The influence of formative assessment on secondary school students' autonomous motivation for biology

Sjoerd S.A. Banken[1]
Utrecht University

This mixed approach study focusses on the influences of teacher's formative assessment (FA) practices on secondary school student's autonomous motivation for biology. Using the five key strategies of FA, *feed-up, feedback, feed-forward, peer-feedback and self-assessment,* as described by the Assessment for Learning discourse, three biology teachers were observed in class on their use of these strategies. A broader questionnaire was also administered to compare the teachers' use of FA to that of the average of Dutch secondary school teachers. To assess developments in student motivation in line with the Self-Determination Theory framework, a Self-Regulation questionnaire on distinct motivational constructs was administered at the start and end of the three month study period in the five observed classes (students aged 12-15). FA-effects mentioned in teacher interviews and student focus groups were coded in terms of being supportive or thwarting of the three basic psychological needs for competence, autonomy and relatedness, required to become autonomously motivated. Results show individual differences between teachers in their use of FA strategies, although *feedback* is clearly used most often by all three. Findings from interviews and focus groups indicate that use of formative assessment strategies in secondary biology education, especially feedback, supports the psychological needs for competence and autonomy. However the students' score on motivational constructs (i.e. controlled vs. autonomous) did not change significantly over a three month time-span. The findings are discussed on their implications for educational practice and future research.

Keywords: Self-determination theory (SDT), autonomous motivation, competence, autonomy, relatedness, biology education, formative assessment (FA) & Assessment for Learning (AfL).

INTRODUCTION

Astronomer Carls Sagan bluntly said: 'Every kid starts out as a natural-born scientist, and then we beat it out of them. A few trickle through the system with their wonder and enthusiasm for science intact' (1996). Students start their school career with strong interest towards science, but engagement and motivation in education has been shown to drop enormously trough secondary school (e.g., Gallup, 2016; Osborne, Simon, & Collins, 2003; Potvin & Hasni, 2014). This is alarming since interest, an aspect of high quality

---

This master's thesis was written by Sjoerd Banken (Utrecht University, Science Education and Communication, student number 4131185) in the academic year 2019-2020 in the context of the course Research Project Science Education and Communication (FI-MSECR30, 30 ECTS). The research work was supervised by Ralph F. G. Meulenbroeks (Utrecht University, Freudenthal Institute) and second examiner was Arthur Bakker (Utrecht University, Freudenthal Institute).

motivation, is a strong predictor of academic success, deep learning, engagement, and other desirable educational outcomes (e.g. Cerasoli, Nicklin, & Ford, 2014; Ryan & Deci, 2000; Vansteenkiste, Sierens, Soenens, Luyckx, & Lens, 2009). Especially in more developed countries the interest towards science is low amongst older students resulting in a decreased motivation towards studying science (van Griethuijsen et al., 2015). This is a source of societal concern since these countries have strong knowledge economies that depend on adequate numbers of graduates in science and technology. Besides this, the goals to increase scientific literacy amongst the public and gaining public support for science in general, has also called for the lack of motivation towards school science to stay on the educational and political agenda (Tytler & Osborne, 2012) .

Attention towards formative assessment

Educational researchers study the effect of teaching strategies and new tools on students' motivation and performance, for example inquiry-based teaching, context-based teaching and 'education outside the classroom' that have all been shown to promote motivation towards science (Bølling, Otte, Elsborg, Nielsen, & Bentsen, 2018; Savelsbergh et al., 2016). In the past 20 years, however, increasing interest has been vested on another variable in this context: classroom assessment (e.g. Black & Wiliam, 2009; Lau, 2016; Pulfrey, Buchs, & Butera, 2011; Shepard, Penuel, & Pellegrino, 2018; Taras, 2009). The use of grades as a normative evaluation of learning (standardized tests) seems to be an essential, even inescapable, aspect of education.

When an assessment is used to summarize learning this is referred to as 'summative assessment'. This type of assessment is needed for selection, certification and monitoring of progress. On the other hand, Hattie and Timperley conclude that current assessments are '*too* often used as external accountability thermometers rather than as feedback devices that are integral to the teaching and learning process' (2007, p. 104). The strong bureaucratic emphasis in education on summative use of (external) assessments means that high stakes are dependent on those assessments for both teachers and students. This subsequently leads to a 'teaching to the test' culture characterized by surface-learning approach, non-desirable behaviour such as cheating and disengagement from anything that is not graded (Bennett, 2011; Sluijsmans & Kneyber, 2016). On top of that, strong emphasis on incentives such as grades has been associated with the adoption of performance-avoidance goals (focus on avoiding failure) (Butler, 2006), related to low ability-related self-esteem, anxiety, shame and reduced autonomous motivation (Pulfrey et al., 2011). Furthermore, Cerasoli *et al*. conclude quality type tasks that require creativity, autonomy and learning should not be directly incentivised with grades because this narrows cognitive focus and increases likelihood of counterproductive behaviours (2014).

An alternative approach to assessment, formative evaluation or assessment, referred to as formative assessment (FA) in this study, is a process aimed at providing feedback about the learning process while it takes place to further enhance learning (Black & Wiliam, 2009). In the field of education a shift in focus is visible from only using summative assessment (SA), that takes place after learning and often leads to questionable quantity type results (Harlen & Crick, 2002), towards embedding FA in that is believed to give more qualitative insight into the students' progress (Penuel & Watkins, 2019; Shepard et al., 2018). FA is the new orthodoxy believed by many to have the potential to improve education; as such it is becoming a standard part of teacher-training programs (e.g. Bennett, 2011; Shepard et al., 2018; Taras, 2009; William, 2014).

Returning to our original concern, could implementation of formative assessment then be a way to stimulate students' motivation towards school science? Formative assessment requires and stimulates active student participation, which could lead to increased motivation (Edward L. Deci & Ryan, 2008; Wiliam, 2011). Yet, so far little evidence has surfaced, since scientific consensus on formative assessment is weak with researchers using different definitions and theories of action in their rationale (Penuel & Watkins, 2019; Taras, 2009). As Bennet (2011, p. 20-21) ends his critical review:

> Formative assessment is both conceptually and practically still a work-in-progress. That fact means we need to be more sensible in our claims about it, as well as in our expectations for it. That fact also means we must continue the hard work needed to realise its considerable promise.

Certain aspects of formative assessment such as the power of using feedback after an assessment and self-assessment are well acknowledged (Hattie & Timperley, 2007). The overall process of FA however has not been adequately empirically substantialized (R. de Kleijn, Prins, Lutz, van Look, & van Tartwijk, 2016; Savelsbergh et al., 2016). Moreover, motivation-specific longitudinal studies of the effect of formative assessment are sparse while anecdotal and theoretical reports are abundant and usually very positive on the potential benefits of formative assessment (Brookhart, Moss, & Long, 2010; Sluijsmans & Kneyber, 2016; Stroet, Opdenakker, & Minnaert, 2015).

Meanwhile focus on FA is an emerging trend in Dutch education: teachers are innovating individually, are coached in professional training programs, or cooperate in school development teams to adopt formative teaching practices; some schools organize large pilot cohorts that allow teachers great freedom to experiment with student-centred education with little graded assessments and more formative assessment practices (Castelijns, J., & Andersen, 2013; R. de Kleijn et al., 2016; Sluijsmans & Kneyber, 2016).

As stated before, it is remains unclear how this development will eventually affect student motivation for school science. Therefore, the purpose of the current research is to study students' motivation in response to formative teaching strategies. This will be done by studying three Dutch secondary school biology teachers that are innovating their teaching by adopting FA practices. Formative assessment will be approached from the Assessment for Learning (AfL) discourse, that focuses on daily practice and in class use of assessment by teachers (Wiliam, 2011).

Research questions

The research question thus becomes: 'What influence does formative assessment have on secondary school students' autonomous motivation for biology?' In order to answer this question three biology teachers and several of their classes were studied over a three month period. To obtain a more comprehensive answer the following sub questions were postulated:

1. How often are the five formative key strategies used by biology teachers?
2. Does the students' motivation pattern for biology, i.e. controlled vs. autonomous, change in a formative teaching setting?
3. What influence does formative assessment have on students' perceived competence, relatedness and autonomy in relation to school biology?
4. What are the teachers' perceived effects of introducing formative assessment in secondary school biology on student students' perceived competence, relatedness and autonomy?

THEORETICAL BACKGROUND

Motivation and SDT

Motivation is complex multifaceted concept. It is defined by Harlen and Crick as 'the drive, incentive or energy to do something', it embraces, for example, effort, self-regulation and interest (2002). One of the leading theories in the field of motivation research is the Self Determination Theory (SDT) (Ryan & Deci 2017). which will be used in this study as a framework to investigate students' motivation for school biology. SDT describes a multidimensional specification of six types of motivation that are on a continuum from amotivation to intrinsic motivation. Along that continuum motivation is increasingly autonomous, that is determined from an internal perceived locus of causality (Deci & Ryan, 1985; Ryan & Deci, 2000). Completely self-determined actions are driven by intrinsic motivation, solely because of interest and enjoyment of a task. According to SDT three psychological needs have to be satisfied in order to support intrinsic motivation or to internalize more extrinsic forms of motivation: competence (the need to produce desired outcomes and to experience success), autonomy (the need to feel ownership of one's behaviour, it being in accordance with one's volition; which is not 'independence') and relatedness (the need to feel connected to others) (Ryan & Deci, 2000). Complete lack of self-determination results in 'amotivation', which means that no personal intention of performed behaviour is experienced.

Extrinsically motivated behaviour is by definition not performed primarily out of interest, but in search of an external reward, a consequence it is thought to be instrumentally linked to. Extrinsic motivation is considered to be crucial for socially prescribed activities, since those are often not inherently interesting or enjoyable (Taylor et al., 2014). Extrinsic motivation is divided into four regulatory styles: external, introjected, identified and integrated, in the order of increasing self-determination (Ryan & Deci, 2000). External regulation occurs when trying to obtain rewards or avoid punishment (Vansteenkiste & Ryan, 2013). For instance studying hard to avoid punishment after failing a test. Introjected regulation refers to avoiding feelings of shame or guilt or wanting to attain success relative to others' expectations to prove one's worth. Identified regulation happens when an individual personally attributes value to an

