Studies in Educational Evaluation 70 (2021) 101027

Contents lists available at ScienceDirect

Studies in Educational Evaluation

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

We agree on what we see: Teacher and student perceptions of formative assessment practice

M.J. Veugen *, J.T.M. Gulikers, P. den Brok

Wageningen University and Research, Education and Learning Sciences, Hollandseweg 1, 6706 KN, Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Teacher evaluation
Evaluation utilization
Formative assessment
Student perceptions

ABSTRACT

This study examines teacher and student perceptions of formative assessment (FA) activities used by teachers in the classroom. These activities are divided into five phases that together comprise the FA cycle: (1) clarifying expectations, (2) eliciting responses, (3) analysing and interpreting responses, (4) communicating about responses and (5) adjusting teaching and learning. Reliable questionnaires were used to measure the perceptions of 96 teachers and 1,095 students with regard to FA practice. Paired t-tests indicated no differences between the perceptions of teachers and students, except for with regard to clarifying expectations. Teachers used FA activities primarily to clarify expectations and elicit student responses, and they were least likely to apply them to adjust teaching and learning. The results suggest that the framework of the FA cycle could be used as an analytical lens for the reliable evaluation of the FA activities of teachers.

1. Introduction

Formative assessment has been recognized as one of the most effective strategies for improving teaching and learning in secondary education, and many scientists and teacher educators argue that it should be a core element of teaching and learning (Black & Wiliam, 2018; OECD, 2005). According to Black and Wiliam (2009), assessments are formative when they gather information about the learning processes of students in order to enhance decision-making about follow-up actions and to adjust teaching and learning to meet student needs. Formative assessment (FA) is an ongoing process of monitoring the learning process in order to continue teaching and promote learning based on the current performance of students (Pat-El, Tillema, Segers, & Vedder, 2015). Despite an increase in the use of FA in classroom practice (Carless, 2017), previous studies have shown that it has yet to become common teaching practice in secondary education (Deneen et al., 2019; Kippers, Wolterinck, Schildkamp, Poortman, & Visscher, 2018). Teachers often find it difficult to translate FA theory into practice (Robinson, Myran, Strauss, & Reed, 2014). Moreover, when FA activities are adopted, they tend to conform to the 'letter' of the concept, while they should be applied in its 'spirit'. According to the 'letter', FA is used as a set of teacher activities, without promoting student autonomy in the process. In contrast, according to the 'spirit', FA reflects an aligned, ongoing process that promotes student autonomy (Marshall & Drummond, 2006; Birenbaum, 2014). Additional empirical research is needed in order to determine what the FA practice of teachers look like, to analyse and explain the FA process in the classroom and to establish how students are involved in this process (Andersson & Palm, 2018; Antoniou & James, 2014). Moreover, Wiliam (2018) argues that more descriptive research is needed in order to identify what teachers do in their FA classrooms before it will be possible to study the effects of FA on student learning. This study aims to contribute additional empirical evidence on the FA practice of teachers, focussing on how teachers and students perceive FA practice as applied in the classroom.

1.1. The formative assessment process

Many theories and frameworks are available for describing and defining formative assessment (Bennett, 2011; Ruiz-Primo & Furtak, 2006; Wiliam, 2011). In addition to sharing a number of common features, each of these frameworks has its own accents in its vision of the FA process. In this study, we used the 'formative assessment cycle' (the FA cycle, Gulikers & Baartman, 2016, 2017) as a conceptual and analytical framework. This framework is based on a comprehensive literature review of FA activities applied by teachers in the classroom. It describes FA as a cyclic process consisting of five phases, with descriptions of specific teacher activities in the classroom for each phase. In the FA cycle, the role of the teacher consists of (1) clarifying expectations in learning

* Corresponding author.
E-mail addresses: marijke.veugen@wur.nl (M.J. Veugen), judith.gulikers@wur.nl (J.T.M. Gulikers), perry.denbrok@wur.nl (P. den Brok).

https://doi.org/10.1016/j.stueduc.2021.101027
Received 23 October 2020; Received in revised form 29 April 2021; Accepted 1 May 2021
Available online 21 May 2021

goals and success criteria; (2) eliciting student responses to collect information about the learning process; (3) analysing and interpreting these responses; (4) communicating with students about the responses, (5) and adjusting teaching and learning by taking follow-up actions (Fig. 1). Each of the five phases distinguished in the FA cycle entails specific FA teaching activities that have been identified in earlier studies as important to the improvement of teaching and learning (Gulikers & Baartman, 2017). The FA cycle could be used to describe both shorter-term, more informal ways of using the FA process in the classroom and longer-term, more formal ways of implementing the FA process (Gulikers & Baartman, 2017).

The FA cycle and activities identified within it focus on the teacher. Although this does not suggest that students have no role in FA, it is the teacher's responsibility to design a stimulating learning environment in which student agency is activated and in which students can actively engage in the assessment process (Boud & Molloy, 2013; Carless & Winstone, 2020). The assumption underlying the FA cycle is that, when teachers are able to align and implement the FA activities of all five phases, and when they are able to involve their students in this process, this improves the monitoring and promotion of student learning, thereby potentially enhancing student achievement and autonomy (Christoforidou, Kyriakides, Antoniou, & Creemers, 2014; Wiliam & Leahy, 2015). In turn, this could foster FA activities that adhere more to the 'spirit' than to the 'letter' of the concept.

1.2. Teacher practices in the FA cycle

To our knowledge, this empirical study is the first to use the comprehensive FA cycle as a lens for analysing the FA activities of teachers in the classroom. Previous studies have examined teacher activities in parts of the FA cycle, thus providing an indication of what teachers find more or less difficult to implement in the classroom. For example, some quantitative and qualitative studies in both primary and secondary education have shown that teachers tend to have difficulty clarifying learning goals and success criteria to their students (Phase 1) (Antoniou & James, 2014; Kippers et al., 2018). As reported in studies by Heritage, Kim, Vendlinski, and Herman (2009) and by Forbes, Sabel, and Biggers (2015), primary education teachers did not take follow-up actions to adjust teaching and learning to the learning needs of their students (Phase 5). Some studies have reported more positive findings with regard to Phases 2–4. For example, teachers used a large variety of activities to gather information about student learning processes (Phase 2), used student responses to assess student comprehension (Phase 3) and provided students with feedback (Phase 4) (Antoniou & James, 2014; Heritage et al., 2009; Volante & Beckett, 2011). Another indication from earlier research is that many secondary education teachers have difficulty encouraging students to use self-assessment and peer-assessment activities as part of the FA process (Hawe & Parr, 2014; Kippers et al., 2018; Panadero, Andrade, & Brookhart, 2018; Volante & Beckett, 2011). Teacher activities related to the active involvement of students have been described for each phase of the FA cycle (see Appendices A and B). By using the FA cycle as an analytic lens, this study aims to provide new insights into difficulties encountered by teachers in the implementation of the FA process in their classrooms and how these difficulties can be overcome.

1.3. Teacher and student perceptions of FA practice

For the FA practice of teachers to have an impact on student learning, both teachers and students need to perceive classroom FA practice in similar ways (Pat-El et al., 2015; Segers & Tillema, 2011). Previous questionnaire studies have nevertheless revealed significant differences in teacher and student perceptions of FA practice, with teachers often overestimating their own FA practice (Dobbelaer, 2019; Pat-El et al., 2015). Such overestimation might be the result of wishful thinking, self-protection against disappointment or incompetence in self-evaluation (Den Brok, Levy, Wubbels, & Rodriguez, 2003; Kruger & Dunning, 1999). In contrast, student evaluations are based on their observations of teaching and their interpretations of FA practice. More importantly, the perceptions that students have of assessment practices can influence the learning strategies that they choose (Segers, Nijhuis, & Gijselaers, 2006). Antoniou and James (2014) call for further research on the alignment between teacher and student perceptions of FA practice, in order to explore how these perceptions might point to a common understanding of FA practice. To this end, the current study investigates the FA classroom practice of teachers by comparing the perceptions of teachers and students, based on information obtained through questionnaires, thereby arriving at a valid representation of the classroom

