ZfE CrossMark

Formative assessment and intrinsic motivation: The mediating role of perceived competence

Annika Lena Hondrich (iD) · **Jasmin Decristan** · **Silke Hertel** ·
Eckhard Klieme

Abstract Formative Assessment (FA) refers to eliciting evidence about students' understanding and using the information to support learning, e.g. via individual feedback. There is evidence that FA fosters students' motivation, but less is known about the underlying processes. The present study investigates direct effects of FA on intrinsic motivation as well as the mediating role of students' perceived competence, drawing on Deci and Ryan's self-determination theory. In a randomized, controlled trial, primary school teachers were either assigned to an FA training ($n = 17$) or to a control group (CG; $n = 11$). All teachers then taught two science units in their classrooms (FA: $n = 319$ students, CG: $n = 232$). Multilevel regression analyses showed a higher perceived competence and a marginally higher intrinsic motivation for FA students after unit 1. After unit 2, both intrinsic motivation and perceived competence were higher in the FA condition, and the impact on intrinsic motivation was significantly mediated by students' perceived competence after the

A. L. Hondrich (✉) · Prof. Dr. E. Klieme
Deutsches Institut für Internationale Pädagogische Forschung (DIPF),
Schloßstraße 29, 60486 Frankfurt, Germany
E-Mail: hondrich@dipf.de

Prof. Dr. E. Klieme
E-Mail: klieme@dipf.de

Prof. Dr. J. Decristan
Institut für Bildungsforschung, Bergische Universität Wuppertal,
Rainer-Gruenter-Straße 21, 42119 Wuppertal, Germany
E-Mail: decristan@uni-wuppertal.de

Prof. Dr. S. Hertel
Institut für Bildungswissenschaft, Universität Heidelberg, Akademiestraße 3, 69117 Heidelberg,
Germany
E-Mail: hertel@ibw.uni-heidelberg.de

first unit. These results confirm and extend previous findings on the effectiveness of formative assessment on motivational outcomes.

Keywords Formative assessment · Intrinsic motivation · Perceived competence · Primary school · Self-determination theory

Formatives Assessment und intrinsische Motivation: die mediierende Wirkung von Kompetenzerleben

Zusammenfassung Formatives Assessment (FA) bezeichnet eine Lernverlaufsdiagnostik, die genutzt wird um das Lernen der Schülerinnen und Schüler (SuS) zu fördern, indem u. A. Rückmeldungen gegeben werden. Bisherige Studien zeigen positive Effekte von FA auf die Motivation, jedoch fehlt Forschung über die zugrundeliegenden Prozesse. Die vorliegende Studie untersucht direkte Effekte von FA auf die intrinsische Motivation von SuS, sowie abgeleitet von Deci und Ryans Selbstbestimmungstheorie, die mediierende Wirkung von Kompetenzerleben. Grundschullehrpersonen wurden randomisiert einer FA-Fortbildung ($n = 17$) oder einer Kontrollgruppe (CG, $n = 11$) zugeordnet und unterrichteten anschließend zwei Einheiten in ihren Klassen (FA: $n = 319$; CG: $n = 232$ SuS). Mehrebenenregressionsanalysen zeigten nach der ersten Einheit ein höheres Kompetenzerleben und eine marginal höhere intrinsische Motivation in der FA Gruppe, sowie nach der zweiten Einheit positive Effekte auf beide Variablen. Der Effekt von FA auf die intrinsische Motivation war zudem mediiert durch das Kompetenzerleben nach Einheit 1. Dies bestätigt und erweitert bisherige Kenntnisse über die motivationsfördernde Wirkung von FA.

Schlüsselwörter Formatives Assessment · Intrinsische Motivation · Kompetenzerleben · Grundschule · Selbstbestimmungstheorie

1 Introduction

During the last decades, formative assessment (FA) has been widely discussed as an effective strategy for supporting students according to their individual learning needs (e. g., Black and Wiliam 1998, 2009). Although effects of formative assessment on students' achievement are often in focus, implementing formative assessment strategies are also perceived as a powerful tool to foster students' motivation (Black and Wiliam 1998; Pat-El et al. 2012). Drawing on Deci and Ryan's self-determination theory, students' perceived competence and self-efficacy are vital for the development of their intrinsic motivation (Deci and Ryan 2000) and could thus play a mediating role. However, empirical studies investigating the impact of formative assessment on students' motivation are scarce, and even less research can be found on potentially mediating factors to explain this impact. In the present study, set in primary school science classes, we seek to evaluate whether formative assessment is successful in fostering students' intrinsic motivation and perceived competence, as well as investigate the mediating function of students' perceived competence.

1.1 Formative assessment

In agreement with authors like Black and Wiliam (2009) or Bell and Cowie (2000), formative assessment is considered a process in which evidence about students' understanding is elicited and subsequently used to support students' learning. The formative usage of the assessment information is essential and requires the assessments to provide detailed information on students' understanding relative to a previously established learning goal, identifying weaknesses, and keeping track of students' learning gains (Shavelson et al. 2008; Black and Wiliam 2009; Maier 2010). The gathered information either serves as feedback for teachers in order to adapt instruction to the identified learning needs; or the information is provided to the students as formative feedback (Hattie and Timperley 2007; Shute 2008). Formative feedback thus places an emphasis on students' actual understanding and their learning progress (thus emphasizing temporal reference norms) as well as on strategies of how to take the next learning steps (Pat-El et al. 2012). In order for formative assessment strategies to be effective, students need to make use of the feedback and instruction provided by teachers. When providing formative assessment, one important goal is therefore to encourage students to take responsibility for their own learning, by making the formative process transparent to students, and creating a constructive learning atmosphere in which mistakes are regarded as valuable information (Black and Wiliam 2009).