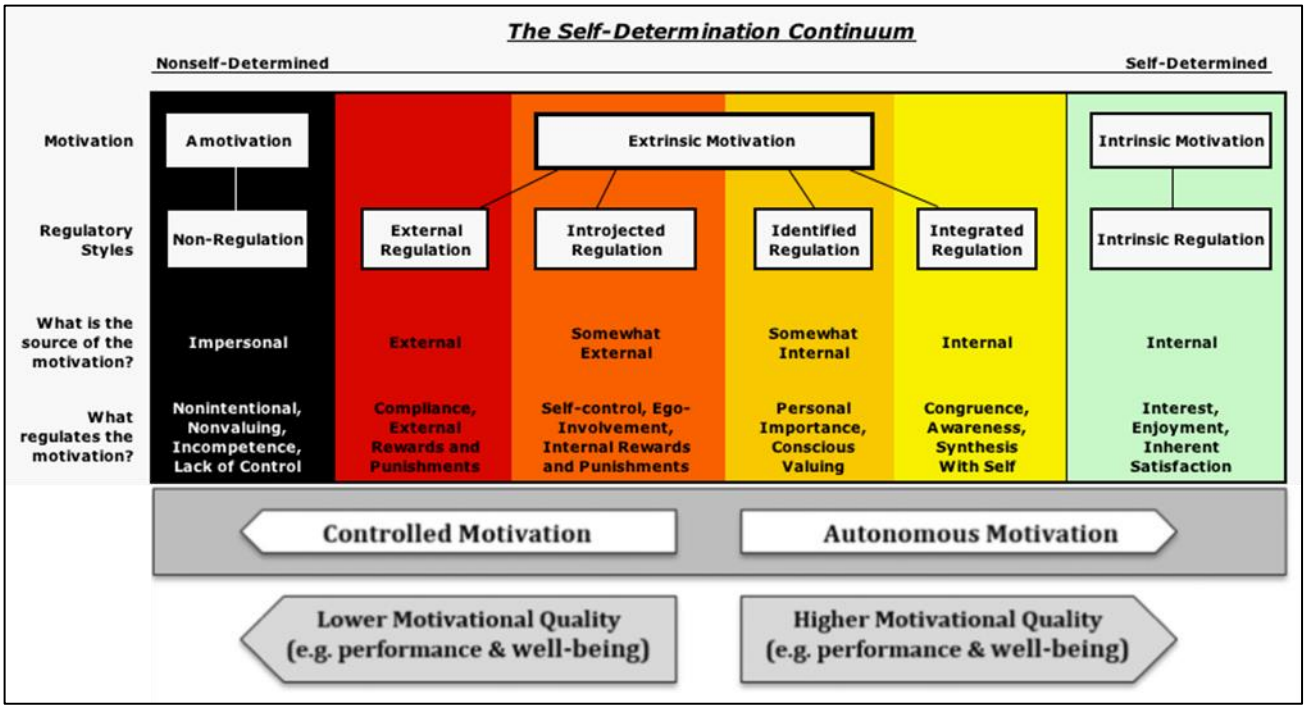

Figure 1: Adapted version of motivation continuum: Organismic Integration Theory Taxonomy of Regulatory Styles (Center for Self-Determination, 2017)

activity, for instance a student rereading a textbook because she believes this will help to understand the subject matter (Ryan & Deci, 2000). Finally, integrated regulation motivates behaviour that is synchronous with an individual values and commitments, e.g. adopting a vegetarian diet because of being an environmental aware person. Figure 1 gives a visual model of the discussed motivation continuum. The motivation continuum can be simplified for practical reasons, in that case external and introjected regulation are grouped into 'controlled motivation' and identified, integrated and intrinsic regulation are taken together as 'autonomous motivation', see figure 1 (e.g. Cerasoli et al., 2014). Each person has a complex profile of intensities of motivation types dependent on the task and circumstances at hand (Ryan & Deci, 2000).

SDT assumes that neighbouring motivations types on the continuum are stronger tied together then the ones further apart. Evidence for this hypothesis however is still somewhat inconclusive (Taylor et al., 2014). Vansteenkiste *et al.* (2009) report that four distinct motivational profiles can be found in high school students based on autonomous and controlled motivation scales (i.e., high– high, high–low, low–high, and low–low). Teachers should try to create a climate where students are not thwarted but stimulated in their feeling of competence, autonomy and relatedness since this will result in 'high quality motivation', relative strong autonomous motivation, that has been linked to deep learning, self-regulated learning, improved well-being and performance and increased engagement (e.g. Ryan & Deci, 2000; Tessier, Sarrazin, & Ntoumanis, 2010; Vansteenkiste & Ryan, 2013; Vansteenkiste *et al*., 2004). Indeed students who display high quality motivation experience their teachers as structuring (competence stimulating), involved (i.e. high relatedness) and autonomy supportive (Vansteenkiste et al., 2009). The use of FA might assist teachers in supporting these basic needs in their students.

Formative assessment

FA might be described as a part of innovative pedagogy as it steps away from traditional lecturing and advocates for active participation from the student (Black & Wiliam, 2018). The traditional and still dominant 'teacher-centred' paradigm is based upon transmission of knowledge, basically "I know and you don't know, and I'm here to communicate it to you and explain it to you" (Osborne, as cited in Hatch, 2018). Instead, scientific literature has started to define learning from a constructivist viewpoint: it is a process that requires students to actively construct their knowledge by interacting with the information, discuss it with their peers and teacher, and compare and connect it with their previous conceptions (Nicol & Macfarlane-Dick, 2006). Active learning and student responsibility for learning are key elements of the 'student-centred' approach to instruction (Lea, Stephenson & Troy, 2003). Formative assessment is an integral part of student-centred learning (Schuld, Kanjee, & White, 2017). FA allows the teacher and students to collect and then act on information about learning that takes place in the classroom. It is defined by Assessment for Learning protagonists Black & Wiliam as:

> Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction [i.e. learning activities] that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (2009, p.9)

Formative assessment can be very formal, for example using portfolios or written tests, but can also be informal when, for instance, using classroom discussion, quizzes or observations. Three processes are essential when using feedback to enhance learning (Hattie & Timperley, 2007; Wiliam & Thompson, 2007)**:** Establishing where the learners are going, establishing where they are in their learning and

establishing what needs to be done to get them there. Five key strategies of FA are derived when the three processes are connected with the different actors in education (teacher, peer, learner) also shown in figure 2 (Wiliam and Thompson 2007):

1. Clarifying and sharing learning intentions and criteria for success (referred to as 'feed-up'.)
2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding (referred to as 'feedback')
3. Providing feedback that moves learners forward (referred to as 'feed-forward')
4. Activating students as instructional resources for one another (referred to as 'peer-feedback')
5. Activating students as the owners of their own learning (referred to as 'self-assessment')

Research indicates that Dutch secondary school teachers use the main formative strategies in roughly one to two out of four of their lessons and self- and peer assessment are rarely used (Kippers, Wolterinck, Schildkamp & Poortman, 2016). Besides, the quality of that feedback can be improved, as teachers often remain unclear/imprecise or do not refer to learning goals in their comments. Successful implementation of FA is dependent on many factors, for instance teachers' pedagogical visions, FA knowledge and skills, understanding of subject-matter and students' self-assessment and feedback skills (Heitink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016).

Effective feedback is descriptive and aimed at the task, learning process or self-regulation level (e.g. 'You have made excellent use of Binas to find information'). Feedback that comprises external controllers (punishments and rewards) or is aimed at the personal level unrelated to specifics of the task such as 'you are great' is ineffective and can have negative effects on engagement, learning and social comparison (Hattie & Timperley, 2007).

<table>
<tr><th></th><th>Where the learner is going</th><th>Where the learner is now</th><th>How to get there</th></tr>
<tr><td>Teacher</td><td rowspan="3">Clarifying, sharing, and understanding learning intentions and success criteria</td><td>Engineering effective discussions, tasks, and activities that elicit evidence of learning</td><td>Providing feedback that moves learning forward</td></tr>
<tr><td>Peer</td><td colspan="2">Activating students as learning resources for one another</td></tr>
<tr><td>Learner</td><td colspan="2">Activating students as owners of their own learning</td></tr>
</table>

Fig. 2. Five Key Strategies of Formative Assessment (from Wiliam & Thompson 2007)

Effect of formative assessment on motivation

The small body of research that focusses on motivational effects of formative assessment has been reviewed by de Kleijn *et al*. (2015) and Gulikers & Baartman (2017). These researchers conclude that half of the studies in secondary education reported positive effects of assessment on motivation (evenly distributed over extrinsic and intrinsic motivation), attitude and engagement. Moreover, they notice that negative feedback should be combined with positive feedback to avoid detrimental effect on perceived competence, and use of digital tools and peer-feedback can stimulate motivation (R. de Kleijn et al., 2016).

In an explorative study on two different types of formative assessment (i.e. group and individual) in a Swedish medical course, students reported feeling pressure to study, growing interest, becoming aware of their own learning and obtaining insight into how their peers reasoned and what kind of questions they could expect in their final exam (Weurlander, Söderberg, Scheja, Hult, & Wernerson, 2012). In a survey study conducted by Pat-El, Tillema, & van Koppen (2012) in Dutch secondary vocational education 'monitoring' (i.e. feedback and stimulation of self-monitoring and self-regulation) and 'scaffolding' (i.e. interaction, feed-up and feed-forward) were found to be predictive for intrinsic motivation. Scaffolding was found to positively influence competence, autonomy and relatedness. Interestingly, monitoring and teacher proximity exclusively stimulated fulfilment of the social psychological need for relatedness (Pat-El *et al*., 2012).

The evidence presented above suggests that different types of formative assessment practices have a varying effect on the basic psychological needs and motivation. Scaffolding seems to be the most powerful way to enhance student motivation, an observation also evident in other research (Hattie & Timperley, 2007; Kirschner & van Merriënboer, 2013).

Hypotheses

The teachers participating in this study were selected for their reported interest, motivation and/or known use of formative assessment in their lesson practice. The researcher therefore expects to see considerable use of the key formative assessment strategies. Subsequently, students are expected to experience reduced controlled motivation for biology because they are more focussed towards understanding subject-content and less focussed towards outcomes of summative assessment (i.e. grades), that often work as a major incentive (Black & Wiliam, 2009; Cerasoli et al., 2014). Autonomous motivation is expected to increase because the process of FA is very student-focussed thus giving attention to their basic psychological needs. More specifically, feelings of relatedness, competence and autonomy are expected to be supported since FA uses (peer) interaction, provides quality informational feedback and feed-forward on performance and enables students to take an active role in their own learning by challenging them (Deci & Ryan, 2016; de Kleijn et al., 2016; Pat-El et al., 2012).

METHODS

In this study we adopt a mixed approach to describe what happens with motivation towards school biology when teachers are implementing the five key FA strategies as described by Wiliam & Thompson (2007). Main methods are observations, questionnaires, teacher interviews and student focus groups. This allows for a triangulation of data that reduces the possibility of observer bias or retrospection problems (De Loof, Struyf, Boeve-de Pauw, & Van Petegem, 2019).

Participants

Three biology teachers, employed at different Dutch higher general secondary schools, and five of their classes (2th grade and 3th grade, respectively K-7 and K-8) were studied over a three month timespan. Teacher A (female, age 30, total 6 years teaching experience), teacher B (female, age 38, total 18 years teaching experience) and teacher C (male, age 36, total 14 years teaching experience) were selected for their reported interest, motivation and/or known use of formative assessment in their lesson practice. Their schools are all public schools and located in semi-urban to urban areas. Only the school of teacher C had a non-traditional educational approach called 'domain-education', which is more student-centred and requires more planning and self-regulated learning of the students. In this school classes consist of

mixed ability levels in the first three years. Two studied biology classes of teacher C were mixed *Higher General Secondary Education* and *pre-university education* 3rdgrade classes ('3 Havo/VWO'). Two studied classes of teacher B were second grade *Higher General Secondary Education* and *pre-university education* ('2 Havo' and '2 VWO'). The studied class of teacher A was as 3rd grade *Higher General Secondary Education* class ('3 Havo').

Student motivation questionnaire

A student motivation questionnaire, an adapted version of the in Dutch translated self-regulation questionnaire Academic (SRQ-A) (Peeters, 2015) was administered a few weeks after the start of the school year ('the pre-test'), and this was repeated at the end of the study period after 14 weeks ('post-test'). The students completed the questionnaires during class time on a voluntary basis and the researcher was present to answer students' questions. Students were informed that the questionnaire concerned motivation toward school biology and that their responses would be handled confidentially and anonymously. A personal code, the last four digits of their phone number, was asked to allow for matching between pre- and post-test whilst keeping the questionnaire anonymous.

