as between emerging and established (50-74% of the lessons); a cluster *'middle'* in which the experience of strategy use was more divided, from (almost) never to established (0-74% of the lessons); and a cluster *'low'* in which students reported that their experience of three of the four AfL strategies was between (almost) never and sporadic (0-24% of the lessons). Sharing learning goals and success criteria was experienced more frequently in this last cluster, between sporadic and emerging (25-49% of the lessons). The cluster analysis showed that the use of sharing learning goals and success criteria was largely addressed in all three clusters. The use of the other three strategies varied. This knowledge may be used when designing a teacher PD program. Taking into account teachers' needs by means of their cluster memberships can enable differentiation within the TPD program.

6.2.2 Cognitive task analysis of the competencies Assessment for Learning requires from teachers

The two research questions that were addressed in the next study (Chapter 3) were: *What knowledge, skills and attitudes does a teacher need to be able to implement AfL in the lesson? and What factors influence the complexity of AfL?* To answer these questions, a cognitive task analysis was carried out, taking expert teacher behavior as a starting point, using classroom observations and interviews with teachers and content experts for data collection. A total of 16 teachers divided over three secondary school subjects (English language, Dutch language and chemistry), and considered to be 'AfL experts' in their teaching practice participated in this study. An overview of the skills teachers need for AfL was presented in a skills hierarchy and complemented by formulating performance objectives. These indicators specify desired task performance. The results of this study showed that – in line with many other aspects of effective teaching – teachers should not

only focus on executing the lesson using AfL strategies, but also need to prepare the entire lesson series beforehand, prepare the lessons within this series, and evaluate lessons to make AfL successful. The insights gained are a valuable contribution to the AfL knowledge base and offer a basis for developing teacher professional development trajectories aimed at equipping teachers with the skills for using AfL in the classroom.

In addition, domain-specific knowledge and pedagogical content knowledge are essential for teachers. To be able to tune AfL to the instructional goals, teachers need knowledge of students' previous learning in relation to the curriculum and the learning goals. In particular, knowledge of students' misconceptions is essential to be able to determine gaps in students' understanding of a particular curriculum topic. Moreover, five factors that influence the complexity of applying AfL in the classroom were identified: (1) student group composition, (2) the extent to which the teacher has control during the lesson, (3) the level of difficulty of the lesson content (goal and topic), (4) the amount of available support, such as instructional material containing suggestions for remediation and remediation materials, and (5) the amount of school support, for example, collaboration with colleagues and professional development.

6.2.3 Developing a teacher professional development program for Assessment for Learning

The focus of Chapter 4 was teacher professional development. The two research questions that were addressed in this study were: *How relevant and usable did teachers find the Assessment for Learning professional development program that was designed on the basis of the 4C/ID model? And What knowledge, skills and attitudes needed for AfL did teachers learn as a result of participating in the AfL professional development program?* The four-component instructional design (4C/ID) model for complex learning was used to design a teacher professional development (TPD) program for developing teachers' AfL competencies. The AfL-TPD program was conducted on a subject-specific basis (English language, Dutch language, and chemistry) and in total ,12 teachers from four different secondary schools participated. The AfL-TPD program focused on workplace-based learning and used participants' daily lessons as learning tasks. These real-life whole tasks presented the full range of variability, required all of the necessary skills, and allowed daily opportunities for practising them. To study teachers' satisfaction with the TPD program and teachers' learning in terms of their development of the knowledge, skills, and attitude necessary for implementing AfL in their practice, a mixed-method approach was utilized. Teachers completed a questionnaire and also participated in focus group interviews for triangulation purposes. Coaches from the TPD program participated in a separate focus group interview (Cohen et al., 2013).

The study yielded valuable lessons about how to support teachers in learning to use AfL in their teaching practice. Using learning tasks in daily teaching practice made the content of the AfL-TPD program applicable and resulted in teacher satisfaction with the relevance and usability of the program. Using a whole-task approach helped teachers transfer newly learned skills to their professional task; in this way, teachers experienced the complexity of implementing AfL while they practiced doing so. Using exemplars and video-recorded lessons to demonstrate expert strategies and stimulate reflective skills supported active and collaborative learning by the teachers, and teachers gained knowledge, skills, and attitudes related to the effective use of AfL. The use of video-recorded lessons of their own teaching practice facilitated reflective discussions between peers (in small groups of three to four teachers) and gave opportunities for giving and receiving feedback. The video-recorded lessons also gave opportunities for comparison to generate inner feedback, which is the new knowledge that learners generate when they compare their current knowledge and competence against some reference information (Nicol, 2020). More generally, the study shows that the 4C/ID model can provide the basis in learning theory for the design of professional development programs for the acquisition of complex teacher competencies that require the integration of skills, knowledge and attitudes.

6.2.4 Effects of teacher professional development for Assessment for Learning on student self-regulation

The research question that was addressed in this fourth and final study (Chapter 5) was: *What is the effect of an AFL-TPD program on students' self-regulation of learning?* In an experimental study, the effect of an intensive, 8-month AfL teacher professional development program on students' ability to self-regulate their learning was investigated. Teachers who involve students in classroom assessment, showing that teachers and students are jointly responsible for implementing the core strategies of AfL, can help their students to become better self-regulated learners (Panadero et al., 2018). A total of 29 schools participated in this study and they were randomly assigned to either the experimental condition (N = 14) or the control condition (N = 15). In total, 31 teachers participated in the experimental condition and followed the teacher professional development program for AfL. Each teacher was asked to focus on AfL in one of their classes and the students (N = 477) in these classes were assigned to the experimental condition. A total of 26 teachers participated in the control group, and their students (N = 298) were assigned to the control condition. This implied that their teachers did not participate in the teacher professional development program for AfL. An online student self-report questionnaire, the Dutch translation of the Motivated Strategies for Learning Questionnaire (MSLQ), was administered as pre-test and post-test. To evaluate the effect of the intervention on students' self-regulation of learning,

the pre-test and post-test scores of students for the experimental and control condition were compared.

The students in the experimental group did not outperform students in the control group in terms of gains in their ability to self-regulate their learning. Although, according to the literature (Andrade & Brookhart, 2020; Heritage, 2018; Panadero et al., 2018), we can assume that AfL (ultimately) promotes the self-regulated learning of students, we were not able to confirm this. One possible explanation may be that the teacher PD was not sufficiently focused on the promotion of students' self-regulated learning during the process of AfL. The timespan may also have been too short for teachers to learn the new skills, apply them well enough in the classroom, incorporate repeated practice, and expect a subsequent effect on student performance. Another aspect that should be mentioned is that teachers who participated in this study had to deal with school closure and distance education due to the Covid-19 pandemic. Schools were closed for multiple months, meaning that students received only distance education via online communication. Teachers faced a new challenge giving online lessons, and using AfL in online lessons might have been an ever-bigger challenge for them. Consequently, teachers probably had less opportunity to stimulate students to engage in AfL activities during the COVID-19 crisis, and students might have had fewer opportunities to develop their self-regulated learning (SRL) skills. Not only the daily practice in the schools was strongly affected, but also the PD trajectory itself. One of the meetings had to be postponed until after the summer holidays and several teachers dropped out.

The study showed how difficult it is to improve student agency in assessment practices, even when teachers are supported in an AfL-TPD program. Not only students' actual strategy use is important, but also students' knowledge about the usefulness of SRL strategies (Why is it beneficial to do this?). Therefore, teachers' knowledge and beliefs about promoting SRL may be pre-conditions for AfL processes reaching their full potential and students becoming self-regulated learners (Brandmo et al., 2020). The AfL-TPD program might need to be focused more on student involvement in AfL practices and on the development of students' self-regulatory skills in order to find the expected effects (Heritage, 2018). For example, the reflection and feedback sessions in the AfL-TPD program could be focused more explicitly on what teachers do to promote SRL. The AfL-TPD program should provide teachers with guided and sustained support as they apply this new knowledge in their practice. In this way, teachers can experience SRL and AfL just as, ideally, their students will experience them due to these innovations in classroom practice.

6.3 CONCLUSIONS

6.3.1 Reflection on the outcomes of the research

Integrating AfL with daily classroom activities is a complex process and teachers need a particular set of competencies to work on this together with their students. Although the theory on AfL is rather well developed with regard to the important strategies in the AfL process (i.e., sharing learning goals and criteria for success, eliciting information on students' learning process, providing feedback to move learning forward and peer- and self-assessment), the results of this study provide a more in-depth practical framework for the implementation of the AfL process in the classroom. The idea that PD can foster improvements in teaching is widely accepted (e.g., Borko, 2004; Darling-Hammond et al., 2017; Desimone, 2009; Guskey, 2002). Despite this widespread agreement about its importance, precisely how PD fosters teacher learning and teaching practice remains elusive (Kennedy, 2016). In response to this challenge, Kennedy (2016) suggested focusing on the underlying theory of learning embraced in the teacher PD. The study presented in this dissertation shows that the 4C/ID model can provide the basis in learning theory for the design of AfL-TPD programs that require the integration of teacher skills, knowledge and attitude. In addition, three important insights were gained that can strengthen teacher PD for AfL.

Implementing AfL demands teacher competencies that cover all phases of instruction

The AfL process is complex and is considered to be a demanding task for many teachers (Cowie & Harrison, 2016). While implementing AfL in their classrooms, teachers must have knowledge about the four phases of instruction and the strategies within them, domain knowledge, pedagogical content knowledge and knowledge of students' misconceptions. Teachers also need to be aware of the factors that make implementing AfL more or less complex. Learning takes time, especially if tasks are complex. Teachers need to have the opportunity to experiment in a supportive context, for example, with support from school leaders who have a clear vision of AfL in the school.

To successfully implement AfL in the classroom, linking the skills teachers need and the different phases of instruction is beneficial. The extensive analysis of data from expert teachers, described in chapter 3, revealed that the competencies teachers need to implement AfL in their lessons cover all phases of instruction. In line with the framework for adaptive teaching by Parsons et al. (2018), implementing AfL requires teacher action during planning, in the midst of teaching, and when reflecting on their instruction. The different teacher skills necessary for AfL can be spread over these phases as depicted in Figure 6.1. In the planning phase, teachers prepare a lesson, based on the evaluation of the previous lesson and based on the preparation of a lesson series for an instructional topic.

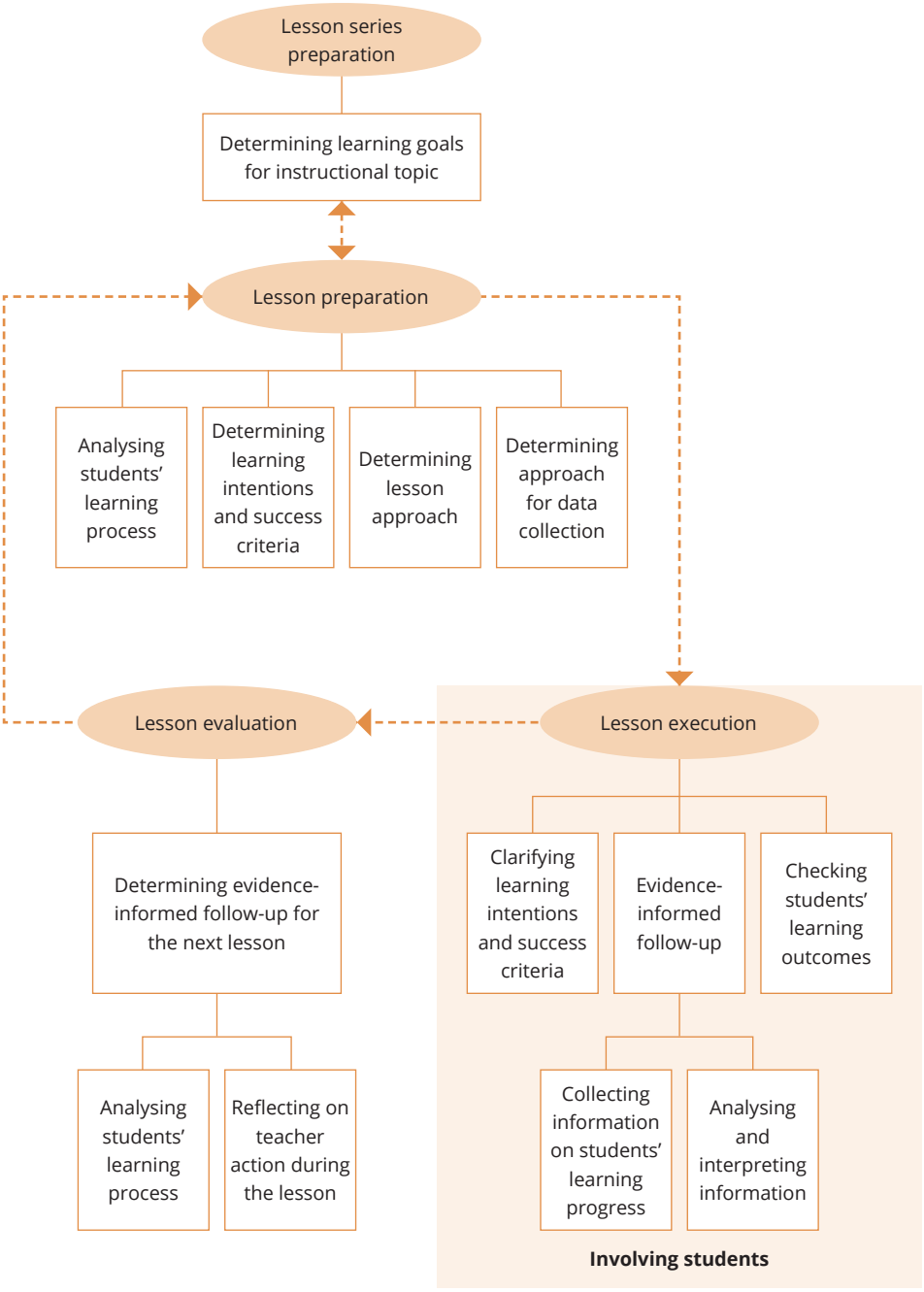


Figure 6.1 Teacher Competencies for Cyclical AfL Process Linked to Different Phases of Instruction

In the midst of teaching, teachers enact the lessons while tailoring instruction to students' learning needs, eliciting evidence of learning and encouraging students to be owners of their learning. During reflection on their instruction and students' learning process, teachers determine follow-up activities for the next lesson, based on the elicited evidence. Teachers use the output of the reflection in the planning phase for the following lesson.
Most models presenting the process of AfL are general models with a focus on the strategies (e.g., Gulikers et al., 2021; Ruiz-Primo & Li, 2013; Wiliam, 2011), and not on what this looks like during the different phases. Teacher PD for AfL can be strengthened when AfL is considered as a cyclical process in which the necessary teacher skills are practiced throughout all phases of instruction, and in which partnership, shared responsibility and interplay between teacher and students and between students among themselves is key.