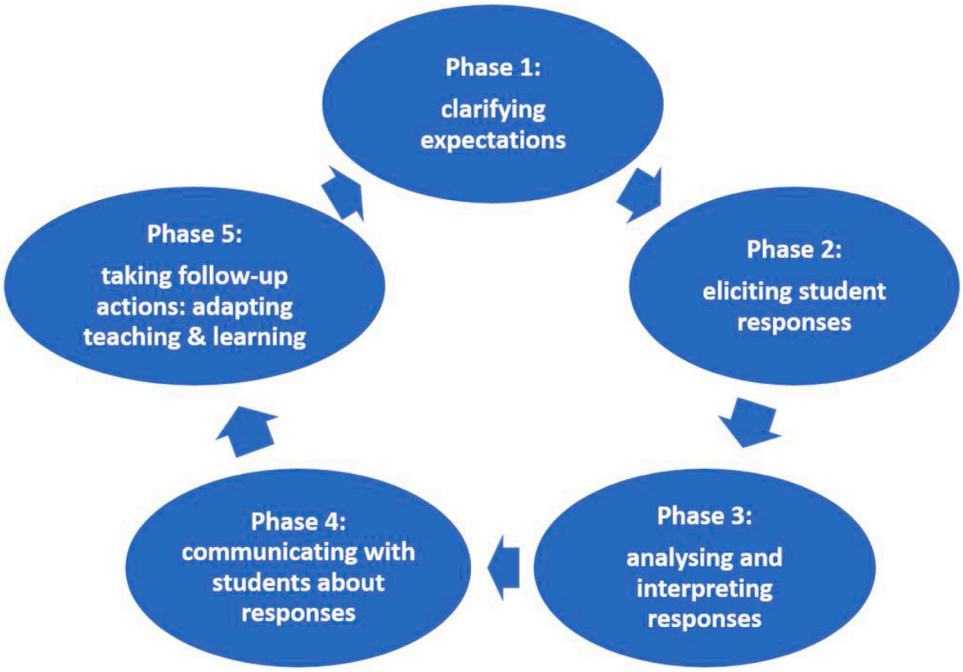

Fig. 1. The Formative Assessment Cycle (FA Cycle) (permission to use from Authors).

practice of teachers (Dobbelaer, 2019; Segers & Tillema, 2011). One advantage of collecting quantitative data from teachers and students regarding their perceptions of classroom FA practice, as compared to collecting data on the perceptions of external observers, is that teachers and students have full knowledge of the classroom context, and their perceptions are based on their own day-to-day experiences (Dobbelaer, 2019; Fraser & Walberg, 1981).

1.4. Background characteristics influencing FA perceptions

Earlier research has shown that differences in the perceptions of students and teachers could be partially explained by differences in background characteristics (Den Brok et al., 2003; Hansen, 2020). The background characteristics of students might influence their perceptions of the FA practice of a teacher. For example, Van der Kleij (2019) found that higher-achieving students were more positive in their perceptions of the feedback given by their teachers than were lower-achieving students. The background characteristics of teachers might also influence the ways in which they perceive their own FA practice. For example, Christoforidou et al. (2014) found that the teaching experience of primary school teachers influenced the ways in which they perceived their assessment skills. More specifically, teachers with more experience perceived themselves as using more FA activities than did teachers with less experience. Havnes, Smith, Dysthe, and Ludvigsen (2012) found that the FA practice of secondary school teachers, as well as their perceptions thereof, could differ by the subject (domain) being taught. For example, teachers and students of vocational school subjects perceived themselves as using more peer assessment when working on hands-on assignments than did those involved with academic subjects. In addition, for courses in mathematics, English and Norwegian as well, students and teachers perceived that feedback was provided at the end of the learning process, while for vocational courses, they perceived that feedback was given more during the learning process. In addition to the characteristics of teachers and students, characteristics of the school system could also lead to variations in perceptions (Goldstein, 1999). Birenbaum (2014) illustrates the complexity of how an assessment culture evolves, due to the intertwinement of factors relating to teachers, students and schools. Schildkamp, Poortman, Luyten, and Ebbeler (2017) demonstrated that collaboration amongst teachers and stimulating factors in the school organization (e.g. a shared vision on how to use assessment data) had a positive influenced on the teachers' use of assessment data to adapt instruction to the needs of their students in the learning process. In this study, the influence of characteristics relating to teachers, students and school are taken into account.

1.5. Research context and aim

In the present study, teachers participated in a professional development programme in the form of a learning network for FA. This four-year network project was designed according to the FA professional development guidelines developed by Wiliam and Leahy (2015). A substantial number of teachers and a school leader from each school participated. The FA cycle formed the theoretical backbone of the learning network, and five meetings were organized each year, all having a similar structure. During these network days, experts assisted the teachers as they collaborated to design subsequent steps in their FA classroom practice, implemented the activities within their schools and engaged in mutual reflection on and evaluation of their practice. As a condition for participating in the learning network, FA was required to be a focal point in the school's vision on education. In an intake interview, the school's vision, implementation approach and support for FA was discussed. The results of these interviews ultimately revealed two broad approaches to the implementation of FA in schools. In one approach, teams of teachers collaboratively implemented FA in the same class (i.e. all teachers of the same class of students used FA). In the other approach, teachers implemented FA on a more individual basis within the classroom of their choice, with the result that students encountered FA more as a matter of coincidence than of structure. Given this clear division in approaches to the implementation of FA, these approaches are taken into account as a school-background variable in the analyses of this study. This study examines the FA practice of teachers during the first phase of participating in the learning network. Based on the issues discussed in the preceding sections, the following three research questions were formulated:

1) To what extent do teachers use FA activities relating to the five phases described in the FA cycle in their classroom practice, according to their *own perceptions*, and how do teacher-background characteristics influence these perceptions?
2) To what extent do teachers use FA activities relating to the five phases described in the FA cycle in their classroom practice, according to their *students' perceptions*, and how do student-background characteristics influence these perceptions?
3) What differences exist between teacher and student perceptions of the FA practice of teachers in the classroom, and can teacher-background characteristics further explain these differences?

2. Method

2.1. Design

This study involved the quantitative investigation of the perceptions of secondary school teachers and students with regard to the FA practice of teachers at the start and within the context of a learning network for FA. The activities of the five phases of the FA cycle were used to develop questionnaires to measure these perceptions. After reviewing the quality of the questionnaire scales, quantitative analyses were conducted in order to answer the research questions, thereby identifying the FA activities that were used more and less often in terms of the five phases of the FA cycle, as well as the relationship between the application of these activities and the background variables.

2.2. Participants

2.2.1. Teachers

In all, 13 different secondary schools participated in this study, with participation from 96 teachers (32 male and 64 female). As mentioned earlier, these schools had two different approaches to the implementation of FA. In some schools, teachers adopted FA in the same classroom, while in others, teachers applied FA in different classrooms (thus constituting a school-level variable). At the teacher level, teachers taught in four different types of education, as is typical for the context of secondary education in the Netherlands: practically oriented pre-vocational education [*Dutch abbreviation: vmbo-bg*], theoretically oriented pre-vocational education [*vmbo-tl*], senior general secondary education [*havo*] and pre-university education [*vwo*] (Government of the Netherlands, 2020). The teachers taught 19 different school subjects, distributed across three over-arching domains: Alpha (languages, such as English), Beta (such as mathematics) and Gamma (such as history). They taught in the upper and lower grades, with wide variations in years of teaching experience. On average, however, they were quite experienced as teachers (M = 13, SD = 10, range = 1–40) (see Table 1).

2.2.2. Students

Each teacher was asked to choose one class in which they used FA to participate in this study. In all, 1,095 students of 59 teachers completed the student FA questionnaire. Of these students, 474 were male, 558 were female and 63 did not indicate their gender. The students' ages varied between 11 and 18 years (M = 13, SD = 1.3). Corresponding to the differences between teachers, the students also differed in terms of grade level, type of education and school attended (see Table 1 for a distribution of these characteristics).

Table 1
Teacher and Student Characteristics.

		Teachers		Students	
		n	(%)	*n*	(%)
Gender	Female	64	66.7	558	51.0
	Male	32	33.3	474	43.3
	Unknown			63	5.8
Teaching upper or lower secondary classes	Upper	27	28.1		
	Lower	18	18.8		
	Both	51	53.1		
Grade	Grade 7			558	50.1
	Grade 8			194	17.7
	Grade 9			252	23.0
	Grade 10			59	5.4
	Grade 11			34	3.1
Education type	Vmbo-bg	17	17.7	161	14.7
	Vmbo-tl	19	19.8	160	14.6
	Havo	45	46.9	439	40.1
	Vwo	15	15.6	282	25.8
	Unknown			53	4.8
Domain[a]	Alpha	35	36.5	631	57.6
	Beta	30	31.3	178	20.7
	Gamma	31	32.3	237	21.6
School	School 1	13	13.5	166	15.2
	School 2	13	13.5	216	19.7
	School 3	7	7.3	93	8.6
	School 4	14	14.6	216	19.7
	School 5	6	6.3	120	10.9
	School 6	6	6.3	26	2.4
	School 7	7	7.3	62	5.7
	School 8	4	4.2	26	2.4
	School 9	4	4.2	29	2.6
	School 10	6	6.3	44	4.0
	School 11	6	6.3	0	0
	School 12	6	6.3	30	2.7
	School 13	4	4.2	67	6.1
FA implementation approach	Shared FA class	20	20.8	259	23.7
	Different FA class	76	79.2	836	76.3
Teacher's years of experience	< 5 years	34	35.4		
	6 – 15 years	32	33.3		
	> 16 years	30	31.3		
Student age	11 years			19	1.7
	12 years			340	31.1
	13 years			247	22.6
	14 years			224	20.5
	15 years			157	14.3
	16 years			64	5.8
	17 years			9	.8
	18 years			1	.1
	unknown			34	3.1

[a] Gamma = Geography, History, Economics, Societal Studies + also Theme Green, Art, Music, Gymnastics; Alpha = Dutch, English, German, French and Spanish; Beta = Physics, Mathematics, Biology, Chemistry, Technology + Theme Green Animals.