Formative assessment, according to the definition given above, can take different forms. On a formality dimension, formative assessment ranges from gathering and using information "on the fly", as the opportunity arises, to curriculum-embedded, preplanned assessments and feedback, inserted at specific junctures of the curriculum when an important learning goal should have been met (Shavelson et al. 2008). In the current study, we focus on curriculum-embedded formative assessment, which may provide "thoughtful, curriculum-aligned, and valid ways of determining what students know", without "leaving the burden of planning and assessing on the teacher alone" (Shavelson et al. 2008, p. 300).

1.2 Intrinsic motivation

Especially in the context of life-long learning, the development of a high learning motivation is crucial and constitutes an important educational goal. According to Deci and Ryan (2000), motivation varies from extrinsic and externally controlled (students engage in a learning task for external reasons like good grades) to self-determined or even intrinsic (students engage in a learning task for inherent reasons like enjoyment or the intellectual challenge). Intrinsic motivation is generally regarded as most adaptive for students' achievement and well-being (Deci and Ryan 2000). Research shows that a high intrinsic learning motivation is a central predictor for students' learning success, especially regarding conceptual understanding (e. g., Flink et al. 1992; Guay et al. 2000; Deci and Ryan 2000). Compared to students who depend solely on extrinsic motivation, students with high intrinsic motivation are more likely to perform better, be more persistent in learning, and use a deep level of processing (Deci and Ryan 2000; Vansteenkiste et al. 2006). However, despite

these benefits of intrinsic motivation, its support and maintenance within the school context is obviously a challenge. Although students start out with rather high levels of intrinsic motivation, their intrinsic motivation typically decreases already during the course of primary school (e.g., Spinath and Spinath 2005; Lepper et al. 2005). Thus, it is a central question as to how intrinsic motivation can be maintained and fostered in instruction.

Similar to the extrinsic—intrinsic dichotomy, achievement goal theories make a distinction between a mastery orientation, meaning that students engage in tasks in order to understand the topic and extend their knowledge and skills, and a performance orientation, meaning that students engage in tasks to outperform others and demonstrate high competence (e.g., Heyman and Dweck 1992). Many authors have drawn a link between intrinsic motivation and a mastery goal orientation (Heyman and Dweck 1992; Deci and Ryan 2000; Gottfried et al. 2001; Spinath and Spinath 2005) and there is evidence that these concepts originating from different theoretical backgrounds in fact may be combined in a common motivational factor (Marsh et al. 2003). In the present paper, we therefore conceptualize intrinsic motivation as the engagement in a task out of enjoyment and interest as well as to extend one's understanding and gain mastery (see also Spinath and Spinath 2005).

1.3 Perceived competence as a precursor of intrinsic motivation

Many motivational theories posit that students' perceived competence or related concepts like self-efficacy have an enormous impact on their intrinsic motivation (Bandura 1997; Deci and Ryan 2000; Wigfield and Eccles 2000). Within Deci and Ryan's self-determination theory, the experience of competence is considered an innate human need (together with autonomy and relatedness), which becomes especially relevant in the school context. If students' basic need to feel competent and effective in their learning activities is fulfilled within the classroom, intrinsic and self-determined forms of motivation are expected to be facilitated (Deci and Ryan 2000). That is, if students feel that they can be successful and gain mastery, their academic curiosity and learning joy are supported; if they feel incompetent and ineffective, they are likely to lose interest in pursuing the tasks and investing effort. Similarly, students' self-efficacy, which refers to the confidence in their own capabilities to complete a task successfully (Bandura 1997), is expected to positively affect students' intrinsic motivation (Bandura 1997; Wigfield and Eccles 2000).

There is empirical evidence from cross-sectional data showing that students' perceived competence is indeed correlated with their intrinsic motivation (e.g., Zisimopoulos and Galanaki 2009). However, the empirical support for a causal relation between perceived competence and intrinsic motivation within longitudinal designs is weak. Of the few existing longitudinal studies, most found no or only a weak influence of perceived competence beliefs on changes in students' intrinsic motivation (Spinath and Spinath 2005; Spinath and Steinmayr 2012). As Deci and Ryan's theoretical assumptions are still widely accepted, this lack of empirical support is an unresolved question. One reason could be that competence has usually been defined as rather stable ability beliefs, either relative to others or compared to a standard criterion, e.g., "I am good at maths" (Harter 1982; Marsh et al. 1999; Spinath and

Steinmayr 2012). As other authors have noted (Spinath and Steinmayr 2012), this may not adequately cover the subjective experience of competence in the sense of Deci and Ryan, which may fluctuate across tasks and, above all, does not necessarily depend on reaching normative standards or on outperforming others. It is therefore important to investigate the influence of perceived competence on intrinsic motivation by focusing on the students' individual experience of competence.

1.4 Fostering students' perceived competence and intrinsic motivation through formative assessment strategies

Deci and Ryan's theory implies that a high intrinsic motivation and perceived competence are not fixed personal traits but are influenced by classroom instruction. Likewise, student's self-efficacy, albeit influenced by individual factors such as students' prior abilities and beliefs, has been shown to develop under the influence of situational factors, e.g., teacher feedback (Schunk and Rice 1991; Schunk and Zimmerman 2007). In the following, we will argue how implementing formative assessment strategies in instruction are expected to foster students' perceived competence and intrinsic motivation.