The original 32-item self-report questionnaire, grounded in the self-determination theory, was developed and validated by Ryan and Connel (1989) to measure both controlled motivation (external regulation and introjected regulation) and autonomous motivation (identified regulation and intrinsic motivation) for doing schoolwork. The questionnaire used in this study can be found in Appendix A. It contains four questions on effort put into school biology such as 'Why do you make your homework for Biology?'. Each question is followed by eight statements, so a total of 32 which represent the four different regulatory scales, for which students score on a four point Likert-scale whether they totally agree (4) or totally disagree (1) or the in-between-options. One of the statements is: 'because it is fun'. This correspond to the intrinsic motivation, the most autonomous scale. Reliability of the scales was assessed, see table 1. Intrinsic motivation, identified regulation and introjected regulation scales had Cronbach's alphas from .78 to .86, indicating high internal consistency. Cronbach's alpha on the extrinsic regulation scale was .66, indicating acceptable internal consistency. Inspection of reliability the analysis did not yield any questions that could be deleted to increase alpha and strengthen the scales.

Table 1

Reliability analysis of motivation regulation scales of Self-Regulation Questionnaire

Scales	Cronbach's Alpha (α)	N of Items
Intrinsic_motivation	,856	7
Identified_regulation	,841	7
Introjected_regulation	,777	9
Extrinsic_regulation	,661	9

Out of the 130 filled in questionnaires at the start of the study period ('pre-test') and the 126 at the end of the study period ('post-test') 119 matching pairs could be made using the four digit code, only these were used in further analysis. Some students were not able to complete both tests due to illness Faulty filled out questionnaires were also eliminated. Some students had misread the instruction above the questionnaire and scored incorrect responses such as circling both the 2 and 3 point answer as to indicate a neutral 2,5 opinion to a statement. These types of values were regarded as missing values in following statistical analysis. Pre- and post-test data were inspected on normality. Figure 3 shows the Q-Q plot of identified regulation scores on the pre-test which suggests non-normality, other Q-Q plots are shown in appendix C. Shapiro-Wilk tests showed a significant departure from normality for intrinsic motivation and identified regulation on both pre- and post-test, while introjected regulation and extrinsic motivation are normally distributed on both tests, see table 2

Table 2

Test of normality (Shapiro-Wilk) on motivation scales

	Test-Statistic (W)	df	Sig.
Pre_Intrinsic_motivation	,965	119	,003
Pre_Identified_regulation	,939	119	,000
Pre_Introjected_regulation	,985	119	,224
Pre_Extrinsic_regulation	,990	119	,530
Post_Intrinsic_motivation	,957	119	,001
Post_Identified_regulation	,963	119	,002
Post_Introjected_regulation	,984	119	,183
Post_Extrinsic_regulation	,984	119	,159

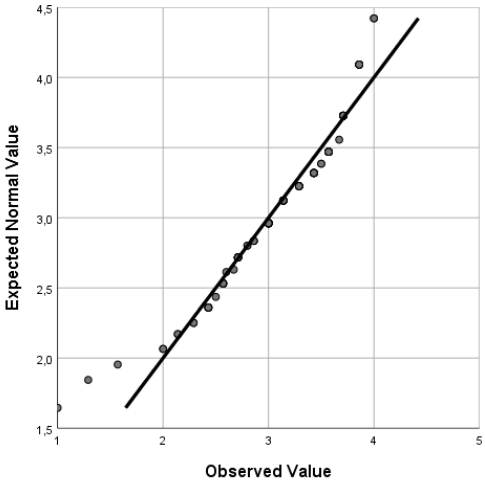

Figure 3. Q-Q plot of normality of the construct identified regulation at the pre-test (N=119).

Observations

Observations were made to gain insight into how often the formative assessment key strategies are used in the lessons of the participating teachers. In between 10 and 20 (biology) lessons of each teacher were observed. The researcher was seated at the back of the classroom and interference with the teacher and students during class was kept as low as possible. Some of the observations were taken in parallel classes of these teachers, classes that did not partake the other aspects of this study (questionnaires and focus groups). The studied biology class of teacher A was also observed in physics classes which were also taught by Teacher A, a necessary adaptation since her biology lessons took place only once a week and were often dropped due to school activities. The first few observations were part of an introductory phase during which a partly bottom-up development of the observation coding scheme took place. The scheme is an elaboration on the five key strategies of the formative assessment cycle as described by Wiliam and Thompson (2007), which were split into different categories of teacher-student interactions, instructions and comments. Each new formative action was counted and if possible some quotes where written down e.g. Teacher A responding to a student answer by saying 'very good explanation' was counted as *Feedback – giving compliment/confirmation*. The coding scheme can be found in Appendix E.

By way of second coding, another science education researcher was briefed on the observation sheet and coded two lessons. A very similar pattern emerged, with only a few counts differentially placed in categories belonging to the same formative key strategy. No formal interrater reliability was calculated, but after a brief discussion, the author and second coder concluded that the coding scheme had sufficient validity. The complete observations after second coding, which amounted to 4 up to 8 lessons per teacher, were used to actually calculate the average frequency occurrence of the different FA strategies per lesson as presented in the results section. Since there was an unequal duration of lessons due to scheduling differences between the schools (60 minutes for Teacher A; 50 minutes for Teacher B; 90 minutes for Teacher C), the occurrences per lesson were al normalized to 60 minute lessons.

Teachers interview and FA questionnaire

All three teachers were asked to fill in the Formative Assessment (*AfL & DBDM)* self-report questionnaire developed and validated by Kippers, Wolterinck, Schildkamp, Poortman, & Visscher (2018). This allows for a limited comparison of the studied teachers to the national average of Dutch prevocational[2] teachers A limited comparison, since the constructs in this questionnaire differ somewhat from the definition of the key Formative Assessment strategies by Wiliam & Thompson that is used in this study. This will elaborated on in the results.

After filling out the questionnaire, the teachers were interviewed for ± 60 minutes each on their use of FA in biology lessons. The interviews focused on the teachers' perceived effects of formative assessment on students. To allow for in depth questioning the structure was kept very open, with all key FA strategies covered in the initial and follow-up questions such as '*Why do you put emphasis on the learning goals as you have just described?'* or '*How do students respond to that?'*. The interviews were audio-taped, transcribed and coded in Nvivo.

Student focus groups

Parents of children in studied classes were informed about the study and asked for consent to let their children partake in a group interview (focus group) with four other students from the same class. The response rate was low, consent was given for roughly a third of the students of each class, only one parent explicitly refused. From each class five children with parental consent were randomly selected and asked to partake in a 30 minute group interview during or after class. All students accepted except for one student who was absent due to sickness and one student who wasn't available after class, these were substituted by subsequently random selected students with parental consent. Participating students were informed that the interview would cover the assessment of knowledge and skills during biology, that their answers would be audio-taped, transcribed anonymously and used for this study. The focus groups took place in a separate room to insure privacy and they were guided by the researcher. It was stressed that all answers and opinions would be welcome and students were stimulated to respond to each other's statements and to elaborate on the effects of assessments in biology. The focus groups were semi-structured in the same way as the teacher interviews to insure that all that all key FA strategies were covered, the focus group interview scheme (in Dutch) can be found in appendix B.

Interview and focus group coding

Responses were coded top-down in Nvivo using the defined Formative key strategies as described by Wiliam & Thompson (2007) and using the rating sheet 'components of need-supportive teaching' that describes competence, autonomy and relatedness supporting and thwarting interactions from Stroet *et al*. (2015, p. 133). To further clarify, the main codes were 'feed-up', 'feedback', 'feed-forward', 'peer-feedback', 'self-assessment' and six influences: negative or positive effect on students' perceived competence, autonomy and relatedness. Only double or multiple coded quotes that contain use of a Formative key strategy combined with an effect on competence, autonomy and/or relatedness are reported in the results section, see table 3 and 4.

To enhance validity and reliability of the coding a large amount of coded transcript was discussed with another researcher. During this process discussion on the definitions helped to sharpen the coding. Only

[2] VMBO

interactions that were explicitly aimed at helping the student progress in his/her learning were coded as 'feed-forward'. Tests, tasks or informative feedback such as compliments about students current achievement, that might still lead to subsequent action, where coded as 'feedback'. When students explicitly monitor, evaluate or redirect their own learning this was coded as 'self-assessment' (Cauley & McMillan, 2010). The code 'peer-feedback' was used broader and more informal for all cases of interactions between students that consisted of helping or correcting each other on learning strategies or biology subject matter. This type of formative assessment is often not initiated or even deliberately facilitated by teachers. To determine reliability of the coding, 44 of the 175 final quotes that were coded with a Formative Assessment strategy label combined with an effect on competence, autonomy and/or relatedness were randomly selected from all teacher interviews and student focus groups. These were 2nd coded by a science education researcher familiar with SDT. Overall, the average interrater reliability yielded a Cohen's kappa value of .90, indicating a very good agreement.

RESULTS

First, the use of Formative Assessment will be described to answer the first research sub question. Subsequently, the motivational regulation styles at the start and end of the study period are presented to answer the second research sub question. Finally, the relevant codes from the student focus groups and teacher interviews are summarised in a table and the most important findings are elaborated to answer the third and fourth research sub questions.

Teacher use of formative assessment strategies

Fig. 4 shows the average observed use of the Formative Assessment strategies by the teachers over several lessons. It is evident that feedback, followed by feed-forward, was the most often used strategy (22 up to 54 times on average per lesson hour). Teacher B used these strategies almost triple fold compared to the other two teachers. Feed-up, peer-feedback and self-assessment were use only several times per lesson (1 up to 6,5 times). The largest relative difference was found for the strategy self-assessment that Teacher C used almost five times more often than Teacher A and two times as often as Teacher B, all absolute frequencies being relatively low.

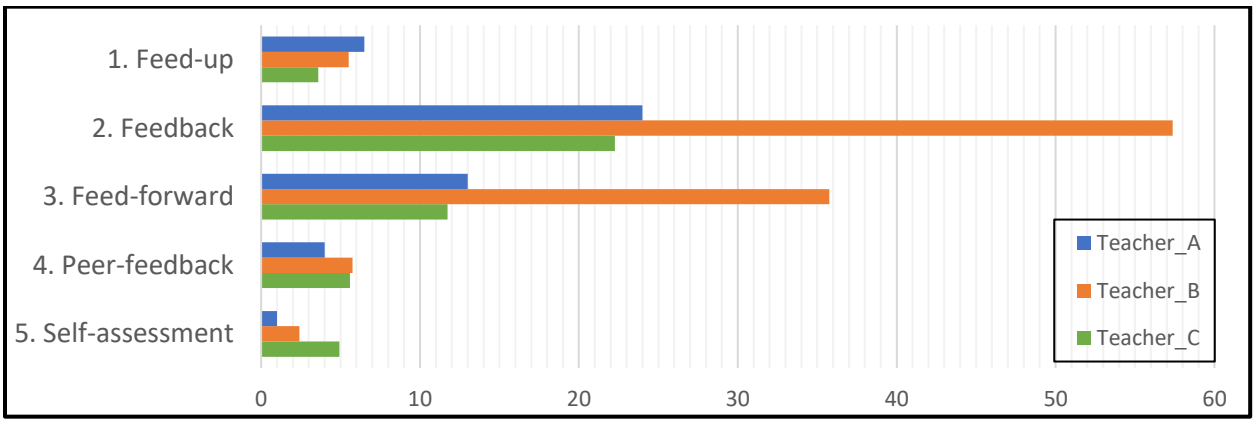

Figure 4. Average observed occurrences of FA-strategy use by teachers over 4 to 8 lessons (corrected to 1 lesson hour (60 min)). Used coding scheme is visualised in appendix E.

