Check for
updates

# Effects of formative assessment on intrinsic motivation in primary school mathematics instruction

**Larissa Aust[1]** · **Birgit Schütze[1]** · **Jan Hochweber[2]** · **Elmar Souvignier[1]**

## Abstract

Empirical studies on effects of formative assessment (FA) on motivational outcomes and underlying mechanisms are scarce. The core elements of FA, feedback and adaptive teaching behavior, might be helpful in promoting the experience of competence, which in turn is a prerequisite for promoting intrinsic motivation. However, instructional impact also depends on the students' perception of teacher behavior. Therefore, this study ($N = 27$ teachers and 613 students) aimed to test whether FA has a positive effect on intrinsic motivation mediated by students' perceived competence support (PCS). In a three-group design, two versions of FA were compared with a control group. In both versions of FA, teachers used a tool for learning progress assessment (LPA), while in one version, teachers received a combination of LPA and additional support consisting of materials for feedback and adaptive instruction (LPA+). A path model for half-longitudinal designs was estimated. Results support the motivational effect of FA. LPA and LPA+ both positively influence students' PCS which is in turn associated with higher intrinsic motivation. A small indirect effect on intrinsic motivation mediated by PCS was shown for both intervention groups, which was only significant for LPA. For LPA+, higher-performing students particularly benefitted from the intervention. Teachers' use of FA practices can foster students' PCS which seems to be a promising way in motivation-enhancing teaching.

---

✉ Larissa Aust
  larissa.aust@uni-muenster.de

  Birgit Schütze
  birgit.schuetze@uni-muenster.de

  Jan Hochweber
  Jan.Hochweber@phsg.ch

  Elmar Souvignier
  elmar.souvignier@uni-muenster.de

[1]  University of Münster, Fliednerstraße 21, 48149 Münster, Germany

[2]  St. Gallen University of Teacher Education, St. Gallen, Switzerland

## Introduction

Formative assessment (FA) is considered an effective framework for promoting learning in school (Black & Wiliam, 2009; Kingston & Nash, 2011; Klute et al., 2017; Lane et al., 2019). Previous research on FA has focused on its effects on student achievement which are well documented. However, in addition to achievement, student motivation is another important goal dimension in education as formulated in school curricula (Department for Education, 2013; Kunter, 2005). At the same time, motivation is an essential prerequisite for successful learning (e.g., Gottfried, 1990; Lemos & Veríssimo, 2014). However, effects of FA on motivation have rarely been studied (for exceptions, see Hondrich et al., 2018; Rakoczy et al., 2019). In particular, little is known about the underlying mechanisms mediating effects of FA on students' intrinsic motivation and about personal learner characteristics moderating these effects. These two specific questions are addressed in this article. A theoretical rationale for hypotheses about mediating processes is provided by the self-determination theory (Ryan & Deci, 2000), postulating that intrinsic motivation arises as a consequence of experiencing competence. When instruction is carried out in the sense of FA (i.e., adapting instructions to the students' achievement level and providing appropriate feedback), students can have positive learning experiences and feel themselves as competent. In this regard, according to the well-known "Offer-and-Use-Model for Teaching Effectiveness" (Fend, 1981; Helmke, 2003) or "Multilevel supply-use model of student learning" (Brühwiler & Blatchford, 2011) students' perceptions and interpretations of instruction ultimately determine outcomes of instruction in school. Linking self-determination theory to the importance of students' perceptions, more precisely, teacher behavior should be perceived by the students as supporting their competencies to affect their intrinsic motivation. Hence, this study assumes that the effect of FA on students' intrinsic motivation is mediated by students' perceived competence support.

### Formative assessment and teacher behavior

FA is a proven concept with adequate evidence of its potential to promote school learning (Klute et al., 2017; Lane et al., 2019). It refers to "…all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (Black & Wiliam, 1998, pp. 7–8). The assessment of students' achievement, the use of this assessment information for individualized feedback, and the adaption of teaching activities to individual learning needs can be identified as central features of FA (Wiliam & Thompson, 2008). At this point, it already becomes evident that teachers play the key role in this process as they decide whether and how to implement FA practices in their day-to-day teaching. Both an intervention study employing classroom recordings (Yin et al., 2008) and meta-analyses (Klute et al., 2017; Lane et al., 2019) show that the effectiveness of FA depends on the specific implementation in the classroom, that is, teachers' use of effective FA practices. Within the related framework of data-based decision-making (Mandinach, 2012), it is also assumed that the effects of data assessment on student learning are mediated by teachers' data interpretation and changes in teaching behavior in terms of adjusting instructions to the students' skills (e.g., Hebbecker et al., 2022).

## Learning progress assessment

Evidence on the concrete design of FA features and their effective use in the classroom, however, is still scarce. Despite the high potential associated with FA (Black & Wiliam, 2009), a successful implementation poses typical challenges. For instance, the assessments need to be easy to integrate into the classroom and provide valuable information about existing misconceptions; teachers should be supported in using the data to make informed decisions about further instruction, and a flexible adaptation of the materials to teachers' routines should be ensured (Hondrich et al., 2016; Yin et al., 2008). One prominent approach to assist diagnostic processes in mathematics education is provided by instruments for progress monitoring, such as learning progress assessment (LPA, Förster & Souvignier, 2015). LPA refers to a standardized assessment of students' learning progress with short parallel tests repeated at intervals of about 3 weeks throughout the school year (e.g., Souvignier et al., 2021). These data collected in the learning process provide teachers with meaningful information about students' achievement level and should inform teachers' instructional decisions. It has been empirically confirmed that making use of LPA contributes to student learning (Förster & Souvignier, 2014, 2015). However, translating assessment information into adaptive support and learning conducive feedback remains a major challenge for teachers (Visscher, 2021). Hence, it might be useful to provide teachers with materials that support them in adaptive teaching and providing feedback. In the present study, we therefore apply two differently structured approaches of FA: The effects of (a) an online-based LPA tool alone and (b) a combination of the LPA tool with prepared material for feedback and adaptive teaching (thereby covering all central FA features) are compared to (c) conventional mathematics instruction.

## Students' intrinsic motivation and experience of competence

Intrinsic motivation is of particular relevance for school learning. It represents the prototype of autonomous behavior and thus provides the basis for self-determined learning. A central characteristic is that the performance of the action is experienced as inherently interesting or enjoyable (Deci & Ryan, 1985; Ryan & Deci, 2000). Intrinsic motivation has positive consequences for student outcomes, including better learning, increased persistence, deeper processing, and increased well-being (Guay et al., 2000; Howard et al., 2021; Vansteenkiste et al., 2006). The question of how intrinsic motivation can be fostered by teacher behavior is therefore of central importance. A starting point for teachers' influence on students' motivation might be students' experience of competence as a facet of intrinsic motivation. According to the self-determination theory (Ryan & Deci, 2000), fulfilling the need for competence (together with autonomy and relatedness) contributes to the development of intrinsic motivation. A recently published meta-analysis (Bureau et al., 2022) shows that competence is the strongest positive predictor of self-determined motivation. Besides SDT, further prominent motivational theories (such as *social-cognitive theory*: Bandura, 1997, or the *expectancy-value model*: Eccles & Wigfield, 2002; Wigfield & Eccles, 2000) posit that students' perceived competence or related concepts like self-efficacy or ability self-concept are prerequisites for experiencing intrinsic motivation.

Thus, these theories provide a theoretical rationale for how FA practices can influence intrinsic motivation via an enhanced experience of competence. For instance, from a theoretical point of view, the feeling of competence can be enhanced by the teacher showing

the discrepancy between current achievement level and learning goal and providing information on how to move forward to reach this goal without feeling under pressure or control (*feedback*, Ryan & Deci, 2002). In addition, matching the instruction to the student's current learning level should lead to a better fit between skills and demands, which should reduce the likelihood of being overchallenged. Following a theoretical paper by Linnenbrink-Garcia et al. (2016) as well as considerations on the basic dimensions of teaching quality, especially student support (Praetorius et al., 2018), teachers' differentiated, high-quality instruction which is well-structured, appropriately paced, and ensures that demands are slightly above students' current achievement level should optimally foster students' feeling of competence (*adaption*).