Formal and explicit use of opportunities for comparison to generate (internal) feedback
The 4C/ID model is the basis for the design of an AfL-TPD program that requires the integration of teacher skills, teacher knowledge and attitude. Besides this basis in learning theory for the design of teacher PD, more characteristic features of PD such as teacher involvement, practice-based and collaborative learning were incorporated in the design as well. The use of collaborative reflective discussions in our design created a rich learning environment for teachers and facilitated acquisition of skill in applying feedback and reflection. Having teachers use video fragments taken from their own recorded lessons, analyse them and compare them with videos made by their peers can initiate an informal, spontaneous and natural feedback process (Nicol, 2020). Teachers generate internal feedback by comparing their current knowledge about 'good' teaching against the information generated from the video fragments, both their own fragments and those of their peers. This process of internal feedback is the new knowledge that learners generate when they compare their current knowledge and competence against some reference information (Nicol, 2020). These opportunities for comparison may be incorporated in teacher PD for AfL more formally and explicitly, to help teachers build their capacity to use their own processes of comparison, so they will be better able to plan, evaluate, develop and regulate their own learning in PD programs. PD has been argued to be more effective if teachers take part as a group (Cordingley et al., 2015; Desimone, 2009; Dunst et al., 2015; Timperley et al., 2007; Walter & Briggs, 2012; Wei et al., 2009). Most often, the requirement for collaboration is formulated as the need to work with multiple peers or a "community of practice". In our TPD program, a team of three teachers working in the same school attended the AfL-TPD sessions, to stimulate opportunities for feedback and reflection on their own teaching practice between the meetings by using these colleagues as peers (Smith, 2016).

Mobilisation of key stakeholders, the students, to become active learners

For AfL processes to be effective, teachers need to involve students actively in generating, processing and responding to feedback information (Carless, 2020), in order to give students the opportunity to develop the capacity to self-regulate their learning (Panadero et al., 2016). Teachers starting to implement AfL might not involve all students in the assessment process (Panadero et al., 2016) and might make insufficient use of peer- and self-assessment in their lessons (Kippers et al., 2018). The analysis of student perceptions of the use of AfL in the classroom, as described in chapter 2, revealed that peer- and self-assessment were the least often used strategies, although these strategies make AfL more explicit for students, and therefore more visible and noticeable. This calls for partnership, shared responsibility and interplay between teachers and students and between students (Carless & Winstone, 2020; Panadero et al., 2018). Teachers may find it more difficult to use these strategies in their mainly teacher-centred teaching practices and focus primarily on other AfL practices, such as clarifying and sharing assessment criteria, designing learning situations, and providing feedback (Jonsson et al., 2015). Therefore, teachers' knowledge and beliefs about promoting the role of students may be contingent for AfL processes reaching their full potential and students becoming self-regulated learners (Brandmo et al., 2020). Desimone and Wilson (2022, p.5) stated that PD frameworks should make more explicit the student role in their own learning: 'If the student changes what they know and do, this is much more likely to be permanent. If we are able to change students' ability to monitor, evaluate and improve their own work, this is likely to exponentially increase the power of instruction'.

Teacher PD for AfL can be strengthened when the role of the student in monitoring, evaluating and improving their own work and learning is acknowledged and students become active learners within the PD intervention. Teacher PD can embrace this component and support can be provided to teachers to learn to do this.

6.3.2 Reflection on the research method

The four empirical studies, which were described in chapters 2, 3, 4 and 5, took place within the context of Dutch secondary education. Although this offered the opportunity to thoroughly investigate these processes in one specific context, these exploratory findings cannot be generalized to other settings and populations (Miles et al., 2018; Yin, 2014). As with any study, although carefully chosen, each research method comes with certain disadvantages and requires certain disclaimers. For this dissertation, student perception data were collected by means of two questionnaires. For the study in chapter 2, students indicated how often different AfL strategies occurred in their lessons. For the study in

chapter 5, students' ability to self-regulate their learning was examined using students' self-perceptions. By using a survey, self-report data were collected and therefore some critical concerns need to be taken into account. These concerns are, among others, that students in secondary education might not have fully developed their abstract thinking skills (Roth et al., 2016) and their perceptions might be influenced by both student characteristics (e.g., ethnicity, student performance; Levy et al., 2003) and teacher variables unrelated to teaching efficacy (e.g., teacher popularity; Fauth et al., 2014). A known risk of surveys is also that the respondents overestimate or underestimate themselves, which might have been the case in chapter 5, where students had to indicate their ability to self-regulate their learning. Nevertheless, recent studies have illustrated how teachers and teaching can benefit from making use of formative student feedback (Rollett et al., 2021a) and how students' ratings can serve as valid measurements of teaching quality (Bijlsma et al., 2022).

The delivery of the AfL-TPD program took place over a period of a year. Teachers reported about positive effects of the intervention on their knowledge, skills and attitude towards AfL use in their lessons. Although, according to the literature, we can assume that AfL (ultimately) promotes students' self-regulated learning (e.g., Brandmo et al., 2020; Heritage, 2018; Panadero et al., 2018), we were not able to confirm this. The study described in chapter 5 focused on student outcomes. The students in the experimental group did not outperform students in the control group in terms of gains in their ability to self-regulate their learning. One possible explanation may be that the duration of the PD trajectory was too short for teachers to learn the new skills, apply them well enough in the classroom, incorporate repeated practice, and thereby accomplish an effect on student performance. Teachers who participated in this study indicated that applying AfL was not yet a daily routine. Teachers pointed to the pitfall of easily falling back on their old routines, due to the hectic pace of the school day.

The challenge of linking student achievement (or other types of outcomes for students) to any type of teacher professional development starts, as Yoon et al. (2007, p. 4) asserted, with the following line of reasoning. Professional development affects student achievement through three steps. First, professional development enhances teacher knowledge and skills. Second, improved knowledge and skills impact classroom teaching. Third, improved teaching raises student achievement. If one link is weak or missing, improved student learning cannot be expected. If a teacher fails to apply new ideas from professional development in classroom instruction, for example, students will not benefit. Future studies of the effect of professional development on both teachers and students may therefore address its direct effect on the quality of teaching and its indirect effect on

student achievement (Kennedy, 2016). Researchers may follow teachers beyond the end of the PD and monitor how they incorporate the new ideas into their ongoing practices. The monitoring of student learning may be performed even 1 or 2 years beyond the close of the PD itself.

6.4 RECOMMENDATIONS FOR PRACTICE AND FUTURE RESEARCH

The design principles applied in this study can serve as an example for other studies and TPD programs for the development of complex teacher skills: whole tasks in daily teaching practice stimulating active and collaborative learning in a substantial amount of time.

The program was developed using the 4C/ID model, including a *whole task approach with real-life learning tasks*. This can help teachers transfer newly learned skills to their professional task, their daily lessons in their own classrooms (Van Merriënboer & Kester, 2008), and in this way teachers experience the complexity of implementing AfL while they practice doing so. Teachers were actively engaged by means of practice-oriented tasks in which learning was stimulated, by connecting what they already knew or could do with what they had learned. The program stimulates *active learning* using exemplars and video-recorded lessons, demonstrating expert strategies and stimulating reflective skills. The use of modelling examples and video fragments from their own teaching practice facilitated the discussions between peers (in small groups) about the performance objectives of AfL and facilitated the process of giving and receiving *feedback* (Lynch et al., 2012). The use of these reflective discussions created a rich learning environment for teachers and facilitated *collaborative teacher learning* (Vangrieken et al., 2015). The dialogic use of exemplars and the use of video-recorded fragments of lessons to stimulate reflective discussions on teaching practice can become a routine within school teams and stimulate a learning culture there. Students could also be involved in these reflective discussions of teaching practice, and student feedback might affect teaching quality (Bijlsma et al., 2019). The program spanned approximately 8 months, which could provide *ample time* to develop the basic teacher competencies needed for AfL. More time is needed to further develop these competencies in their own teaching practice (Darling-Hammond et al., 2017; Van Veen & Zwart, 2012). Although the AfL-TPD program and design decisions used are promising for other studies and TPD programs, some improvements may be considered.

6.4.1 Scaling-up and measuring the effect of the AfL-TPD program on teachers' use of the AfL strategies and teaching quality

One focus for future research may be the question of how to scale-up time-intensive TPD programs and what the effect will be if the training is rolled out on a larger scale. Professional development takes time and effort, and strains teachers' already mostly overburdened schedules. The AfL-TPD program described in this dissertation had a relatively low number of participants (range between 12 and 112 teachers) and therefore did not constitute a systemic change. Through scaling up, more teachers could be reached to learn about AfL. Preliminary results of redesigning the program with a blended delivery format, with the goal of making it scalable so that a larger group of secondary school teachers are able to gain access to the program, have shown promise (Oosterheert, 2021). In this way it can be equally effective and more cost-effective, more flexible, and more sustainable, and this format offers teachers and schools more independence compared to face-to-face learning (Owston et al., 2008; Yurtseven Avci et al., 2020). For example, the use of a video-based professional learning platform can support providing peer-to-peer feedback to allow reflective discussions online between the meeting sessions, and to engage teachers as learning resources for one another. An important focus for large-scale programs is to ensure the expertise of the trainers, for example, their own personal experience and expertise with AfL, experience with working with teachers, and familiarity with teachers and the problems they face (Kennedy, 2016).

Future studies of the effect of professional development for teachers' AfL competencies may explicitly address its direct effect on the quality of teaching and its indirect effect on student achievement. The new skills can only lead to improved student achievement, the ultimate goal of professional development, if they are applied in practice (Guskey, 2002). Researchers could follow teachers beyond the end of the PD to monitor how teachers incorporate the new ideas into their ongoing practices and look at whether teachers apply what they have learned and whether they do it well. Despite the fact that in this study teachers became more knowledgeable about AfL, they indicated that applying AfL was not yet a daily routine. The AfL-TPD program enabled teachers to start applying AfL in their teaching practice, but learning and coaching should continue in the school, supported by students, colleagues and school leaders (Schildkamp et al., 2020). For example, the teachers who participated in the AfL-TPD program can stimulate their own learning processes and transfer their knowledge to other colleagues within the school (e.g., through teacher collaboration in professional learning communities), in order to foster and expand professional development (Prenger et al., 2019). The literature states that successful professional development programs typically last longer than 1 year (Houtveen & Van de

Grift, 2007; Houtveen & Van de Grift, 2012). The monitoring of improvement in student learning may be performed even 1 or 2 years beyond the termination of the PD itself. In this dissertation, no effects of the AfL-TPD program on students' self-regulation of learning were found. Follow-up studies could look at the academic achievements of students, because an effect on academic performance would be expected first; more time is needed for an effect on self-regulated learning.

6.4.2 Mapping student knowledge, skills, and attitudes needed to become active learners in AfL processes

This study focused on teacher knowledge, skills and attitude with regard to AfL, but students also play a crucial role in AfL (Gulikers et al., 2021). Students can be stimulated to take an active role in AfL, assessing their own learning. By comparing their current learning status with the learning goals and success criteria they are able to make judgements about their goal attainment (Hattie & Timperley, 2007; Nicol & MacFarlane-Dick, 2006; Sadler, 1989). These activities can activate students' cognitive and motivational capacities, focus students on their learning goals, and provide feedback and strategies that they can use to help them reach their goals (Panadero et al., 2018). Our study shows how difficult it is to improve student agency in assessment practices, even when teachers are supported in this by a PD program. Not only students' actual strategy use is important, but also students' knowledge about the usefulness of SRL strategies (Why is it beneficial to do this?). Therefore, teachers need to be knowledgeable about promoting SRL in their lessons, and teachers need to believe that SRL is necessary for AfL processes to reach their full potential. Only then can it be expected that students will learn to appreciate SRL and subsequently become self-regulated learners themselves (Brandmo et al., 2020). The AfL-TPD program might need to be focused more on student involvement in AfL practices and on the development of students' self-regulatory skills to achieve the intended effects (Heritage, 2018). Therefore, it would be interesting to perform a cognitive task analysis (CTA) focused on students who are experienced users of AfL with their teachers. The skills, knowledge and attitudes from these above-average students may be systematically analysed, and the complexity factors they experience can also be mapped. The steps for a CTA described in this dissertation could serve well for a follow-up study: what skills and knowledge does AfL require from students? This knowledge can be used to enrich TPD programs to make teachers aware of the shared responsibility and need for interplay between teachers and students and between students in AfL processes in classrooms.

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

Andrade, H., & Brookhart, S. M. (2016). *The Role of Classroom Assessment in Supporting Self-Regulated Learning. July 2016*, 293–309. https://doi.org/10.1007/978-3-319-39211-0_17

Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. *Assessment in Education: Principles, Policy and Practice, 27(4)*, 350–372. https://doi.org/10.1080/0969594X.2019.1571992

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/s11092-013-9188-4

Artelt, C. (2000). Wie prädiktiv sind retrospektive Selbstberichte über den Gebrauch von Lernstrategien für strategisches Lernen? *Zeitschrift Für Pädagogische Psychologie, 14(2/3)*, 72–84. https://doi.org/10.1024//1010-0652.14.23.72

Bailey, A. L., & Heritage, M. (2018). *Self-Regulation in Learning: The Role of Language and Formative Assessment.* ERIC.

Beguin, A., & Ehren, M. (2011). Aspects of accountability and assessment in the Netherlands. *Zeitschrift Für Erziehungswissenschaft, 13*, 25–36.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18(1)*, 5–25. https://doi.org/10.1080/0969594X.2010.513678

Bijlsma, H. J. E., Glas, C. A. W., & Visscher, A. J. (2022). Factors related to differences in digitally measured student perceptions of teaching quality. *School Effectiveness and School Improvement, 0(0)*, 1–21. https://doi.org/10.1080/09243453.2021.2023584

Bijlsma, H. J. E., Visscher, A. J., Dobbelaer, M. J., & Veldkamp, B. P. (2019). Does smartphone-assisted student feedback affect teachers' teaching quality? *Technology, Pedagogy and Education, 28(2)*, 217–236. https://doi.org/10.1080/1475939X.2019.1572534

Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice, 22(1)*, 161–177. https://doi.org/10.1080/0969594X.2014.999643

Black, P., & Wiliam, D. (1998a). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5(1)*, 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5(1)*, 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29(5)*, 623–637. https://doi.org/10.1080/0141192032000133721

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

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92(1)*, 81–90. https://doi.org/10.1177/003172171009200119

Blom, S., & Severiens, S. (2008). Engagement in self-regulated deep learning of successful immigrant and non-immigrant students in inner city schools. *European Journal of Psychology of Education, 23(1)*, 41–58. https://doi.org/10.1007/BF03173139

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7(2)*, 161–186.