In order to put these results in some perspective, Figure 5 shows the result of the AfL-DBDM self-report questionnaire on how much teachers use formative assessment (i.e. Assessment for Learning and Data-based decision making) in their classroom practice. The average score from 479 teachers from Dutch schools, from the original study, is incorporated as a reference point (Kippers et al., 2018). It is important to remember that that the constructs in this AfL-DBDM questionnaire differ from the definition of the key Formative Assessment strategies by Wiliam & Thompson that is used in this study. For instance 'feedback' in the AfL-DBDM questionnaire has a very narrow teacher-centred focus on using formal (written) test results to provide high quality feedback/feed-forward to students or adjust future instruction. This is used only 'sporadic' by Teacher B, but often by Teachers A and C (between 'emerging' and 'established', 50% - 75%). The construct 'Asking questions and classroom discussions' in AfL-DBDM entails informal direct in class interactions that mainly correspond to feedback, feedforward and peer-feedback in this study. Contrary to the 'feedback' construct, Teacher B's score for 'asking questions and classrooms discussions' is much higher ('established' 75%) then the average teacher ('emerging' 50%), while Teacher A has a much lower score (just above 'sporadic', 25%). On the construct 'sharing learning intentions and success criteria', which mainly corresponds to the feed-up strategy, all three teachers scored high resemblance to the average teacher. Remarkably, on the construct 'peer- and self-assessment' Teacher C scored twice as high compared to the average teacher, while Teacher A and B scored less than half. This seems to be somewhat resembled in the observations (figure 4).

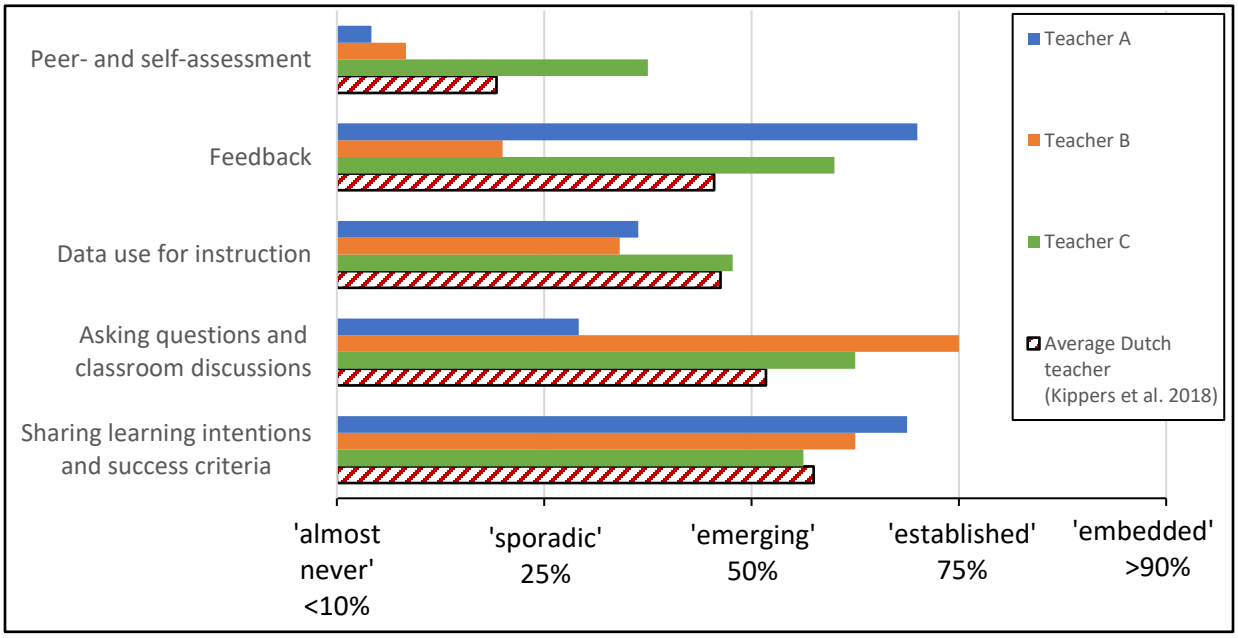

Figure 5. Teachers response on formative assessment use (AfL-DBDM) questionnaire based on Kippers *et al.* (2018)

Student motivation questionnaire

All participants' (N=119) pre- and post-test mean scores of the motivation regulation styles for school biology are shown in Figure 6. The results split per studied class are visualised in appendix D. Autonomous motivation seems to have slightly decreased while controlled motivation slightly increased. The largest difference however, for identified regulation, is just -0.09 points difference between the pre-test (M = 3.09, SD = 0.57) and post-test (M = 3.00, SD = 0.57). Students' self-reported motivation did not change significantly over the 3 month study period as indicated by a Wilcoxon signed-rank test (Wilcoxon, 1945) for intrinsic motivation (Z = -0.865, p = 0.387) and identified regulation (Z = -1.405, p = 0.160) and by a dependent samples T-test for introjected regulation ($t(118)$ = -0.37, p = 0.715) and extrinsic regulation ($t(118)$ = -0.91, p = 0.37). These results suggest students' motivation for biology did not change in the three month period during which the influence of formative assessment used by their biology teachers was studied.

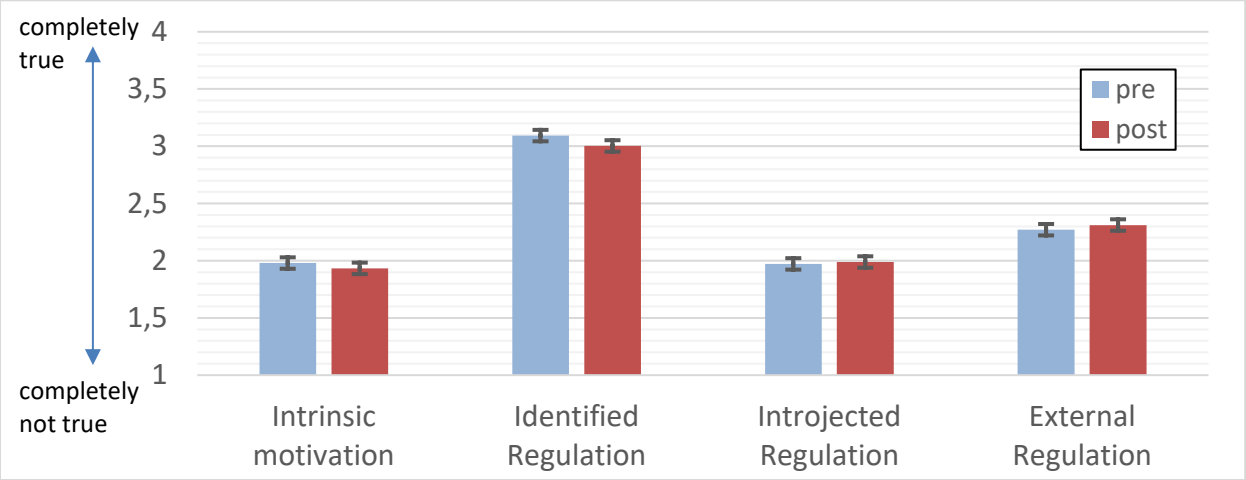

Figure 6. Students' (N=119) response to the Academic Self-Regulation Questionnaire (32 items on 4-point Likert scales) on motivation towards school biology at the start (*pre)* and end (*post*) of the three month study period.

Teacher interviews and student focus groups

All 175 quotations that indicated use of a formative assessment strategy combined with an effect of competence, autonomy or relatedness are summarised in table 3 and 4. A relatively similar pattern emerged from both the teacher perceived effects as the student reported effects. A positive effect of feedback on perceived competence is mentioned most often. To a lesser degree FA strategies are mentioned with a positive effect on autonomy, especially feed-up in the teachers' perception. Competence and autonomy thwarting effects are also indicated, though 3 to 5 times less than the positive effects. Students especially brought up feedback practices with a negative effect on competence, while teachers experience relatively stronger autonomy impairment due FA implementation. Interestingly, almost all the quotes regarding feedback or feedforward that supports relatedness stem from Teacher C and his students. Negative effects on relatedness are barely mentioned. The following section elaborates on the described mechanisms.

Table 3
Summary of coded quotes from teacher interviews that describe use of Formative Assessment (FA) strategies with an effect on competence, autonomy or relatedness. *In some cases a quote describes two strategies being combined with a single effect or vice versa, this is corrected for in the totals.

Teacher perceived	Competence +	Autonomy +	Relatedness +	Competence -	Autonomy -	Relatedness -	Total FA strategy
1: Feed-up	6	10	0	1	2	0	16*
2: Feedback	16	9	4	2	3	1	33*
3: Feed-forward	10	3	4	1	2	0	19*
4: Peer-feedback	3	0	0	1	1	0	4*
5: Self-assessment	1	3	0	1	1	0	5*
Total CAR-effect	31*	25	8	6	8*	1	71

Table 4
Summary of coded quotes from student focus groups that describe use of Formative Assessment (FA) strategies with an effect on competence, autonomy or relatedness (CAR). *In some cases a quote describes two strategies being combined with a single effect or vice versa, this is corrected for in the totals.

Student reported	Competence +	Autonomy +	Relatedness +	Competence -	Autonomy -	Relatedness -	Total FA strategy
1: Feed-up	11	4	1	1	1	0	16*
2: Feedback	30	9	4	10	2	0	50*
3: Feed-forward	9	6	4	1	1	0	17*
4: Peer-feedback	6	5	1	2	0	1	12*
5: Self-assessment	7	2	0	1	3	0	13
Total CAR-effect	59*	26	9*	15	7	1	104

Teacher perceived effects

Competence

Most quotes are related to feedback and feed-forward strategies that promote clarity, gives students a chance to perform or aids them to reach success.

Teachers' focus on student understanding through questioning and observing leads to specific, tailored guidance (scaffolding) as they are able to identify when and to which students they should give extra instruction. Furthermore, teachers report providing success experiences and boosting student's confidence by giving positive feedback on good performances, making students aware of what they have accomplished (e.g. learning goals) and providing clues that help students proceed (feed-forward).

> *Teacher_C: ".... The point is that the student eventually realizes 'I can do this'. And if that student and I agree that he/she can do it, then we're done. And then you no longer have... then I think you need less formal tools. …I assume, or hope, that it is a boost in motivation. By talking about the content, I also stimulate a bit of curiosity, … to think about it.''*

Teachers C and A describe how clear feed-up, by explicating and providing learning goals per lesson/paragraph, stimulates awareness in students on what they should do and when they have achieved success. Teacher A reports that when she hasn't put the learning goals on the online learning environment, students ask for them to check themselves in preparation of the final test.

Comments on how FA strategies may also be competence-thwarting are relatively few and seemed to stem from faulty implementations of FA (e.g. no personal interaction or assistance due to time shortage, emphasis on bad performance, or insufficiently structured).

Autonomy

Most quotes in this category are related to feed-up and feedback strategies that evoke a more self-determined locus of causality of behaviour for biology by stressing relevance of instruction or organizing it in a way that promotes fun.