2.3. Instruments

2.3.1. Teacher and student FA questionnaires

The teacher questionnaire was constructed especially for the present study. It contained 44 items describing the FA activities representing the five phases of the FA cycle, as identified in a review of 106 empirical studies describing the FA activities of teachers in the classroom (Gulikers & Baartman, 2016, 2017). These activities were transformed into items that were formulated in terms of activities that could be observed by students (e.g. 'I discuss the learning goals with the students' [Phase 1]) and non-observable activities (e.g. 'I take time to analyse student responses' [Phase 3]). For each item, teachers were asked to indicate the extent to which they used that FA activity in their teaching, based on a five-point Likert scale ranging from 'almost never' (1) to 'almost always' (5). In addition to the FA items, the questionnaire included items about teacher-background characteristics: years of teaching experience, type of education taught, level of classes taught (upper or lower), school subject taught, gender and the school's approach to the implementation of FA. The student questionnaire, which was also constructed especially for this study, contained 23 items about the FA activities of their teachers in the five phases of the FA cycle. These items were similar to the observable items in the teacher questionnaire, but described from a student perspective (e.g. 'The teacher discusses the learning goals with us' [Phase 1]). Other questions included student-background characteristics: gender, age, grade level and type of education. Several example items are presented in Table 2, and the questionnaires are provided in Appendices A and B for the questionnaires. All items were scored along a five-point Likert scale ranging from 'almost never' (1) to 'almost always' (5).

2.3.2. Validity and reliability

The researchers developed the questionnaires in several rounds, building on the FA activities that had been identified in the literature study as being representative of the five FA phases (Gulikers & Baartman, 2017). A pilot test of the teacher questionnaire was conducted with a selection of the teachers participating in the learning network, in order to check for comprehensibility and face validity. The student questionnaire was also pilot tested with 12 students who did not participate further in the study. The students varied according to age, gender, grade and type of education. All of the students, regardless of grade level or type of education, remarked that they found the questionnaire understandable and not too lengthy. To further increase validity, a teacher instruction manual for the student questionnaire was added, in which teachers were asked to discuss certain FA concepts (e.g. 'learning goal') with their students before having them complete the questionnaire. Several minor, final adjustments were made to the questionnaires before the final version was completed. The internal consistency of the subscales (i.e. representing the five phases) was calculated based on the data of all participating teachers and students for both questionnaires using Cronbach's alpha, with values ranging between .72 and .82 for the teacher questionnaire and between .64 and .87 for the student questionnaire (see Table 2). These values indicate acceptable to good reliability (Lattin, Carroll, & Green, 2003), and they did not increase remarkably when any item was removed. For this reason, all items were retained in the questionnaires. To check for discriminant validity between the scales, Pearson correlations between the five scales were calculated and examined. Given that all of the scales measured elements of the overall FA cycle construct, they should be related to each other, but not too closely (De Jong & Westerhof, 2001). The Pearson correlations for the five scales varied between .36 and .73 (see Table 3), thus indicating acceptable discriminant validity between the scales (Aluri & Fraser, 2019; Borsboom, Mellenbergh, & van Heerden, 2004). Confirmatory factor analyses (CFA) were used to further test construct reliability (Lattin et al., 2003): the five scales loaded significantly ($p < .05$, $df = 5$, with loadings between .68 and .88) on one construct, which was interpreted as total FA practice, for both the teacher and student questionnaire. The CFA results for the teacher questionnaire showed mostly appropriate fit indices (GFI = 0.94, RMR = .02, AGFI = .82, NFI = .93, CFI = .95), although some indices were not fully acceptable (TLI = .90, RMSEA = 0.15, with a lower bound of 0.07 and an upper bound of 0.23) (Hu & Bentler, 1999). The results for the student questionnaire led to similar conclusions, revealing mostly appropriate fit indices (GFI = 0.97, RMR = .024, AGFI = .91, NFI = .97, CFI = .97), along with some that were not fully acceptable (TLI = .94, RMSEA = 0.12, with a lower bound

Table 2
Cronbach's Alpha's of the Five Scales of the Teacher and Student Questionnaires.

Scale	Teacher Questionnaire ($n = 96$)			Student Questionnaire ($n = 1095$)		
	Number of items	Cronbach's Alpha	Example items	Number of items	Cronbach's Alpha	Example items
Phase 1: Clarifying expectations	12	.72	I discuss the learning goals of the lesson with students. I know the most common misconceptions in my subject/domain.	5	.79	The teacher discusses the learning goals with us. The teacher uses examples to show us what the task should look like.
Phase 2: Eliciting student responses	8	.75	I give students the opportunity to show what they have learned. I use informal FA methods, like asking deeper questions and observations.	3	.64	The teacher gives us the opportunity to show what we have learned. The teacher lets us explain our answers to each other.
Phase 3: Analysing and interpreting responses	9	.80	I compare students' work to the criteria for the task. I let students compare their own work to the criteria for the task.	3	.69	The teacher compares our work to the criteria for the task. The teacher lets us compare our own work to the criteria for the task.
Phase 4: Communicating with students about responses	9	.76	Based on the analyses of Phase 3, I discuss the strengths with individuals/ small groups I let students think about their own strengths and points for improvement.	9	.87	The teacher discusses my strengths with me. The teacher lets us think about the strengths and points for improvement with each other.
Phase 5: Taking follow-up actions: adapting teaching & learning	6	.82	I help the students to come up with the next step in learning. Based on my previous analyses, I consciously differentiate within my group.	3	.79	The teacher helps us to come up with the next step in learning. The teacher lets us come up with our next step in learning with each other.

Table 3
Pearson Correlations of the Five Scales of the Teacher and Student Questionnaires.

	Phase 1		Phase 2		Phase 3		Phase 4		Phase 5	
	Teachers $n =$ 96	Students $n =$ 1095	Teachers $n =$ 96	Students $n =$ 1095	Teachers $n =$ 96	Students $n =$ 1095	Teachers $n =$ 96	Students $n =$ 1095	Teachers $n =$ 96	Students $n =$ 1095
Phase 1	–	–	.60	.56	.56	.49	.46	.63	.51	.56
Phase 2	.60	.56	–	–	.60	.58	.36	.59	.50	.51
Phase 3	.56	.49	.60	.58	–	–	.62	.62	.64	.51
Phase 4	.46	.63	.36	.59	.62	.62	–	–	.62	.73
Phase 5	.51	.56	.50	.51	.64	.51	.62	.73	–	–

of 0.10 and an upper bound of 0.15) (Hu & Bentler, 1999). A reason for the lower TLI score could be that the model retained a complex but less restrictive model that did not reveal a much better trade-off compared to the empty model (Schmermelleh-Engel, Moosbrugger, & Müller, 2003). The lower RMSEA score could be due to the presence of many free parameters (10, $df = 5$) in the model, thus indicating a lack of fit with the population covariance matrix.

2.4. Procedure

All of the participating teachers and the parents of the participating students were asked to complete an approval form, which stated that their privacy and anonymity would be guaranteed. After consent was obtained, the teachers and their students were asked to complete the FA questionnaire on a digital device through a secured weblink between December 2018 and March 2019. At this point, the teachers had been participating in the learning network (which started April 2018) for several months, and they were becoming familiar with the FA cycle. This increased the validity of their responses to the questionnaires (Kruger & Dunning, 1999). During the months of participating in the network, the teachers had started to reflect on and experiment with FA cycle activities in their teaching. In the learning-network days during this period, the teachers engaged in collaborative reflection on their own FA practice based on the FA cycle, focussing predominantly on the first two phases. The students completed the questionnaire for specific teachers at approximately the same time that these teachers completed the questionnaire. All data were collected using the *Qualtrics* program and were exported to *IBM SPSS Statistics 25*.

2.5. Analyses

2.5.1. Teacher and student perceptions, and background variables

Multilevel analyses were used to explore the first two research questions, in order to do justice to the hierarchical structure of the data (Goldstein, 1999; Kreft & de Leeuw, 1998). First, the total amount of variance was established at the student, class and school level for each of the five FA-phase variables, as well as for the variable *total FA practice* in empty models. Second, a more elaborate model was tested, in which all background variables were added. We then determined which variables had a significant influence on perceptions of FA practice. Third, the significant variables were analysed in new models, distinguishing between the explained and unexplained variance. We compared the deviances between the new and empty, using the analysis of variance (ANOVA) feature of the *nlme* package in *R,* in order to determine whether the new models showed a statistically significant better fit than the empty models. These new models, with the statistically significant

variables included, are described in the results section.