Boekaerts, M., & Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in self-regulation? *Educational Psychology Review, 18(3)*, 199–210. https://doi.org/10.1007/s10648-006-9013-4

Borgen, F. H., & Barnett, D. C. (1987). Applying cluster analysis in counseling psychology research. *Journal of Counselling Psychology, 34(4)*, 456.

Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain. *Educational Researcher, 33(8)*, 3–15. https://doi.org/10.3102/0013189X033008003

Brandmo, C., Panadero, E., & Hopfenbeck, T. N. (2020). Bridging classroom assessment and self-regulated learning. *Assessment in Education: Principles, Policy & Practice, 27*. https://doi.org/10.1080/0969594X.2020.1803589

Bryant, D. A., & Carless, D. R. (2010). Peer assessment in a test-dominated setting: empowering, boring or facilitating examination preparation? *Educational Research for Policy and Practice, 9(1)*, 3–15. https://doi.org/10.1007/s10671-009-9077-2

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice.* Routledge.

Carless, D. (2020). From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes. *Active Learning in Higher Education.* https://doi.org/10.1177/1469787420945845

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43(8)*, 1315–1325.

Carless, D., & Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assessment & Evaluation in Higher Education, 42(6)*, 930–941. https://doi.org/10.1080/02602938.2016.1211246

Carless, D., & Winstone, N. (2020). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education, 0(0)*, 1–14. https://doi.org/10.1080/13562517.2020.1782372

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24(2)*, 205–249. https://doi.org/10.1007/s10648-011-9191-6

Clark, R, Feldon, D., Van Merriënboer, J. J. G., Yates, K., & Early, S. (2008). Cognitive task analysis. In *Handbook of Research on Educational Communications and Technology.* (pp. 577–591).

Clark, R. (2014). Cognitive task analysis for expert-based instruction in healthcare. In *Handbook of Research on Educational Communications and Technology* (pp. 541–551). Springer New York. https://doi.org/10.1007/978-1-4614-3185-5_42

Coburn, C. E., & Turner, E. O. (2012). The practice of data use: An introduction. *American Journal of Education, 118(2)*, 99–111. https://doi.org/10.1086/663272

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112(1)*, 155.

Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education*. routledge.

Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B. and, Saunders, L., & Coe, R. (2015). Developing great teaching: Lessons from the international reviews into effective professional development. *Teacher Development Trust, 44* (February), 1–21. educationnext.org

Cowie, B., & Harrison, C. (2016). Classroom processes that support effective assessment. *Handbook of Human and Social Conditions in Assessment*, 335–350.

Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the Motivated Strategies for Learning Questionnaire. *Learning and Individual Differences, 21(4)*, 337–346.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39(3)*, 124–130.

Crisp, G. T. (2012). Integrative assessment: Reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education, 37(1)*, 33–43. https://doi.org/10.1080/02602938.2010.494234

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective Teacher Professional Development. *Palo Alto, CA: Learning Policy Institute, June*.

Den Brok, P., Brekelmans, M., & Wubbels, T. (2006). Multilevel issues in research using students' perceptions of learning environments: The case of the Questionnaire on Teacher Interaction. *Learning Environments Research, 9(3)*, 199.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38(3)*, 181–199. https://doi.org/10.3102/0013189X08331140

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society and Education, 7(3)*, 356–369. https://doi.org/10.25115/psye.v7i3.516

Desimone, L. M., & Stuckey, D. (2014). Sustaining professional development. *Handbook of Professional Development in Education: Successful Models and Practices, Pre K-12, 12*, 467–482. https://www.c-sail.org/sites/default/files/DesimoneStuckeyChapter.pdf

Desimone, L. M., & Wilson, J. (2022). Dimensions of teacher professional development: Lessons from the formative assessment literature [paper presentation]. *AERA 2022: San Diego, California, United States*, 1–15.

Dignath, C., Buettner, G., & Langfeldt, H.-P. (2008). How can primary school students learn self-regulated learning strategies most effectively?: A meta-analysis on self-regulation training programmes. *Educational Research Review, 3(2)*, 101–129.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A

meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3(3)*, 231–264.

Dignath, C., & Büttner, G. (2018). Teachers' direct and indirect promotion of self-regulated learning in primary and secondary school mathematics classes – insights from video-based classroom observations and teacher interviews. *Metacognition and Learning, 13(2)*, 127–157. https://doi.org/10.1007/s11409-018-9181-x

Dignath, C., & Veenman, M. V. J. (2021). The Role of Direct Strategy Instruction and Indirect Activation of Self-Regulated Learning—Evidence from Classroom Observation Studies. *Educational Psychology Review, 33(2)*, 489–533. https://doi.org/10.1007/s10648-020-09534-0

Dobbelaer, M. J. (2019). *The quality and qualities of classroom observation systems* [University of Twente]. https://doi.org/10.3990/1.9789036547161

Dockterman, D. M. (2017). Discrepancies between students' and teachers' ratings of instructional practice: A way to measure classroom intuneness and evaluate teaching quality. In *ProQuest Dissertations and Theses*. https://escholarship.org/uc/item/1sr500x5

Donahue, J. M. (1994). *Student perceptions of their teachers, their school, and themselves as learners.* Iowa State University.

Dunst, C. J., Bruder, Mary, B., & Hamby, Deborah, W. (2015). Metasynthesis of in-service professional development research: Features associated with positive educator and student outcomes. *Educational Research and Reviews, 10(12)*, 1731–1744. https://doi.org/10.5897/err2015.2306

Eggen, T., & Sanders, P. F. (1993). Psychometrie in de praktijk. *Arnhem: Cito Instituut Voor Toetsontwikkeling.*

Elwood, J. (2006). Formative assessment: Possibilities, boundaries and limitations. *Assessment in Education: Principles, Policy and Practice, 13(2)*, 215–232. https://doi.org/10.1080/09695940600708653

Everitt, B. (1980). Cluster Analysis. *Quality and Quantity, 14*, 75–100.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Field, A. (2000). *Cluster Analysis*. http//www.discoveringstatistics.com/docs/cluster.pdf

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.

Fletcher-Wood, H. (2018). *Responsive teaching: cognitive science and formative assessment in practice.* Routledge.

Fletcher, A., & Shaw, G. (2012). How does student-directed assessment affect learning? Using assessment as a learning process. *International Journal of Multiple Research Approaches, 6(3)*, 245–263. https://doi.org/10.5172/mra.2012.6.3.245

Frerejean, J., Van Geel, M., Keuning, T., Dolmans, D., Van Merriënboer, J. J. G., & Visscher, A. J. (2021). Ten steps to 4C/ID: training differentiation skills in a professional development program for teachers. *Instructional Science, 49(3)*, 395–418. https://doi.org/10.1007/s11251-021-09540-x

Frerejean, J., Van Merriënboer, J. J. G., Kirschner, P. A., Roex, A., Aertgeerts, B., & Marcellis, M. (2019).

Designing instruction for complex learning: 4C/ID in higher education. *European Journal of Education, 54(4)*, 513–524.

Frick, T. W., Chadha, R., Watson, C., & Zlatkovska, E. (2010). Improving course evaluations to improve instruction and complex learning in higher education. *Educational Technology Research and Development, 58(2)*, 115–136. https://doi.org/10.1007/s11423-009-9131-z

Garcia, T., & Pintrich, P. R. (1996). Assessing students' motivation and learning strategies in the classroom context: The Motivated Strategies for Learning Questionnaire. In *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 319–339). Springer.

George, P. S. (2005). A rationale for differentiating instruction in the regular classroom. *Theory Into Practice, 44(3)*, 185–193. https://doi.org/10.1207/s15430421tip4403_2

Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. *National Comprehensive Centre for Teacher Quality.*

Gopher, D. (2006). Emphasis Change as a Training Protocol for High-Demand Tasks. In &A. K. (eds. . A.F. Kramer, D.A. Wiegman (Ed.), *Attention From Theory to Practice* (pp. 209–224). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195305722.003.0015

Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education, 23(5)*, 531–557. https://doi.org/10.1007/s10972-012-9278-z

Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of Chiropractic Medicine, 5(3)*, 101–117.

Gulikers, J. T. M., & Baartman, L. (2017). *Doelgericht professionaliseren. Formatief toetsen met effect! Wat DOET de docent in de klas?: Eindrapport NRO-PPO overzichtsstudie dossiernummer 405-15-722.*

Gulikers, J., Veugen, M., & Baartman, L. (2021). What are we Really Aiming for? Identifying Concrete Student Behavior in Co-Regulatory Formative Assessment Processes in the Classroom. *Frontiers in Education, 6(October)*, 1–14. https://doi.org/10.3389/feduc.2021.750281

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8(3)*, 381–391. https://doi.org/10.1080/135406002100000512

Haertel, E. H. (2013). Reliability and validity of inferences about teachers based on students test scores. In *Educational Testing Service*.

Hamodi, C., López-Pastor, V. M., & López-Pastor, A. T. (2017). If I experience formative assessment whilst studying at university, will I put it into practice later as a teacher? Formative and shared assessment in Initial Teacher Education (ITE). *European Journal of Teacher Education, 40(2)*, 171–190.

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

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*, 101–111. https://doi.org/10.1016/j.tate.2013.07.008

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J., & Clarke, S. (2018). *Visible learning: feedback*. Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77(1)*, 81–112. https://doi.org/10.3102/003465430298487

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review, 17*, 50–62. https://doi.org/10.1016/j.edurev.2015.12.002

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89(2)*, 140–145.

Heritage, M. (2016). *Assessment for learning: Co-regulation in and as student–teacher interaction* (pp. 327–343). https://doi.org/10.1007/978-3-319-39211-0_19

Heritage, M. (2018). Assessment for learning as support for student self-regulation. *The Australian Educational Researcher, 45(1)*, 51–63. https://doi.org/10.1007/s13384-018-0261-3

Hill, H. C., Umland, K., Litke, E., & Kapitula, L. R. (2012). Teacher quality and quality teaching: Examining the relationship of a teacher assessment to practice. *American Journal of Education, 118(4)*, 489–519. https://doi.org/10.1086/666380

Hopfenbeck, T. N., & Stobart, G. (2015). Large-scale implementation of Assessment for Learning. *Assessment in Education: Principles, Policy and Practice, 22(1)*, 1–2. https://doi.org/10.1080/0969594X.2014.1001566

Houtveen, A. A. M., & van de Grift, W. J. C. M. (2007). Effects of metacognitive strategy instruction and instruction time on reading comprehension. *School Effectiveness and School Improvement, 18(2)*, 173–190. https://doi.org/10.1080/09243450601058717

Houtveen, T., & van de Grift, W. (2012). Improving reading achievements of struggling learners. *School Effectiveness and School Improvement, 23(1)*, 71–93. https://doi.org/10.1080/09243453.2011.600534

Hubbard, L., Datnow, A., & Pruyn, L. (2014). Multiple initiatives, multiple challenges: The promise and pitfalls of implementing data. *Studies in Educational Evaluation, 42*, 54–62. https://doi.org/10.1016/j.stueduc.2013.10.003

Huisman, B., Saab, N., Van den Broek, P., & Van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: a Meta-Analysis. *Assessment and Evaluation in Higher Education, 44(6)*, 863–880. https://doi.org/10.1080/02602938.2018.1545896

Istance, D., & Dumont, H. (2010). Future directions for learning environments in the 21st century. *The Nature of Learning: Using Research to Inspire Practice*, 317–338.

Jonsson, A., Lundahl, C., & Holmgren, A. (2015). Evaluating a large-scale implementation of assessment for learning in Sweden. *Assessment in Education: Principles, Policy & Practice, 22(1)*, 104–121. https://doi.org/10.1080/0969594X.2014.970612

Kane, T., & Staiger, D. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. MET Project. *Bill & Melinda Gates*

Foundation. http://eric.ed.gov/?id=ED540960

Kennedy, M. M. (2016). How Does Professional Development Improve Teaching? *Review of Educational Research, 86(4)*, 945–980. https://doi.org/10.3102/0034654315626800

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30(4)*, 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

Kingston, N., & Nash, B. (2015). " Formative assessment: A meta-analysis and a call for research": Erratum. *Educational Measurement: Issues and Practice, 34(2)*, 55–55. https://doi.org/10.1111/emip.12075

Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education, 75*, 199–213. https://doi.org/10.1016/j.tate.2018.06.015

Klenowski, V. (2009). Assessment for Learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice, 16(3)*, 263–268. https://doi.org/10.1080/09695940903319646

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119(2)*, 254.

Lai, M. K., & Schildkamp, K. (2013). Data-based decision making: An overview. In *Data-based Decision Making in Education* (pp. 9–21). Springer Netherlands. https://doi.org/10.1007/978-94-007-4816-3_2

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63(3)*, 18–24.

Lee, I. (2011). Bringing innovation to EFL writing through a focus on assessment for learning. *Innovation in Language Learning and Teaching, 5(1)*, 19–33. https://doi.org/10.1080/17501229.2010.502232

Levy, J., Den Brok, P., Wubbels, T., & Brekelmans, M. (2003). Students' perceptions of interpersonal aspects of the learning environment. *Learning Environments Research, 6(1)*, 5–36. https://doi.org/10.1023/A:1022967927037

Lynch, R., McNamara, P. M., & Seery, N. (2012). Promoting deep learning in a teacher education programme through self-and peer-assessment and feedback. *European Journal of Teacher Education, 35(2)*, 179–197.

Lysaght, Z., & O'Leary, M. (2013). An instrument to audit teachers' use of assessment for learning. *Irish Educational Studies, 32(2)*, 217–232. https://doi.org/10.1080/03323315.2013.784636

Mandinach, E., & Gummer, E. (2016). *Data literacy for educators: Making it count in teacher preparation and practice* (WestEd (ed.)). Teachers College Press.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record, 114(11)*, 1–48.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). Making sense of data-driven decision making in education. *Education*, 1–15.