On the one hand, direct positive effects of formative assessment on intrinsic motivation are plausible. As formative assessment highlights the value of improving one's capabilities and learning from one's mistakes rather than evaluating academic results compared to others, direct positive effects on students' mastery goal orientation can be expected, which we define as an aspect of intrinsic motivation. On the other hand, formative assessment is hypothesized as supporting students' perceived competence as a mediator of intrinsic motivation. First, the envisioned adaptation of the instructional input to students' learning needs should lead to a better fit between students' current understanding and task requirements, so that students feel more competent and successful in their learning activities. Second, formative practices lead to a strong emphasis on students' existing competencies and their learning progress (Ryan and Deci 2000; Shute 2008). Instead of drawing comparisons to other students and making summative judgments on a general level, formative assessments and feedback evaluate students' understanding with respect to learning goals and to students' previous performance, thus making students' learning progress more explicit (e.g., Black and Wiliam 2009; Hattie and Timperley 2007). Therefore, implementing the principles of formative assessment are hypothesized to help students achieve their learning gains and foster their experience of competence, which should in turn positively influence their intrinsic motivation.

A number of empirical studies provide evidence that formative assessment can indeed be effective in fostering students' intrinsic motivation: research on formative feedback shows that informative, elaborative feedback has a positive impact on students' intrinsic motivation or related variables like interest (Rakoczy et al. 2008; Shute 2008). However, studies which investigate motivational effects of formative assessment within an ecologically valid, regular classroom environment are still scarce; in the existing studies, positive effects of formative assessment on motivation were not always found (e.g., Yin et al. 2008). Even less empirical evidence is present on the effects of formative assessment on students' perceived competence

or self-efficacy. So far, the evidence is mostly based on studies with either low ecological validity (for a review see Miller and Lavin 2007), or lacking an appropriate experimental design with a control condition. Therefore, more research is needed to examine the impact of formative assessment on intrinsic motivation and perceived competence within a regular educational setting.

1.5 Research aim

The present study investigated curriculum-embedded formative assessment within an ecologically valid setting in primary school science classes, evaluating the effects on students' motivational outcomes. Three hypotheses were specified:

1. *Curriculum-embedded formative assessment fosters students' intrinsic motivation.*
2. *Curriculum-embedded formative assessment fosters students' perceived competence.*
3. *The effect of formative assessment on students' intrinsic motivation is mediated by students' perceived competence.*

2 Method

2.1 Design

The present study was part of a research initiative which set out to evaluate different teaching strategies for inquiry-based science education (Project "Individual support and adaptive learning environments in primary school" (IGEL); Decristan et al. 2015; Hardy et al. 2011; Hondrich et al. 2016). Teachers in this study were randomly assigned on school level to two conditions, formative assessment and control group. Teachers of both groups took part in professional development workshops on the topic of floating and sinking, while the formative assessment group additionally received training in formative assessment. All teachers then taught two curriculum units on floating and sinking in their classrooms. The formative assessment group teachers were expected to implement formative assessment- in the first unit, supported by predesigned materials, and in the second unit by devising their own materials. Students' perceived competence and motivation were assessed via questionnaires before the intervention (pre-test), after the first unit (post 1) and after the intervention was completed (post 2).

2.2 Sample

The sample underlying the present study consisted of $N = 28$ German primary school teachers with his or her third grade science class each (in all, $N = 551$ students). 17 teachers and 319 students participated in the formative assessment condition, 11 teachers and 232 students in the control condition. All of the schools were located in central Germany, including both rural (57% of classes) and urban areas. Class size varied between 10 and 26 students, with a mean of 20 students. The

participating teachers were mostly female (86%), had a mean age of 43.4 years (*SD* = 9.8) and an average teaching experience of 15.8 years (*SD* = 9.8). All of the teachers had taught science during the past five years. The students (48% female) were 8.8 years old on average (*SD* = 0.5). Students from immigrant families (i.e., they reported that either one or both of their parents were not born in Germany) made up 38% of the sample. Data were collected in the academic year 2010/2011.

2.3 Treatment

2.3.1 The curriculum

Our study was based on two third grade units on floating and sinking adapted from Möller and Jonen (2005). The overarching learning goal of the first unit was to understand and apply the concept of relative density, which was subdivided into four learning steps: (1) disproving common misconceptions on floating and sinking; (2) understanding floating and sinking as a property of material; (3) appreciating density as relevant property of material; and finally, (4) comparing the density of water with the density of objects to predict their floating or sinking. The second unit focused on the concepts of buoyancy force and displacement in order to build an integrated conception of floating and sinking, again subdivided into four steps: (1) experiencing displacement and disproving common misconceptions; (2) realizing volume as the determining variable for displacement; (3) experiencing buoyancy force; (4) understanding the causal connection between displacement and buoyancy to predict the swimming or sinking of objects. Within both units, each learning step was implemented using an inquiry-based approach. Starting with a research question, students' hypotheses were collected and student experiments or teacher demonstrations were planned, conducted and discussed. Finally, the findings were applied using differentiated worksheet tasks. Teachers could freely choose and assign tasks from three sets, i.e. complex transfer tasks, consolidation tasks and basic repetition tasks (these often assigned additional student experiments challenging specific misconceptions).

Each unit consisted of 9 lessons lasting 45 min each, combinable as double lessons of 90 min. According to the standard educational schedule, each unit was expected to span slightly more than two weeks.

2.3.2 Curriculum-embedded formative assessment

For the treatment group, we embedded elements of formative assessment into the first unit, creating a formative assessment version to be compared to the baseline (control) version (see also Hondrich et al. 2016). Our program of curriculum-embedded formative assessment included (a) short written tasks to assess students' current conceptual understanding, (b) individual, written, semi-standardized feedback and (c) the adaptation of instruction including the assignment of differentiated worksheet tasks. When implementing the formative assessment elements in their classrooms, teachers were asked to emphasize the formative purpose of the assess-

ments and feedback and draw a connection to the students' activity as "researchers", who constantly probe and revise their ideas to improve their understanding.

