



Key stakeholder voices: Investigating student perceptions of teachers' use of assessment for learning

Christel H. D. Wolterinck-Broekhuis¹ · Cindy L. Poortman¹ ·
Kim Schildkamp¹ · Adrie J. Visscher¹

Received: 25 July 2022 / Accepted: 30 December 2023
© The Author(s) 2024

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

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, 1998). Assessment has two important functions: formative, to provide support for future learning, and summative, to

✉ Christel H. D. Wolterinck-Broekhuis
c.h.d.wolterinck@utwente.nl

¹ Faculty of Behavioral Management, and Social Sciences, University of Twente, Section ELAN, P.O. Box 217, 7500AE Enschede, The Netherlands

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. (2016), 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:

RQ: To what extent do students experience Assessment for Learning strategy use in their classroom?

2 Theoretical framework

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, 1998; Klenowski, 2009). Black and Wiliam (2010) identified five core strategies for AfL practice in the classroom. Each of these core strategies is described below.

2.1.1 Clarifying, sharing and understanding learning intentions 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, 1998; Crisp, 2012; Heitink et al., 2016; Wiliam & Leahy, 2015). The term *learning intentions* 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 intentions 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).

2.1.2 Eliciting evidence of student learning

This strategy involves creating opportunities to gather evidence of student learning through (informal) assessment (Black & Wiliam, 1998; 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).

2.1.3 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).

2.1.4 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, 1998; 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 and are enthusiastic about it. *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 (William & 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; Baartman & Gulikers, 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 (William & Leahy, 2015).

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).

3 Method

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 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

Table 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

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 random 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.

3.2 Instruments

3.2.1 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 intentions 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 clusters 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 intentions 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 (0.70-0.80; (Field, 2013); see Appendix B.

3.2.2 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 analyzed 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).

4 Results

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. The highest mean score was for teachers' sharing of learning intentions and success criteria, at 3.07 ($N=604$; $SD=0.86$), which can be interpreted as emerging (happening in approximately 50% of the lessons). For example, the mean score for "My teacher matches success criteria with learning intentions" was 3.03, and the mean for "My teacher 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')" was 3.13, based on student reports. Nearly three-quarters of the students (71.0%) indicated that sharing learning intentions 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.

Students reported experiencing teachers' use of the strategy of eliciting evidence less frequently than sharing learning intentions and success criteria, with a mean score of 2.58 ($N=586$; $SD=0.79$), which can be interpreted as between sporadic and emerging (25–50% 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. 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).

Based on the questionnaire responses, the mean score for teachers' use of the strategy of feedback was 2.33 ($N=544$; $SD=0.95$), which can be interpreted as between sporadic and emerging (25–50% of the lessons). 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

Table 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%)

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

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. 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.

Peer- and self-assessment was the least often used strategy according to students, with a mean score of 1.94 ($N=526$; $SD=0.83$), which can be interpreted as between (almost) never and sporadic (happening in less than 25% of the lessons). 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 stars and a wish)” was scored 2.18 on average. 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.

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 3) indicated that the difference between the scores given by male and female students ($(M=3.07, SD=0.81$ and $M=2.89, SD=0.81$, respectively) for sharing learning intentions and success criteria was statistically significant: $t(519)=2.75, p<0.01$. The effect size, $d=0.22$, represents a small effect (Cohen, 1992; Field, 2013). In addition, the difference between the scores given by male and female students ($(M=2.65, SD=0.92$ and $M=2.35, SD=0.94$, respectively) for feedback was statistically significant: $t(474)=3.68, p<0.01$. The effect size, $d=0.32$, represents a small 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.

4.3 Cluster analysis

The cluster analysis revealed three different clusters based on the extent of AfL strategy use students reported experiencing (Table 4):

1. Cluster 1 (high); students ($N=82$) who belong to the first cluster reported that they experienced all four AfL strategies used 50–74% of their lessons.
2. Cluster 2 (middle); students belonging to the second cluster ($N=183$) reported that they experienced the AfL strategy of sharing learning intentions and suc-

Table 3 Results of independent samples t-tests for gender and subjects

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

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

Table 4 Clusters based on perceived extent of AfL strategy use

AfL strategies	Cluster 1: High (<i>N</i> =82)		Cluster 2: Mid- dle (<i>N</i> =183)		Cluster 3: Low (<i>N</i> =177)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
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

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

- cess 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.
3. Cluster 3 (low); students who belong to cluster three (*N*=177) experienced the use of sharing learning intentions 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.

5 Discussion and conclusions

The importance of AfL has been advocated by researchers for many years (e.g., Wiliam, 2011), but not many studies have investigated students' perceptions of the extent to which AfL is implemented in the classroom. 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.