Marshall, B., & Drummond, J. M. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education, 21(2)*, 133–149. https://doi.org/10.1080/02671520600615638

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50(3)*, 43–59. https://doi.org/10.1007/bf02505024

Merrill, M. D. (2012). *First principles of instruction.* John Wiley & Sons.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative data analysis: A methods sourcebook*. Sage publications.

Molenaar, I., Horvers, A., & Baker, R. S. (2021). What can moment-by-moment learning curves tell about students' self-regulated learning? *Learning and Instruction*, *72*. https://doi.org/10.1016/j.learninstruc.2019.05.003

Nicol, D. (2020). *The power of internal feedback: exploiting natural comparison processes*. https://doi.org/10.1080/02602938.2020.1823314

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31(2)*, 199–218. https://doi.org/10.1080/03075070600572090

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment and Evaluation in Higher Education, 39(1)*, 102–122. https://doi.org/10.1080/02602938.2013.795518

Nugteren, M. L., Jarodzka, H., Kester, L., & Van Merriënboer, J. J. G. (2018). Self-regulation of secondary school students: Self-assessments are inaccurate and insufficiently used for learning-task selection. *Instructional Science, 46(3)*, 357–381. https://doi.org/10.1007/s11251-018-9448-2

O'Leary, M., Lysaght, Z., & Ludlow, L. (2013). A measurement instrument to evaluate teachers' assessment for learning classroom practices. *The International Journal of Educational and Psychological Assessment, 14(2)*, 40–60. https://doi.org/10.13140/2.1.1289.0241

Oosterheert, J. A. (2021). *From teachers to schools: scaling-up professional development for assessment learning* [University of Twente]. http://essay.utwente.nl/85554/

Owston, R., Wideman, H., Murphy, J., & Lupshenyuk, D. (2008). Blended teacher professional development: A synthesis of three program evaluations. *Internet and Higher Education, 11(3–4)*, 201–210. https://doi.org/10.1016/j.iheduc.2008.07.003

Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: a roadmap of where we are, how we got here, and where we are going. *Australian Educational Researcher, 45(1)*, 13–31. https://doi.org/10.1007/s13384-018-0258-y

Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self- and co-regulated learning: the role of evaluative judgement. In *European Journal of Psychology of Education* (Vol. 34, Issue 3). https://doi.org/10.1007/s10212-018-0407-8

Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of

known unknowns and potential directions. *Educational Psychology Review, 28(4)*, 803–830.

Panadero, E., Jonsson, A., Strijbos, J.-W. (2016). Scaffolding Self-Regulated Learning Through Self-Assessment and Peer Assessment: Guidelines for Classroom Implementation. In: Laveault, D., Allal, L. (eds) *Assessment for Learning: Meeting the Challenge of Implementation. The Enabling Power of Assessment, vol 4*. Springer, Cham. https://doi.org/10.1007/978-3-319-39211-0_18

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research, 60(6)*, 723–735. https://doi.org/10.1080/00313831.2015.1066436

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., Pierczynski, M., & Allen, M. (2018). Teachers' Instructional Adaptations: A Research Synthesis. *Review of Educational Research, 88(2)*, 205–242. https://doi.org/10.3102/0034654317743198

Pekrul, S., & Levin, B. (2007). Building student voice for school improvement. In *International handbook of student experience in elementary and secondary school* (pp. 711–726). Springer.

Petersen-Brown, S., Lundberg, A. R., Ray, J. E., Dela Paz, I. N., Riss, C. L., & Panahon, C. J. (2019). Applying spaced practice in the schools to teach math vocabulary. *Psychology in the Schools, 56(6)*, 977–991. https://doi.org/10.1002/pits.22248

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31(6)*, 459–470.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82(1)*, 33.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63(2)*, 167–199.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq). *Educational and Psychological Measurement, 53(3)*, 801–813. https://doi.org/10.1177/0013164493053003024

Pintrich, P. R., & Zusho, A. (2002). Student Motivation and Self-Regulated Learning in the College Classroom. In *Higher education: Handbook of theory and research (pp. 55–128)*. Springer. https://doi.org/10.1007/978-94-010-0245-5_2

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

Popham, W. J. (2008). *Transformative assessment*. ASCD.

Popham, W. J. (2011). *Transformative assessment in action: An inside look at applying the process.* ASCD.

Prenger, R., Poortman, C. L., & Handelzalts, A. (2019). The effects of networked professional learning communities. *Journal of Teacher Education, 70(5)*, 441–452. https://doi.org/10.1177/0022487117753574

Rollett, W., Bijlsma, H., & Röhl, S. (2021). Student Feedback on Teaching in Schools: Current State of

Research and Future Perspectives. *Student Feedback on Teaching in Schools*, 259.

Romesburg, H. C. (1990). Cluster analysis for researchers Robert E. *Florida: Krieger Pub. Co*.

Roth, A., Ogrin, S., & Schmitz, B. (2016). Assessing self-regulated learning in higher education: a systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability, 28(3)*, 225–250. https://doi.org/10.1007/s11092-015-9229-2

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

Ruiz-Primo, M., & Li, M. (2013). Examining formative feedback in the classroom context: New research perspectives. *Handbook of Research on Classroom Assessment*, 215–232.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education, 35(5)*, 535–550. https://doi.org/10.1080/02602930903541015

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144. https://doi.org/10.1007/BF00117714

Sanders, P. (2011). Toetsen op school. *Cito,* 1–198. papers://53a564bf-cd27-4bf5-aaf7-76e13e37c82d/Paper/p25

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research, 61(3)*, 257–273. https://doi.org/10.1080/00131881.2019.1625716

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.

Scriven, M. (1967). The Methodology of Evaluation. In *In R.W. Tyler, R.M. Gagne, & M. Scriven (EDS.), Perspectives of Curriculum Evaluation* (p. Chicago (39-83)).

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78(1)*, 153–189. https://doi.org/10.3102/0034654307313795

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

Sluijsmans, D. M. A., Joosten-ten Brinke, D., & der Vleuten, C. P. M. (2013). Toetsen met leerwaarde. Een reviewstudie naar de effectieve kenmerken van formatief toetsen [Assessment with learning value: A review study into the characteristics of effective formative assessment] NOW-PROO 411-11-697. Den Haag, The Netherlands: NWO.

Smith, K. (2016). Cooperative learning about assessment for learning. *Assessment for Learning: Meeting the Challenge of Implementation*, 181–197.

Smith, K., & Engelsen, K. S. (2013). Developing an assessment for learning (AfL) culture in school: The voice of the principals. *International Journal of Leadership in Education, 16(1)*, 106–125. https://doi.org/10.1080/13603124.2011.651161

Spörer, N., & Brunstein, J. C. (2006). Erfassung selbstregulierten Lernens mit Selbstberichtsverfahren: Ein Überblick zum Stand der Forschung. *Zeitschrift Für Pädagogische Psychologie, 20(3)*, 147–160.

Stalmeijer, R. E., McNaughton, N., & Van Mook, W. N. K. A. (2014). Using focus groups in medical education research: AMEE Guide No. 91. *Medical Teacher, 36(11)*, 923–939. https://doi.org/10.3109/0142159X.2014.917165

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.

Stoll, L., & Myers, K. (1998). *No quick fixes: Perspectives on schools in difficulty*. Psychology Press.

Stringfield, S. (1995). Attempts to enhance students' learning: A search for valid programs and highly reliable implementation techniques. *School Effectiveness and School Improvement, 6(1)*, 67–96.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration*. Ministry of Education.

Topping, K. (2017). Peer assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology, 1(1)*, 1–17.

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education, 38(3)*, 323–342. https://doi.org/10.1080/03054985.2012.689693

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94(1)*, 3.

Van der Kleij, F., & Adie, L. (2020). Towards effective feedback: an investigation of teachers' and students' perceptions of oral feedback in classroom practice. *Assessment in Education: Principles, Policy and Practice, 27(3)*, 252–270. https://doi.org/10.1080/0969594X.2020.1748871

Van der Kleij, F. M., Adie, L. E., & Cumming, J. J. (2019). A meta-review of the student role in feedback. *International Journal of Educational Research, 98*(June), 303–323. https://doi.org/10.1016/j.ijer.2019.09.005

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

Van Geel, M., Keuning, T., Frèrejean, J., Dolmans, D., Van Merriënboer, J. J. G., & Visscher, A. J. (2019). Capturing the complexity of differentiated instruction. *School Effectiveness and School Improvement, 30(1)*, 51–67. https://doi.org/10.1080/09243453.2018.1539013

Van Merriënboer, J. J. G., & Dolmans, D. (2015). Research on instructional design in the health sciences: From taxonomies of learning to whole-task models. *Researching Medical Education*, 193–206.

Van Merriënboer, J. J. G. (2007). Alternate models of instructional design: Holistic design approaches and complex learning. *Trends and Issues in Instructional Design and Technology*, 72–81.

Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development, 50(2)*, 39–61. https://doi.

org/10.1007/BF02504993

Van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology, 20(3)*, 343–352. https://doi.org/10.1002/acp.1250

Van Merriënboer, J. J. G., & Kirschner, P. A. (2017). Ten steps to complex learning. In *Ten Steps to Complex Learning*. Routledge. https://doi.org/10.4324/9781315113210

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38(1)*, 5–13. https://doi.org/10.1207/S15326985EP3801_2

Van Merriënboer, J. J. G., & Kester, L. (2008). Whole-task models in education. *Handbook of Research on Educational Communications and Technology, 3*, 441–456.

Van Veen, K., Zwart, R., & Meirink, J. (2012). What makes teacher professional development effective? A literature review. *Teacher Learning That Matters*, 23–41.

Van Veen, K., & Zwart, R. (2012). What makes teacher professional developmen effective. In M. Kooy & K. Van Veen (Eds.), *Teacher learning that matters* (pp. 3–21). Routledge.

Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review, 15*, 17–40. https://doi.org/10.1016/j.edurev.2015.04.002

Veenman, M. (2011). *Alternative assessment of strategy use with self-report instruments : a discussion*. 205–211. https://doi.org/10.1007/s11409-011-9080-x

Verhaeghe, G., Vanhoof, J., Valcke, M., & van Petegem, P. (2010). Using school performance feedback: Perceptions of primary school principals. *School Effectiveness and School Improvement, 21(2)*, 167–188. https://doi.org/10.1080/09243450903396005

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* (October 2020), 101027. https://doi.org/10.1016/j.stueduc.2021.101027

Walter, C., & Briggs, J. (2012). What professional development makes the most difference to teachers. *Oxford University Press., 27* (October), 23.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63(3)*, 249–294.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). Professional learning in the learning profession. A status report on teacher development in the U.S. and abroad. National Staff Development Council, 1–162. Stanford University. www.nsdc.org

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37(1)*, 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001

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

Wiliam, D., & Leahy, S. (2018). *Formatieve assessment: integreren in de praktijk*. Bazalt.

Wiliam, D., & Thompson, M. (2008). *Integrating assessment with learning: What will it take to make it work?* Routledge.

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52(1)*, 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wolterinck, C. H. D., Van Geel, M., Schildkamp, K., Visscher, A. J. (2021). *A cognitive talk analysis of the skills that assessment for learning requires from teachers.* [Manuscript submitted for publication]. Faculty of Behavioural, Management and Social Sciences, University of Twente.

Wolterinck, C., Poortman, C., Schildkamp, K., & Visscher, A. (2022). Assessment for Learning: developing the required teacher competencies. *European Journal of Teacher Education, 00(00),* 1–19. https://doi.org/10.1080/02619768.2022.2124912

Wolters, C. (2011). Regulation of motivation: Contextual and social aspects. *Teachers College Record, 113(2)*, 265–283.

Wolters, C. A. (2003). Regulation of motivation: Evaluating an underemphasized aspect of self-regulated learning. *Educational Psychologist, 38(4)*, 189–205.

Wolters, C. A., & Benzon, M. B. (2013). Assessing and predicting college students' use of strategies for the self-regulation of motivation. *The Journal of Experimental Education, 81(2)*, 199–221.

Wolters, C. A., & Pintrich, P. R. (1998). Contextual differences in student motivation and self-regulated learning in mathematics, English, and social studies classrooms. *Instructional Science, 26*(1–2), 27–47.

Yin, R. K. (2014). *Case study research design and methods.* Thousand Oaks, CA: Sage.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement. issues \& answers. rel 2007-no. 033. *Regional Educational Laboratory Southwest (NJ1)*.

Yurtseven Avci, Z., O'Dwyer, L. M., & Lawson, J. (2020). Designing effective professional development for technology integration in schools. *Journal of Computer Assisted Learning, 36(2)*, 160–177. https://doi.org/10.1111/jcal.12394

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81(3)*, 329–339. https://doi.org/10.1037/0022-0663.81.3.329

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82(1),* 51.

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives.* Routledge.

SAMENVATTING (DUTCH SUMMARY)

Zelfs wanneer instructie met grote zorg wordt gepland, ervaren leraren elke les opnieuw dat wat zij hun leerlingen wilden leren niet per se overeenkomt met wat de leerlingen daadwerkelijk hebben geleerd. Daarom is het verzamelen van informatie die inzicht geeft in het denken, kennen en kunnen van leerlingen essentieel, als we de kwaliteit van het leerproces van leerlingen willen verbeteren. Dat doen we door middel van toetsing. Toetsing is essentieel binnen het onderwijs, omdat op basis van de verkregen informatie conclusies kunnen worden getrokken over de vorderingen van leerlingen. Toetsen kunnen niet alleen helpen om de vorderingen van leerlingen en de effectiviteit van het onderwijs in kaart te brengen, maar kunnen ook dienen als instrument om onderwijs-op-maat aan te bieden. Informatie uit toetsen kan bovendien door zowel leraren als leerlingen worden gebruikt om onderwijs- en leeractiviteiten aan te passen aan de leerbehoeften van leerlingen, met als doel het leerproces van leerlingen waar nodig bij te sturen en zo lesgeven en leren te verbeteren. Deze werkwijze wordt formatief toetsen genoemd.