2.5.2. Teacher models

For the teacher models, the first level included the variables related to the teachers themselves (N = 96), and the second level included a school-level variable. The following teacher-background variables were included at the teacher level: *domain*, *type of education*, *upper or lower secondary classes*, *gender* and *years of teaching experience*. There was only one teacher-background variable at the school level: the school's approach to the implementation of FA. Some variables were transformed into dummies before they were added in the analyses (Goldstein, 1999). The *domain* taught was divided into three categories: Alpha, Beta and Gamma, with Beta being included as the baseline. *Education type* was categorized into four dummies: *vmbo-bg, vmbo-tl, havo* and *vwo*, with *vmbo-tl* as the baseline. Teaching *upper or lower secondary classes* was divided into three dummy variables: upper, lower and both, with upper as the baseline. For *gender*, boys were coded as 0, and girls were coded as 1. *Years of experience* was divided into three categories: 0–5 years, 6–15 years and more than 15 years. For *FA-implementation approach*, teachers teaching in the same class were coded as 0, and teachers teaching in different classes were coded as 1.

2.5.3. Student models

For the student models, the first level represented the student level, with the second level representing the teacher level and the third level representing the school level. Students with incomplete data were removed from the analysis, leaving a total of 1,042 students. The student-background variables included gender, type of education, domain taught by the teacher, age, grade and the FA-implementation approach used by the student's school. The variables *gender, education type, domain* and *FA-implementation approach* were coded in a manner similar to that used in the teacher analyses, except with Gamma as the baseline in *domain*. *Age* was coded from 0 (11 years) to 7 (18 years), and *grade level* was coded from 0 (Grade 7) to 4 (Grade 11).

2.5.4. Comparing the perceptions of teachers and students perceptions

To explore the third research question, paired *t*-tests were used to compare the perceptions of teachers and students. The 23 observable FA items of the questionnaires were used for this analysis, as they were scored by both the students and their teachers. In all, we were able to compare the scores of 59 classes to those of their teachers. Teacher identification numbers were used to match classes to their teachers, thus making it possible to compare the aggregated class mean to the teacher mean for each phase in the paired *t*-test. To further explore differences between teachers and students, we calculated the percentages of teachers overestimating and underestimating their own FA practice, or whose mean scores fell within a range of one averaged standard deviation from those of their students (Mara & Cribbie, 2012). The average standard deviation was calculated by first adding the average standard deviation of the student classes to the average standard deviation of teachers, and then dividing by two. The average standard deviation was used to set the equivalence interval, which was used to assess whether the difference between the means of teachers and their student class means fell within or outside of this interval. This made it possible to determine whether teachers overestimated or underestimated their FA own practice, or whether their mean scores were similar to those of their student classes ($\mu_1 - \mu_2 > \sigma$, $\mu_1 - \mu_2 > -\sigma$ or $\mu_1 - \mu_2 < \sigma$). In the next step, we established the percentages of teachers whose perceptions overestimated, underestimated or were similar to those of their student classes. Finally, cross-tabulation analyses with chi-square tests were used to investigate whether the teacher-background variables that were found to have a significant influence could explain the equivalence differences between teacher and student perceptions for each FA phase.

3. Results

3.1. Teacher perceptions of FA practice (Research Question 1)

The empty models of the multilevel analyses showed that teachers (*n* = 96) had a mean perception of 3.31 (*SE* = .06) for their total FA practice (see Table 4). The teachers perceived that the FA activities that they used the most were those aimed at *eliciting student responses* (Phase 2; *M* = 3.59, *SE* = .06) and at *clarifying expectations* (Phase 1; *M* = 3.52, *SE* = .06), followed by activities aimed at *communicating with students about responses* (Phase 4; *M* = 3.28, *SE* = .06) and at *analysing* and *interpreting student responses* (Phase 3; *M* = 3.21, *SE* = .06). The teachers perceived that the FA activities that they used the least were those aimed at *taking follow-up actions: adapting teaching and learning* (Phase 5; *M* = 3.11, *SE* = .09). Most of the variance for total FA practice and for each of the five phases was located at the teacher level, ranging from 78.9 % (Phase 5) to almost 100 % (Phase 3) (see Table 4). The differences between schools thus appeared to be limited. When the background variables were included in the models, two variables were statistically significantly related to the teachers' perceptions of *total FA practice*: *FA-implementation approach* (school level) and practically oriented pre-vocational education as *type of education* (teacher level) (see Table 5). The FA-implementation approach explained all of the school-level variance and 14.7 % of the total variance. The practically oriented pre-vocational education variable explained 12.9 % of the school-level variance and 11.8 % of the total variance. The FA-implementation approach apparently made a difference, especially in the teachers' perceptions of activities aimed at *clarifying expectations* (Phase 1) and at *eliciting student responses* (Phase 2). This finding indicates that teachers in schools where teachers taught in shared classed were significantly more positive in their perceptions than were teachers in schools where they taught in different classes. For both phases, this variable explained all of the school-level variance and 13.7 % of the total variance for Phase 1 and 17.8 % of the total variance in Phase 2. The type of education made a difference, especially in the teachers' perceptions of activities aimed at *taking follow-up actions: adapting teaching and learning* (Phase 5), thus indicating that teachers who taught in the practically oriented vocational type of education were significantly less positive in their perceptions regarding their use of this phase than were teachers who taught in other types of education. Some variables had statistically significant associations with teacher perceptions concerning only one specific phase. For example, one of these variables, the *domain* in which a teacher taught, had a significant influence on the teachers' perceptions of their activities aimed at *taking follow-up actions: adapting teaching and learning* (Phase 5). The analyses showed that teachers who taught Beta subjects were less positive in their perceptions with regard to their use of this phase than were teachers who taught subjects in other domains. The variables *domain* and *education type* together explained 11.2 % of the total variance for this phase. Finally, the teachers' *years of experience* seemed to be significantly associated with their perceptions of their use of activities aimed at *communicating with students about responses* (Phase 4). More specifically, teachers with more experience were more positive in their perceptions with regard to this phase than were teachers with less experience. This variable explained almost all of the school-level variance and 6.8 % of the total variance for this phase.

3.2. Student perceptions of FA practice (Research Question 2)

The empty models of the multilevel analyses for the student data showed that the students (N = 1,042) had a mean perception of 3.30 (*SE* = .04) for the total FA practice of their teachers (see Table 6). Students perceived the FA activities of their teachers in the same rank order as the teachers did. The highest scores were for activities aimed at *eliciting student responses* (Phase 2, *M* = 3.68, *SE* = .05), followed by those aimed at *clarifying expectations* (Phase 1; *M* = 3.51, *SE* = .05), *communicating about student responses* (Phase 4; *M* = 3.21, *SE* = .05) and *analysing and*

Table 4
Empty Models of Two-level Analysis of Teachers' (n = 96) Background Characteristics Influences on their FA Perspectives.

		Total FA Estimates	Phase 1 Estimates	Phase 2 Estimates	Phase 3 Estimates	Phase 4 Estimates	Phase 5 Estimates
Fixed part							
Intercept		3.31 (.06)	3.52 (.06)	3.59 (.06)	3.21 (.06)	3.28 (.06)	3.12 (.09)
Variance part %							
Unexplained	Teacher level	80.7	82.8	79.1	~ 100	92.3	78.9
	School level	19.3	17.2	20.9	< .1	7.7	21.1
−2*Loglikelihood		123.03	137.39	141.14	161.27	174.43	214.95

interpreting student responses (Phase 3; $M = 3.17$, $SE = .05$), with the lowest scores being assigned for activities aimed at *taking follow-up actions: adjusting teaching and learning* (Phase 5; $M = 2.94$, $SE = .06$). Most of the variance in the empty models was at the student level for all phases, as well as for total FA practice, ranging from 65.4 % (Phase 2) to 73 % (Phase 3). The remainder of the variance was almost always at the teacher level, except for Phase 2, in which 8.5 % of the variance was at the school level (see Table 6). In the elaborated models, only *age* and *gender* were significantly associated with the students' perceptions of the FA activities of their teachers (see Table 7). For the variable *total FA practice*, only age was significantly associated with FA perceptions, and it explained only 1.3 % of the total variance. Older students were slightly more positive in their perceptions of the FA practice of their teachers. Age explained 1.7 % of the total variance in perceptions of activities aimed at *analysing and interpreting student responses* (Phase 3) and, together with *gender*, it explained 1.5 % of the variance in perceptions of activities aimed at *taking follow-up actions: adapting teaching and learning* (Phase 5). The variable *gender* was significantly related to activities aimed at *clarifying expectations* (Phase 1), *communicating with students about responses* (Phase 4) and *taking follow-up actions: adjusting teaching and learning* (Phase 5). At the same time, however, the variable explained from 0 % (Phase 1) to only 0.3 % (Phase 4) of the total variance in the models with only *age* added. Nevertheless, the negative significant estimate of *gender* indicated that boys were more positive in their perceptions of their teachers' FA activities than were girls.