First empirical evidence also suggests positive effects of teachers' FA use on intrinsic motivation, feelings of competence, or related constructs (Faber et al., 2017; Hondrich et al., 2018; Miller & Lavin, 2007; Rakoczy et al., 2019). For instance, Hondrich et al. (2018) investigated the impact of FA in third-grade science education. The researchers found that students whose teachers were trained in how to realize FA (including assessment tasks, feedback, and adaption of instruction) had a stronger feeling of competence in terms of perceived learning gains and reported higher intrinsic motivation than students in a control condition. In an exploratory study, Miller and Lavin (2007) found that students benefit from a wide range of techniques associated with FA (e.g., improved questioning techniques) in terms of self-esteem and beliefs about their competence. Likewise, in a cluster randomized field trial, Rakoczy et al. (2019) investigated the impact of FA in ninth-grade mathematics instruction revealing that compared to a regular mathematics instruction, teachers' use of FA practices (using assessment tasks and feedback including hints for improvement) had a positive effect on students' change in self-efficacy.

Some studies also investigated the effect of specific teacher behavior (e.g., feedback *or* adaptive teaching) instead of the application of a range of FA practices. Positive effects of feedback on motivational outcomes have been shown in most of these studies (e.g., Harks et al., 2014; Rakoczy, Klieme, Bürgermeister, & Harks, 2008; Wisniewski et al., 2019; for exceptions, see, for example, Drost & Todorovich, 2017). Results of survey studies indicate that students who perceive their teachers to frequently apply formative feedback report higher feelings of competence and intrinsic motivation (Leenknecht et al., 2020; Pat-El et al., 2012). Further, a differentiated teaching style in girls' physical education classes can positively influence their development of intrinsic motivation (Goudas et al., 1995) and reduce negative consequences of the big-fish-little-pond effect regarding low-achieving students' academic self-concept (Roy et al., 2015).

## Students' perceived competence support

To achieve desired outcomes in school (here, increased intrinsic motivation), students' perception of the learning environment appears to be central (Kunter et al., 2007). The critical role of students' perception is also postulated in well-known models of instructional impact (Brühwiler & Blatchford, 2011; Fend, 1981; Helmke, 2003). Thus, it might be relevant whether students subjectively *feel supported* in their development of competencies through the instructional decisions made by their teachers. If students perceive that they are supported in their competencies by the teacher's behavior, an increase in their experience of competence and subsequently in intrinsic motivation seems likely. In this context, the focus lies on the effect of students' perception of competence support on intrinsic motivation, while the possible intermediate step of an increased experience of competence is neglected.

However, the influence of students' *perceived competence support* (PCS) on intrinsic motivation has so far rarely been examined. Initial evidence for secondary school mathematics instruction suggests that PCS can positively influence subsequent self-determined motivation (Rakoczy, Klieme, & Pauli, 2008). Further, results of a laboratory experiment in ninth-grade mathematics classes showed that process-oriented feedback had a positive effect on the development of interest compared to social-comparative feedback which was mediated by PCS (Rakoczy et al., 2013). The present study therefore focuses on the influence of FA on intrinsic motivation mediated by PCS.

## Achievement level as a moderating variable

In intervention studies, the question of differential effects is of particular interest (Fuchs & Fuchs, 2019). Not least, against the background that the FA approach aims at creating an optimal fit between learning requirements and learning opportunities, differential effects at the edges of the achievement spectrum seem conceivable. Assessment information might make it easier for teachers to pay attention to differentiation in the classroom, especially in the case of particularly high-achieving or low-achieving students, and to allow these children to work on suitable tasks. Such an increased fit through FA might also lead to positive effects on motivational variables. However, such effects have rarely been studied to date. In an exploratory analysis on effects of FA on self-esteem and self-competence of Miller and Lavin (2007), changes in the two constructs were significant for lower and higher ability group members, but not for middle ability group members. However, experimental control was rather limited in this study. Slightly more evidence on potential moderation effects of FA is available for student achievement as outcome. Some of these few studies show differential effects in favor of high-achieving children (Faber et al., 2017), some in favor of low-achieving children (Bokhove & Drijvers, 2012; Koedinger et al., 2010), while Lee et al. (2020) did not find evidence for differential effects regarding different classroom types in their systematic review. Even when the individual components of FA are considered, no consistent pattern emerges. Studies suggest that the learning benefits of various types of feedback differ depending on the student's level of achievement (for an overview, see Shute, 2008). Regarding differentiated instruction, a large-scale intervention study in which teachers participated in a professional development training on adaptive instruction in mathematics classrooms showed significant, small effects on mathematics achievement in one of two cohorts (Prast et al., 2018). Effects were equal for students at different achievement levels. The limited research base nevertheless does not allow us to formulate directed hypotheses about the differential effectiveness of FA.

## The present study

In the present study (see Fig. 1), we investigate the impact of (a) LPA and (b) LPA in combination with additional support consisting of material for feedback and adaptive instruction (= LPA+) on students' motivational development compared to business-as-usual instruction. More specifically, we aim at investigating the influence of students' PCS on intrinsic motivation as well as the influence of teachers' use of the aforementioned FA practices on students' PCS. Further, we will analyze whether effects of FA on intrinsic motivation are mediated by students' PCS. While the focus here is on students' perceptions of competence support and their influence on intrinsic motivation as
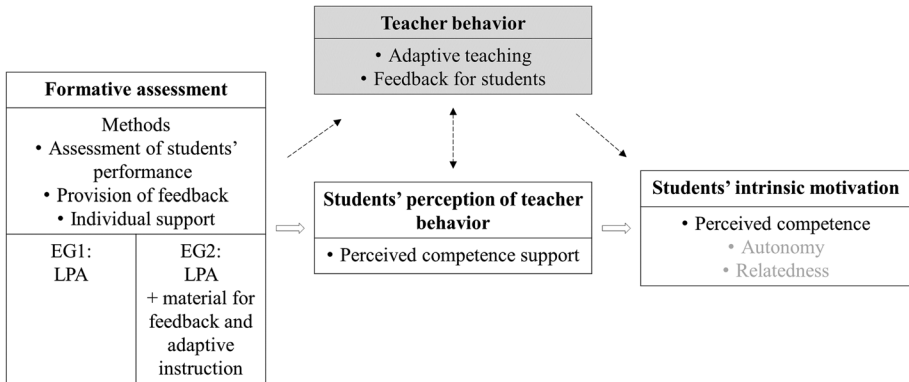
**Fig. 1** Assumptions regarding the effect of FA practices on students' intrinsic motivation via students' perceived competence support. *Note. EG* experimental group, *LPA* learning progress assessment

a final desired outcome in school, students' experience of competence is not assessed. In addition, differential effects depending on students' achievement level will be examined.

The following hypotheses will be addressed:

Hypothesis 1: We expect PCS to positively influence intrinsic motivation.
Hypotheses 2a, b: We expect (a) LPA and (b) LPA+ to have positive effects on PCS compared to a control group.
Hypothesis 2c: The positive effect on PCS is higher for students in the LPA+ group compared to the LPA group.
Hypotheses 3a, b: We expect (a) LPA and (b) LPA+ to have an indirect effect on intrinsic motivation mediated by PCS compared to a control group.
Hypothesis 3c: The positive indirect effect on intrinsic motivation via PCS is higher for students in the LPA+ group compared to the LPA group.

As an additional exploratory question, we analyze if the treatment effects are moderated by students' achievement level.