Een belangrijke vorm van formatief toetsen is Assessment for Learning (AfL). AfL kan onderdeel uitmaken van de dagelijkse lespraktijk van leraren en leerlingen en gaat over een manier van lesgeven waarbij instructie en toetsen met elkaar verweven zijn. Voortdurend wordt informatie verzameld over waar leerlingen zich in het leerproces bevinden en hoe ze zich verhouden ten opzichte van de gestelde leerdoelen. Tijdens een les kunnen verschillende activiteiten gericht zijn op het verzamelen van informatie, zoals bijvoorbeeld het bespreken van huiswerkopdrachten, een onderwijsleergesprek tussen docent en leerlingen, observaties, of de inzet van diagnostische vragen. Op basis van de verzamelde informatie kan het onderwijs vervolgens worden aangepast aan de leerbehoeften van leerlingen en feedback worden gegeven over hoe leerlingen verder kunnen, om zo hun leren te verbeteren. Een cruciaal kenmerk is dat leraren en leerlingen samen verantwoordelijk zijn voor het leerproces. Leerlingen moeten daarbij verantwoordelijkheid nemen voor het eigen onderwijs en daarvoor moeten ze goed weten wat en hoe ze leren, hoe ze zichzelf kunnen beoordelen en hoe ze hun leerproces kunnen sturen, zodat ze hun talenten ontwikkelen.

Hoewel AfL kan leiden tot betere leerresultaten van leerlingen, blijkt de toepassing ervan in het Nederlandse onderwijs nog beperkt te zijn. AfL is een complexe docentvaardigheid en docenten vinden het lastig om het in hun klas te implementeren. Hierdoor leidt AfL niet altijd tot betere leerling prestaties. Leraren hebben behoefte aan professionele ontwikkeling die hen ondersteunt bij het ontwikkelen en implementeren van AfL in hun

klas. De overkoepelende onderzoeksvraag in deze studie was daarom: *Hoe kunnen leraren ondersteund worden bij het ontwikkelen van hun competenties voor formatief toetsen?*

In **Hoofdstuk 2** wordt ingegaan op de vraag in welke mate leerlingen het gebruik van AfL strategieën ervaren in hun lessen? Om deze onderzoeksvraag te beantwoorden werd een leerling vragenlijst afgenomen op 12 scholen voor voortgezet onderwijs. In totaal hebben 685 leerlingen uit de vierde en vijfde klas van de bovenbouw van het voortgezet onderwijs de vragenlijst ingevuld. De vragenlijst bevatte vier afzonderlijke, onafhankelijke schalen, gebaseerd op de belangrijkste AfL strategieën.

Verduidelijken, delen en begrijpen van leerdoelen en succescriteria.
Leerdoelen geven weer wat een leraar wil dat de leerling leert. Succescriteria worden gebruikt om te bepalen waar leerlingen zich in hun leerproces bevinden ten opzichte van de leerdoelen. Het is voor zowel de leraar als de leerling belangrijk om te weten waar de leerling naartoe werkt.

Informatie ontlokken bij leerlingen over begrip van de leerstof.
Deze strategie gaat over het opzetten van klassikale discussies en andere leertaken, om bewijs te verzamelen van de leervorderingen. Door bijvoorbeeld diagnostische vragen te stellen, kunnen misvattingen van leerlingen met betrekking tot de leerstof opgespoord worden. Er bestaat een grote variatie aan technieken, die zowel de leraar als de leerling inzicht kunnen geven in waar de leerlingen zich in hun leerproces bevinden en hoe ze zich verhouden t.o.v. de leerdoelen en bijhorende succescriteria.

Feedback geven die het leerproces stimuleert.
Door het geven van hoogwaardige formatieve feedback, op basis van de verzamelde informatie, kunnen leerlingen zich bewust worden van hun huidige status in het leerproces en hun leren effectiever sturen. Wanneer feedback tijdig en uitgebreid wordt gegeven door de leraren of door de leerlingen aan elkaar, kunnen leerlingen ook gestimuleerd worden om hun eigen leerproces zelf te controleren en te verbeteren.

Leerlingen activeren als bron voor leren, voor zichzelf en voor elkaar.
Deze strategie betreft het activeren van leerlingen als eigenaren van hun eigen leren en als peers voor elkaar. Leerlingen kunnen bijvoorbeeld hun eigen taak of die van hun medeleerlingen beoordelen aan de hand van een rubric (een instrument om de opdrachten te evalueren of beoordelen). Door leerlingen actief te betrekken in het leerproces gaan ze het beoordelingsproces beter begrijpen en wordt het voor hen gemakkelijker om hun

leren te voorspellen, plannen, monitoren en evalueren.

In het onderzoek werd aan de leerlingen gevraagd om de vragenlijst in te vullen voor één vak, ofwel voor Engelse taal, ofwel voor wiskunde. Op deze wijze kon de mate waarin AfL-strategieën werden toegepast in de lespraktijk worden onderzocht. De bevindingen toonden aan dat er geen verschil was tussen de vakken Engelse taal en wiskunde. Leerlingen ervoeren 'het delen van leerdoelen en succescriteria' en 'het verzamelen van bewijs over de leervorderingen van leerlingen' als de meest gebruikte strategieën in hun lessen. Ze gaven aan dat 'feedback' en 'peer- en zelfevaluatie' de minst gebruikte strategieën waren, terwijl deze strategieën expliciet inzetten op de actieve betrokkenheid van leerlingen, en dus zichtbaarder en merkbaarder zijn voor leerlingen.

Uit de resultaten van deze studie kunnen we concluderen dat er aanzienlijke ruimte voor verbetering is met betrekking tot de integratie van AfL in de dagelijkse activiteiten in de klas. Uit de clusteranalyse die in deze studie werd uitgevoerd, kwamen drie verschillende clusters naar voren van ervaren mate van AfL strategiegebruik: een cluster 'hoog' waarin leerlingen het gebruik van alle vier de AfL strategieën rapporteerden als 'opkomend' of 'gevestigd' (50% - 74% van de lessen); een cluster 'midden' waarin het ervaren strategiegebruik meer verdeeld was, van '(bijna) nooit' tot 'gevestigd' (0 - 74% van de lessen); een cluster 'laag' waarin leerlingen rapporteerden dat ze drie van de vier AfL strategieën '(bijna) nooit' of 'sporadisch' gebruikten (0-24% van de lessen). Alleen de strategie 'delen van leerintenties en succescriteria' werd als 'sporadisch' of 'opkomend' (25% - 49% van de lessen) gescoord in dit laatste cluster. Uit de clusteranalyse bleek dat het gebruik van de strategie 'delen van leerintenties en succescriteria' in alle drie de clusters grotendeels aan bod komt. Het gebruik van de andere drie strategieën varieert. Deze kennis kan gebruikt worden bij het ontwerpen van een professionaliseringsprogramma voor leraren gericht op AfL. Door rekening te houden met de behoeften van leraren aan de hand van de clusters wordt maatwerk binnen het programma mogelijk.

De twee onderzoeksvragen die in **Hoofdstuk 3** aan de orde kwamen waren: welke kennis, vaardigheden en houding heeft een leraar nodig om AfL in de les te kunnen implementeren? En welke factoren beïnvloeden de complexiteit van AfL? Om deze vragen te beantwoorden werd een cognitieve taakanalyse uitgevoerd, waarbij het gedrag van expert-leraren m.b.t. AfL als uitgangspunt werd genomen. Voor het verzamelen van de data werd gebruik gemaakt van lesobservaties en interviews met leraren en inhoudsdeskundigen. In totaal namen 16 leraren, verdeeld over drie vakken in het voortgezet onderwijs (Engelse taal, Nederlandse taal en scheikunde) deel aan deze studie. Een overzicht van de vaardigheden

die leraren nodig hebben voor AfL wordt in dit proefschrift weergegeven door middel van een vaardighedenhiërarchie, aangevuld met prestatie-indicatoren. Deze indicatoren specificeren de gewenste taakprestatie voor betreffende vaardigheid. De resultaten van deze studie toonden aan dat om AfL succesvol te maken leraren zich niet alleen moeten richten op het uitvoeren van de les met behulp van AfL strategieën, maar ook de hele lessenserie van te voren moeten voorbereiden, de lessen binnen deze serie moeten voorbereiden en lessen moeten evalueren. De verkregen inzichten zijn een waardevolle bijdrage aan de AfL-kennisbasis en bieden een basis voor het ontwikkelen van professionaliseringstrajecten gericht op het toerusten van leraren met de vaardigheden die nodig zijn voor het gebruik van AfL in de les.

Daarnaast zijn vakkennis en pedagogische kennis essentieel voor leraren. Om AfL te kunnen afstemmen op de instructiedoelen, hebben leraren kennis nodig van het eerdere leren van leerlingen in relatie tot het curriculum en de leerdoelen van het vak. In het bijzonder is kennis van de misconcepties van leerlingen essentieel om hiaten in het begrip van een bepaald curriculumonderwerp te kunnen bepalen. Bovendien werden vijf factoren geïdentificeerd die de complexiteit van de toepassing van AfL in de klas beïnvloeden: 1) de samenstelling van de leerlingengroep, 2) de mate waarin de leraar controle heeft tijdens de les, 3) de moeilijkheidsgraad van de lesinhoud (doel en onderwerp), 4) de hoeveelheid beschikbare ondersteuning, zoals instructiemateriaal met suggesties voor remediëring en remediëringsmateriaal, en 5) de hoeveelheid schoolondersteuning, bijvoorbeeld samenwerking met collega's en professionele ontwikkeling.

In **Hoofdstuk 4** staat de professionele ontwikkeling van leraren voor AfL centraal. De twee onderzoeksvragen die in deze studie aan bod kwamen waren: Hoe relevant en bruikbaar vonden leraren het AfL-professionaliseringsprogramma dat is ontworpen op basis van het 4C/ID (Four Component Instructional Design)? En welke kennis, vaardigheden en attitudes, die nodig zijn voor AfL hebben leraren geleerd als resultaat van deelname aan het AfL-professionaliseringsprogramma? Het 4C/ID model werd gebruikt om een professionaliseringsprogramma voor leraren te ontwerpen voor de ontwikkeling van hun AfL-competenties. Het programma werd uitgevoerd op een vakspecifieke basis (Engelse taal, Nederlandse taal en scheikunde) en in totaal namen 12 leraren van vier verschillende scholen voor voortgezet onderwijs deel. Het professionaliseringsprogramma was gericht op werkplekleren en gebruikte de dagelijkse lessen van de deelnemers als leertaken. Deze 'real-life' leertaken waren gebaseerd op hele taken die een beroep doen op kennis, vaardigheden en attitudes die nodig zijn voor het uitvoeren van taken in de dagelijkse lespraktijk. Ook boden de leertaken dagelijkse mogelijkheden om ze te oefenen.

Om de onderzoeksvragen te beantwoorden werden docenten en de coaches van het professionaliseringsprogramma geïnterviewd en vulden docenten een vragenlijst in.

Het onderzoek leverde waardevolle lessen op over hoe leraren ondersteund kunnen worden bij het leren gebruiken van AfL in hun onderwijspraktijk. Het gebruik van leertaken die aansloten bij de dagelijkse lespraktijk maakte de inhoud van het programma toepasbaar en resulteerde in tevredenheid van docenten over de relevantie en bruikbaarheid van het programma. Het gebruik van een hele-taak benadering hielp leraren om nieuw geleerde vaardigheden over te dragen naar hun dagelijkse professionele werkzaamheden. Leraren ervoeren terwijl ze AfL probeerden te implementeren ook de complexiteit van het implementeren van AfL. Het gebruik van voorbeelden en video-opnames van lessen, om expertstrategieën te demonstreren en reflectieve vaardigheden te stimuleren, ondersteunde het actief en samenwerkend leren door de leraren, en leraren verwierven op die manier kennis, vaardigheden en positieve attitudes met betrekking tot het effectief gebruik van AfL.

De onderzoeksvraag die in **Hoofdstuk 5** werd behandeld luidde: Wat is het effect van een AFL professionaliseringsprogramma op de zelfregulatie van hun leren door leerlingen? In een experimentele studie werd het effect onderzocht van een intensief, 8 maanden durend AfL professionaliseringsprogramma op het vermogen van leerlingen om hun eigen leren zelf te reguleren. Leraren die leerlingen betrekken bij de beoordeling van de leervorderingen van leerlingen in de klas, en daarbij laten zien dat leraren en leerlingen samen verantwoordelijk zijn voor het uitvoeren van de kernstrategieën van AfL, kunnen hun leerlingen helpen om betere zelfregulerende leerlingen te worden. In totaal maakten 31 leraren deel uit van de experimentele conditie waarin ze het AfL professionaliseringsprogramma volgden. Zesentwintig leraren namen deel aan de controleconditie, deze leraren namen niet deel aan het professionaliseringsprogramma. Een online zelfrapportage vragenlijst, de Nederlandse vertaling van de Motivated Strategies for Learning Questionnaire (MSLQ), werd afgenomen als pre-test en post-test, om het effect van de interventie op de zelfregulatie van leren te evalueren, De pre-test en post-test scores van de studenten in de experimentele (N = 477) en de studenten in de controle conditie (N = 298) werden met elkaar vergeleken.

De leerlingen in de experimentele groep presteerden niet beter dan de leerlingen in de controlegroep als het gaat om de verbetering van hun vermogen tot zelfregulatie van hun leren. Hoewel in de literatuur gesteld wordt dat AfL (uiteindelijk) het zelfregulerend leren van leerlingen bevordert, hebben we dit dus niet kunnen bevestigen. Een mogelijke verklaring hiervoor zou kunnen zijn dat het professionaliseringsprogramma nog onvoldoende gericht

was op het bevorderen van zelfregulerend leren van leerlingen tijdens het proces van AfL. De tijdspanne kan ook te kort geweest zijn voor de leraren om de nieuwe vaardigheden te leren, ze goed genoeg toe te passen in de klas en herhaalde oefening in te bouwen. Een ander aspect dat vermeld moet worden, is dat de leraren die aan deze studie deelnamen, te maken hadden met de sluiting van scholen en afstandsonderwijs ten gevolge van de Covid-19 pandemie. Scholen waren gedurende meerdere maanden gesloten, wat betekende dat leerlingen alleen online afstandsonderwijs konden volgen. Leraren stonden voor een nieuwe uitdaging bij het geven van online lessen, en het gebruik van AfL tijdens online lessen was voor hen wellicht een nog grotere uitdaging. Bijgevolg hadden de leraren tijdens de COVID-19-crisis waarschijnlijk minder gelegenheid om leerlingen te stimuleren om zich bezig te houden met AfL-activiteiten en hadden de leerlingen misschien ook minder mogelijkheden om hun vaardigheden op het vlak van zelfregulerend leren (SRL) te ontwikkelen.