Diagnostic assessments   The assessments were developed for the study, partly adapted from materials by Möller and Jonen (2005) which had already been positively evaluated. Open as well as multiple-choice answer formats were used, assessing the target conceptions as well as common misconceptions of students. Following Shavelson et al. (2008), the assessments (four in all) were embedded after each learning step. They were placed at the end of the lessons so that teachers could evaluate and document students' answers after school and use them for adapting instruction and providing feedback (see below). All assessments evaluated the conceptions on floating and sinking students argued with. Additionally, assessments 2, 3, and 4 assessed how well students applied the conception introduced in the respective lesson. We provided teachers with a guideline on how to interpret students' results as well as with a table for documenting students' conceptions and levels of understanding throughout the unit. Students were classified into three levels of understanding: those students who could apply the new concept very well (level 3); those who showed some understanding but still made mistakes in complex tasks (level 2); and those who were still unable to apply the new concept (level 1). These levels constituted the basis for the further formative usage of the information that had been gathered.

Adaptive instruction   The assessment information was used to adapt instruction to students' individual learning needs. The first assessment, focusing on students' preconceptions, allowed teachers to prepare classroom discussions and assemble teams for experimental tasks. After assessments 2, 3 and 4, teachers were instructed to assign the available differentiated tasks according to the three levels of students' understanding: complex transfer tasks were given to students who could apply the new concept very well (level 3); consolidation tasks were assigned to level 2 students, and basic repetition tasks to level 1 students.

Formative feedback   Our feedback concept drew on research on effective formative feedback (Hattie and Timperley 2007) and should inform students how well they had understood the targeted concept, including feedback on specific problems or misconceptions, if present, and provide them with a strategy to improve. Teachers were instructed to provide complete formative feedback twice, after assessments 2 and 4. In order to help teachers realize formative feedback as intended, we provided teachers with feedback templates. Teachers had to fill in specific problems the students had faced and were encouraged to add additional information whenever necessary.

2.3.3 Professional development workshops

A series of five professional development workshops were held for the treatment and control groups, each taking 4.5 h. Two workshops addressed the curriculum on floating and sinking, including pedagogical content knowledge and content knowledge on the concepts of density (workshop 1), as well as buoyancy force and displace-

ment (workshop 5). These workshops were held by the same training team for both groups of teachers.

In between, teachers in the formative assessment condition attended three workshops on formative assessment. To parallelize contact to both groups, control group teachers took part in three workshops on parental counseling instead. In the treatment group, workshop 2 focused on the concept of formative assessment and its impact on students' learning and motivation. Workshop 3 dealt with implementing formative assessment within the first curriculum unit on floating and sinking. Finally, workshop 4 served to prepare teachers for realizing formative assessment in teaching practice and transferring it to other topics.

All teachers received standardized materials (ranging from worksheets to materials for experiments) and a detailed manual for teaching the curriculum. For the first unit, teachers in the formative assessment group received an adapted version of the manual which included the formative assessment materials described above—on top of the tools and teaching guidelines delivered to control group teachers as well. In the second unit, these teachers needed to transfer the formative assessment concept of their own accord.

2.4 Adherence checks

To make sure that teachers implemented the curriculum units as planned, we evaluated teachers' implementation fidelity of the curricular content, which should be high in both groups, as well as the usage of curriculum-embedded formative assessment strategies, which are expected to be high in the treatment group and low in the control group.

We conducted classroom observations of one double lesson (90 min) for all teachers in the sample—either video-based ($n = 20$) or live for those teachers and classes who did not agree to be filmed ($n = 8$), evaluating the occurrence or nonoccurrence of the critical treatment components in the respective lesson, as defined in the teacher's manual (Ruiz-Primo 2006; Gresham 2009). The adherence scores were computed to reflect the percentage of implemented elements relative to the intended elements. For further information on the adherence scores, see Decristan et al. (2015); Hondrich et al. (2016).

As expected, teachers in both groups showed good adherence to the curricular content: in the first unit, mean scores were $M = 87.12\%$ in the formative assessment group ($SD = 19.19$) and $M = 85.22\%$ in the control group ($SD = 10.30$); in the second unit, $M = 73.16\%$ ($SD = 17.92$) in the treatment group and $M = 78.51$ ($SD = 17.54$) in the control group. Scores did not significantly differ across groups and units ($p \geq 0.45$).

In the first curriculum unit with material-based support, treatment group teachers showed high adherence to the formative assessment intervention with a mean score of 95.43% ($SD = 11.15$). In the second unit, implementation of formative assessment elements was considerably lower, but still present ($M = 27.94\%$; $SD = 27.79$). In contrast, none of the curriculum-embedded formative assessment elements (e.g. written feedback) were observed in the control group, in neither unit ($M = 0$; $SD = 0$). The difference between the two groups was significant in both units, tested with

the Mann-Whitney U-test due to non-normal distribution of variables ($z \leq 3.13$; $p < 0.01$; $d \geq 1.31$; see also Hondrich et al. 2016). These results show that despite lower adherence in the second unit, the induction of the treatment was successful.