3.3. Differences between the FA perceptions of teachers and students (Research Question 3)

To measure differences between the perceptions of teachers and students with regard to the FA practice of teachers, the means of student classes ($n = 59$) were compared to the mean scores of their teachers ($n = 59$). This was done for total FA practice, as well for each of the five FA phases, including only the observable FA activities (23 items). According to the paired *t*-test results (see Table 8), the only significant differences between the two groups had to do with their perceptions of *clarifying expectations* (Phase 1), with students perceiving their teachers as performing activities in this phase more often than the teachers themselves perceived ($t(59) = -2.63$, $p < .05$). For the other four phases of the FA cycle and for total FA practice, no significant differences were revealed between the perceptions of teachers and those of students (see Table 8 and Fig. 2).

To further describe how the perceptions of teachers differed from those of their students, we calculated the percentage of teachers who overestimated and underestimated their own FA practice or whose mean scores fell within a range of one averaged standard deviation from those of their students (see Fig. 3). The results showed that, for total FA practice, 58 % of the teachers had perceptions similar to those of their students. The remaining teachers were almost equally divided over those who overestimated (22 %) and underestimated (20 %) their own FA practice. For all five phases, more than half of the teachers had perceptions similar to those of their students. Teachers were least likely to overestimate their own FA practice with regard to the phase of *clarifying expectations* (12 %), and they were the mostly likely to overestimate their own FA practice with regard to *communicating with students about responses* (25 %). For the phase of *taking follow-up actions: adapting teaching and learning*, more teachers (31 %) underestimated their own FA practice than was the case for the other phases, whereas only 12 % underestimated their own FA practice with regard to the phase of *eliciting student responses*. As indicated by the chi-square tests in the cross-tabulation analyses, teachers who taught in practically oriented vocational education were more likely to underestimate their own FA practice more often than did teachers who taught in other types of education ($X^2 = 6.61$, $df = 2$, $p < .05$). Other background variables (i.e. school FA-implementation approach, domain, type of education, teaching experience) did not influence overestimation or underestimation, either for total FA practice or for any of the five independent phases.

4. Discussion

4.1. The FA cycle as a clear construct for describing the FA practice of teachers

The objective of this study was to describe the use of FA activities by teachers in the classroom, based on the perceptions of teachers and students and using the FA cycle as a conceptual model. To this end, the conceptual model was translated into a teacher and a student questionnaire to conduct initial measurements of the use of FA activities by teachers in classroom practice. Analysis of these questionnaire data confirmed that the FA cycle presents a clear construct consisting of five reliable, coherent but distinct constructs of FA. While previous studies have provided insight into activities that are commonly used by teachers (e.g. gathering information about the learning processes of students; Heritage et al., 2009) and those that are hardly ever used in the classroom (e.g. clarifying learning goals and success criteria; Kippers et al., 2018), this study described FA activities as coherent activities related to the five phases of the FA process. The use of the FA cycle as an analytical framework could thus potentially offer a range of opportunities for studying FA classroom practice in a more comprehensive way. The results from the questionnaires reflected congruent teacher and student perceptions of the FA activities of teachers in the classroom, with regard to both total FA practice and each of the five phases. This finding is quite unusual, as comparisons of teacher and student perceptions usually suggest that teachers are more positive in their perceptions of their own practice than their students are (Dobbelaer, 2019; Pat-El et al., 2015). Congruence in the perceptions of teachers and students with regard to FA is nevertheless important to arriving at a common understanding of FA, thus making it possible to create an effective FA process and improve student learning (Antoniou & James, 2014; Pat-El et al., 2015). The congruent perceptions found in our study could potentially be explained by the fact that the FA activities were clearly described in the FA questionnaires, which made it easier for both teachers and students to recognize the FA activities that had been used. Such clear descriptions of FA activities make the FA cycle a useful conceptual model for capturing what FA practice looks like in the classroom for both students and teachers. Another explanation for the overlapping perceptions of teachers and students could be that the teachers were better able to

Table 5
Two-level Analysis of Teachers' (n = 96) Background Characteristics Influences on their FA Perspectives.

		Total FA *FA implementation approach*		Total FA *Practice based vocational education*		Phase 1 *FA implementation approach*		Phase 2 *FA implementation approach*		Phase 4 *Teacher experience*		Phase 5 *Practice based vocational education + Beta*	
		Estimates	Effect size p-value	Estimates	Effect size p-value	Estimates	Effect size p-value	Estimates	Effect size p-value	Estimates	Effect size p-value	Estimates	Effect size p-value
Fixed part													
Intercept		3.24 (.05)		3.35 (.05)		3.45 (.06)		3.51 (.06)		3.13 (.09)		3.31 (.08)	
Level													
	Years of experience									.16 (.07)	.04		
	Vmbo-bg			-.27 (.12)	.03							-.45 (.20)	.03
	Beta											-.43 (.16)	.01
School	FA implementation approach	0.28 (.11)	.03			.32 (.12)	.02	.41 (.12)	.01				
Variance part %													
Explained		14.7		11.8		13.7		17.8		6.8		11.2	
Unexplained	Teacher level	85.3		83.9		86.3		82.2		93.1		75.6	
	School level	< .1		4.3		<.1		< .1		.1		13.2	
Significant tests													
−2*Loglikelihood		118.12		118.96		131.27		132.02		169.88		202.36	
Difference with *df*		−4.91, *df* = 5		−4.06, *df* = 5		6.13, *df* = 5		9.12, *df* = 5		4.55, *df* = 5		12.59, *df* = 6	
p-value, with α = .05		.03		.04		.01		< .01		.03		< .01	

Table 6
Empty Models of Two-level Analysis of Students (n = 1042) Background Characteristics Influences on their FA Perspectives.

		Total FA Estimates	Phase 1 Estimates	Phase 2 Estimates	Phase 3 Estimates	Phase 4 Estimates	Phase 5 Estimates
Fixed part							
Intercept		3.30 (.04)	3.51 (.05)	3.68 (.05)	3.17 (.05)	3.21 (.05)	2.94 (.06)
Variance part %							
Unexplained	Student level	68.5	67.0	65.4	73.0	69.6	70.3
	Teacher level	31.5	33.0	26.1	27.0	30.4	29.7
	School level	< .1	< .1	8.5	< .1	< .1	< .1
−2*Loglikelihood		2094.38	2364.32	2367.01	2666.48	2445.07	2875.26

Table 7
Two-level Analysis of Students' (n = 1042) Background Characteristics Influences on their FA Perspectives.

		Total FA *Age*		Phase 1 *Gender (Female = 1)*		Phase 3 *Age*		Phase 4 *Gender (Female = 1)*		Phase 5 *Age + gender (Female = 1)*	
		Estimates	Effect size p-value	Estimates	Effect size p-value	Estimates	Effect size p-value	Estimates	Effect size p-value	Estimates	Effect size p-value
Fixed part											
Intercept		3.17 (.07)		3.58 (.06)		3.01 (.07)		3.31 (.06)		2.86 (.11)	
Level											
Teacher	Gender			−.14 (.05)	< .01			−.19 (.05)	< .01	−.24 (.06)	< .01
	Age	.06 (.02)	.02			.07 (.03)	.02			.09 (.03)	< .01
Variance part %											
Explained		1.3		0		1.7		.3		1.5	
Unexplained	Student level	68.4		67.0		73.0		69.0		69.6	
	Teacher level	30.3		33.0		25.3		30.7		28.9	
	School level	< .1		< .1		< .1		<.1		<.1	
Significant tests											
−2*Loglikelihood		2088.74		2355.10		2666.01		2429.62		2851.89	
Difference with *df*		5.64, *df* = 6		9.22, *df* = 6		5.41, *df* = 6		15.46, *df* = 6		23.37, *df* = 7	
p-value, with α = .05		.02		< .01		.02		< .01		< .01	

Table 8
T-Values per Activity of the Paired Samples T-tests of all Phases Between Teachers and Student Classes.

	Teachers *n* = 59[a]		Student Classes *n* = 59				
	M	*SD*	*M*	*SD*	*df*	*t*	Sig.[b]
Total FA practice	3.28	.52	3.31	.35	58	.34	.74
Phase 1: clarifying expectations	3.28	.60	3.51	.41	58	−2.63	.01
Phase 2: eliciting student responses	3.83	.59	3.68	.33	58	1.73	.09
Phase 3: analysing and interpreting responses	3.31	.74	3.19	.37	58	1.27	.21
Phase 4: communicating with students about responses	3.27	.60	3.21	.42	58	.67	.50
Phase 5: taking follow-up actions: adapting teaching and learning	2.70	.90	2.94	.51	58	−1.79	.08

[a] Number of teachers = 59, since these teachers were paired to their student class, resulting in new mean scores.
[b] With α = .05, two tailed.

self-assess their FA practice, having spent several months practicing with FA in the classroom before they completed the questionnaires (Kruger & Dunning, 1999). The students may also have been more familiar with FA practice in their classes, as their teachers had experimented with FA for a while before they were asked to complete the questionnaire.