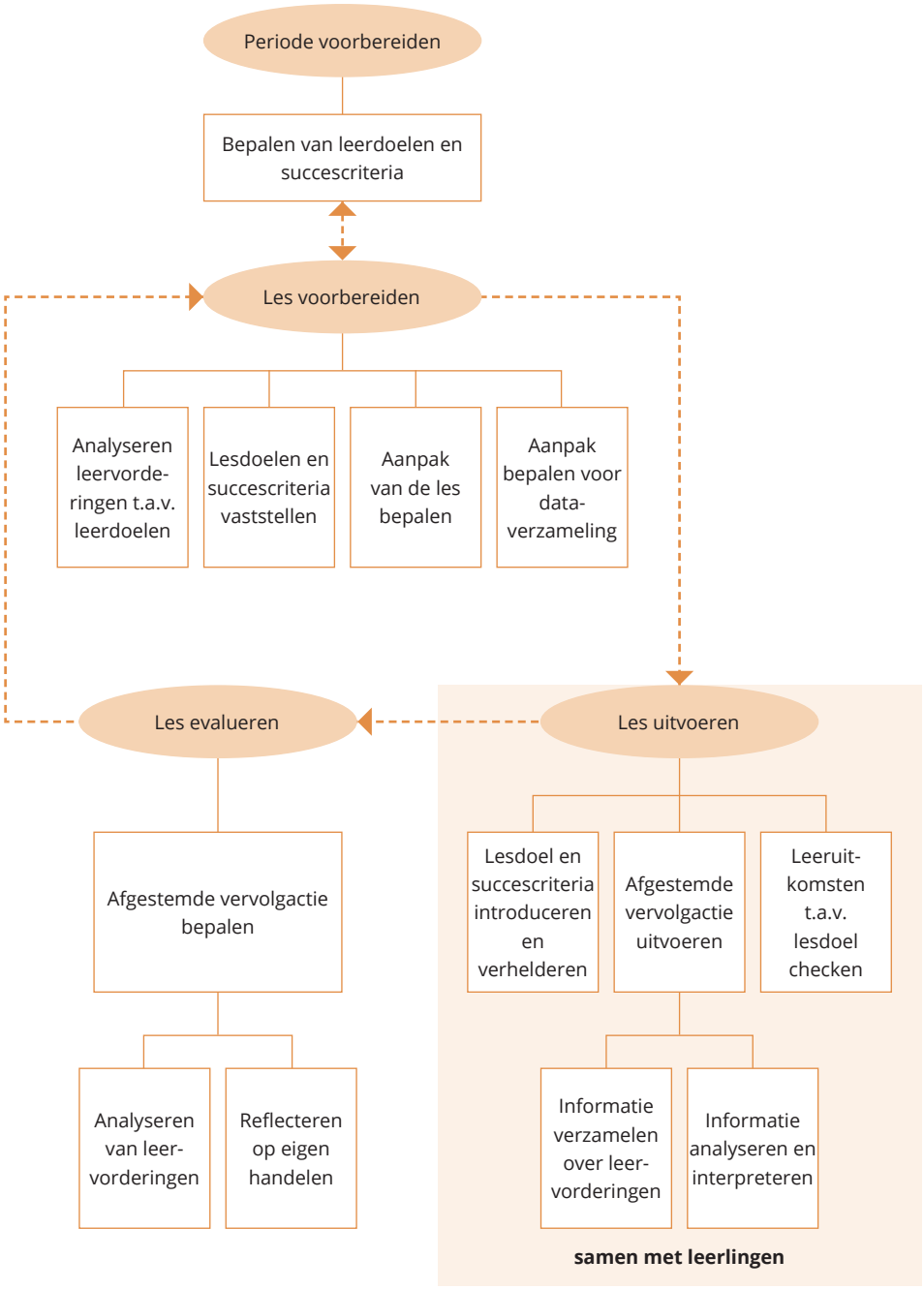
De studie liet zien hoe moeilijk het is om het vermogen van leerlingen om hun leren zelf te reguleren te verbeteren, zelfs wanneer hun leraren worden ondersteund in een AfL-professionaliseringsprogramma. Niet alleen het feitelijke strategiegebruik van leerlingen is belangrijk, maar ook de kennis van leerlingen over het nut van zelfregulatie-strategieën (waarom is het zinvol om dit te doen?). Daarom kunnen de kennis en overtuigingen van leraren over het bevorderen van zelfregulerend leren voorwaarden zijn voor AfL-processen die hun volledige potentieel bereiken en leerlingen zelfregulerende leerlingen maken. Het AfL-professionaliseringsprogramma moet misschien nog meer gericht zijn op betrokkenheid van leerlingen bij AfL praktijken en op de ontwikkeling van zelfregulerende vaardigheden van leerlingen om de verwachte effecten te vinden. Zo zouden de reflectie- en feedbacksessies in het AfL-professionaliseringsprogramma meer expliciet gericht kunnen worden op wat leraren doen om zelfregulatie van leren te bevorderen. Het AfL-professionaliseringsprogramma zou leraren begeleide en aanhoudende ondersteuning moeten bieden wanneer zij deze nieuwe kennis toepassen in hun praktijk. Op die manier kunnen de leraren zelf zelfregulatie van leren en AfL ervaren, net zoals, hopelijk, hun leerlingen dit zullen ervaren als gevolg van deze innovaties in de lespraktijk.

Samenvattend laten de resultaten van de studies in deze dissertatie zien dat er aanzienlijke ruimte voor verbetering is met betrekking tot de integratie van AfL met leeractiviteiten in de dagelijkse lespraktijk van het voortgezet onderwijs in Nederland. Om deze verbeteringen teweeg te brengen is effectieve professionele ontwikkeling van leraren nodig. Het 4C/ID-model kan de basis bieden voor het ontwerp van AfL-professionaliseringsprogramma's waarin de integratie van vaardigheden, kennis en attitude worden vereist. Bovendien

werden vijf belangrijke inzichten verworven die de professionele ontwikkeling met betrekking tot AfL kunnen versterken.

Integratie van AfL met de verschillende fasen van de instructie

Om AfL met succes in de les te implementeren, is het nuttig om de vaardigheden die leraren nodig hebben en de verschillende instructiefasen met elkaar in verband te brengen. Uit de uitgebreide analyse van gegevens van expertleraren, beschreven in hoofdstuk 3, bleek immers dat de competenties die leraren nodig hebben om AfL in hun lessen te implementeren alle fasen van de instructie bestrijken. Implementatie van AfL vereist actie van de leraar tijdens de planningsfase waarin de les wordt voorbereid, tijdens het lesgeven en ook na de les bij het reflecteren op hun instructie. De verschillende vaardigheden die nodig zijn voor AfL kunnen worden gespecificeerd voor elk van deze fasen, zoals afgebeeld in figuur S.1. In de planningsfase bereiden leraren een les voor, gebaseerd op de evaluatie van de vorige les en op de voorbereiding van een lessenserie rondom een inhoudelijk thema. Tijdens de les voeren de leraren de voorbereide leeractiviteiten uit, waarbij ze de instructie afstemmen op de leerbehoeften van de leerlingen, bewijs van leren uitlokken en de leerlingen aanmoedigen om eigenaar van hun leerproces te zijn. Tijdens de reflectie op hun instructie en op het leerproces van de leerlingen, bepalen de leraren, gebaseerd op de verzamelde informatie, welke gerichte vervolgacties er voor/in de volgende les nodig zijn. Leraren gebruiken de resultaten van de reflectie in de planningsfase voor de volgende les, zodat een cyclisch AfL-proces ontstaat.



Figuur S.1 Integratie van AfL-vaardigheden met Fasen van Instructie

Interne feedback genereren met peers

Het 4C/ID model was de basis voor het ontwerp van het AfL-TPD programma dat de integratie van vaardigheden, kennis en attitude vereiste. Het toepassen van reflectieve discussies met peers in ons ontwerp zorgde voor een rijke leeromgeving voor leraren en vergemakkelijkte het verwerven van vaardigheden in het toepassen van feedback en reflectie. Door leraren videofragmenten van hun eigen opgenomen lessen te laten gebruiken, deze te laten analyseren en te vergelijken met video's van hun peers, kon een informeel, spontaan en natuurlijk feedbackproces op gang komen. Leraren genereerden interne feedback, door hun huidige kennis over 'goed' lesgeven af te zetten tegen de informatie die verkregen werd uit de videofragmenten, zowel hun eigen fragmenten als die van hun peers. Dit proces van interne feedback, waarbij leraren hun huidige kennis en competenties vergelijken met bepaalde referentie-informatie, levert nieuwe kennis op waarmee leraren hun lespraktijk kunnen verbeteren. Deze mogelijkheden tot vergelijking kunnen explicieter worden opgenomen in het professionaliseringsprogramma voor AfL. Dit kan leraren helpen om vaardig te worden in het gebruik van vergelijkingsprocessen, zodat ze beter in staat zijn om hun eigen leren te plannen, te evalueren, te ontwikkelen en te reguleren.

Mobilisatie van de voornaamste betrokkenen, de leerlingen

Om AfL-processen in de les effectief te laten zijn, moeten leraren hun leerlingen actief betrekken bij het genereren, verwerken en reageren op feedbackinformatie. Dit geeft leerlingen de kans om het vermogen tot zelfregulatie van hun leren te ontwikkelen. Uit de analyse van de percepties van leerlingen m.b.t. het gebruik van AfL in de klas, zoals beschreven in hoofdstuk 2, bleek dat 'leerlingen activeren als bron voor leren, voor zichzelf en voor elkaar' de minst vaak gebruikte strategieën waren, hoewel deze strategieën AfL explicieter maken voor leerlingen en daardoor zichtbaarder en merkbaarder. Dit vraagt om partnerschap, gedeelde verantwoordelijkheid en samenspel tussen leraren en leerlingen en ook tussen leerlingen onderling. De kennis en overtuigingen van leraren over het bevorderen van de rol van de leerling zijn voorwaardelijk voor AfL. Als de leerlingen in staat zijn om hun eigen werk te monitoren, evalueren en te verbeteren, zal dit waarschijnlijk de kracht van instructie vergroten.

Opschaling door 'Blended Learning'

Professionele ontwikkeling kost tijd en moeite en belast de toch al overvolle agenda's van leraren. Het AfL-professionaliseringsprogramma dat in dit proefschrift beschreven wordt had een relatief klein aantal deelnemers per school (van elke school 3 leraren) wat schaalvergroting binnen de school lastig maakt. Door bijvoorbeeld een mengvorm

te gebruiken van face-to-face bijeenkomsten en ICT-gebaseerde onderwijsactiviteiten, leermaterialen en instrumenten, het zogenaamd 'Blended Learning', zouden meer leraren over AfL kunnen leren. Op deze manier kan professionele ontwikkeling kosten effectiever, flexibeler en duurzamer zijn, en deze vorm biedt leraren en scholen meer onafhankelijkheid ten opzichte van face-to-face leren. Het gebruik van een video-platform kan bijvoorbeeld de online reflectieve discussies mogelijk maken, en het leren samen met peers faciliteren.

Kennis, vaardigheden en attitudes van leerlingen

Deze studie richtte zich op de kennis, vaardigheden en houding van leraren met betrekking tot AfL, maar leerlingen spelen ook een cruciale rol in AfL. Daarom zou het interessant zijn om een cognitieve taakanalyse (CTA) uit te voeren, gericht op leerlingen die veel ervaring hebben met het gebruik van AfL, samen met hun leraren. De vaardigheden, kennis en attitudes van deze qua AfL-gebruik bovengemiddelde leerlingen kunnen dan systematisch worden geanalyseerd en ook de complexiteitsfactoren die zij ervaren kunnen in kaart worden gebracht. De in dit proefschrift beschreven stappen van een CTA zouden goed kunnen dienen voor een vervolgstudie gericht op de beantwoording van de volgende vraag: welke vaardigheden en kennis vraagt AfL van leerlingen? De resulterende kennis zou dan gebruikt kunnen worden om professionaliseringsprogramma's te verrijken en om leraren bewust te maken van de gedeelde verantwoordelijkheid en de noodzaak van samenspel bij AfL, zowel tussen leraren en leerlingen als tussen leerlingen onderling.

APPENDIX A: SELF-REPORT RESPONSE SCALE FOR STUDENTS

Rating Scale	Interpretation	Score point
Embedded	It happens in 90% of the lessons. According to the student, the practice described in the statement is a firmly established feature of the teacher's pedagogy that occurs routinely in day-to-day classroom practice.	5
Established	It happens in 75% of the lessons. According to the student, the practice described in the statements is a customary feature of the teacher's pedagogy that occurs frequently in day-to-day classroom practice.	4
Emerging	It happens in 50% of the lessons. According to the student, the practice described in the statements is a nascent feature of the teacher's pedagogy, and as such it is an approach with which the teacher and students are beginning to engage.	3
Sporadic	It happens in 25% of the lessons. According to the student, the practice described in the statements is an irregular feature of the teacher's pedagogy, and as such it occurs intermittently in day-to-day classroom practice.	2
(almost) Never	It happens in less than 10% of the lessons. According to the student, the practice, as described, is incongruent with day-to-day classroom practice; as such, it is a practice that has either been attempted and abandoned for some reason or not engaged in at all.	1
Don't understand	I don't understand what the statement means According to the student, they either do not understand the terminology used in the statement to describe the practice and/or the practice is foreign to the respondent.	