2.5 Indicators of students' perceived competence and intrinsic motivation

Students' intrinsic motivation and perceived competence were assessed via questionnaires before the treatment (pre-test), after the first curriculum unit (post 1) and after the completion of the second unit (post 2). Both constructs were assessed by scales adapted from Blumberg (2008) and are included in the supplementary online material. The intrinsic motivation scale consisted of 5 items ($\alpha \geq 0.77$), covering both joy in learning as well as a learning goal orientation, in the sense of an overarching intrinsic motivation factor (Marsh et al., 2003), e.g.: *"Why did you put effort in the class on floating and sinking? Because I wanted to know a lot about the things covered by this class"*. Perceived competence included 4 items ($\alpha \geq 0.87$) focusing on the individual experience of learning progress; e.g., *"In the class on floating and sinking, I learned a lot"*. The items were rated on a four-point Likert-scale ranging from 1 = strongly disagree to 4 = strongly agree. Items at pre-test were formulated to refer to science class in general, while items at both post-tests referred to the curriculum on floating and sinking. Otherwise, the items were parallel. All instructions and items were read aloud during class-wide testing. Intra class correlations were investigated to establish the amount of variance originating between classes as compared to within classes (intrinsic motivation: pre-test = 0.01; post 1 = 0.04; post 2 = 0.09; Perceived competence: pre-test = 0.03; post 1 = 0.07; post 2 = 0.04).

2.6 Data analyses

All hypotheses were tested with multilevel regression analyses using MPlus 7 (Muthén and Muthén 2014). We specified a two-level model of students (level 1) nested within classes (level 2). Treatment condition as predictor was entered as a dummy coded variable (0 = control group, 1 = formative assessment) on classroom level. All continuous variables were z-standardized before entering regression. We accounted for missing data using the FIML approach (Olinsky et al. 2003). One-tailed hypothesis testing was performed at the 0.05 α-level for all analyses.

In order to investigate the direct effects of formative assessment on intrinsic motivation and perceived competence (hypotheses 1 and 2), we performed multilevel regression analyses of post-test scores of intrinsic motivation and perceived competence on treatment condition. Both dependent variables were tested after the first unit (post 1) as well as after the second (post 2). In order to control for students' pre-test scores of intrinsic motivation and perceived competence, we included these scores as covariates on the student-level. To account for variance at both levels, they were grandmean-centered (Enders and Tofighi 2007).

To test a multilevel mediation hypothesis (hypothesis 3), two possible approaches are described in the literature: the cross-level approach recommended by Pituch and Stapleton (2010, 2012) and the cluster level approach described by Preacher et al. (2010). As our treatment is introduced on class level while both our mediator and

outcome variables are conceptualized on the individual level, and as the mediating variable is hypothesized to exert influence via absolute scale values rather than via relative standing within a class, we chose the cross level approach of Pituch and Stapleton (2010, 2012) as it most adequately covers our theoretical model[1]. In this approach, the mediating variable on the individual level (perceived competence) is centered at the grandmean to represent students' scale values as compared to all other students in the sample. The hypothesized intra-psychical effect of students' perceived competence on their intrinsic motivation is thus assumed to be captured on this level.

Moreover, in line with Pituch and Stapleton (2012), we separately modeled the compositional effect by adding the aggregated class mean of students' perceived competence as classroom level mediator. Hereby, we wanted to rule out possible effects of class composition, that is, whether being in a class with rather high or low levels of perceived competence has an impact on students' intrinsic motivation over and beyond their individual values of perceived competence. When testing the mediation effect, we controlled for the grandmean-centered pre-test scores of perceived competence and intrinsic motivation as covariates on the individual level as well as their aggregated counterparts on the classroom level (see the supplementary online material for the Mplus syntax code of the cross-level mediation analysis).

3 Results

3.1 Hypothesis 1—impact on students' intrinsic motivation

Descriptive results (see Table 1) show a generally high intrinsic motivation of the students across both groups and all three points of measurement. At pre-test, scores in the treatment and control group were very similar with no significant differences in multilevel regression analysis (motivation: $\beta = 0.07$; $p = 0.51$; perceived competence: $\beta = 0.15$; $p = 0.16$). At both post-tests, students in the treatment group reported higher intrinsic motivation than in the control group. Multiple regression analysis controlling for pre-test scores showed that at post 1, the difference between the two groups was only marginally significant ($\beta = 0.14$, $p = 0.07$). At post 2, however, intrinsic motivation in the formative assessment group was significantly higher ($\beta = 0.28$, $p = 0.03$).

3.2 Hypothesis 2—impact on students' perceived competence

Students' reported perceived competence was also fairly high in both groups (see Table 1). At pre-test, scores were slightly higher in the control group, although the difference was not significant. At both post-tests, students in the treatment group

---

[1] In contrast to Pituch and Stapleton (2012), Preacher et al. (2010) argue that class-level interventions can only exert influence on the variance between classes, so that any mediating effect should only be assessed at class-level (which is also referred to as "cluster only mediation"). When following this approach, our results do not change.

Table 1 Perceived competence and intrinsic motivation in the formative assessment and control group, before the intervention (Pre), after the first curriculum unit (Post 1) and after the second curriculum unit (Post 2)

		Formative Assessment	Control
Individual level			
Perceived competence *M* (*SD*)	Pre	3.54 (0.49)	3.57 (0.44)
	Post 1	3.65 (0.50)	3.49 (0.64)
	Post 2	3.64 (0.49)	3.51 (0.62)
Intrinsic motivation *M* (*SD*)	Pre	3.60 (0.47)	3.53 (0.54)
	Post 1	3.56 (0.52)	3.45 (0.59)
	Post 2	3.56 (0.63)	3.30 (0.78)
Classroom level			
Perceived competence *M* (*SD*)	Pre	3.52 (0.14)	3.57 (0.12)
	Post 1	3.64 (0.20)	3.52 (0.17)
	Post 2	3.59 (0.17)	3.50 (0.18)
Intrinsic motivation *M* (*SD*)	Pre	3.57 (0.11)	3.53 (0.15)
	Post 1	3.56 (0.19)	3.45 (0.15)
	Post 2	3.50 (0.26)	3.32 (0.27)

M mean; *SD* standard deviation (in parentheses)

reported higher perceived competence than in the control group. Controlling for pre-test scores, multiple regression analysis showed that the formative assessment condition indeed had a significant positive effect on students' perceived competence at post 1 ($\beta = 0.27$, $p < 0.05$) and at post 2 ($\beta = 0.27$, $p < 0.01$).