# Method

## Sample

Our sample consisted of $N = 27$ third-grade mathematics classes from 18 schools recruited in Germany. Participation in the study was voluntary. Participating teachers had a mean age of $M = 40.35$ years ($SD_{Age} = 10.08$, $Min_{Age} = 28$, $Max_{Age} = 62$). The majority of teachers were female (87%) and had studied mathematics (72.7%). A total of $N = 613$ students participated in the study. At pre-test, students were on average $M = 8.74$ years old ($SD_{Age} = 0.47$). 52.5% of the students were female and 92.0% were born in Germany.
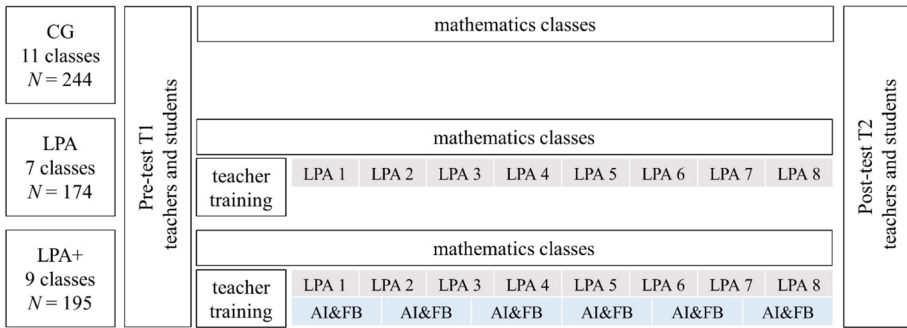
**Fig. 2** Study design. *Note. CG* control group, *LPA* learning progress assessment, *LPA+* learning progress assessment plus support, *AI* adaptive instruction, *FB* feedback

## Design

The study was conducted as a quasi-experimental field experiment with three conditions. FA was realized using two different approaches: Teachers in the experimental group one (EG1 = LPA) used a digital LPA tool, whereas teachers in the experimental group two (EG2 = LPA+) were provided with additional support consisting of materials for feedback and differentiated instruction in their classrooms. Teachers in the control group (CG) conducted their business-as-usual instruction. In all three groups, the regular curriculum for the third school year was taught, with the intervention focussing on arithmetic and math text problems. Assignment to a condition was made at the class level, making sure that classes from the same school were in the same condition. Teachers in both experimental conditions participated in teacher trainings prior to the study. The assessment of both students' mathematics competencies and motivational variables as well as teacher surveys was conducted at the beginning (September) and the end (June) of the school year (see Fig. 2).

## Treatment

Teachers of both treatment groups embedded components of FA in their mathematics classes. The individual components of the interventions are briefly explained below.

## LPA group

**Quop** The online-based assessment tool "quop" which monitors students' math progress was applied (Souvignier et al., 2021). The program includes eight short (8–15 min.), computer-based math tests in multiple-choice format that are taken by the students at intervals of about 3 weeks over the course of half a school year (see Fig. 2, LPA1-LPA8). The math tests include arithmetic (sample item $135 + 17 = ?$, response options 142; 152; 155; 118), geometry and calculation with units, are scored by computer and can be inspected by both teachers and students immediately after the tests have been administered. Teachers can inspect test results on both individual and class level. Results are reported separately for the three areas of competence, with the proportion of correctly solved tasks being indicated. In addition, students' individual test results are compared with a norm sample. Thus, teachers receive objective assessment data that can be used to make instructional decisions and adjustments (for further information on psychometric properties, see Souvignier et al., 2021).

## LPA+ group

The LPA tool quop was also used by the LPA+ group. Additionally, a more differentiated assessment and material for feedback and adaptive instruction were provided. Examples are presented in Table 1.

**Table 1** Implementation of formative assessment activities in the LPA+ group (for semi-written subtraction in the number range up to 1000 with hundreds)

---

**Supplementary test**

*On the next pages you will find math tasks where you have to subtract a number from another number. Write your way of calculation under the task and write the solution in the little box.*
$500 - 100 =$

**Adaptive instruction**
(Reri = learning mascot, graphics omitted)

Explanation card
*Reri looks scared at the following task:*
$900 - 300 = ?$
*Then an idea comes to him: I was not given any ones and any tens, so all I have to do is to subtract the hundreds from each other:*
9 hundreds − 3 hundreds = 6 hundreds.
*I can do that easily in my head or count it on my fingers.*

Practice card
*The following calculations are fun for Reri because he really likes the numbers. Go ahead, calculate with him:*
$800 - 300 = ?$

Solution card
$800 - 300 = 500$

**Feedback**
(Individual implementation by the teachers, supported by strategy cards, feedback guide, visualization (learning map))

Strategy cards
(Including tips for possible clarifications matching the students' explanation cards)
*In this area, the focus is on subtraction with hundreds in the number range up to 1000. The students should understand that they can easily calculate "large numbers" with the methods they have learned so far, since only zeros are added. For this purpose, the place value system is also discussed with the help of the place value table.*

Exemplary feedback given in the feedback guide (group setting)
*You wrote a short test not long ago. Do you remember? I looked at your answers and I saw that you already mastered a lot of tasks. You understood well what minus is and how to calculate to 100 minus. For all tasks that you have solved without problems, you get a green sticker* (visualization on learning map). *There are certain tasks where you can still improve. For such tasks you will get a blue sticker. I have seen that you can still improve in calculating up to 1000 with hundreds. Today, let's practice these tasks. Now I would like to explain to you how you can become better at calculating up to 1000 with hundreds.* (Explanations based on the strategy card will follow).

---

Formulations used in the materials are written in italics

**Additional assessment by differentiated evaluation of quop results and supplementary tests** Going beyond the standard evaluation of quop results, teachers in EG2 were provided with a more differentiated report of quop results focussing on basic arithmetic operations. The percentage of correctly completed tasks per basic arithmetic operation type (addition, subtraction, multiplication, division) was displayed for each student. This information should help teachers at recognizing which students had difficulties *with which* basic arithmetic operation and provide the basis for the selection of one supplementary test. Supplementary tests were used to identify individual strengths and weaknesses *within one* specific type of basic arithmetic operations. For each basic arithmetic type, tests were available for semi-written procedures, written procedures (addition and subtraction only), and math text problems. Teachers selected one supplementary test per student in advance (e.g., semi-written subtraction). Due to time constraints, the tests were administered six times during the school year following quop tests. All tests covered different domains of difficulty (e.g., for semi-written subtraction: subtraction in the number range up to 20, subtraction in the number range up to 100). The feedback on test results for each student included the indication of which domains of difficulty (e.g., subtraction in the number range up to 20) of the respective supplementary test (e.g., semi-written subtraction) were mastered, as well as a support recommendation consisting of two domains of difficulty that were not mastered (indicated by a test result of less than 50%)[1].

**Adaptive instruction** Teachers were supported in their individualized instruction for students. To implement adaptive teaching, they were provided with support toolboxes. For each domain of difficulty of each supplementary test (e.g., semi-written subtraction in the number range up to 20), there was an explanation card as well as practice and solution cards. Particularly high-performing students were given tasks designed for promoting highly talented mathematic students (Käpnick, 2016) to work on. The support usually took place within the framework of a double lesson after each supplementary test. Each student was expected to work on the two domains of difficulty not yet mastered assigned to him or her. Students worked on the support toolbox in achievement-homogeneous or achievement-heterogeneous learning tandems (depending on the teacher's preference).

**Feedback** The goal of the oral feedback was to provide students with information about their individual achievement level and assistance in their learning process. Teachers could decide on the feedback setting (individual vs. in groups) and were asked to follow guidelines of learning promotive feedback. Especially, they were required to address the three central feedback questions: "How am I going?", "Where am I going?", and "Where to next?" (Hattie & Timperley, 2007), in terms of reporting strengths and weaknesses, setting learning goals, and providing appropriate strategies to achieve the learning goals. Other characteristics related, for example, to temporal proximity or attribution. By the use of a feedback guide and strategy cards for each domain of difficulty, teachers were supported in giving feedback. A learning map was used to visualize oral feedback and learning progress.