APPENDIX B: FACTOR ANALYSIS RESULTS FOR THE AFL STUDENT QUESTIONNAIRE (*N* = 685)

Items	**Rotated factor loadings**			
My teacher	LISC	EE	FB	PSA
Uses leaning goals that are stated in words that emphasize knowledge, skills, concepts and/or attitudes, i.e., what the students are learning, NOT what they are doing.	0.56			
Reminds me about the links between what we are learning and the big learning picture (e.g., "We are learning to count money so that when we go shopping, we can check our change").	0.39			
Matches success criteria, related to learning intentions, to pupils.	0.54			
Uses child-friendly language to share learning intentions with students (e.g., "We are learning to make a good guess (prediction) about what is likely to happen next in the story").	0.56			
Uses assessment techniques to facilitate class discussion (e.g., brainstorming).		0.46		
Uses questions to elicit students' prior knowledge on a topic.		0.37		
Encourages students to share the questioning role with the teacher during lessons (e.g., the teacher routinely invites pupils to question their peers' contributions to discussions).		0.49		
Stimulates students to explain to others what they are learning (e.g., if a visitor came to the classroom, I could articulate what I'm learning in terms that identify the knowledge, skills, concepts and/or attitudes being developed).		0.40		
Uses written feedback on pupils' work that goes beyond the use of grades and comments such as "well done" to specify what students have achieved and what they need to do next.			0.39	
Uses teacher-made tests diagnostically to identify strengths and needs in the learning of the student (e.g., identifying common mistakes in the addition of fractions).			0.46	
Uses tests diagnostically to tailor the instruction to the needs of the students by taking into account the strengths and needs of students (e.g., extra lessons on adding fractions).			0.50	
Uses the information on the progress of students to provide feedback to students.			0.38	
Gives students an opportunity to indicate how challenging they anticipate the learning will be at the beginning of a lesson or activity (e.g., by using traffic lights).				0.48
Encourages students to record their progress using, for example, learning logs.				0.65
Stimulates students to assess and comment on each other's work (e.g., they are taught how to use the success criteria for a lesson to judge another pupil's piece of work).				0.74

Items	**Rotated factor loadings**			
My teacher	LISC	EE	FB	PSA
Encourages students to use a range of assessment techniques to review their own work (e.g., rubric, traffic lights, thumbs up/down, two stars and a wish).				0.52
Maintains a visual record of students' progress to celebrate students' learning and to show areas of/for development (e.g., a bulletin board displaying progress in story writing over a term).				0.67
Sets time aside during lessons to allow for self- and peer-assessment.				0.48
Items (*N*)	4	4	4	6
Cronbach's α	.72	.72	.79	.79

Note: LISC = sharing learning intentions and success criteria; EE = eliciting evidence; FB = feedback; PSA = peer- and self-assessment.

APPENDIX C: CODING OF THE TEACHING SITUATIONS LINKED TO THE AFL KEY STRATEGIES

Description of teaching situation	Coded description of teaching situation
The teacher starts by stating the lesson goals for this section, "students learn to master plural forms". The teacher names five words (singular form) and the students write down these words on small whiteboards (in pairs) and complete the plurals. Then all students show their boards to the teacher. The teacher looks and says something about the answers, explains what is good and discusses briefly why (e.g., 's). The students have to look at the right answer projected on the whiteboard in front of the classroom and keep track of how much they got right. The teacher summarises that if students got everything right (5 words), they have mastered the basic rules of plurals. This is repeated twice with another row of 5 words in the singular. At the end, the teacher askes the students to add up all their points. Out of a total of 15 words, the teacher notes the number of points each student obtained. Students scoring less than 10 points have to work with repetition exercises on plural forms. Pupils scoring 10 points or more have to work on an exploratory assignment about the use of the 's' in the Dutch language.	The teacher starts by stating the lesson goals for this section, "students learn to master plural forms" (**KS1**). The teacher names five words (singular form) and the students write down these words on small whiteboards (in pairs) and complete the plurals. Then all students show their boards to the teacher (**KS2**). The teacher looks and says something about the answers, explains what is good (**KS3**) and discusses briefly why (e.g., 's) (**KS4**). The students have to look at the right answer projected on the whiteboard in front of the classroom and keep track of how much they got right (**KS3**). The teacher summarises that if students got everything right (5 words), they have mastered the basic rules of plurals (**KS1, KS4**). This is repeated twice with another row of 5 words in the singular (**KS2, KS3**). At the end, the teacher askes the students to add up all their points (**KS3, KS4**). Out of a total of 15 words, the teacher notes the number of points each student obtained (**KS2**). Students scoring less than 10 points have to work with repetition exercises on plural forms (**KS5**). Pupils scoring 10 points or more have to work on an exploratory assignment about the use of the 's' in the Dutch language (**KS5**).

Note: KS1 = clarifying, sharing and understanding learning goals and criteria for success; KS2 = eliciting evidence of student learning (including self- and peer-assessment); KS3 = analysing student responses; KS4 = communicating about results (including feedback); KS5 = taking concrete actions to adjust teaching and/or learning.

APPENDIX D: PERFORMANCE OBJECTIVES FOR ASSESSMENT FOR LEARNING

Skill	Performance objective
Determining learning goals for instructional topic	The teacher decides which part of the curriculum should be covered and translates the subject content given in the curriculum for the subject into learning goals. The teacher prepares the lesson series, focusing on determining which learning goals should be mastered at the end of the period (usually for a period of four to eight weeks). The teacher formulates learning goals based on a critical reflection of the instructional method, curriculum and student levels. The learning goals meet the following criteria: 1) specific; 2) ambitious; and 3) suitable for the target group.
Analyzing students' learning process	The teacher analyzes the available information about students' learning using the evaluation of the previous lesson, the intended learning goals and determines the level of differentiation needed to meet the needs of the diverse learners.
Determining learning goals and criteria for success	The teacher determines the learning goals for the lesson and describes the teaching needs related to these goals. The teacher formulates learning goals in understandable language for the entire group of students. Based on the learning goals and analysis, the teacher formulates (differentiated) criteria for success.
Determining lesson approach	The teacher formulates organizational and didactic approaches, following logically from the preceding analysis, and related to the learning goals and criteria for success that were formulated. The teacher describes what materials will be used and how within-lesson transitions will take place. The teacher determines how to share learning goals and criteria for success with students and how students will be engaged in these activities in class.
Determining approach for data collection	The teacher determines how and when information about students' learning will be collected during the lesson, in relation to learning goals and criteria for success.
Sharing learning goals and criteria for success	The teacher shares learning goals and criteria for success with students and makes sure they are clear for the students by calling attention to the goals and criteria at different moments during the lesson and encouraging students to be owners of their learning.
Evidence informed follow up	The teacher stimulates students' learning processes by tailoring instruction to the students' learning needs, giving tips, and/or asking questions, to help students correcting errors and misconceptions. *Collecting information on students' learning progress:* during the lesson, the teacher collects information about students' learning related to the learning goals and criteria for success. The teacher shows that mistakes are allowed: this is about insight into learning rather than summative evaluation or grading (realizing a safe and stimulating learning climate). *Analyzing and interpreting information:* the teacher (preferably together with students) analyzes and interprets the information collected during the lesson, using the intended learning goals and criteria for success, to address mistakes and misconceptions. The conclusions from these analyses are shared with students in order to inform follow-up actions.

Skill	Performance objective
Checking students' learning outcomes	The teacher together with the students evaluates the lesson by eliciting information to what extent the learning goals of the lesson were achieved.
Determining informed follow up for the next lesson	The teacher reflects on the lesson in relation to the learning goals and criteria for success – whether the goals were achieved or not – and determines what is needed for the next lesson. *Analyzing students' learning process:* the teacher analyzes and interprets the information collected to determine students' learning progression. *Reflection on teacher action during the lesson:* the teacher reflects on their acting in relation to the lesson preparation and intended learning goals for students.

APPENDIX E: OVERVIEW OF THE INTRODUCTORY MEETING OF THE AFL-TPD PROGRAM

Introductory meeting

Half-day group meeting, involving a series of learning tasks to kick off the TPD program. In general, the learning tasks are meant to provide a shared mental model of the whole cycle of AFL with its phases and principles and to address possible misconceptions at the start of the program. During this session, the coach introduces the emphasized aspects of task class 1 (enacting a lesson: sharing learning goals and criteria for success)

Learning task	Description
1.1: Shared mental model	A card game for developing a shared mental model of AfL. Teachers discus the concept of AfL based on their prior knowledge about formative assessment
1.2: Misconceptions	Using a set of prepared statements related to frequent misconceptions concerning AfL, the coach engages the group in a discussion to resolve differing perspectives on what AfL is and how it should be used.
1.3: Domain knowledge	Teachers compare two theoretical models describing AfL and indicate the advantages and disadvantages of both models.
1.4: Dialogic use of exemplars	Teachers watch video fragments from three different examples showing how an expert teacher shares learning goals and success criteria with the students, followed by ranking the exemplars and having a short group discussion.
1.5: Formative assessment	The four main skills of the AfL process (preparing a lesson series, preparing a lesson, lesson enactment, lesson evaluation) and successive skills/performance objectives are presented. Teachers indicate which skills they have already mastered well and which skills require attention.
1.6: Modelling example	Teachers watch a fragment from a modelling example showing how an expert teacher shares learning goals and success criteria with the students, followed by a group discussion.

Supporting information

(Gulikers & Baartman, 2017; Leahy, Lyon, Thompson, & Wiliam, 2005)

APPENDIX F: OVERVIEW OF TASK CLASS 1 IN THE AFL-TPD PROGRAM

Task class 1	
In the first task class, teachers practice individually in their daily teaching practice, focusing on only a subset of the skills: prepare a number of lessons that are illustrative for your way of using assessment for learning in teaching practice. Record one of the lessons on video. Select a snippet of up to 5 minutes in which you show your way of sharing learning goals and criteria for success with students. Upload the video clip to the online learning environment at Pitch2Peer.	
Learning tasks	For six weeks, teachers practice the newly learned skills in their daily teaching practice. It is essential that teachers learn how to apply the whole cycle for effective AfL while executing these skills.
Supporting information	*Cognitive feedback* The coach observes the recorded lessons and provides individual cognitive feedback on teachers' performance by using an online tool for uploading video recordings and providing feedback. The feedback focuses on sharing learning goals and criteria for success with students. *Reflection session* The task class ends with a group meeting in which teachers share their experiences and reflect on their improvements and personal learning goals. During this session, the coach also introduces the emphasized aspects of the next task class.

PUBLICATIONS AND PRESENTATIONS

Scientific publication

Accepted manuscripts

Wolterinck, C. H. D., Poortman, C. L., Schildkamp, K., Visscher, A. J. (2022). Assessment for Learning: Developing the Required Teacher Competencies. *European Journal of Teacher Education, 1 – 19.* doi: 10.1080/02619768.2022.2124912

Manuscripts submitted for publication

Wolterinck, C. H. D., Poortman, C. L., Schildkamp, K., Visscher, A. J. (submitted). Key Stakeholder Voices: Investigating Student Perceptions of Teachers' Use of Assessment for Learning

Wolterinck, C. H. D., Van Geel, M., Schildkamp, K., Visscher, A. J. (submitted). A Cognitive Task Analysis of the Skills that Assessment for Learning Requires from Teachers.

Wolterinck-Broekhuis, C. H. D., Poortman, C.L., Schildkamp, K., Visscher, A.J. (submitted). Strengthening Student Self-regulation: Investigating the Effect of Teacher Professional Development for Assessment for Learning.

Professional Publications

Wolterinck, C. H. D. (2022). Formatief Toetsen in het Voortgezet Onderwijs: Onderzoek en Professionaliseringsprogramma. *Beter Begeleiden, 12(6),* 24-27

International Presentations

Wolterinck C. H. D., De Vries, J. A., Poortman, C. L., Schildkamp, K., Visscher, A. J., (2021, March). *Teacher Professional Development Programmes in Assessment for Learning*. Paper presented at International Congress for School Effectiveness and Improvement, ICSEI 2021, online.

Wolterinck C. H. D., Poortman, C. L., Schildkamp, K., Visscher, A. J., (2020, January). *The Effects of an Assessment for Learning Professional Development Program.* Paper presented at International Congress for School Effectiveness and Improvement, ICSEI 2020, Marrakesh, Morocco.

Wolterinck C. H. D., De Vries, J. A., Poortman, C. L., Schildkamp, K., Visscher, A. J., (2020, January). *The Effects of a TPD Program on Formative Assessment in the Classroom: The Development of an Observation Scheme.* Round table session at International Congress for

School Effectiveness and Improvement, ICSEI 2020, Marrakesh, Morocco.

Wolterinck C. H. D., Schildkamp, K., Van Merriënboer, J. J. G., Visscher, A. J., (2019, January). Assessment for Learning: Assessment for Learning: *A Cognitive Task Analyses in Secondary Education.* Paper presented at International Congress for School Effectiveness and Improvement, ICSEI 2019, Stavanger, Norway.

Wolterinck C. H. D., Morskieft, T., De Kuijer, M., Schildkamp, K., (2019, January). *Experience the Implementation of Formative Assessment from a Policy, Practice and Research Perspective.* Paper presented at International Congress for School Effectiveness and Improvement, ICSEI 2019, Stavanger, Norway.

Wolterinck C. H. D., Schildkamp, K., Van Merriënboer, J. J. G., Visscher, A. J., (2018, January). Assessment for Learning: *A Cognitive Task Analyses of the Explicit and Implicit Knowledge and Skills of Teachers in Secondary Education.* Paper presented at International Congress for School Effectiveness and Improvement, ICSEI 2018, Singapore.

Wolterinck, C.H.D., Adamska, M. I., Schildkamp, K., Visscher, A. J., (2016, November). *Formative Assessment from the Perspective of Teachers and Students*. Poster presented at ICO's International Fall School 2016, Bad Schussenried, Germany.

Wolterinck, C. H. D., Kippers, W. B., Schildkamp, K., Poortman, C. L. (2016, April). *Factors Influencing the Use of Formative Assessment in the Classroom.* Paper presented at American Educational Research Association meeting, AERA 2016, Washington, United States.

National Presentations

Wolterinck, C. H. D., Schildkamp, K., (2021, November). Evidence-informed Formatief Toetsen: van Theorie naar Praktijk. Workshop gepresenteerd tijdens 4th conferentie Formatief Evalueren 2021, online.

Wolterinck, C. H. D., Trimbos B., (2020, November). InformED, een Neus voor Kwaliteit Ontwikkelen: Professionaliseringstraject voor Docenten in het VO. Workshop gepresenteerd tijdens 3rd conferentie Formatief Toetsen in de Klas 2020, online.

Wolterinck, C. H. D., Schildkamp, K., (2020, November). *Tips om Formatief Toetsen in te Zetten in de Les.* Workshop gepresenteerd tijdens 3rd conferentie Formatief Toetsen in de Klas 2020, online.

Wolterinck, C. H. D., Schildkamp, K. (2020). *Tips om Formatief Toetsen in te Zetten in de Les.* Workshop gepresenteerd tijdens docentencongres van Stichting Schoolinfo 2020, online.

Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., Visscher, A. J., Van Merriënboer, J. J. G., (2019, November). *Tools om zicht te krijgen op de formatieve cultuur in school.* Workshop gepresenteerd tijdens 2nd conferentie Formatief Toetsen in de Klas 2019, Nieuwegein.