4.2. The FA practice of teachers

This study demonstrates that a sample of secondary school teachers, all of whom had participated the same learning network, made use of FA activities within all five phases of the FA cycle in their classroom practice. Results from both the teachers and their students indicated that the teachers had used FA activities primarily to *clarify expectations* (Phase 1) and to *elicit student responses* (Phase 2), and that they had made the least use of activities aimed at *taking follow-up actions: adapting teaching and learning* (Phase 5) the least. In essence, this means that teachers were able to put most of the FA cycle activities into practice, but that they were still in need of further development in the implementation of the activities of the five phases in order to achieve the full implementation of FA practice in their day-to-day teaching, paying special attention to FA activities related to taking follow-up actions. The finding that the activities of Phases 1 and 2 were applied most might strengthen the

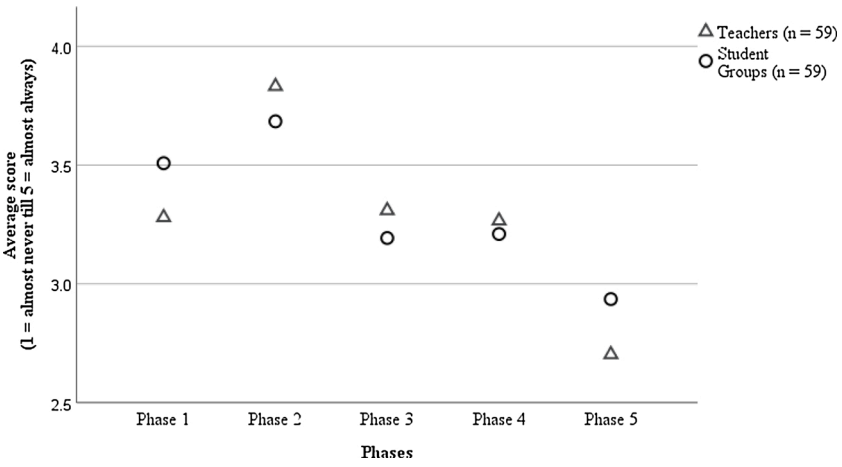

Fig. 2. Paired *t*-test results of teachers (n = 59) and students (n = 59) mean scores.

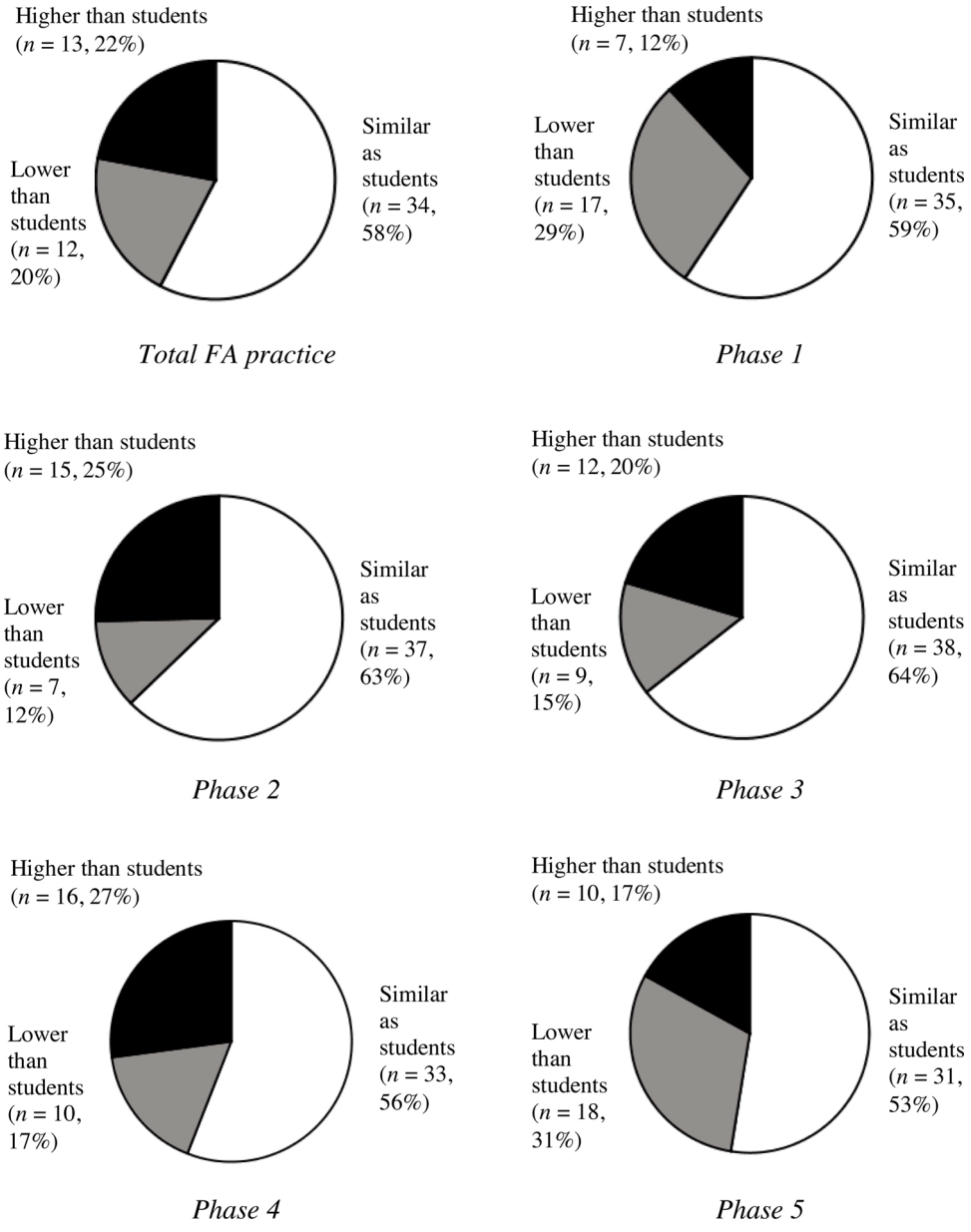

Fig. 3. Percentages of teachers under-, overestimating and having the same means per phase and of the total FA practice.

effects of participating in the learning network, as these two phases had received the most explicit attention before the questionnaires were administered. While the teachers had studied and were able to implement all of the phases, this might explain why they had made less use of activities associated with Phases 3, 4 and 5 at that time. The result that teachers experienced the most difficulties with regard to using FA activities relating to the last phase is in line with the findings of previous studies (Gulikers & Baartman, 2017; Forbes et al., 2015; Heritage et al., 2009). Many teachers have difficulty adjusting their teaching to the needs of their students, and they often use strategies such as repeating their instructions or pacing (Gulikers & Baartman, 2017). As identified in previous studies, reasons for the difficulty associated with this phase include that teachers either did not feel capable of adapting their teaching to the needs of their students or did not dare to deviate from prescribed curriculum plans (Forbes et al., 2015). Without activities that can actually be used to adapt teaching or learning in response to student data, it is unlikely that the FA process will be able to advance learning (Heritage et al., 2009; Wiliam, 2006). This final step of taking follow-up actions (Phase 5) is also what differentiates the FA process from summative assessment. Whereas the first four phases *could* also be used in summative assessments—for example, feedback (Phase 4) can also be given within the context of summative assessment—Phase 5 is used to *continue* this process, resulting in better decisions, based on the former phases, in order to improve teaching and learning. This finding also highlights the difference between providing feedback (Phase 4) and feed-forward (Phase 5). Teachers need to create opportunities that allow their students to process feedback (Phase 4) in order to improve their learning (Phase 5) (e.g. by preparing follow-up tasks; Carless & Winstone, 2020). The insight that feedback and taking follow-up actions are two different activities, both of which are essential in the FA process, might help teachers to define their activities more clearly within these two phases. For this reason, Phase 5 requires special attention in both professional development and research on the implementation of FA in the classroom. Given that this study is based on the FA activities of teachers during the start-up stage of the learning network, it can be concluded that subsequent professional-development activities should devote explicit attention to specific and in-depth activities, design, experimentation and reflection in relation to Phase 5. This finding also emphasizes that it can be difficult for teachers to act according to the 'spirit' of FA when implementing the full FA process (Andersson & Palm, 2018; Marshall & Drummond, 2006). Teachers need to see the activities within the FA cycle as a comprehensive and ongoing process for enacting the 'spirit' of FA. Another interesting finding with regard to stimulating the 'spirit' of FA can be observed at the item level of the questionnaires (see Appendices A and B). Activities related to teachers promoting the active involvement of students in the FA process in terms of self-assessment (e.g. *I let students compare their own work to the criteria for the task; Phase 3*) or peer-assessment (e.g. *I let students think about each other's strengths and points for improvement; Phase 4*) were perceived less frequently than were other more teacher-centred activities. This finding is in line with those of previous studies, in which teachers were also found to use self-assessment and peer-assessment activities the least (Hawe & Parr, 2014; Kippers et al., 2018; Panadero et al., 2018; Volante & Beckett, 2011). Given that FA is a highly interactive and dialogical process, it is imperative for teachers to learn how to design their FA environments in such a way that their students will indeed be actively involved in all phases of the FA process (Carless & Winstone, 2020). Professional-development programmes should therefore include additional training to help teachers implement activities of the FA cycle that promote student engagement in the FA process.