### Professional development trainings

Teachers of the two experimental groups participated in professional development trainings prior to the study. The trainings were held separately at university but by the same training

---

[1] This cut-off criterion was raised during the course of the study because, otherwise, too many students would have mastered each domain of difficulty.

teams. In both groups, the general project procedure and the project goals were presented. Afterwards, the computer-based LPA tool quop was introduced, and the implementation of the tool in the classroom was discussed. Instructions for the implementation, evaluation, and interpretation of the learning process data were given. In EG1 (LPA; duration of the training: 90 min), the teachers were given no specific information on how to implement feedback and adaptive instruction but were instructed to carry out these two steps at their own discretion. Teachers in EG2 (LPA+; duration of the training: 120 min) were provided with additional information on the more differentiated quop results, the supplementary tests, the use of the support toolbox, how to provide feedback, and recommendations on how to implement feedback and support material in their instruction.

## Measures

Data were collected from both teachers and students at the beginning and end of the school year. At pre-test, teachers' demographic information was collected. At post-test, teachers in the experimental groups were asked to provide information on the implementation of progress monitoring with quop (EG1 and 2) and additional support (EG2, see "Treatment fidelity").

On student level, math achievement was assessed at pre-test; PCS and intrinsic motivation were assessed at pre- and post-test. Both PCS and intrinsic motivation were assessed with self-report ratings. Items were rated on a four-point Likert scale ranging from 1 = "does not apply" to 4 = "applies exactly." Items at pre-test referred to general practice in math classrooms, whereas items at post-test were related to the mathematics lessons of the last 3 months. Apart from that, items were identical. Due to practical reasons and internal consistency issues, the number of items differed between pre- and post-test. After item selection, all scales displayed satisfactory to good internal consistency.

**Perceived competence support** PCS was assessed using an adapted version of the perceived competence scale of the "IGEL"-project (Decristan et al., 2015; Hardy et al., 2011) based on Kunter (2005) and Prenzel et al. (1996). Six items ($\alpha = 0.76$) were included at pre-test, and eight items ($\alpha = 0.82$) at post-test, e.g., "In math classes, I was supported in better understanding the topics that were covered in class."

**Intrinsic motivation** The intrinsic motivation scale was also based on the IGEL-project and was adapted from Blumberg (2008) and Bos et al. (2005). It consisted of four items ($\alpha = 0.68$) at pre-test and six items ($\alpha = 0.74$) at post-test, e.g., "Why do you put effort into math classes? Because I want to understand more about the subject."

**Math achievement** Math achievement was assessed using the subtests arithmetic and math text problems of the DEMAT 2+ (Krajewski et al., 2004). Twenty-four items were employed ($\alpha = 0.89$, sample item "Take the double! 5 ➔ X").

## Treatment fidelity

Our strategy for assessing treatment fidelity was twofold: In both experimental groups, we used self-report ratings of teaching practices. In the LPA+ group, teachers' feedback practices were additionally assessed at random by trained observers.

**Self-report ratings** At the end of the school year, teachers were asked to rate their implementation fidelity on a four-point Likert scale ranging from 1 = "does not apply" to 4 = "applies exactly." In both groups, the use of quop was assessed with four items. Teachers in EG2 additionally provided information on their use of the supplementary tests (five items), their feedback given to students (seven items), and adaptive instruction using the support material (four items). Implementation fidelity varied across teachers with the majority of the teachers reporting a satisfactory level of implementation fidelity. Nonetheless, some teachers indicated low implementation fidelity for LPA+. One teacher (1) did not participate in the professional development training and reported very low implementation fidelity. For another teacher (2), no information on implementation fidelity was available. We decided to conduct the analyses with all participating classes. At the same time, we conducted analyses after excluding either class (1) or classes (1) and (2). The results proved to be stable. We provide information on treatment fidelity and the results of the other analyses in the supplemental material in the Open Science Framework (https://osf.io/pk5rm/).

**Observer ratings** Teachers' feedback practices were rated once by two trained observers if teachers agreed to the observation. The rating scheme covered 19 aspects of promotive feedback, e.g., identifying strengths/weaknesses, providing strategies, and setting learning goals and attribution. In each feedback situation, the presence or absence of these characteristics were rated (0 = "characteristic not available," 1 = "characteristic partially available," 2 = "characteristic available"). In case of inconsistency, observers agreed on a common rating. Interrater reliability across all characteristics was satisfactory ($K = 0.71$). Overall, implementation fidelity was satisfying except for providing (detailed) strategies. Detailed information is provided in the supplemental material.

## Data analysis

Our study design is consistent with cross-lagged panel models for half-longitudinal designs (Cole & Maxwell, 2003; Preacher, 2015). In these models, the effect of the treatment on the mediator at T2 and the effect of the mediator at T1 on the outcome at T2 are estimated controlling for pre-test differences in both the mediator and the outcome. Given the assumption of stationarity, implying that the effect of the mediator on the outcome is stable over time, the indirect effect can be estimated as the product of the two path coefficients (see Fig. 3).

All variables were grand-mean-centered at T1. Motivation and PCS at T2 were specified as outcome variables. We estimated two separate path models to examine hypotheses one to three (model 1) and the exploratory research question (model 2), using the lavaan-package (Rosseel, 2012) in R (R Core Team, 2022). For both models, we entered the treatment variables as predictors using two dummy-coded variables with the control condition representing the reference category (LPA vs. CG: 0 = CG, 1 = LPA; LPA+ vs. CG: 0 = CG, 1 = LPA+). We estimated the effects of the treatment variables and PCS at T1 on both outcomes. In addition, the effects of achievement at T1 on PCS at T2 and the effect of motivation at T1 on motivation at T2 were included in the model. To test hypothesis 1, we examined the effect of PCS at T1 on motivation at T2 (path c, cf. Fig. 3). To test hypotheses 2a, b, we examined the effects of the treatment variables on PCS at T2 (paths a and b). The indirect effects of the treatment variables on motivation at T2 via PCS (hypotheses 3a, b) were tested based on the products of the path coefficients $a \times c$ and $b \times c$, respectively. The differences between the treatment variables effects' were tested on statistical
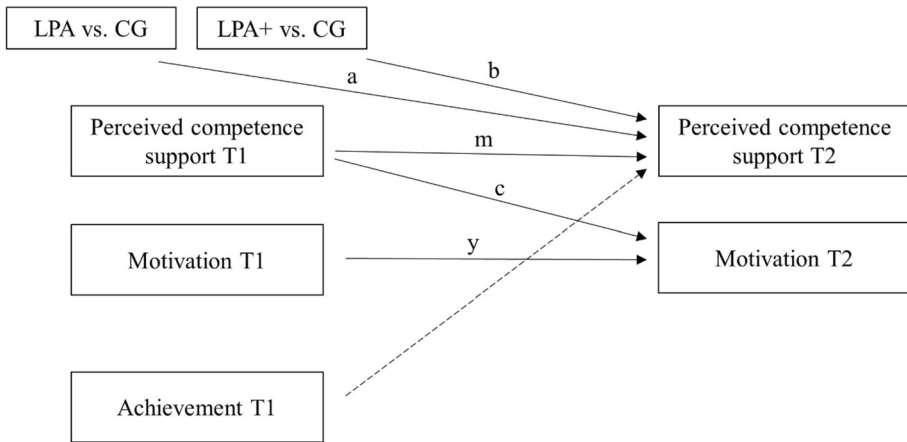
**Fig. 3** Illustration of the half-longitudinal mediation model. *Note.* The indirect effects are calculated by multiplying a × c and b × c, respectively. For better clarity, the direct effects of group membership on motivation are omitted in this figure

significance (hypothesis 2c, effects on PCS at T2; hypothesis 3c, indirect effects on motivation at T2). For model 2, to investigate the exploratory research question, we additionally specified the interaction effects (a) treatment variables × achievement at T1 on PCS at T2 and (b) PCS at T1 × achievement at T1 on motivation at T2. In both models, we used robust maximum likelihood estimation. For the direct effects, we calculated Cohen's $f^2$ (Cohen, 1988; conventions small = 0.02, moderate = 0.15, large = 0.35). For the indirect effects, we report confidence intervals for the bias-corrected bootstrap (DiCiccio & Efron, 1996). To adequately account for the hierarchical data structure and obtain correct standard errors, the variance estimation procedure "cluster" (Rosseel et al., 2022) was applied in both models. Except for the exploratory research question, all hypotheses were tested one-tailed at the 0.05 α-level.