3.3 Hypothesis 3—mediating effect of perceived competence

The results of the cross-level mediation analysis are presented in Fig. 1.

We found a significant indirect effect of formative assessment on intrinsic motivation mediated by perceived competence on the individual level ($\beta(a*b_1) = 0.24$ $p < 0.01$). The indirect effect via perceived competence on cluster level was smaller and did not reach significance ($\beta(a*b_2) = 0.09$, $p = 0.07$). In all, the total indirect effect amounted to $\beta = 0.33$, $p < 0.01$, while the direct effect c of formative assessment on intrinsic motivation when controlling for perceived competence as mediator was not significant ($\beta = 0.11$ $p = 0.22$).

These results support our hypothesis that students' perceived competence (as compared to all other students in the sample) indeed functions as a mediator for the positive effect of formative assessment on students' intrinsic motivation. As perceived competence on the individual level was grandmean-centered, the non-significant effect on the classroom level indicates that there is no major effect of class composition in perceived competence operating beyond the effect of the students' individual perceived competence.
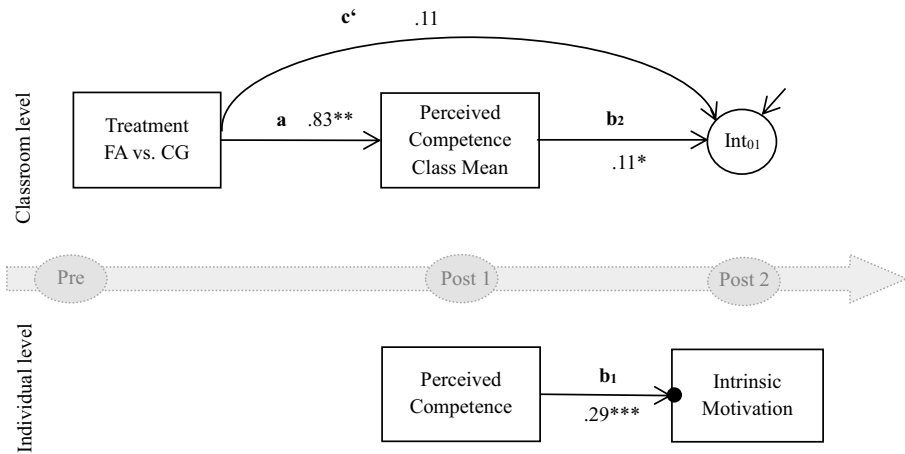
Fig. 1 Cross-level mediation analysis investigating the treatment effects (formative assessment vs. control group) on intrinsic motivation via perceived competence as mediator. *Note*. Standardized regression weights. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$; one-tailed-test, respectively. *FA* Formative Assessment; *CG* control group. The dot represents the random intercept; Int01: classroom-level random intercept. Indirect effect individual level $a*b_1 = 0.24**$. Indirect effect classroom level $a*b_2 = 0.09$. Total indirect effect $a*(b_1 + b_2) = 0.33**$. Total effect $(a*b_1) + (a*b_2) + c = 0.45**$. We included prior perceived competence and prior intrinsic motivation at the individual and at the classroom level of analysis as covariates in this model (here not depicted for improved clarity)

4 Discussion

4.1 Summary and interpretation of results

The aim of the present study was to investigate the effects of curriculum-embedded formative assessment on students' intrinsic motivation (hypothesis 1) and perceived competence (hypothesis 2), as well as investigating the mediation hypothesis that formative assessment fosters intrinsic motivation via a higher perceived competence of students (hypothesis 3).

Regarding hypothesis 1, our results showed significant positive effects of formative assessment on students' intrinsic motivation. Although intrinsic motivation of students in the formative assessment group was only marginally higher than in the control group at the first post-test (after the first teaching unit of approximately two weeks), we found significant effects at the second post-test after the second unit and approximately four weeks of intervention. This supports the view that formative assessment is effective in fostering students' motivation (e.g., McMillan et al. 2010; Pat-El et al. 2012), but also indicates that these effects may not show immediately but require a certain period of time to develop. It is interesting that the motivational benefit for the treatment group was more pronounced in the second unit, in which teachers' implementation of formative assessment was considerably lower. This could be explained by lasting effects of the implementation of formative assessment during the first unit, which would be in accordance with our mediation hypothesis. Moreover, as teachers tend to adapt teaching programs to their needs

and conditions (Garet et al. 2001; Tierney 2006; Desimone 2009), another expla-
nation is that teachers increasingly used other formative assessment strategies in
the second, transfer unit, which were not captured in the implementation fidelity
measures—e.g., on-the-fly strategies like oral questioning and feedback.

It also has to be noted that, in accordance with findings of other authors (e.g.,
Spinath and Spinath 2005), control group students' reported intrinsic motivation
decreased during the course of the intervention. The positive effect of the treatment
compared to regular teaching was thus expressed not in an absolute increase of
intrinsic motivation, but in its stability.

Regarding our second hypothesis, the results of both post-tests support the view
that formative assessment fosters students' perceived competence. Compared to stu-
dents in the control condition, formative assessment students reported that they had
learned and understood more during both units. This indicates that implementing
formative assessment in instruction changes the process of teaching and learning
in a way that makes instruction more supportive of students' need for competence.
These results are in accordance with other findings showing that formative assess-
ment strategies like formative feedback foster students' self-efficacy and perceived
competence (Rakoczy et al. 2008; Pat-El et al. 2012).