Teacher quotes also relate to formative assessment as a way to stimulate students' choice and control over the learning process. Teacher C reports on using very explicit learning goals hand-outs per chapter. These serve as a self-reflection tool as students have to check the learning goals when they believe they have achieved them. This approach leads to more discussion about the understanding of the concepts instead of discussion about restrictions such a regulations and deadlines, they seems more eager to proceed. Besides, students are quite free to work at their own pace, with or without peers, and finish each chapter by passing an non-graded end-test. Teacher C uses this system to minimise 'test-pressure' for his students and he reports that students have a relatively relaxed attitude towards biology and appear more aware of their learning progress. Teacher A challenges students by putting them back to work to check and improve their own answer instead of correcting them, thereby stimulating responsibility for their own learning. Moreover, she enables resitters to work on next-level material and practice-tests. Teacher B tries to foster curiosity for biology by asking relatable questions such as: 'Why does your heart beat faster when you exercise?'.

Eight comments on FA use were classified as autonomy thwarting. In those cases the formative assessment seemed insufficiently relevant to the students or forces them too far out of their comfort zone e.g., teachers A and B stated that at the start of the year they often have to force performance-avoiding students to answer questions and elaborate on their thoughts. In addition, teacher B applies some pressure to make students analyse their end-of-chapter test (self-assessment), because students do not see the use of this.

Relatedness

Only Teacher C clearly states that he fosters a sense of connectedness by being interested in the personal situation of the student when having feedback/feed-forward conversations and also creating opportunities to give support to his students:

> *''…. sometime in class I give assignments that they have to do in groups, then I try to raise the level a bit.* [So] *That they have to think deeper, be more creative in searching for a solution and will need more feedback from me.* [students will ask:] *'I can't figure this out, how should I do this?' and then I can jump in and solve it together with them.''*

Student reported effects

Competence

The largest part of students' quotes interpreted as affecting competence are associated with feedback strategies (30 supportive and 10 competence thwarting coded quotes). Assessment questions and feedback primarily provide high quality information to the students on what they should be capable of, what their current level of understanding is or where they should focus on e.g.:

> "R22: *.... a practice test that is always handy, because those are sort of the same questions as you can expect in the test. So you know whether you are ready or whether you have to learn even more .. "*

Students also describe that formative assessments are often smaller than summative assessment. Besides, they are administered directly or soon after an instruction, which makes them easier to complete. Cooperation with peers leads to assistance and more accurate judgement over produced work (peer-feedback). Students of teacher B describe how the teacher tailors her help on assignments (feed-forward) by first directing her students to read the instruction in the textbook and only assisting when students really struggle. Students appreciate how they are forced to work out the problem by themselves and enjoy subsequent success, which they believe enhances their learning.

Competence thwarting effects are perceived when FA is timed inappropriately, unclear or unaligned (too difficult or easy) for the student. One students of Teacher A elaborates on the evaluative impact of feedback at the level of the person that seems to result in a competitive and fixed view on learning:

> *R23: Suppose she [Teacher] asks two people and the other person does not know* [the answer]*, then you are like 'ah, I know it and you don't' you feel a bit smarter than the rest.*

Interestingly two students from Teacher C report a lack of opportunity show mastery due to the absence traditional formal (summative) tests in his class, an exemplary quote:

> R3: *I therefore think that a real hard test, not necessarily for a real grade, but a test would be useful. Because then you can know: 'Okay, this I understand very well, this I don't get'.*
> R5: *Yes! (…) And then you can get more motivated, because then you can also get higher grades, because now it seems that a 'sufficient'* [grade] *is the highest that has been given for all these things, although some probably understand it very well.*

Autonomy

Mainly students from the two classes of Teacher C report strong autonomy supportive effects from used FA practices (15 of 26 quotes). These students report that they feel respected by Teacher and state that they really appreciate the freedom the teacher gives when he enables them to decide to what extent they require extra instruction, which assessments are necessary and what the planning will be like. Giving peer-feedback was highly appreciated (i.e. described as 'fun' several times). Another mechanism is clarified in the following quote:

> *R18: Well then* [during in frontal teaching] *... it is like this: you only pay a little attention, because you don't have to know it all by yourself, but in a practical you have to .. you want to understand it completely and..*
>
> *R19: It's nicer too*
>
> *R18: Yes certainly.. because you like doing that* [experiments]. *You can't.. do it without understanding it.*

These quotes illustrate that FA can engage and activate students (e.g.: no longer *'bored by only listening'*).

Some FA practices (e.g. filling out a test-analysis after an end-of-chapter test) were described as unhelpful or experienced as forced. Also assignments that are too difficult or not properly checked by the teacher or peers are regarded a waste of time.

Relatedness

Teacher C's students report that he is frequently checking up on them (feedback), available to answer questions (feed-forward) and keeps into account how the students want to work, for instance by letting them work with peers of their choice. They feel that teacher C is really supportive, caring and committed. This shows high similarities with the statements Teacher C made.

DISCUSSION

The aim of this study was to answer following research question:

> 'What influence does formative assessment have on secondary school students' autonomous motivation for biology?'

Results obtained in the study will be discussed per postulated sub question (last two taken together) to reach a final conclusion. Thereafter several limitations and implications of this research are noted.

How often are the five formative key strategies used by biology teachers?

Observations revealed that the three teachers used all Formative Assessment strategies as defined by Black & Wiliam (2009), especially feedback practices, followed by feed-forward (figure 4). Teacher B used these strategies more than twice as often as Teacher A and C, assumingly mainly due to pedagogical differences. Teacher B's preference for frontal whole class instruction might be a contributing factor for her 'established' use of the construct 'asking questions and classroom discussions' (75% of lessons). It is not surprising that feedback and feedforward were observed with high frequencies considering the highly iterative nature of formative assessment during instruction (i.e. a constant interaction between eliciting information, evaluation and providing feedback to move learning forward) which is very common to teachers' daily practice (Kippcers et al., 2018).

The 'peer- and self-assessment' construct on the questionnaire was scored relatively low by Teacher A and B, and relatively high by Teacher C. The higher use of self-assessment by Teacher C is visible in the observations and the - in interview - mentioned use of the learning goal sheets as a reflection tool. Reluctance towards using peer- and self-assessment by Teachers A and B is exemplified by the following quote:

> Teacher B: *'We've done this in the past, keeping a log, it really didn't work! You have to remind yourself that every lesson and you have to let them fill it in every lesson and they don't really know what to fill in themselves…'*

Contrasting to the questionnaires, observations did reveal frequent use of peer-feedback (between 4 and 6 instances per lesson hour). This strategy is facilitated by teachers when assigning or allowing groups work that leads to discussion about instruction material. This informal way of using peers as learning resources for each other is not incorporated in other self-report questionnaires e.g., by Kippers et al. (2018). It proved to be very hard to reliably quantify this type of informal peer-feedback by tallying teacher behaviour, as the teacher most of the time simply 'starts' and 'ends' a time-period that allows for these interactions instead of giving clear instructions.

All in all, the teachers made variable use of the key formative assessment strategies. Only teacher C can be considered to be somewhat advanced in using FA strategies, especially with self-assessment.

This suggest that, unlike the hypothesis, teachers that actively try to use FA still find it hard to embed this into their daily teaching practice.

Does the students' motivation pattern for biology, i.e. controlled vs. autonomous, change in a formative teaching setting?

Results show that there was no significant change in the students autonomous and controlled motivation between start and end of the three month study period, as assessed by measuring motivation regulation styles (i.e. intrinsic, identified, introjected and external) using an adapted Academic Self-regulation Questionnaire (Peeters, 2015). The lack of significant motivational change in studied biology classes can be interpreted as a positive finding because of the expected decrease in autonomous motivation for school in this age group (e.g., Gallup, 2016; Osborne, Simon, & Collins, 2003; Potvin & Hasni, 2014).

What influences do teachers perceive and students report formative assessment has on students' perceived competence, relatedness and autonomy?

The teacher interviews and student focus groups revealed an overall strikingly similar pattern of how FA, in particular feedback, can support students' needs for competence and autonomy and in some cases also support relatedness or thwart their competence and autonomy, see table 3 & 4. The mechanisms that teachers and students describe show that investing in clarifying the learning goals, eliciting evidence of learning and providing feedback that moves learning forward will often provide students with a sense of confidence and willingness to participate in learning. The FA's focus on student understanding often leads to very tailored and specific guiding by teachers and peers, which can be identified as 'scaffolding', that has been shown to be effective to support students psychological needs (e.g. Kirschner & van Merriënboer, 2013; Pat-El et al., 2013).

Inappropriate timing, unalignment and unclear instruction or feedback are the main reported ways how formative assessments strategies can have competence and autonomy reducing effects, reconfirming the importance for teachers to check how FA is received by the students (Hattie & Timperley, 2007). Time constraints play a major role here.

The pedagogical choice to focus on informal and personal feedback and feed-forward is crucial for FA to have a positive effect on students' need for relatedness. Interestingly, several students of teacher C reported a lack of opportunity to 'really perform' due to the absence of traditional summative testing for biology. This was interpreted as a competence thwarting effect and indicates that some students demonstrate a need to have their work objectively evaluated in order to show mastery, referring to performance-goal orientation, hinting to introjected regulation.

Conclusion

Taken together, the results indicate that use of formative assessment strategies in secondary biology education, especially feedback, can support the psychological needs for competence and autonomy, required to become autonomously motivated. However, students' autonomous motivation for biology did not significantly change over the three month study period, suggesting that teachers' average or slightly advanced use of Formative Assessment practices is not sufficient to significantly stimulate autonomous motivation. On the other hand, the strategies deployed by the teachers did not result in a decrease in motivation for school biology, which is often reported in early adolescents (Potvin & Hasni, 2014).

Limitations

Several limitations of the current study come to mind. The first being the used observations to measure teachers use of formative assessment. Although this method gives an unique opportunity to compare self-reported use of FA with actual occurrences in the classroom, it also has some downsides. Because each formative interaction that could be overheard and categorized was tallied into a key FA-strategy, a strong bias towards audible and interactive formative practices is likely. In other words, a teacher that uses frequent classroom questioning and discussions will appear very at good formative assessment compared to a teacher that uses silent FA, for instance observations, written feedback or automatic responses in an online tool. Furthermore, this method requires the observer to be able to hear all teacher-student interactions in class also when the teachers has individual conversations with students, which is frankly impossible without an amplification or recording device. However, second coder agreement was satisfactory.

Another potential limitation of this research is that the distinction of FA strategies and STD need-supportive teaching effects is very contextual and subject to a high degree interpretation (Reeve, 2006). Especially the distinction between 'feedback' and 'feed-forward' is often unclear as these strategies are strongly connected. Though we tried to increase reliability by using second coders, a certain degree of subjectivity is inevitable when studying need-supportive teaching and formative assessment in classrooms.

Lastly, the mixed method approach and partly longitudinal nature of this study allows for a unique combination of data sources to describe in depth what influences of formative assessment on students autonomous motivation for biology can be distilled. The findings are quite specific to the relatively small group of participating teachers and students. Therefore generalisability of the findings is questionable and should be tested in future research. Furthermore, this study is very descriptive and was not designed to measure the precise effect sizes on motivation or underlying constructs. The latter would be interesting to determine the actual effectiveness of using FA-strategies.