Missing data analysis revealed that the percentage of missing data was below 15% for all constructs (PCS T1, 9.8%; PCS T2, 12.6%; motivation T1, 10.9%; motivation T2, 10.4%; achievement T1, 6.9%). The two-step procedure suggested by Jamshidian and Jalal (2010) and Jamshidian et al. (2014) was used for analyzing the missing data pattern. The assumption of Missing Completely At Random (MCAR) was not rejected (Hawkin's test, $p < .001$; non-parametric test of homoscedasticity, $p = .689$). Thus, we considered it justified to deal with missing data using the conventional full information maximum likelihood procedure (FIML) requiring the even weaker assumption that values are missing at random (MAR; Enders, 2001; Lüdtke et al., 2007).

## Results

Means and standard deviations of all examined variables are presented in Table 2.

Both proposed models displayed very good model fit (model 1: $\chi^2(2) = 3.76$, RMSEA = 0.05, SRMR = 0.02, CFI = 0.99, TLI = 0.95; model 2: $\chi^2(4) = 6.75$, RMSEA = 0.04, SRMR = 0.01, CFI = 0.99, TLI = 0.96; robust fit-indices are reported).

**Table 2** Means and standard deviations for all variables

| | CG | | | LPA | | | LPA+ | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | M | SD | N | M | SD | N | M | SD |
| PCS T1 | 226 | 3.62 | 0.39 | 160 | 3.41 | 0.59 | 184 | 3.48 | 0.55 |
| PCS T2 | 216 | 3.30 | 0.60 | 161 | 3.38 | 0.47 | 179 | 3.39 | 0.52 |
| IMOT T1 | 225 | 3.49 | 0.54 | 158 | 3.35 | 0.65 | 185 | 3.46 | 0.51 |
| IMOT T2 | 217 | 3.41 | 0.44 | 161 | 3.38 | 0.52 | 179 | 3.42 | 0.46 |
| ACH T1 | 226 | 14.96 | 5.98 | 160 | 13.86 | 6.05 | 185 | 16.30 | 5.30 |

*CG* control group, *LPA* learning progress assessment, *LPA+* learning progress assessment plus support, *PCS* perceived competence support, *IMOT* intrinsic motivation, *ACH* achievement (max. 24 points), *T1* pre-test, *T2* post-test

Results of the path model for research questions 1–3 (model 1) are provided in Table 3 (for an illustration, see Fig. 4). Consistent with hypothesis 1, PCS at T1 significantly predicted intrinsic motivation at T2. Regarding research question 2, significant, yet very small, positive effects on PCS at T2 were found for both LPA (hypothesis 2a) and LPA+ (hypothesis 2b). Contrary to hypothesis 2c, the positive effect on PCS at T2 was not higher for LPA+ compared to LPA (EG2 β = −0.024, p = .430). The indirect effect of the treatment variables on intrinsic motivation via PCS (hypotheses 3a, b) was calculated by multiplying the corresponding path coefficients. A significant indirect effect was found for LPA (β = 0.062, 90% CI [.007, 0.117]) but not for LPA+ (β = 0.057, 90% CI [−.009, 0.122]). The indirect effects (hypothesis 3c) did not differ between the two treatments (β = −0.005, p = .429). In this model, the explained variance was $R^2 = 0.15$ for both PCS at T2 and intrinsic motivation at T2.

To examine potential moderator effects of students' achievement, the interactions between achievement at T1 and the treatment variables as well as achievement and PCS at T1 were integrated in a second path model (model 2; see Table 4 and Fig. 5, respectively). A significant interaction effect was found for LPA+, indicating that the higher the level of math achievement at T1, the greater the positive effect of LPA+ on PCS at T2.

**Table 3** Results of the path model for research questions 1–3 (model 1)

| Dependent variable | Predictor | β | SE | z | p | $f^2$ |
|---|---|---|---|---|---|---|
| Motivation T2 | LPA vs. CG | 0.054 | 0.11 | 0.52 | .302 | |
| | LPA+ vs. CG | 0.084 | 0.08 | 1.04 | .150 | |
| | Motivation T1 | 0.171 | 0.06 | 3.09 | .001 | |
| | PCS T1 | 0.213 | 0.04 | 4.88 | <.001 | .08 |
| PCS T2 | LPA vs. CG | 0.290 | 0.14 | 2.05 | .020 | .01 |
| | LPA+ vs. CG | 0.266 | 0.16 | 1.67 | .048 | .01 |
| | Achievement T1 | −0.106 | 0.05 | −2.16 | .016 | |
| | PCS T1 | 0.406 | 0.05 | 7.53 | <.001 | |

One-tailed testing was applied when estimating the effects; $f^2$ was calculated for direct effects indicated in the research questions

*CG* control group, *LPA* learning progress assessment, *LPA+* learning progress assessment plus support, *PCS* perceived competence support, *T1* pre-test, *T2* post-test
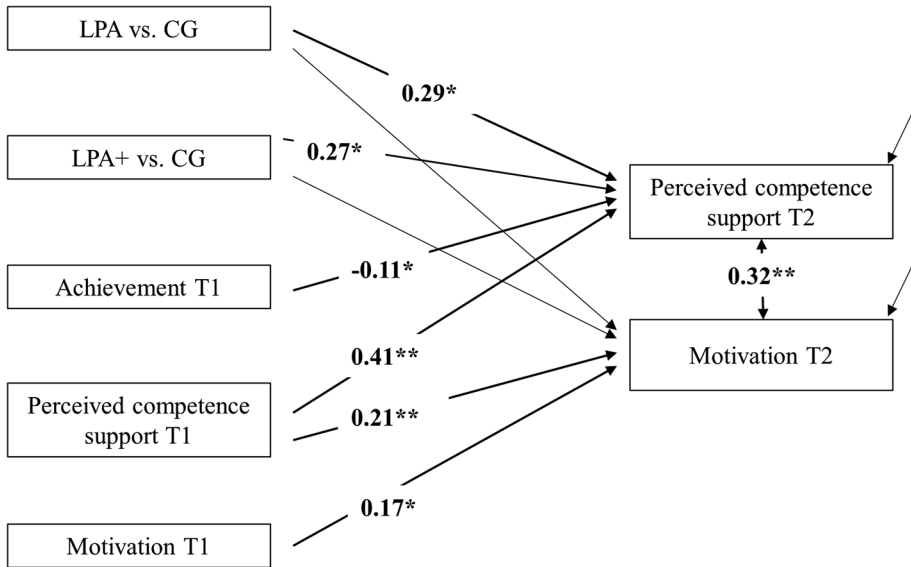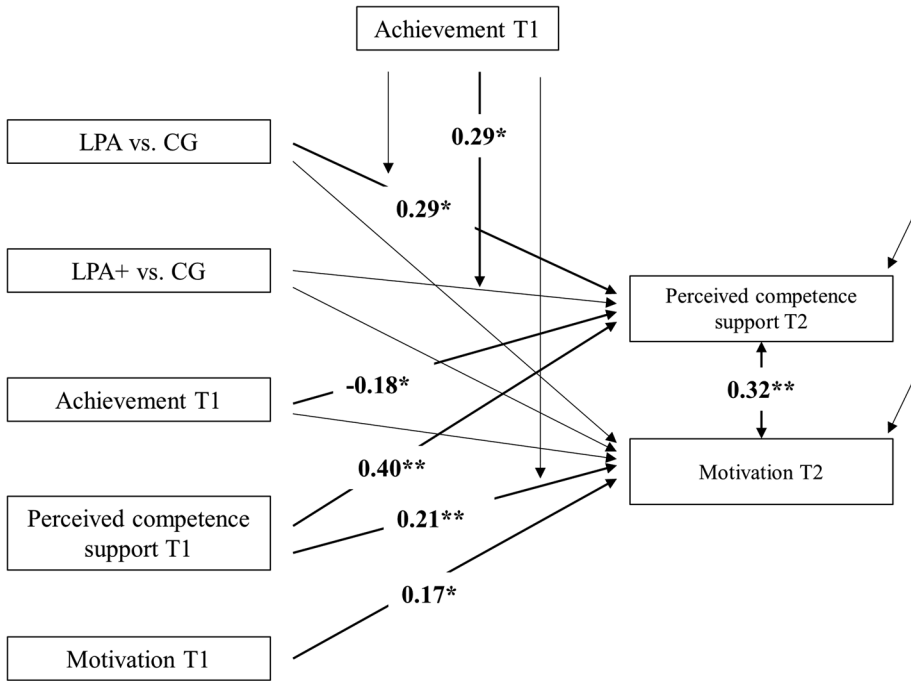
**Fig. 4** Half-longitudinal mediation for treatment effects on motivation via PCS (model 1). *Note. CG* control group, *LPA* learning progress assessment, *LPA+* learning progress assessment plus support, *T1* pre-test, *T2* post-test, for better clarity, coefficients that did not reach statistical significance were omitted, and presented coefficients were rounded to the second decimal place. *$p < .05$; **$p < .001$