Investigating the mediating role of students' perceived competence for the impact
of formative assessment on intrinsic motivation (hypothesis 3), we found that stu-
dents' higher perceived competence at post-test 1 mediated the effect of formative
assessment on students' intrinsic motivation at post-test 2. In fact, when controlling
for perceived competence, the treatment effect on students' intrinsic motivation was
no longer significant. This underscores the important role of students' perceived
competence for positive motivational outcomes: if students experience success in
their learning, they become more intrinsically motivated, meaning that they enjoy
their learning activities more and are more oriented towards mastering the content.
Our findings thus provide empirical evidence for Deci and Ryan's postulation that
fulfilling students' basic need for competence is a predictor of the development of
self-determined and intrinsic motivation (Deci and Ryan 2000). We conceptualized
the mediating effect of perceived competence on the individual, student level, be-
cause according to self-determination theory, it is the individual, absolute level of
perceived competence which influences students' intrinsic motivation through in-
ternal psychological processes. Our results support this view. Although we found
a significant effect of class mean perceived competence on intrinsic motivation, the
over-all indirect effect on this level did not reach statistical significance. However,
further research, ideally with more power to detect compositional effects on class
level, is necessary to support these findings.

4.2 Practical implications

Our results show that formative assessment indeed constitutes a valuable strategy
of fostering students' perceived competence and intrinsic motivation within instruc-
tion. Therefore, an emphasis should be placed on supporting the implementation of
formative assessment within classroom instruction—this includes using assessments
as a means to keep track of the individual learning progress, adapting instruction to

keep tasks challenging but manageable, and providing detailed feedback with information on how to improve. The present study as well as previous research shows that the use of formative assessment strategies can be enhanced by teacher professional development (e. g., Torrance and Pryor 2001; William et al. 2004). However, the lower implementation fidelity in the second, transfer unit also points to possible obstacles: as planning for formative assessment and preparing materials is time-consuming and demanding (Tierney 2006; Bennett 2011), it is not surprising that teachers reduced the frequency of formative assessment elements. One way to address this problem is to provide teachers with predesigned curricula and materials (e. g., Desimone 2009; Bennett 2011). Still, the positive effects after the second unit indicate that there had been changes to teachers' instructional practice, and that even with the lower rate of formative assessment in the second unit, the treatment was obviously sufficient for students to feel more competent and intrinsically motivated than in regular instruction. Further research will need to investigate more in-depth which aspects and forms of formative assessment are most vital for enhancing students' perceived competence and intrinsic motivation, and whether and how instruction changes in more subtle ways than in using obviously visible assessment tasks or feedback sheets.

4.3 Limitations and directions for future research

The present study has limitations which should be acknowledged. Above all, the sample size on class level ($N = 28$) is rather small, so that the power to detect small effects especially on classroom level is reduced (Pituch and Stapleton 2012) and more sophisticated analyses (e. g., latent modeling of variables) were not possible.

Moreover, it is important to note that mediation analyses are based on correlations, so that causal interpretations must be made with caution, even though the longitudinal design we used limits the amount of possible interpretations. In future, studies with larger sample sizes on cluster level could investigate and rule out possible reverse causal effects by using cross-lagged panel analyses.

Another limitation is the treatment group teachers' reduced implementation fidelity of formative assessment elements in the second teaching unit. Further research is necessary to fully understand why the effects of formative assessment were still present with a considerably lower number of embedded formative assessment elements.

In the present study, we chose to focus on students' perceived competence to predict their intrinsic motivation, so that the basic needs of autonomy and relatedness (Deci and Ryan 2000) were not included in this study—in future research, it would be interesting to investigate students' autonomy and relatedness as well and compare their relative impact on intrinsic motivation. Moreover, further mediating variables outside Deci and Ryan's theoretical framework are conceivable, e. g. students' temporal or individual reference norm orientation (e. g., Dickhäuser et al. 2017).

It also has to be noted that we only assessed one aspect of perceived competence, students' perceived learning gains, in order to focus on the individual experience of competence rather than on comparisons with peers or on rather stable beliefs of

normative competence. When assessing perceived competence in retrospective, we see this aspect as most closely related to the idea of competence in the sense of Deci and Ryan, as it represents a summary of the students' emotional experiences during the teaching unit rather than a comparison with peers or beliefs of normative competence. Still, it would be interesting in future studies to investigate students' competence experience on an even less abstract level, e.g. immediately after a task has been completed.

As we designed a comprehensive program with the aim of fully realizing the formative assessment principles in instruction, our study does not provide information on the differential effectiveness of specific formative assessment elements provided separately (e.g., assessments, feedback, making formative assessment principles evident to students). More research is needed to investigate more in-depth the effects of specific formative assessment elements and, related to these elements, investigate further potential mediating variables.

5 Conclusion

The present study contributes to the understanding of how students' perceived competence and intrinsic motivation can be fostered in classroom instruction. Although the primary focus of research on formative assessment is set on students' learning gains, our results show that formative assessment is also successful in fostering students' motivational outcomes within a regular educational setting. Still, more research is necessary to replicate these findings and gain further knowledge about the way how formative assessment strategies effect students' motivation.

References

Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Freeman.

Bell, B., & Cowie, B. (2000). The characteristics of formative assessment in science education. *Science Education*, *85*, 536–553.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, *18*, 5–25.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*, 7–74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*, 5–31.