Wolterinck, C. H. D. (2019, September). Formatief Toetsen in de Klas, van Theorie naar Praktijk. Workshop gepresenteerd tijdens de conferentie Twents Meesterschap 2019, Universiteit Twente, Enschede.

Wolterinck, C. H. D., Poortman, C. L. (2018, November). *Formatief Toetsen in de Klas, van Theorie naar Praktijk*. Workshop gepresenteerd tijdens 1st conferentie Formatief Toetsen in de Klas (2018), Nieuwegein.

Wolterinck, C. H. D., Schildkamp, K., Van Merriënboer J. J. G., Visscher, A. J. (2018, June). *Formatief Toetsen: Een Cognitieve Taakanalyse naar de Kennis en Vaardigheden van Docenten in het Voortgezet Onderwijs. Paper gepresenteerd op de Onderwijs Researchdagen* (ORD) 2018, Nijmegen.

Wolterinck, C. H. D., Kippers, W. B. (2018, January). *Formative Assessment in Classroom Practice*. Workshop gepresenteerd tijdens, ResearchED 2018, Amsterdam.

Wolterinck, C. H. D., Kippers, W. B. (2018, January). *Formatief Toetsen, Wat Vraagt Dat van Docenten en Leerlingen?* Workshop gepresenteerd tijdens de conferentie Twents Meesterschap 2018, Universiteit Twente, Enschede.

Wolterinck, C. H. D., Kippers, W. B., Schildkamp, K., (2017). *Formatief Toetsen voor het ontwikkelen van een feedbackcultuur*. Workshop gepresenteerd op Schoolleidingendag Stichting Carmelcollege 2017,

Wolterinck, C. H. D., Kippers, W. B., Schildkamp, K., Poortman, C. L. (2016, May). *Formatief Toetsen in de Klas: Bevorderende en Belemmerende Factoren. Paper gepresenteerd op de Onderwijs Researchdagen* (ORD) 2016, Rotterdam.

ICO PHD DISSERTATION SERIES

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the ICO Partner Universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, Radboud University Nijmegen, University of Amsterdam, University of Antwerp, University of Ghent, KU Leuven, Université Catholique de Louvain, University of Groningen, University of Twente, Utrecht University, Vrije Universiteit Amsterdam, and Wageningen University, and formerly Tilburg University (until 2002).

458. Vaessen, B.E. (08-01-2021). *Students' perceptions of assessment and student learning in higher education courses.* Eindhoven: Eindhoven University of Technology
459. Maureen, I.Y. (15-01-2021). *Story time in early childhood education: designing storytelling activities to enhance (digital) literacy development.* Enschede: University of Twente
460. Van Alten, D.C.D. (19-03-2021). *Flipped learning in secondary education history classrooms: what are the effects and what is the role of self-regulated learning.* Utrecht: Utrecht University
461. Gestsdóttir, S.M. (08-03-2021). *Observing history teaching: historical thinking and reasoning in the upper secondary classroom.* Amsterdam: University of Amsterdam
462. Chim, H.Q. (30-03-2021). *Physical Activity Behavior and Learning in Higher Education.* Maastricht: Maastricht University
463. Krijnen, E. (15-04-2021). *Family literacy in context: Exploring the compatibility of a family literacy program with children's homes and schools.* Rotterdam Erasmus University Rotterdam
464. Stolte, M. (07-05-2021). *(In)attention for creativity: Unraveling the neural and cognitive aspects of (mathematical) creativity in children.* Utrecht: Utrecht University
465. Rathé, S. (12-05-2021). *Focusing on numbers – An investigation of the role of children's spontaneous focusing on Arabic number symbols in early mathematical development.* Leuven: KU Leuven
466. Theelen, H. (12-05-2021). *Looking around in the classroom. Developing preservice teachers' interpersonal competence with classroom simulations.* Wageningen: Wageningen University
467. De Jong, L.A.H. (20-05-2021). *Teacher professional learning and collaboration in secondary schools.* Leiden: Leiden University
468. Sincer, I. (20-05-2021). *Diverse Schools, Diverse Citizens? Teaching and learning citizenship in schools with varying student populations.* Rotterdam: Erasmus University

Rotterdam

469. Slijkhuis, E.G.J. (20-05-2021). *Fostering active citizenship in young adulthood.* Groningen: University of Groningen
470. Groothuijsen-Vrancken, S.E.A. (02-06-2021). *Quality and impact of practice-oriented educational research.* Utrecht: Utrecht University
471. Hingstman, M. (07-06-2021). *Supporting struggling students: prevention and early intervention with Success for All.* Groningen: University of Groningen
472. Gerdes, J. (14-06-2021). *All inclusive? Collaboration between teachers, parents and child support workers for inclusive education in prevocational schools.* Amsterdam: Vrije Universiteit Amsterdam
473. Bai, H. (18-06-2021). *Divergent thinking in young children.* Utrecht: Utrecht University
474. Wijnker, W. (23-06-2021). *The Unseen Potential of Film for Learning: Film's Interest Raising Mechanisms Explained in Secondary Science and Math.* Utrecht: Utrecht University
475. Brummer, L. (24-09-2021). *Unrooting the illusion of one-size-fits-all feedback in digital learning environments.* Groningen: University of Groningen
476. Veldman, M.A. (01-07-21). *Better together, social outcomes of cooperative learning in the first grades of primary education.* Groningen: University of Groningen
477. Wang, J. (06-07-2021). *Technology integration in education: Policy plans, teacher practices, and student outcomes.* Leiden: Leiden University
478. Zhang, X. (06-07-2021). *Teachers' teaching and learning motivation in China.* Leiden: Leiden University
479. Poort, I.C. (02-09-2021). *Prepared to engage? Factors that promote university students' engagement in intercultural group work.* Groningen: University of Groningen
480. Guo, P. (07-09-2021). *Online project-based higher education Student collaboration and outcomes.* Leiden: Leiden University
481. Jin, X. (21-09-2021). *Peer feedback in teacher professional development.* Leiden: Leiden University
482. Atherley, E.N. (27-09-2021). *Beyond the struggles: Using social-developmental lenses on the transition to clinical training.* Maastricht: Maastricht University
483. Martens, S.E. (15-10-2021). *Building student-staff partnerships in higher education.* Maastricht: Maastricht University
484. Ovbiagbonhia, R. (08-11-2021). *Learning to innovate: how to foster innovation competence in students of Built Environment at universities of applied science.* Wageningen: Wageningen University
485. Van den Boom-Muilenburg , S.N. (11-11-2021). *The role of school leadership in schools that sustainably work on school improvement with professional learning communities.*

Enschede: University of Twente

486. Sachishal, M.S.M. (11-11-2021). *Science interest - Conceptualizing the construct and testing its predictive effects on current and future behavior.* Amsterdam: University of Amsterdam
487. Meeuwissen, S.N.E. (12-11-2021). *Team learning at work: Getting the best out of interdisciplinary teacher teams and leaders.* Maastricht: Maastricht University
488. Keijzer-Groot, A.F.J.M. (18-11-2021). *Vocational identity of at-risk youth – Tailoring to support career chances.* Leiden: Leiden University
489. Wolthuis, F. (25-11-2021). *Professional development in practice. Exploring how lesson study unfolds in schools through the lens of organizational routines.* Groningen: University of Groningen
490. Akkermans-Rutgers, M. (06-12-2021). *Raising readers. Stimulating home-based parental literacy involvement in the first, second, and third grade.* Groningen: University of Groningen
491. Hui, L. (06-12-2021). *Fostering Self-Regulated Learning: The Role of Perceived Mental Effort.* Maastricht: Maastricht university
492. Jansen, D. (08-12-2021). *Shadow education in the Netherlands: The position of shadow education in the educational landscape and students' school careers.* Amsterdam: University of Amsterdam
493. Kamphorst, F. (15-12-2021). *Introducing Special Relativity in Secondary Education.* Utrecht: Utrecht University
494. Eshuis, E.H. (17-12-2021). *Powering Up Collaboration and Knowledge Monitoring: Reflection-Based Support for 21st-Century Skills in Secondary Vocational Technical Education.* Enschede: University of Twente

DANKWOORD (ACKNOWLEDGEMENTS)

De bestemming is bijna bereikt! Hierdoor ben ik blij dat ik nu zover ben dat ik het dankwoord kan gaan schrijven. In de afgelopen jaren heb ik veel opzij gezet om naast mijn baan op SG Marianum een promotieonderzoek te volbrengen. Van verschillende kanten heb ik de kans gekregen om mijn rugzak goed te vullen. Mijn praktijkervaring als docent en schoolleider in het voortgezet onderwijs kwamen me goed van pas bij het onderzoek dat ik onder handen had: *Hoe kunnen docenten vaardig worden in formatief toetsen in de les?* Het was een leerzame reis die me bracht naar verschillende scholen in Nederland en naar conferenties in Singapore, Washington, Noorwegen en Marokko. Zo leerde ik nieuwe mensen kennen, leerde ik mezelf kennen. Regelmatig moest ik uit mijn comfortzone stappen, maar waarop ik vervolgens trots kon terugkijken. Een promotieonderzoek doe je niet alleen, dat doe je samen met een team. Daar wil ik meerdere mensen voor bedanken.

Het was **Kim Schildkamp** die mij in 2015 voorstelde om een promotieonderzoek te gaan doen. We moesten maar eens gaan praten! Onze gedeelde interesse was datagebruik door leraren samen met leerlingen in de les om te komen tot instructieverbetering en het uiteindelijke doel: het leren van leerlingen verbeteren. Het thema voor de reis ontstond: 'Formatief Toetsen in de les'. Kim, wat heb ik ontzettend veel van je kunnen leren: een onderzoeker worden, een netwerker zijn, groeien als schrijver van wetenschappelijke artikelen. Naast Kim wil ik **Cindy Poortman** en **Adrie Visscher** bedanken als mijn vaste begeleiders in het gehele traject. Cindy, ik heb bewondering voor de manier waarop je mij keer op keer voorzag van de nodige (kritische) feedback, maar altijd opbouwend en altijd op zoek naar een betere versie. Ik heb me daardoor niet alleen professioneel, maar ook persoonlijk enorm kunnen ontwikkelen. Adrie, ook jij hebt in dit leerzame traject een belangrijke rol gehad. Kritisch – ja, maar opbouwend – meestal met humor, dat ook. En altijd respectvol. Bedankt!

Stichting Carmelcollege heeft mij de ruimte gegeven om twee dagen in de week als onderzoeker op de UT aan de slag te gaan. Er ging een wereld voor mij open. Tussen de onderzoekers op de UT kon ik veel halen en als praktijkdeskundige had ik ook het een en ander te brengen. Hiervoor wil **Tom Morskieft** en **Jan Put** bedanken. Jullie hebben voor mij binnen Carmel en binnen Marianum een weg gebaand. Op mijn school Marianum heb ik steun gehad van veel collega's om het uitdagende traject te combineren met mijn overige werkzaamheden. Dankjewel **Gitti Burema** voor je support. En dankjewel **Nicky Konings**, jij hebt mij durende het hele traject op de voet kunnen volgen en op sommige momenten moed kunnen inpraten. Wat ben ik trots op de foto die jij maakte en die nu de voorkant van mijn proefschrift siert!

Als buitenpromovenda was ik slechts twee dagen in de week op de UT, maar desondanks heb ik veel collegialiteit ervaren op de vijfde verdieping. De vele lunchwandelingen, etentjes en koffietafelgesprekken had ik niet willen missen. **Anne Tappel**, wat ben ik blij dat ik je heb leren kennen. We zijn collega's op de UT, bij Carmel en op Marianum en ik hoop dat we dat nog lang blijven. Ik bewonder je nuchtere kijk, je positieve houding, je humor. **Jitske de Vries**, samen vormden wij het Formatief-Toetsenduo. We gingen samen naar de trainingsbijeenkomsten voor docenten, naar conferenties en we verzorgden diverse presentaties en workshops op scholen. Fijn dat we dit jaar allebei ons promotietraject hebben afgerond. **Elske van den Boom - Muilenburg**, bedankt voor het sparren over onze onderzoeken waarbij we allebei heel nadrukkelijk de wetenschap en de praktijk hebben weten te verbinden. **Wilma Kippers (†)**, bedankt voor onze samenwerking. Je blijft altijd in mijn gedachten.

Ik wil alle leerlingen en docenten bedanken die hebben bijgedragen aan mijn onderzoek door de tijd nemen om vragenlijsten in te vullen, of interviews te geven. Speciale dank gaat uit naar de docenten die deelnamen aan de taakanalyse en aan de trainingen die we hebben gegeven. **Bas Trimbos**, **Inge Jansen** en **Emiel de Kleijn** van SLO, hartelijk dank voor jullie deskundigheid en collegialiteit tijdens de ontwikkeling van het trainingstraject en het trainen van de docenten. Zonder jullie en zonder alle leerlingen en docenten was dit proefschrift er niet geweest.

Lieve **familie Broekhuis** en **familie Wolterinck**, jullie vormen een belangrijke uitvalbasis. Ik wil jullie bedanken voor jullie interesse in mijn werk en voor de steun die ik heb ervaren, in periodes waarin het mee zat en ook in de periodes waarin het tegen zat. In de afgelopen jaren is veel vrije tijd gaan zitten in het promotietraject. Nu dat klaar is, beloof ik jullie dat het nu eerst tijd is voor andere dingen!

Lieve **Lisanne**, **Eline** en **Femke**. Mama is eindelijk klaar met haar studie! Wat ben ik geweldig trots op jullie, op wie jullie zijn, hoe jullie je ontwikkelen, hoe jullie het doen. Ik wil jullie bedanken voor de steun en het geduld dat jullie hebben gehad als ik weer eens niet gezellig op de bank kon komen zitten. Tot slot, lieve **Han**. Zonder de onvoorwaardelijke steun van jou had ik dit niet kunnen volbrengen. Bedankt voor de ruimte die je me hebt gegeven, dat je er altijd voor me bent. Bij jou voel ik me thuis en kan ik zijn wie ik ben. Samen met de meiden zijn we een hecht team!

Christel Wolterinck
Oktober 2022

PHOTOCOVER

‘The assignment’

Pupils of SG Marianum model their self-portrait.
The assignment is unambiguous, the output very diverse.
The pupils: all alike, but so different.
An assignment for the school.

FOTO OMSLAG

‘De opdracht’

Leerlingen van SG Marianum boetseren hun zelfportret.
De opdracht is eenduidig, de opbrengst heel divers.
De leerlingen: allen gelijk, maar zo verschillend.
Een opdracht voor de school.