**Table 4** Results of the path model for research question 4 (model 2)

| Dependent variable | Predictor | β | SE | z | p | $f^2$ |
|---|---|---|---|---|---|---|
| Motivation T2 | LPA vs. CG | 0.054 | 0.10 | 0.52 | .303 | |
| | LPA+ vs. CG | 0.074 | 0.08 | 0.95 | .172 | |
| | Motivation T1 | 0.168 | 0.06 | 3.07 | .001 | |
| | PCS T1 | 0.212 | 0.04 | 4.84 | <.001 | |
| | Achievement T1 | 0.027 | 0.04 | 0.74 | .230 | |
| | Achievement T1 X PCS T1 | 0.008 | 0.03 | 0.29 | .769 | .00 |
| PCS T2 | LPA vs. CG | 0.291 | 0.14 | 2.07 | .019 | |
| | LPA+ vs. CG | 0.214 | 0.16 | 1.37 | .086 | |
| | PCS T1 | 0.403 | 0.05 | 7.74 | <.001 | |
| | Achievement T1 | −0.184 | 0.08 | −2.29 | .011 | |
| | LPA vs. CG X Achievement T1 | 0.060 | 0.09 | 0.65 | .514 | .00 |
| | LPA+ vs. CG X Achievement T1 | 0.287 | 0.11 | 2.58 | .010 | .01 |

Two-tailed testing was applied when estimating interaction effects; one-tailed testing was applied when estimating all other effects; $f^2$ was calculated for direct effects indicated in the research questions

*CG* control group, *LPA* learning progress assessment, *LPA+* learning progress assessment plus support, *PCS* perceived competence support, *T1* pre-test, *T2* post-test

**Fig. 5** Half-longitudinal mediation for treatment effects on motivation via PCS with interaction effects (model 2). *Note.* *CG* control group, *LPA* learning progress assessment, *LPA+* learning progress assessment plus support, *T1* pre-test, *T2* post-test, for better clarity, coefficients that did not reach statistical significance were omitted, and presented coefficients were rounded to the second decimal place. *$p < .05$; **$p < .001$

No significant interaction effects were found for LPA and achievement at T1 and PCS and achievement at T1. The explained variances for PCS at T2 and motivation at T2 were $R^2 = 0.16$ and $R^2 = 0.15$, respectively.

## Discussion

The aims of this study were manifold: first, we examined whether students' PCS is positively associated with intrinsic motivation. Second, we analyzed if teachers' use of FA practices has a positive effect on students' PCS and whether there is an indirect effect of FA on intrinsic motivation mediated by students' PCS. Further, we investigated whether the effects on motivation depend on students' achievement level. As a particular strength of the study, two differently structured approaches of FA were realized: While teachers in EG1 received results from a digital LPA tool, teachers in EG2 were provided with the combination of LPA results and additional support consisting of materials for feedback and adaptive instruction.

As expected, we found PCS to significantly predict intrinsic motivation (hypothesis 1). For both LPA (hypothesis 2a) and LPA with an additional support component (hypothesis 2b), significant, small positive effects on PCS were shown compared to the control group. Contrary to our expectations, these effects did not differ between the two experimental

conditions (hypothesis 2c). There are indications of small indirect effects for LPA (hypothesis 3a) and LPA+ (hypothesis 3b) on intrinsic motivation mediated by PCS. However, statistical significance was only reached for LPA. There was no difference in the indirect effect between LPA and LPA+ (hypothesis 3c). Regarding possible differential effects depending on students' achievement level, a significant interaction effect for LPA+ emerged. The higher the students' achievement level, the stronger the positive effects of the intervention. For LPA, in contrast, no interaction effect was found. The effect of PCS on intrinsic motivation is not related to students' achievement level.

Our results highlight the importance of students' perception of supportive teacher behavior for their motivational development: The more students feel supported in the development of their competencies by their teachers' adaptive behavior, the greater their intrinsic motivation. Given the positive outcomes associated with intrinsic motivation (Howard et al., 2021), this finding highlights the need for promoting the feeling of competence support in the classroom. Both experimental groups were associated with increased PCS compared to the control group. This suggests that FA practices are effective in changing teaching behavior towards an enhanced competence support in the mathematics classroom. This finding is consistent with previous studies on the relationship between the use of FA practices and perceived competence (Leenknecht et al., 2020; Miller & Lavin, 2007; Pat-El et al., 2012).

The very small effects of LPA and LPA+ may be due to the fact that in these conditions, regular mathematics classes were enriched by teachers' use of FA practices, but not necessarily in all lessons adaptive instruction and feedback were implemented. Likewise, only very weak indirect effects were found for both groups (which were significant only for LPA). It appears that the influence of the implemented methods might not be strong enough to achieve greater effects on intrinsic motivation mediated by students' PCS. The assumed mechanism that students' PCS strengthened by teachers' use of FA practices is beneficial for their intrinsic motivation could nevertheless be confirmed. A possible reason for the lack of greater indirect effects might be a low frequency of FA application. For instance, whereas the present study extended over a long period of time, but with a comparatively low frequency of FA, the implementation of FA in a quite similar study by Hondrich et al. (2018) occurred over a relatively short period of time but was very intensive (two units of 9 h each in slightly over 2 weeks). Thus, a higher dosage of FA application may be needed to achieve the desired effects.