Blumberg, E. (2008). *Multikriteriale Zielerreichung im naturwissenschaftsbezogenen Sachunterricht der Grundschule – Eine Studie zum Einfluss von Strukturierung in schülerorientierten Lehr-Lernumgebungen auf das Erreichen kognitiver, motivationaler und selbstbezogener Zielsetzungen*. Inaugural-Dissertation zur Erlangung des akademischen Grades des Doktors in den Erziehungswissenschaften an der Westfälischen Wilhelms-Universität Münster. Münster. urn:nbn:de:hbz:6-42569418514.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry*, *11*, 227–268.

Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., & Hardy, I. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *The Journal of Educational Research*, *108*, 358–370.

Desimone, L. (2009). Improving impact studies of teachers' professional development: toward better conceptualizations and measures. *Educational Researcher*, *38*, 181–199.

Dickhäuser, O., Janke, S., Praetorius, A.-K., & Dresel, M. (2017). The effects of teachers' reference norm orientations on students' implicit theories and academic self-concepts. *Zeitschrift für Pädagogische Psychologie*, *31*, 205–219.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*, 121–138.

Flink, C., Boggiano, A. K., Main, D. S., Barrett, M., & Katz, P. H. (1992). Children's achievement-related behaviours: the role of extrinsic and intrinsic motivated behaviours. In A. K. Boggiano & T. S. Pittman (Eds.), *Achievement and motivation: a social-developmental perspective* (pp. 189–214). Cambridge: University Press.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: a longitudinal study. *Journal of Educational Psychology*, *93*, 3–13.

Gresham, F. M. (2009). Evolution of the treatment integrity concept: current status and future directions. *School Psychology Review*, *38*, 533–540.

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: the Situational Motivation Scale (SIMS). *Motivation and Emotion*, *24*, 175–213.

Hardy, I., Hertel, S., Kunter, M., Klieme, E., Warwas, J., Büttner, G., & Lühken, A. (2011). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. *Zeitschrift für Pädagogik*, *57*, 819–833.

Harter, S. (1982). The perceived competence scale for children. *Child Development*, *53*, 87–97.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112.

Heyman, G. D., & Dweck, C. S. (1992). Achievement goals and intrinsic motivation: their relation and their role in adaptive motivation. *Motivation and Emotion*, *16*, 231–247.

Hondrich, A. L., Hertel, S., Adl-Amini, K., & Klieme, E. (2016). Implementing curriculum-embedded formative assessment in primary school science classrooms. *Assessment in Education: Principles, Policy & Practice*, *23*, 353–376.

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: age differences and academic correlates. *Journal of Educational Psychology*, *97*, 184–196.

Maier, U. (2010). Formative Assessment—Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, *13*(2), 293–308.

Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, *34*, 155–167.

Marsh, H. W., Craven, R. G., Hinkley, J. W., & Debus, R. L. (2003). Evaluation of the big-two-factor theory of academic motivation orientation: an evaluation of jingle-jangle fallacies. *Multivariate Behavioral Research*, *38*, 189–224.

McMillan, J. H., Cohen, J., Abrams, L., Cauley, K., Pannozzo, G., & Hearn, J. (2010). *Understanding secondary teachers' formative assessment practices and their relationship to student motivation*. ERIC document reproduction service No. ED507712.

Miller, D., & Lavin, F. (2007). "But now I feel I want to give it a try": formative assessment, self-esteem and a sense of competence. *The Curriculum Journal*, *18*, 3–25.

Möller, K., & Jonen, A. (2005). *Die KiNT-Boxen-Kinder lernen Naturwissenschaft und Technik. Paket 1: Schwimmen und Sinken*. Essen: Spectra.

Muthén, L. K., & Muthén, B. O. (2014). *Mplus user's guide* (6th edn.). Los Angeles: Muthén & Muthén.

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, *151*, 53–79.

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*, 449–454.

Pituch, K. A., & Stapleton, L. M. (2010). Hierarchical linear and structural equation modeling approaches to mediation analysis in randomized field experiments. In M. Wiliams & W. P. Vogt (Eds.), *Innovation in social research methods* (pp. 590–619). Los Angeles: SAGE.

Pituch, K. A., & Stapleton, L. M. (2012). Distinguishing between cross- and cluster-level mediation processes in the cluster randomized trial. *Sociological Methods & Research*, *41*, 630–670.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209–233.

Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction. *Zeitschrift für Psychologie*, *2*, 111–124.

Ruiz-Primo, M. A. (2006). *A multi-method and multi-source approach for studying fidelity of implementation*. CSE report 677.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78.

Schunk, D. H., & Rice, J. M. (1991). Learning goals and progress feedback during reading comprehension instruction. *Journal of Reading Behavior*, *23*, 351–364.

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading and Writing Quarterly*, *23*, 7–25.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M., & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: a collaboration between curriculum and assessment developers. *Applied Measurement in Education*, *21*, 295–314.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *178*, 153–189.

Spinath, B., & Spinath, F. M. (2005). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction*, *15*, 87–102.

Spinath, B., & Steinmayr, R. (2012). The Roles of Competence Beliefs and Goal Orientations for Change in Intrinsic Motivation. *Journal of Educational Psychology*, *104*, 1135–1148.

Tierney, R. D. (2006). Changing practices: influences on classroom assessment. *Assessment in Education: Principles, Policy & Practice*, *13*, 239–264.

Torrance, H., & Pryor, J. (2001). Developing formative assessment: using action research to explore and modify theory. *British Educational Research Journal*, *27*, 615–631.

Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: another look at the quality of academic motivation. *Educational Psychologist*, *41*, 19–31.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, *11*, 49–65.

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., & Furtak, E. M. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied measurement in education*, *21*, 335–359.

Zisimopoulos, D. A., & Galanaki, E. P. (2009). Academic intrinsic motivation and perceived academic competence in Greek elementary students with and without learning disabilities. *Learning Disabilities Research & Practice*, *24*, 33–43.