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 intentions 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). 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).

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 AfL, 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 *high* cluster in which students reported the use of all four AfL strategies as between emerging and established (50–74% of the lessons); a *middle* cluster in which the strategy use experienced was more spread,

from (almost) never to established (0–74% of the lessons); and a *low* cluster 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 intentions 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).

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 someone's teaching can be assessed efficiently multiple times (equal to the number of students in a class at one and the same moment 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, Baartman and Gulikers (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 be an interesting focus, in order to increase student engagement in AfL practices and strengthen their self-regulated learning.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11092-024-09428-7>.

Data availability The data that support the findings of this study are available from the corresponding author, [Wolterinck-Broekhuis, C.H.D.], upon reasonable request.

Declarations

Competing interests The authors report there are no competing interests to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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. 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>
- Baartman, L., & Gulikers, J. (2017). Doelgericht professionaliseren: Formatieve toetspraktijken met effect! Wat DOET de docent in de klas? [Goal-oriented professionalisation: Formative assessment practices with effect! What does the teacher do in the classroom?]. 1–109. Retrieved 2018 April 12 from <https://www.nro.nl/nro-projecten-vinden/?projectid=405-15-722-doeelgericht-professionaliseren-formatieve-toetscompetenties-met-effect>.

- Bailey, A. L., & Heritage, M. (2018). *Self-Regulation in Learning: The Role of Language and Formative Assessment*. ERIC.
- 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>
- 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. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- 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>
- Borgen, F. H., & Barnett, D. C. (1987). Applying cluster analysis in counseling psychology research. *Journal of Counseling Psychology*, 34(4), 456.
- 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., & 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>
- 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.
- 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>
- 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.
- 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. UCLA. ProQuest ID: Dockterman_ucla_0031D_16031. Merritt ID: ark:/13030/m5b32qqc. Retrieved from <https://escholarship.org/uc/item/Isr500x5>
- Donahue, J. M. (1994). *Student perceptions of their teachers, their school, and themselves as learners*. Iowa State University.
- 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. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.
- Field, A. (2000). Cluster analysis. <https://www.discoveringstatistics.com/docs/cluster.pdf>. Accessed 12 April 2018
- 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>
- 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>

- 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/educ.2021.750281>
- Haertel, E.H. (2013). *Reliability and validity of inferences about teachers based on student test scores*. Center for Research on Human Capital and Education, Research and Development, Educational Testing Service, Princeton, The National Press Club, Washington DC. <https://www.ets.org/Media/Research/pdf/PICANG14.pdf>
- 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., & 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. (2016). Assessment for learning: Co-regulation in and as student–teacher interaction. In *Assessment for learning: Meeting the challenges of implementation* (pp. 327–343). https://doi.org/10.1007/978-3-319-39211-0_19
- 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>
- Hubbard, L., Datnow, A., & Prunyn, 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>
- Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Bill & Melinda Gates Foundation. <https://files.eric.ed.gov/fulltext/ED540960.pdf>
- 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>
- Lai, M. K., & Schildkamp, K. (2013). Data-based decision making: An overview. In *Data-based Decision Making in Education* (pp. 9–21). Netherlands, Springer. 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>
- 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>
- Ministry of Education, Culture & Science. (2016). <http://www.scholenopdekaart.nl>, <http://www.statline.cbs.nl>. Accessed 12 April 2018
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education. Evidence from recent RAND research*. Retrieved from Santa Monica, CA: RAND Corporation. https://www.rand.org/content/dam/rand/pubs/occasional_papers/2006/RAND_OP170.pdf

- Marshall, B., & Drummond, M. Jane. (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>
- 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>
- 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>
- 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. *European Journal of Psychology of Education*, 34(3). <https://doi.org/10.1007/s10212-018-0407-8>
- 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.
- Rollett, W., Bijlsma, H., Röhl, S. (2021a). Student feedback on teaching in schools: Current state of research and future perspectives. In: W. Rollett, H. Bijlsma, & S. Röhl (Eds.), *Student feedback on teaching in schools*. Springer, Cham. https://doi.org/10.1007/978-3-030-75150-0_16
- Rollett, W., Bijlsma, H., & Röhl, S. (2021b). Student feedback on teaching in schools: Using student perceptions for the development of teaching and teachers. Springer, Cham. <https://doi.org/10.1007/978-3-030-75150-0>
- 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>
- 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.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.
- 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>
- 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, 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>
- 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>
- 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*. West Palm Beach, FL: Learning Sciences International.
- 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>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.