A strength of the study relates to the comparison of two different approaches of formative assessment. Unexpectedly, a comparison of the effects of LPA and LPA+ failed to yield an advantage of the additional support component. Therefore, the provision of progress monitoring information appears to be a key element in promoting PCS. This finding seems surprising in view of both feedback (Krijgsman et al., 2019; Rakoczy et al., 2013) and adaptive teaching (Goudas et al., 1995; Roy et al., 2015) being associated with the experience of competence or related constructs. As a possible explanation, results from feedback research (Harks et al., 2014; Rakoczy et al., 2019) can be considered, suggesting that the effect of feedback (in this case, on interest in math) is mediated by its perceived usefulness. It may be that feedback was not perceived as more useful by students in the LPA+ group. Classroom observations which indicated that teachers in the LPA+ condition rarely taught students appropriate strategies for achieving the learning goals support this notion. It may therefore be assumed that, despite the support components teachers were provided with, the quality of feedback did not substantially differ between the experimental groups (for the challenge of improving teachers' feedback practice in mathematics instruction, see Schütze et al., 2017). Possibly, this is also true for adaptive instruction. It seems

that teachers in both groups used the assessment information in a similar way for adaptive teaching, with the provision of differentiation options through the LPA+ materials having no substantial additional effect. The potentially fundamental role of LPA in the context of FA was also discussed in a study conducted by Hebbecker and Souvignier (2018). In reading classrooms, the researchers found additional effects of feedback or support components on neither achievement nor motivational variables and interpreted this result (in addition to considerations of a possibly limited implementation of the components) as an indication that LPA had already a sufficiently high stimulative nature to implement the concept of FA. However, more information on how teachers in the LPA group dealt with the information on learning progress would be useful to explain the missing additional effect through feedback and adaptive instruction. Another explanation could relate to the duration of the teacher trainings. For practical reasons, the training sessions differed by only 30 min between the intervention groups. Although teachers in the LPA+ group were provided with feedback material and support over the course of the school year, more time for professional development might have been helpful to get better acquainted with the material.

Examining differential effects, it was found that the effect on PCS in the LPA group was similar for students of different achievement levels, but in the LPA+ group, high-achieving students appeared to particularly benefit from the intervention. It is possible that the high-achieving students feel that their potential is seen and supported to a greater extent than in the regular classroom, especially as a result of the tasks designed for promoting highly talented mathematics students. The ready-made material obviously made it easier for the teachers to offer tasks to the high-achieving students in particular that showed a good fit with their abilities. Similarly, it is conceivable that the high-achieving students did particularly well with the learning setting, that is, learning in tandems and working independently with new materials.

## Limitations

This study makes a valuable contribution to a better understanding towards the motivational effects of FA practices. Still, some limitations should be acknowledged when interpreting the results. First, the rather small sample size at the class level ($N = 27$) did not allow for analyses of the half-longitudinal model in a multilevel path model. Estimation of the effects therefore occurred at the individual level, correcting for standard errors given the nested data structure. The analysis of mediation effects by half-longitudinal designs, moreover, implies further restrictions (Cole & Maxwell, 2003). While it is possible to test whether PCS functions as a partial mediator of group membership on intrinsic motivation, a direct test of a complete mediation is not possible. Further, the assumption of stationarity cannot be tested without at least three waves of data. If this assumption is violated, estimates of the indirect effect will be biased. A confirmation of the observed effects in a longitudinal design with at least three waves of data would therefore be promising. In terms of implementing assessment-based instruction, the supplementary tests need further improvement. At the beginning of the study, too many students mastered each domain of difficulty, so the cut-off criterion had to be strengthened again during the course of the study. This, in turn, may have led to many children working with tasks that were too challenging for them. However, when high-achieving children are correctly identified, the materials seem particularly conducive to motivation. Another limitation relates to the lack of control of the actual classroom activities. Even if the self-reports of the teachers allow for an approximate evaluation of treatment

fidelity, we do not know how exactly the individual components were implemented in the classroom and did not have the opportunity to monitor teachers' classroom activities more objectively, for instance, based on video observations or expert ratings. An approximation of these external judgments is provided by the observation of teachers in the LPA+ group in one feedback situation by two independent observers. However, not all teachers agreed to this. At the same time, the self-reports of some classes show a rather low implementation fidelity, especially for LPA+. Not all teachers implemented the core elements of feedback and adaptive instruction to a desirable extent. The study can therefore help to draw conclusions about the effect of the intervention when the components are implemented "as conducted" rather than "as intended" (Century & Cassata, 2016).

## Implications

Overall, the study enables us to better understand the underlying mechanisms of motivational effects of FA. Although only small effects were found, the study suggests that FA is a powerful approach to enhancing students' PCS in the classroom, which leads to an increase in intrinsic motivation. The implementation of FA practices in everyday school life should thus be further supported. In particular, the dosage and quality of FA should be increased to achieve stronger effects. The results give reason to reflect on the role of each feature of FA. It seems that LPA can play a key role in promoting adaptive teaching behavior and therefore making students feel supported in their competencies, which is why the dissemination of practicable tools for LPA should be given priority. The differential effect for high-achieving children in the LPA+ condition indicates that elaborated materials can support the implementation of adaptive instruction. However, to implement the concepts presented in this study on a broader level, some adjustments are still needed. Intensive professional development trainings to support teachers in implementing FA can play a crucial role here. The study provides clear evidence that students' perception of supportive teacher behavior has a motivating effect. Consequently, teachers need to be assisted in being able to continuously demonstrate this behavior in the classroom.

## Declarations

**Competing interests** The authors declare no competing interests.

# References

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

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

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

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 [Multi-criterial goal attainment in elementary school science education] [Doctoral dissertation, University of Münster].* https://miami.uni-muenster.de/Record/d767660e-c351-46ca-9828-7075ef173a3c. Accessed 16 Nov 2023

Bokhove, C., & Drijvers, P. (2012). Effects of a digital intervention on the development of algebraic expertise. *Computers & Education, 58*(1), 197–208. https://doi.org/10.1016/j.compedu.2011.08.010

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., Voss, R., & Walther, G. (2005). *IGLU. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente [IGLU. Scale manual for the documentation of the survey instruments].* Waxmann.

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction, 21*(1), 95–108. https://doi.org/10.1016/j.learninstruc.2009.11.004

Bureau, J. S., Howard, J. L., Chong, J. X. Y., & Guay, F. (2022). Pathways to student motivation: A meta-analysis of antecedents of autonomous and controlled motivations. *Review of Educational Research, 92*(1), 46–72. https://doi.org/10.3102/00346543211042426

Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what, how, why, where, and who. *Review of Research in Education, 40*(1), 169–215. https://doi.org/10.3102/0091732X16665332

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum. https://doi.org/10.4324/9780203771587

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 558–577. https://doi.org/10.1037/0021-843X.112.4.558

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press. https://doi.org/10.1007/978-1-4899-2271-7

Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., Lühken, A., Adl-Amini, K., Djakovic, S.-K., Mannel, S., Naumann, A., & Hardy, I. (2015). Impact of additional guidance in science education on primary students' conceptual understanding. *The Journal of Educational Research, 108*(5), 358–370. https://doi.org/10.1080/00220671.2014.899957

Department for Education. (2013). *The national curriculum in England. Key stages 1 and 2 framework document.* https://assets.publishing.service.gov.uk/media/5a81a9abe5274a2e8ab55319/PRIMARY_national_curriculum.pdf. Accessed 10 Mar 2023

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11*(3), 189–212. https://doi.org/10.1214/ss/1032280214

Drost, D. K., & Todorovich, J. R. (2017). Perceived competence and skill development in physical education: The effect of teacher feedback. *Journal of Sports Science, 5*(6), 291–304. https://doi.org/10.17265/2332-7839/2017.06.001

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(1), 128–141. https://doi.org/10.1207/S15328007SEM0801_7

Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education, 106*, 83–96. https://doi.org/10.1016/j.compedu.2016.12.001

Fend, H. (1981). *Theorie der Schule [School theory]*. Urban & Schwarzenberg.