4.3. Background variables have an influence on perceptions of FA

Consistent with previous studies (Christoforidou et al., 2014; Van der Kleij, 2019), this study demonstrates that the background characteristics of teachers and students (e.g. years of teaching experience or the gender of students) have a slight influence on their perceptions of FA practice. In line with Heitink, van der Kleij, Veldkamp, Schildkamp, and Kippers (2016), our results suggest that the approach that a school adopts with regard to the implementation of FA matters. More specifically, teachers applying FA in shared classes perceived themselves as using more FA activities than did teachers applying FA in individually selected classes. One possible explanation for this finding could be that teachers who apply FA in the same classroom are more likely to have a shared vision on FA and to support each other more by sharing problems and successes, while such collaboration is likely to be missing for teachers who apply FA in their classrooms on a more individual basis (Andersson & Palm, 2018; Schildkamp et al., 2017). Finally, in line with Havnes et al. (2012), this study found that the FA perceptions of teachers differed slightly according to the subject or domain taught. At the same time, however, whereas teachers of Beta subjects perceived themselves as using fewer activities aimed at taking follow-up actions (Phase 5) than did their colleagues who taught other subjects, students in Beta classes did not perceive any fewer FA activities associated with Phase five than did students in classes in other subject domains. Additional qualitative research is therefore needed in order to investigate whether and exactly how the practice of teachers with regard to the FA cycle differs between subjects or domains.

4.4. Limitations and implications for future research

This study demonstrates that the FA cycle can be helpful for developing a more comprehensive conceptualization of what FA activities in the classroom look like by analysing the FA process in five coherent phases. This perspective could be useful for future empirical FA research, as well as for promoting the implementation of the 'spirit' of FA. Given that FA should be a comprehensive and ongoing process in the day-to-day classroom practice of teachers, the FA cycle could further promote its implementation. Despite the utility of its findings, one limitation of this study is that it was based solely on quantitative questionnaires measuring perceptions of FA classroom practice. Given that perceptions do not fully represent actual practice, however, the congruence between the perceptions of teachers and students identified in this study do suggest a common understanding of FA in the classroom. This suggests that these results reflect a relatively valid assessment of the actual FA practice of teachers in the classroom (Hansen, 2020). A more in-depth overview of actual FA activities, as well as of the ways in which teachers align the various phases of FA in their classrooms would require more qualitative data (e.g. classroom observations and interviews with teachers and students). Additional qualitative research could also shed light on the use of formal and informal types of FA by teachers within the FA cycle (Shavelson et al., 2008). Future studies could also investigate the precise role of technology in FA practice. Another avenue for future research could involve investigating how participation in a learning network could promote and affect the ways in which teachers use the FA cycle over time. Finally, future studies could explore differences in the actual classroom practice of teachers working in schools with a collaborative approach to the implementation of FA, as compared to those teaching in schools with a more individual implementation approach. Another limitation of this study is that it focusses only on teacher activities, thus ignoring student FA practice. This study investigated the FA activities of teachers, as they are the ones responsible for creating the learning environments in which students are encouraged to engage in the FA process (Boud & Molloy, 2013; Carless & Winstone, 2020). As indicated by the item-level results of the questionnaire, however, teachers apparently experience greater difficulty implementing activities relating to self-assessment and peer-assessment. If students do not actively take part in the FA process, the ultimate effectiveness of the intended FA activities for improving student learning will remain unclear (Carless & Winstone, 2020; Wiliam, 2011). Future research is therefore needed in order to investigate exactly how teachers vary in the application of activities aimed at promoting student engagement in FA

and the ways in which students respond to such activities in terms of engagement, behaviour and learning. For example, future studies could investigate whether teachers who successfully implement such activities use different FA cycle practice and are able to use FA more according to its 'spirit' than are teachers who are less successful in enacting these activities and who use FA more according to the 'letter' of the concept. An additional avenue for research concerns the effects of the various FA practices on the learning and behaviour of students.

Funding

This work was supported by the Government of the Netherlands, Ministry of Education, Culture and Science, [grant number 113214].

Author statement file

M.J. Veugen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Visualization, Roles/Writing - original draft. J.T.M. Gulikers: Conceptualization, Methodology, Resources, Supervision, Validation, Writing - review & editing. P. den Brok: Conceptualization, Methodology, Resources, Supervision, Validation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Teacher FA Questionnaire with mean scores

Teacher FA Questionnaire, and mean scores of all Teachers (n = 96) on FA activities

Activity	*M*	*SD*
Phase 1		
1. I formulate learning goals for the next lesson.	3.92	1.00
2. I formulate learning goals for a longer period (lesson plans or annual targets).	3.79	1.10
3. I formulate the criteria for a well-elaborated assignment (success criteria).	3.49	.89
4. When I formulate learning goals, I take the curriculum into account.	4.08	.83
5. I know the most common misconceptions in my subject/domain.	3.88	.79
6. I discuss the learning goals of the lesson with students.	3.71	1.02
7. I discuss the learning goals for a longer period (lesson plans/annual targets) with students.	3.52	1.23
8. I discuss the importance of the learning goals with students.	3.27	1.02
9. I check whether the students understand the learning goals.	3.48	.88
10. I let students create their own learning goals.	2.15	1.02
11. I discuss the criteria for a well-elaborated assignment (success criteria) with students.	3.57	.98
12. I use examples to show students what the task should look like.	3.31	1.06
Phase 2		
13. I use different FA methods to determine what students know and are capable of doing.	3.45	.87
14. I consciously choose FA methods that fit the learning goals.	3.25	.92
15. I use formal FA methods, like pre-testing and interim testing, reports and portfolio discussions.	3.15	1.03
16. I use informal FA methods, like asking deeper questions and observations.	4.02	.83
17. I use a variety of FA methods to identify the students' misconceptions.	3.20	.82
18. I give students the opportunity to show what they have learned.	3.92	.79
19. I ask students to explain their answers.	4.20	.69
20. I let students explain their answers to each other.	3.54	.85
Phase 3		
Based on the information gathered in Phase 2…		
21. … I identify the strengths of the whole group.	3.47	.86
22. … I identify the points for improvement of the whole group.	3.59	.89
23. … I identify the strengths of individual students.	3.24	.93
24. … I identify the points for improvement of individual students.	3.35	.91
25. … I identify the students' misconceptions with regard to the learning topic.	3.38	.84
26. I take time to analyse the students' reactions.	2.86	.88
27. I compare students' work to the criteria for the task.	3.41	.99
28. I let students compare their own work to the criteria for the task.	3.29	.96
29. I let students compare each other's work to the criteria for the task.	3.22	1.01
Phase 4		
Based on the analyses of Phase 3…		
30. … I discuss the strengths of the whole group.	3.25	.97
31. … I discuss the points for improvement of the whole group.	3.45	.94
32. … I discuss the strengths with individuals/small groups	3.17	1.02
33. … I discuss the points for improvement with individuals/small groups.	3.36	1.02
34. I let students think about their own strengths and points for improvement.	2.85	1.11
35. I let students think about each other's strengths and points for improvement.	2.64	1.12
36. I give students feedback when they are working on an assignment.	4.03	.77
37. I let students improve their work after receiving feedback.	3.63	.97
38. I provide students with tools to give each other feedback.	3.16	1.23
Phase 5		
39. I deviate from the learning material/planned targets when I discover that the learning goals have not yet/already been reached.	3.80	1.03

(*continued on next page*)

(*continued*)

Activity	*M*	*SD*
40. I deviate from the agreed lesson plans when I discover that the learning goals have not yet/already been reached.	3.84	.99
41. I help the students to come up with the next step in learning.	2.82	1.10
42. I let students come up with their next step in learning by themselves.	2.73	1.12
43. I let students come up with their next step in learning with each other.	2.25	.97
44. Based on my previous analyses, I consciously differentiate within my group.	3.09	1.02

Appendix B. Student FA Questionnaire, with mean scores of teachers and students

Student FA Questionnaire, with mean scores for Teachers (n = 59) and Student Classes (n = 59) on FA Activities.