Implications

It appears that more research and better teacher support is required to reveal how powerful formative assessment strategies can be to support student motivation. Although, the three participating biology teachers in this study were selected based on their reported interest and motivation to use FA, they are not able to completely embed the formative assessment practices as described by Kippers et al. (2018). Some problems that the teachers mentioned were time shortage, students' dependency on grades, and their reluctance to fully engage in FA, especially performing serious peer-feedback and self-assessment. Indeed, these teachers are trying to implement FA while receiving limited school support, which has been identified as an important prerequisite for FA (Heitink et al., 2016; Kippers et al., 2018). Essentially the teachers operate alone in their FA development, which is also a possible explanation why students struggle to adapt and engage in FA: they are used to different teaching strategies and working for grades in other classes.

To stimulate FA school policies and professional teacher development programs should focus on long lasting cooperation within the school context, allowing teachers to exchange good practices, (video)coach each other and focus on direct applicability, by for instance creating rubrics (Bronkhorst & van Tartwijk, 2016). Furthermore, the implementation of these type of interventions that stimulate FA use by teachers could provide interesting semi-experimental settings to study the effect size of FA-strategies on motivation and underlying constructs. We also recommend future studies to use a more

inclusive scope by tracking all types of FA (i.e. also the ones that are invisible or take place outside of the classroom), studying multiple teachers of the same group of students and to take a longer timeframe. This will allow for better analysis of long-term effects and increase ecological validity.

Aside from training teachers to harness the indicated need-supportive effects of FA, focus might also be put towards student skills, attitudes and training needed to effectively participate in FA. Since formative assessment requires a high level self-regulation and goal setting from students, their metacognitive skills have to be able to facilitate them to actively steer their own learning process. How teachers can optimally assist students to self-regulate their learning is an important subject to explore in future research (Zimmerman & Kitsantas, 2005).

Furthermore, future research should clarify what high quality FA looks like and how it can be effectively measured across a larger group of teachers. As discussed in the limitations tallying of FA provided an unique insight into how often the teachers are using FA as described by the AfL discourse, but little is known about the relative motivational impact of each type of formative interaction. Knowledge on this would increase feasibility of identifying teachers that are expert at using high quality FA in their lessons, who could subsequently be studied to analyse the true potential influence of FA.

Finally, we conclude and agree with Bennet (2011) that formative assessment remains a work-in-progress in both teaching practice and education research. It does seem an endeavour worth undertaking as our research suggests that students' need for competence, autonomy, and under specific conditions relatedness are mainly supported by FA. These mechanism might be further unravelled and exploited to foster autonomous motivation for biology and restore students' wonder and enthusiasm for science.

Acknowledgements

Special thanks to the three teachers who together with their biology classes willingly participated in this study. Many thanks also to my supervisor Ralph Meulenbroeks for his always positive and warm attitude and sound advice.

References

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

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability(Formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5. 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

Bølling, M., Otte, C. R., Elsborg, P., Nielsen, G., & Bentsen, P. (2018). The association between education outside the classroom and students' school motivation: Results from a one-school-year quasi-experiment. *International Journal of Educational Research*, *89*, 22–35. https://doi.org/https://doi.org/10.1016/j.ijer.2018.03.004

Bronkhorst, L., & van Tartwijk, J. (2016). Professionaliseren in het geven van feedback. In *Toetsrevolutie: naar een feedbackcultuur in het voortgezet onderwijs* (pp. 135–149). Culemborg: Phronese.

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

Butler, R. (2006). Are mastery and ability goals both adaptive? Evaluation, initial goal construction and the quality of task engagement. *British Journal of Educational Psychology*, *76*(3), 595–611. https://doi.org/10.1348/000709905X52319

Castelijns, J., & Andersen, I. (2013). *Beoordelen om te leren. Leerlingen als mede-beoordelaars van hun eigen leerproces.* 's-Hertogenbosch: KPC Groep in opdracht van het ministerie van OCW.

Cauley, K. M., & McMillan, J. H. (2010). Formative Assessment Techniques to Support Student Motivation and Achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*(1), 1–6. https://doi.org/10.1080/00098650903267784

Center for Self-Determination. (2017). The motivation continuum: Organismic integration theory taxonomy of regulatory styles.

de Kleijn, R., Prins, F., Lutz, C., van Look, K., & van Tartwijk, J. (2016). Formatief toetsen en leerlingmotivatie. In *Toetsrevolutie: Naar een feedbackcultuur in het voortgezet onderwijs.* Culemborg: Phronese.

De Loof, H., Struyf, A., Boeve-de Pauw, J., & Van Petegem, P. (2019). Teachers' Motivating Style and Students' Motivation and Engagement in STEM: the Relationship Between Three Key Educational Concepts. *Research in Science Education*. https://doi.org/10.1007/s11165-019-9830-3

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E.L., & Ryan, R. M. (2016). Optimizing Students' Motivation in the Era of Testing and Pressure: A Self-Determination Theory Perspective. In R. R. Liu W., Wang J. (Ed.), *Building Autonomous Learners* (pp. 9–29). Singapore: Springer. https://doi.org/10.1007/978-981-287-630-0_2

Deci, Edward L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie Canadienne*, *49*(3), 182–185. https://doi.org/10.1037/a0012801

European commission. (2004). *Europe needs More Scientists: Report by the High Level Group on Increasing Human Resources for Science and Technology.* Brussels.

Gallup. (2016). *Gallup Student Poll*. Washington. Retrieved from https://www.gallup.com/file/services/189863/GSP_2015KeyFindings.pdf

Gulikers, J., & Baartman, L. (2017). *"doelgericht professionaliseren. Formatief toetsen met effect: wat doet de docent in de klas?"*

Harlen, W., & Crick, D. R. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning*. London.

Hatch, J. (2018). Better teachers are needed to improve science education. *Nature*, *562*(7725), S2–S4. https://doi.org/10.1038/d41586-018-06830-2

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/https://doi.org/10.1016/j.edurev.2015.12.002

Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., & Poortman, C. L. (2016). Strategieën voor formatief toetsen in de lespraktijk: onderzoek en concrete voorbeelden. In R. Kneyber & D. Sluijsmans (Eds.), *Toetsrevolutie: naar een feedbackcultuur in het voortgezet onderwijs* (pp. 113–125). Phronese.

Kippers, Wilma 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/https://doi.org/10.1016/j.tate.2018.06.015

Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do Learners Really Know Best? Urban Legends in Education. *Educational Psychologist*, *48*(3), 169–183. https://doi.org/10.1080/00461520.2013.804395

Kleijn, R. A. M. de, Prins F. J., Lutz, C., van Look, K. C. J., & van Tartwijk, J. (2015). *Impact of assessment on motivation: effects and mechanisms in different contexts*. Den Haag.

Lau, A. M. S. (2016). 'Formative good, summative bad?' – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, *40*(4), 509–525. https://doi.org/10.1080/0309877X.2014.984600

Lea, S. J., Stephenson, D., & Troy, J. (2003). Higher Education Students' Attitudes to Student-centred Learning: Beyond "educational bulimia"? *Studies in Higher Education*, *28*(3), 321–334. https://doi.org/10.1080/03075070309293

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

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079. https://doi.org/10.1080/0950069032000032199

P Cerasoli, C., Nicklin, J., & T Ford, M. (2014). *Intrinsic Motivation and Extrinsic Incentives Jointly Predict Performance: A 40-Year Meta-Analysis*. *Psychological bulletin* (Vol. 140). https://doi.org/10.1037/a0035661

Pat-El, R. J., Tillema, H., Segers, M., & Vedder, P. (2013). Validation of Assessment for Learning Questionnaires for teachers and students. *British Journal of Educational Psychology*, *83*(1), 98–113. https://doi.org/10.1111/j.2044-8279.2011.02057.x

Pat-El, R., Tillema, H., & van Koppen, S. W. M. (2012). Effects of formative feedback on intrinsic motivation: Examining ethnic differences. *Learning and Individual Differences*, *22*(4), 449–454. https://doi.org/https://doi.org/10.1016/j.lindif.2012.04.001

Peeters, W. (2015). Motivatie meten: 2 vragenlijsten. Retrieved August 10, 2019, from

https://www.vernieuwenderwijs.nl/motivatie-meten-2-vragenlijsten/

Penuel, W. R., & Watkins, D. A. (2019). Assessment to Promote Equity and Epistemic Justice: A Use-Case of a Research-Practice Partnership in Science Education. *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 201–216. https://doi.org/10.1177/0002716219843249

Potvin, P., & Hasni, A. (2014). Analysis of the Decline in Interest Towards School Science and Technology from Grades 5 Through 11. *Journal of Science Education and Technology*, *23*(6), 784–802. https://doi.org/10.1007/s10956-014-9512-x

PT Staff. (1996). Carl Sagan, author interview. *Psychology Today. Stress: It's Worse Than You Think.* Retrieved from https://www.psychologytoday.com/intl/articles/199601/carl-sagan?page=3

Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology*, *103*(3), 683–700. https://doi.org/10.1037/a0023911

Reeve, J. (2006). Teachers as Facilitators: What Autonomy-Supportive Teachers Do and Why Their Students Benefit. *Elementary School Journal - ELEM SCH J*, *106*, 225–236. https://doi.org/10.1086/501484

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Savelsbergh, E. R., Prins, G. T., Rietbergen, C., Fechner, S., Vaessen, B. E., Draijer, J. M., & Bakker, A. (2016). Effects of innovative science and mathematics teaching on student attitudes and achievement: A meta-analytic study. *Educational Research Review*, *19*, 158–172. https://doi.org/https://doi.org/10.1016/j.edurev.2016.07.003

Schuld, N., Kanjee, A., & White, T. (2017). Tinder or flint: igniting grade 2 teachers' understanding of learning, teaching and assessment. *Journal of Education*, *17*, 109–132.

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using Learning and Motivation Theories to Coherently Link Formative Assessment, Grading Practices, and Large-Scale Assessment. *Educational Measurement: Issues and Practice*, *37*(1), 21–34. https://doi.org/10.1111/emip.12189

Sluijsmans, D., & Kneyber, R. (2016). *Toetsrevolutie: naar een feedbackcultuur in het voortgezet onderwijs*. Culemborg: Phronese. Retrieved from http://toetsrevolutie.nl/wp-content/uploads/2016/11/Toetsrevolutie-WEB.pdf

Stroet, K., Opdenakker, M.-C., & Minnaert, A. (2015). What motivates early adolescents for school? A longitudinal analysis of associations between observed teaching and motivation. *Contemporary Educational Psychology*, *42*, 129–140.

Taras, M. (2009). Summative assessment: the missing link for formative assessment. *Journal of Further and Higher Education*, *33*(1), 57–69. https://doi.org/10.1080/03098770802638671

Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S., & Koestner, R. (2014). A self-determination theory approach to predicting school achievement over time: the unique role of intrinsic motivation. *Contemporary Educational Psychology*, *39*(4), 342–358. https://doi.org/https://doi.org/10.1016/j.cedpsych.2014.08.002

Tessier, D., Sarrazin, P., & Ntoumanis, N. (2010). The effect of an intervention to improve newly qualified teachers' interpersonal style, students motivation and psychological need satisfaction in sport-based physical education. *Contemporary Educational Psychology*, *35*(4), 242–253. https://doi.org/https://doi.org/10.1016/j.cedpsych.2010.05.005

Tytler, R., & Osborne, J. (2012). Student Attitudes and Aspirations Towards Science BT - Second International Handbook of Science Education. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.) (pp. 597–625). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-9041-7_41

van Griethuijsen, R. A. L. F., van Eijck, M. W., Haste, H., den Brok, P. J., Skinner, N. C., Mansour, N., … BouJaoude, S. (2015). Global Patterns in Students' Views of Science and Interest in Science. *Research in Science Education*,

45(4), 581–603. https://doi.org/10.1007/s11165-014-9438-6

Vansteenkiste, M., & Ryan, R. (2013). On psychological growth and vulnerability: basic psychological need satisfaction and need frustration as an unifying principle. *JOURNAL OF PSYCHOTHERAPY INTEGRATION*, *3*, 263–280. https://doi.org/10.1037/a0032359

Vansteenkiste, Maarten, Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational Profiles From a Self-Determination Perspective: The Quality of Motivation Matters. *Journal of Educational Psychology*, *101*, 671–688. https://doi.org/10.1037/a0015083

Vansteenkiste, Maarten, Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology*, *87*(2), 246–260. https://doi.org/10.1037/0022-3514.87.2.246

Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education*, *37*(6), 747–760. https://doi.org/10.1080/02602938.2011.572153

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., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *Future of assessment: Shaping teaching and learning* (pp. 53–82). NJ: Lawrence Erlbaum Associates.