	Teachers $n = 59$		Student Classes $n = 59$	
Activity	*M*	*SD*	*M*	*SD*
Phase 1				
1. The teacher discusses the learning goals with us.	3.87	1.00	4.06	.60
2. The teacher discusses the importance of the learning goals with us.	3.41	.97	3.47	.51
3. The teacher checks whether we understand the learning goals.	3.63	.81	3.64	.51
4. The teacher lets us create our own learning goals.	2.21	1.02	2.52	.54
5. The teacher uses examples to show us what the task should look like.	3.28	1.00	3.85	.45
Phase 2				
6. The teacher gives us the opportunity to show what we have learned.	3.90	.80	3.67	.44
7. The teacher asks us to explain our answers.	4.12	.67	4.05	.40
8. The teacher lets us explain our answers to each other.	3.48	.88	3.32	.47
Phase 3				
9. The teacher compares our work to the criteria for the task.	3.43	.91	3.28	.45
10. The teacher lets us compare our own work to the criteria for the task.	3.28	.98	3.33	.42
11. The teacher lets us compare each other's work to the criteria for the task.	3.22	1.05	2.97	.54
Phase 4				
12. The teacher discusses the strengths of the whole group.	3.33	.92	3.43	.93
13. The teacher discussed the points to improve of the whole group.	3.43	.93	3.52	.41
14. The teacher discusses my strengths with me.	3.19	1.02	2.83	.62
15. The teacher discusses my points for improvement with me.	3.26	.99	3.03	.59
16. The teacher lets me think of my own strengths and points for improvement.	2.85	1.08	2.82	.54
17. The teacher lets us think about the strengths and points for improvement with each other.	2.63	1.08	2.63	.63
18. The teacher gives us feedback when we are working on an assignment.	4.03	.74	3.71	.48
19. The teacher lets us improve our work after receiving feedback.	3.60	.89	3.67	.52
20. The teacher provides us with tools to give each other feedback.	3.07	1.17	3.42	.51
Phase 5				
21. The teacher helps us to come up with the next step in learning.	2.94	1.06	2.97	.57
22. The teacher lets us come up with our next step in learning by ourselves.	2.81	1.09	3.07	.46
23. The teacher lets us come up with our next step in learning with each other.	2.36	1.03	2.76	.59

References

Aluri, V. L. N., & Fraser, B. J. (2019). Students' perceptions of mathematics classroom learning environments: Measurement and associations with achievement. *Learning Environments Research, 22*, 409–426. https://doi.org/10.1007/s10984-019-09282-1.

Andersson, C., & Palm, T. (2018). Reasons for teachers' successful development of a formative assessment practice through professional development – A motivation perspective. *Assessment in Education: Principles, Policy & Practice, 25*(6), 576–597. https://doi.org/10.1016/j.learninstruc.2016.12.006.

Antoniou, J., & 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.

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.

Birenbaum, M. (2014). Conceptualizing assessment culture in school (2014). In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.). *Designing assessment for quality learning* (vol. 1, pp. 285–302). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-5902-2.

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

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 6*(38), 698–712. https://doi.org/10.1080/02602938.2012.691462.

Carless, D. (2017). Scaling up assessment for learning: Progress and prospects. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (vol. 5, pp. 3–17). Singapore: Springer. https://doi.org/10.1007/978-981-10-3045-1.

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

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

De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research, 4*, 51–85. https://doi.org/10.1023/A:1011402608575.

Den Brok, P., Levy, J., Wubbels, T., & Rodriguez, M. (2003). Cultural influences on students' perceptions of videotaped lessons. *International Journal of Intercultural Relations, 27*, 355–374. https://doi.org/10.1016/S0147-1767(03)00016-6.

Deneen, C. C., Fulmer, G. W., Brown, G. T. L., Tan, K., Leong, W. S., & Tay, H. Y. (2019). Value, practice and proficiency: Teachers' complex relationship with assessment for learning. *Teaching and Teacher Education, 80*, 39–47. https://doi.org/10.1016/j.tate.2018.12.022.

Dobbelaer, M. J. (2019). *The quality and qualities of classroom observation systems (Doctoral dissertation)*. Retrieved on 19 March, 2020 from: (pp. 81–85) Enschede, the Netherlands: University of Twente https://ris.utwente.nl/ws/portalfiles/portal/88769970/Dobbelaer_The_quality_and_qualities_of_classroom_observation_systems.pdf.

Forbes, C. T., Sabel, J. L., & Biggers, M. (2015). Elementary Teachers' Use of Formative Assessment to Support Students' Learning About Interactions Between the Hydrosphere and Geosphere. *Journal of Geoscience Education, 63*(3), 210–221. https://doi.org/10.5408/14-063.1.

Fraser, B., & Walberg, H. (1981). Psychosocial learning environment in science classrooms: A review of research. *Studies in Science Education, 8*(1), 67–92. https://doi.org/10.1080/03057268108559887.

Goldstein, H. (1999). *Multilevel statistical models* (internet version) Retrieved on 16 September, 2019 from: (2nd edition). London: Arnold Publishers, Whiley. https://stats.idre.ucla.edu/wp-content/uploads/2016/02/goldstein-1.pdf.

Government of the Netherlands. (2020). *Secondary education*. April 23, Retrieved from https://www.government.nl/topics/secondary-education.

Gulikers, J., & Baartman, L. (2016). Teachers' formative assessment practices in the classroom: A literature review. In *Paper presented at the Compentence Conference*. Retrieved on May 10th 2021, from https://www.slideshare.net/JudithGulikers/teachers-formative-assessment-practices-in-the-classroom-a-literature-review-using-the-formative-assesssment-cycle.

Gulikers, J. T. M., & Baartman, L. K. J. (2017). *Doelgericht professionaliseren: formatieve toetspraktijken met effect! Wat DOET de docent in de klas? [Targetted professonial development: formative assessment practices with effect! What the teacher DOES in the classroom]. NRO review report number 405-15-722*. the Netherlands: Wageningen University. Retrieved on 10 May, 2021 from: https://www.nro.nl/sites/nro/files/migrate/Inhoudelijke-eindrapport_NRO-PPO-405-15-722_DEF.pdf.

Hansen, G. (2020). Formative assessment as a collaborative act. Teachers' intention and students' experience: Two sides of the same coin, or? *Studies in Educational Evaluation, 66*. https://doi.org/10.1016/j.stueduc.2020.100904.

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation, 38*, 21–27. https://doi.org/10.1016/j.stueduc.2012.04.001.

Hawe, E., & Parr, J. (2014). Assessment for learning in writing classroom: an incomplete realisation. *The Curriculum Journal, 25*(2), 210–237. https://doi.org/10.1080/09585176.2013.862172.

Heitink, M. C., van der Kleij, F., 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., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28*(3), 24–31. https://doi.org/10.1111/j.1745-3992.2009.00151.x.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A multidisciplinary journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

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*(October), 199–213. https://doi.org/10.1016/j.tate.2018.06.015.

Kreft, I. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage Publications. https://doi.org/10.4135/9781849209366.

Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037//0022-3514.77.6.1121.

Lattin, J. M., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Pacific Grove, CA: Thomson Brooks/Cole. https://doi.org/10.1111/j.1745-3992.2009.00151.x.

Mara, C. A., & Cribbie, R. A. (2012). Paired-samples tests of equivalence. *Communication is Statistics – Simulation and Computation, 41*, 1928–1943. https://doi.org/10.1080/03610918.2011.626545.

Marshall, B., & Drummond, 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.

OECD [The Organization for Economic Cooperation and Development]. (2005). *Formative assessment. Improving learning in secondary classrooms [Adobe Digital Editions version]. Retrieved on 19 March, 2020 from*. https://www.oecd-ilibrary.org/education/formative-assessment_9789264007413-en.

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. *The Australian Educational Researcher, 45*(1), 13–31. https://doi.org/10.1007/s13384-018-0258-y.

Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2015). Multilevel predictors of differing perceptions of Assessment for Learning practices between teachers and students. *Assessment in Education: Principles, Policy & Practice, 22*(2), 282–298. https://doi.org/10.1080/0969594X.2014.975675.

Robinson, J., Myran, S., Strauss, R., & Reed, W. (2014). The impact of an alternative professional development model on teacher practices in formative assessment and student learning. *Teacher Development, 18*(2), 141–162. https://doi.org/10.1080/13664530.2014.900516.

Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3-4), 205–235. https://doi.org/10.1080/10627197.2006.9652991.

Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. *School Effectiveness and School Improvement, 28*(2), 242–258. https://doi.org/10.1080/09243453.2016.1256901.

Schmermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 28–74. Retrieved on 16 September from https://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf.

Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation, 37*, 49–54. https://doi.org/10.1016/j.stueduc.2011.03.008.

Segers, M., Nijhuis, J., & Gijselaers, W. (2006). Redesigning a learning and assessment environment: The influence on students' perceptions of assessment. *Studies in Educational Evaluation, 32*, 223–242. https://doi.org/10.1016/j.stueduc.2006.08.004AND.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., et al. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education, 21*(4), 295–314. https://doi.org/10.1080/08957340802347647.

Van der Kleij, F. M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and association with individual student characteristics. *Teaching and Teacher Education, 85*, 175–189. https://doi.org/10.1016/j.tate.2019.06.010.

Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education/Revue Canadienne De L'éducation, 34*(2), 239–255. Retrieved on 23 March, 2020 from www.jstor.org/stable/canajeducrevucan.34.2.239.

Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment, 11*(3-4), 283–289. https://doi.org/10.1080/10627197.2006.9652993.

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

Wiliam, D. (2018). *Formative assessment: Confusions, clarifications & prospects for consensus [Seminar presentation]*. February 28. United Kingdom: Oxford https://www.dylanwiliam.org/Dylan_Wiliams_website/Presentations.html.

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