William, D. (2014). *Cijfers geven werkt niet.* Meppel: Ten Brink Uitgevers.

Zimmerman, B., & Kitsantas, A. (2005). The Hidden Dimension of Personal Competence: Self-Regulated Learning and Practice. *Handbook of Competence and Motivation*.

SUPPLEMENT

Appendix A

Academic Self-Regulation Questionnaire (SRQ-A) (identical to post-test)

Universiteit Utrecht

Vragenlijst motivatie voor biologie

Hieronder volgen vier vragen met telkens acht stellingen. Geef per stelling aan wat je er van vindt door een cijfer te omcirkelen. De antwoorden zijn anoniem en worden nooit aan iemand anders doorgegeven. Over drie maanden vul je om deze vragenlijst opnieuw in. Daarom vragen we om de laatste vier cijfers van je telefoonnummer, om die vragenlijst aan deze te kunnen koppelen.
Alvast bedankt voor je deelname.

Laatste 4 cijfers (mobiele) telefoon: _____ **Leeftijd:** _____ **Jongen** ☐ **Meisje** ☐

Klas: _____ **Docent:** _____

A. Waarom maak je huiswerk voor biologie?	**Helemaal niet waar**			**Helemaal waar**
1. Omdat ik wil dat de docent denkt dat ik een goede leerling ben.	1	2	3	4
2. Omdat ik problemen krijg als ik het niet doe.	1	2	3	4
3. Omdat het leuk is.	1	2	3	4
4. Omdat het mij anders een slecht gevoel over mijzelf geeft.	1	2	3	4
5. Omdat ik het onderwerp wil begrijpen.	1	2	3	4
6. Omdat ik dat hoor te doen.	1	2	3	4
7. Omdat ik het leuk vindt mijn huiswerk te maken.	1	2	3	4
8. Omdat het belangrijk is om mijn huiswerk te maken.	1	2	3	4
B. Waarom werk je in de klas?				
9. Zodat de docent niet boos op mij wordt.	1	2	3	4
10. Omdat ik wil dat docent mij een goede leerling vindt.	1	2	3	4
11. Omdat ik nieuwe dingen wil leren.	1	2	3	4
12. Omdat ik mij schaam als ik het niet af krijg.	1	2	3	4
13. Omdat het leuk is.	1	2	3	4
14. Omdat het de regel is.	1	2	3	4
15. Omdat ik het leuk vind om te werken in de klas.	1	2	3	4
16. Omdat het belangrijk is om te werken in de klas.	1	2	3	4

C. Waarom probeer je klassikaal moeilijke vragen te beantwoorden tijdens de les?	**Helemaal niet waar**			**Helemaal waar**
17. Omdat ik wil dat andere leerlingen mij slim vinden.	1	2	3	4
18. Omdat ik mij schaam als ik het niet probeer.	1	2	3	4
19. Omdat ik het leuk vind om moeilijke vragen te beantwoorden.	1	2	3	4
20. Omdat ik dat hoor te doen.	1	2	3	4
21. Om er achter te komen of ik het goed of fout heb.	1	2	3	4
22. Omdat het leuk is om moeilijke vragen te beantwoorden.	1	2	3	4
23. Omdat het voor mij belangrijk is om te proberen moeilijke vragen te beantwoorden.	1	2	3	4
24. Omdat ik wil dat de docent aardige dingen over mij zegt.	1	2	3	4
D. Waarom doe je je best voor biologie?				
25. Omdat ik dat hoor te doen.	1	2	3	4
26. Zodat mijn docent denkt dat ik een goede leerling ben.	1	2	3	4
27. Omdat ik het leuk vind om mijn schoolwerk goed te doen.	1	2	3	4
28. Omdat ik in de problemen kom als ik het niet goed doe.	1	2	3	4
29. Omdat ik erg slecht over mijzelf denk als ik dat niet doe.	1	2	3	4
30. Omdat het voor mij belangrijk is om mijn best te doen voor school.	1	2	3	4
31. Omdat ik erg trots op mijzelf ben als ik het goed doe.	1	2	3	4
32. Omdat ik een beloning kan krijgen als ik het goed doe.	1	2	3	4

Self-Regulation Questionnaires-Academic (SRQ-A) door Ryan and Connell (1989).
Op basis van vertaling door Wessel Peeters (Vernieuwenderwijs, 2018). Onderzoek door Sjoerd Banken

Appendix B
Interview protocol Focus Groups

Hoi goedemiddag, fijn dat jullie hier zitten. Dit groepsinterview gaat over het testen van kennis en vaardigheden bij biologie. Ik heb een aantal vragen voor jullie. Er zijn geen goede of foute antwoorden, ik ben benieuwd naar jullie mening. Jullie mogen op elkaar reageren, maar probeer niet door elkaar heen te praten. Ik zou jullie eerst willen vragen om je eigen naam te zeggen en te bevestigen dat je het goed vindt om deel te nemen aan dit onderzoek. Dat betekent dat je toestemming geeft om geluidsopname te maken van je antwoorden die anoniem verwerkt zullen worden.

1. **Wat vinden jullie van het vak biologie? Waarom? [de opwarmer]**

2. **Wat voor toetsen krijgen jullie in de biologieles? Op welke manier wordt getest wat jullie weten en kunnen?**
    a. **Krijg je altijd een cijfer deze toetsen? Tellen alle toetsen mee?**
    b. **Wat vinden jullie daarvan?**

3. **Alle opdrachten en testen die niet mee tellen bij biologie, dus waar je geen cijfer/beoordeling voor krijgt, wat vinden jullie daarvan?**
    a. **Idee dat je daar meer of minder van leert?**

4. **Is het altijd duidelijk wat je gaat leren in de biologieles? Hoe weet je dat?**
    a. **Hebben jullie daar invloed op?**
    b. **Weten jullie wanneer je het goed genoeg doet of genoeg begrijpt?**
    c. **Wat vinden jullie daarvan?**

5. **Hoe helpt de docent jullie om erachter te komen of je de stof echt begrijpt?**
    a. **Voorbeeld?**
    b. **Wat vinden jullie ervan als biologiedocent laat merken of je op de goede weg bent?**

6. **Stellen jullie vragen aan de docent als je iets niet snapt?**
    a. **Waarom?**

7. **Doet jullie biologiedocent wel eens iets extra voor leerlingen die het moeilijk vinden?**

8. **Kijk je wel eens je eigen werk na of dat van je buurman?**
    a. **Waar kijk je dan naar?**
    b. **Wat vind je daarvan?**

9. **Vraagt de docent jullie wel eens om zelf na te denken over wat er wel en wat er niet goed gaat bij biologie?**
    a. **Waar zou je dan op letten? Wat denkt de rest?**
    b. **Wat vind je daarvan?**

10. **Is er nog iets anders dat je zou willen zeggen over toetsen bij biologie?**

Appendix C
statistical analyses of normality of SRQ-A constructs

Table C1

Tests of Normality of motivation regulation scales of Self-Regulation Questionnaire

construct	Shapiro-Wilk Statistic	df	Sig.
Pre_Intrinsic_mot	,965	119	,003
Pre_Identified_reg	,939	119	,000
Pre_Introjected_reg	,985	119	,224
Pre_Extrinsic_reg	,990	119	,530
Post_Intrinsic_mot	,957	119	,001
Post_Identified_reg	,963	119	,002
Post_Introjected_reg	,984	119	,183
Post_Extrinsic_reg	,984	119	,159

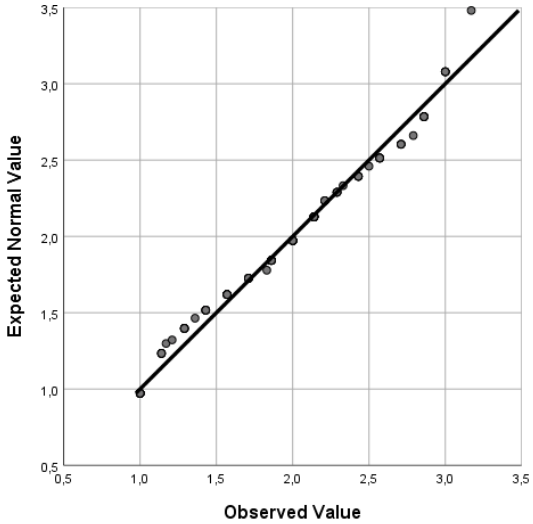

Figure C2. Q-Q plot of normality of the construct Intrinsic motivation at the pre-test (N=119).

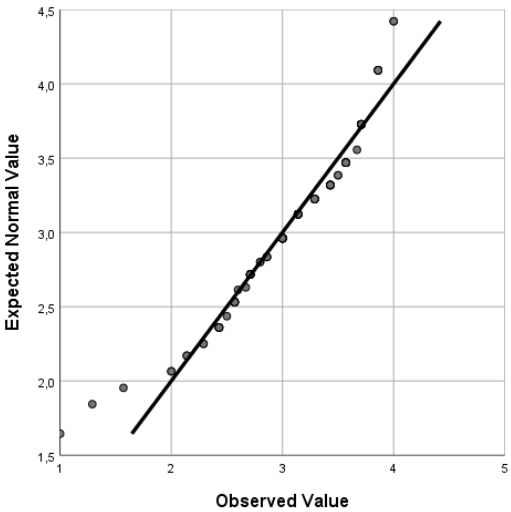

Figure C3. Q-Q plot of normality of the construct identified regulation at the pre-test (N=119).

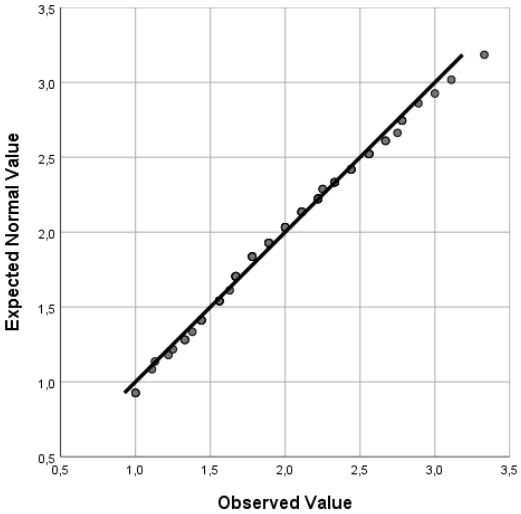

Figure C4. Q-Q plot of normality of the construct introjected regulation at the pre-test (N=119).

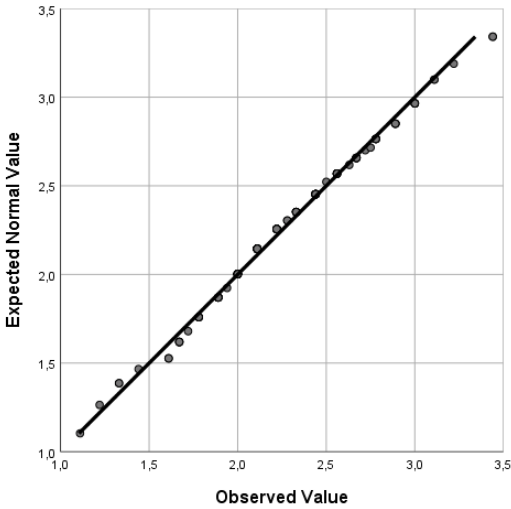

Figure C5. Q-Q plot of normality of the construct extrinsic regulation at the pre-test (N=119).

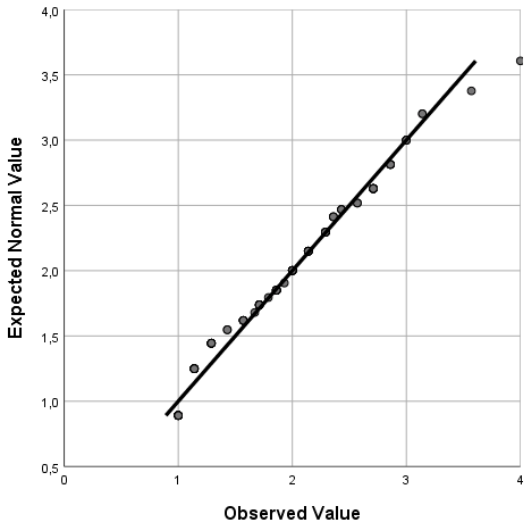

Figure C6. Q-Q plot of normality of the construct intrinsic motivation at the post-test (N=119).

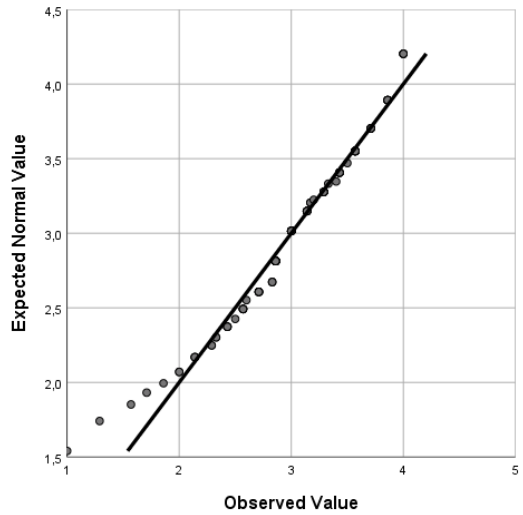

Figure C7. Q-Q plot of normality of the construct identified regulation at the post-test (N=119).

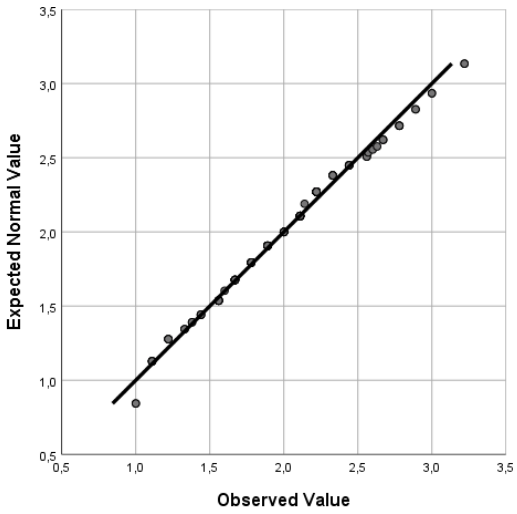

Figure C8. Q-Q plot of normality of the construct introjected regulation at the post-test (N=119).

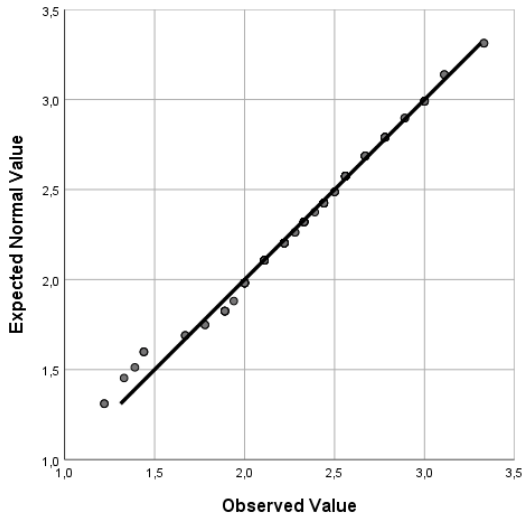

Figure C9. Q-Q plot of normality of the construct extrinsic regulation at the post-test (N=119).

Appendix D

Motivation regulation styles average per studied class

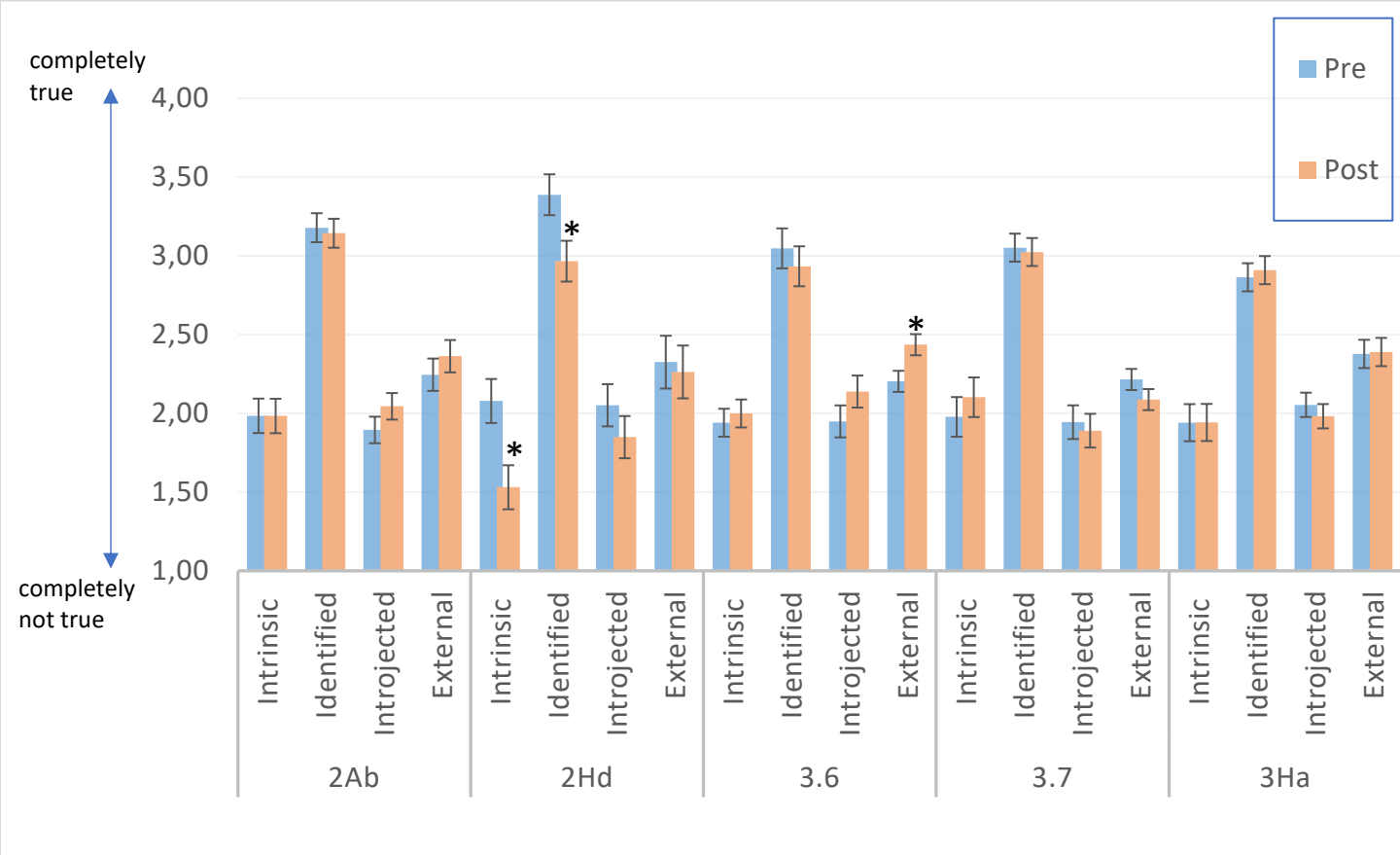

Figure D1. Average motivation per class (N= between 18 and 30) per regulation type (32 items on a 4-point Likert scale) before and after 3-month study period.

Appendix E

Observation form for formative practices by teachers

Observatie # – Klas # – datum

Hieronder volgt een schema met links de 5 hoofdstrategieën van formatief toetsen die de docent kan inzetten.

De strategieën zijn gespecificeerd in enkele onderdelen/varianten. De bedoeling is om te scoren hoe vaak elke variant van een strategie wordt gebruikt middels turven in de kolom 'frequentie'. Er is ruimte in de kolom 'uitwerking' om citaten of observaties te noteren die het gebruik van een bepaalde variant bewijzen, bijvoorbeeld bij Feedup -> Leerdoel -> Uitwerking: 'Docent: Ik vind het dus belangrijk dat je het schema [glucose regulatie] op blz 24 begrijpt.'

Het gewicht kan per score sterk kan verschillen. Het kan één opmerking zijn, maar ook het initiëren van een werkvorm die langere tijd duurt. In het tweede geval beter om de tijdsduur te noteren, bijvoorbeeld 'groepswerk, van 11:50 tot 12:22'.

Strategie	**Code**	**Uitwerking**	**Frequentie**
geen	Werkwijze benoemen *Basis structuur van les en opdracht*		
Feedup: Clarifying and sharing learning intentions and criteria for success;	Leerdoel (wat is het doel, *waarom doen we dat?)*		
	Voordoen / voorbeeld geven *Expert practice*		
	Succescriteria (*hoe doe je het goed?*)		
Feedback: Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding;	Theorie-terugvragen		
	Nieuwe theorie-vraag		
	Quiz/test / product opleveren		
	Benoemen/ herhalen		
	Compliment /bevestiging		
	Correctie inhoud		
	Observeren		
	(in)checken Proces-vraag		
	Meta-opmerking soort proces feedback		

Providing feedback that moves learners **forward;** *scaffolding*	**Aanvullende Instructie** Hulp bieden	-	
	Guiding question **Suggestieve vraag**		
	aansporing verder na te denken (Tegenvragen)		
	Tips – ideeën - nieuwe uitdagingen aanbieden		
Activating students as instructional resources for one another; **peer-feedback**	Groepswerk		
	Uitlegbeurt* *Expliciet verzoek docentrol opnemen*		
	Product-test-materiaal door elkaar laten **controleren** (ook klassikaal)		
	Overleg klassikaal (Denken-delen-uitwisselen)		
	Smoezelen – overleg tussendoor		
Activating students as the owners of their own learning. **Self-assessment**	Zelfinschatting - theorie		
	jezelf controleren		
	Zelf-regulatie – proces (hoe heb je gepland? Wat had je beter kunnen doen?)		