Förster, N., & Souvignier, E. (2014). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction, 32*, 91–100. https://doi.org/10.1016/j.learninstruc.2014.02.002

Förster, N., & Souvignier, E. (2015). Effects of providing teachers with information about their students' reading progress. *School Psychology Review, 44*(1), 60–75. https://doi.org/10.17105/SPR44-1.60-75

Fuchs, D., & Fuchs, L. S. (2019). On the importance of moderator analysis in intervention research: An introduction to the special issue. *Exceptional Children, 85*(2), 126–128. https://doi.org/10.1177/0014402918811924

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology, 82*(3), 525–538. https://doi.org/10.1037/0022-0663.82.3.525

Goudas, M., Biddle, S., Fox, K., & Underwood, M. (1995). It ain't what you do, it's the way that you do it! Teaching style affects children's motivation in track and field lessons. *The Sport Psychologist, 9*(3), 254–264. https://doi.org/10.1123/tsp.9.3.254

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*(3), 175–213. https://doi.org/10.1023/A:1005614228250

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 [Adaptive learning opportunities in elementary school]. *Zeitschrift Für Pädagogik, 57*(6), 819–833. https://doi.org/10.25656/01:8783

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology, 34*(3), 269–290. https://doi.org/10.1080/01443410.2013.785384

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

Hebbecker, K., Förster, N., Forthmann, B., & Souvignier, E. (2022). Data-based decision-making in schools: Examining the process and effects of teacher support. *Journal of Educational Psychology, 114*(7), 1695–1721. https://doi.org/10.1037/edu0000530

Hebbecker, K., & Souvignier, E. (2018). Formatives assessment im Leseunterricht der Grundschule – Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts [Formative assessment in third-grade reading instruction—Implementation and effectiveness of a modular concept including prepared teaching-material]. *Zeitschriftf ür Erziehungswissenschaft, 21*(4), 735–765. https://doi.org/10.1007/s11618-018-0834-y

Helmke, A. (2003). *Unterrichtsqualität: Erfassen, Bewerten, Verbessern [Teaching quality: Assessment, evaluation, improvement]*. Kallmeyer.

Hondrich, A. L., Decristan, J., Hertel, S., & Klieme, E. (2018). Formative assessment and intrinsic motivation: The mediating role of perceived competence. *Zeitschrift für Erziehungswissenschaft, 21*(4), 717–734. https://doi.org/10.1007/s11618-018-0833-z

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*(3), 353–376. https://doi.org/10.1080/0969594X.2015.1049113

Howard, J. L., Bureau, J., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science, 16*(6), 1300–1323. https://doi.org/10.1177/1745691620966789

Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika, 75*(4), 649–674. https://doi.org/10.1007/s11336-010-9175-3

Jamshidian, M., Jalal, S., & Jansen, C. (2014). MissMech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *Journal of Statistical Software, 56*(6), 1–31. https://doi.org/10.18637/jss.v056.i06

Käpnick, F. (2016). *Mathe für kleine Asse: Empfehlungen zur Förderung mathematisch interessierter und begabter Kinder im ersten und zweiten Schuljahr [Math for little aces: Recommendations for*

*the support of mathematically interested and gifted children in the first and second school year].* Cornelsen.

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

Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017). *Formative assessment and elementary school student academic achievement: A review of the evidence. REL 2017-259.* Regional Educational Laboratory Central.

Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Educational Computing Research, 43*(4), 489–510. https://doi.org/10.2190/EC.43.4.d

Krajewski, K., Dix, S., & Schneider, W. (2004). *DEMAT 2+. Deutscher Mathematiktest für zweite Klassen [DEMAT 2+. German mathematics test for second grade].* Hogrefe.

Krijgsman, C., Mainhard, T., van Tartwijk, J., Borghouts, L., Vansteenkiste, M., Aelterman, N., & Haerens, L. (2019). Where to go and how to get there: Goal clarification, process feedback and students' need satisfaction and frustration from lesson to lesson. *Learning and Instruction, 61*, 1–11. https://doi.org/10.1016/j.learninstruc.2018.12.005

Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht [Multiple goals in mathematics instrcution].* Waxmann.

Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction, 17*(5), 494–509. https://doi.org/10.1016/j.learninstruc.2007.09.002

Lane, R., Parrila, R., Bower, R., Bull, R., Cavanagh, M., Forbes, A., Jones, T., Leaper, D., Khosronejad, M., Pellicano, L., Powell, S., Ryan, M., & Skrebneva, I. (2019). Literature review: Formative assessment evidence and practice. In *Australian Institute for Teaching and School Leadership (AITSL).* https://www.lpofai.edu.au/media/u5ahfia0/literature-review.pdf. Accessed 21 Apr 2023

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education, 33*(2), 124–140. https://doi.org/10.1080/08957347.2020.1732383

Leenknecht, M., Wijnia, L., Köhlen, M., Fryer, L., Rikers, R., & Loyens, S. (2020). Formative assessment as practice: The role of students' motivation. *Assessment & Evaluation in Higher Education, 46*(2), 236–255. https://doi.org/10.1080/02602938.2020.1765228

Lemos, M. S., & Veríssimo, L. (2014). The relationships between intrinsic motivation, extrinsic motivation, and achievement, along elementary school. *Procedia - Social and Behavioral Sciences, 112*, 930–938. https://doi.org/10.1016/j.sbspro.2014.01.1251

Linnenbrink-Garcia, L., Patall, E. A., & Pekrun, R. (2016). Adaptive motivation and emotion in education: Research and principles for instructional design. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 228–236. https://doi.org/10.1177/2372732216644450

Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung [Handling of missing data in psychological research: Problems and solutions]. *Psychologische Rundschau, 58*(2), 103–117. https://doi.org/10.1026/0033-3042.58.2.103

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist, 47*(2), 71–85. https://doi.org/10.1080/00461520.2012.667064

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*(1), 3–25. https://doi.org/10.1080/09585170701292109

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/10.1016/j.lindif.2012.04.001

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM, 50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Prast, E. J., van de Weijer-Bergsma, E., Kroesbergen, E. H., & van Luit, J. E. H. (2018). Differentiated instruction in primary mathematics: Effects of teacher professional development on student achievement. *Learning and Instruction, 54*, 22–34. https://doi.org/10.1016/j.learninstruc.2018.01.009

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology, 66*, 825–852. https://doi.org/10.1146/annurev-psych-010814-015258

Prenzel, M., Kristen, A., Dengler, P., Ettle, R., & Beer, T. (1996). Selbstbestimmt motiviertes und interessiertes Lernen in der kaufmännischen Erstausbildung [Self-determined motivated and interested learning in the initial commercial training]. In K. Beck & H. Heid (Eds.), *Lehr-Lern-Prozesse in der*

*kaufmännischen Erstausbildung: Wissenserwerb, Motivierungsgeschehen und Handlungskompeten-zen: Zeitschrift für Berufs- und Wirtschaftspädagogik* (13th ed., pp. 108–127). Steiner.

R Core Team. (2022). *R: A language and environment for statistical computing [Computer software]*. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/. Accessed 28 May 2023.

Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction, 27*, 63–73. https://doi.org/10.1016/j.learninstruc.2013.03.002

Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction: Grading and feedback in mathematics classrooms. *Zeitschrift Für Psychologie, 216*(2), 111–124. https://doi.org/10.1027/0044-3409.216.2.111

Rakoczy, K., Klieme, E., & Pauli, C. (2008). Die Bedeutung der wahrgenommenen Unterstützung motiva-tionsrelevanter Bedürfnisse und des Alltagsbezugs im Mathematikunterricht für die selbstbestimmte Motivation [The impact of the perceived support of basic psychological needs and of the perceived relevance of contents on students' self-determined motivation in mathematics instruction]. *Zeitschrift Für Pädagogische Psychologie, 22*(1), 25–35. https://doi.org/10.1024/1010-0652.22.1.25

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction, 60*, 154–165. https://doi.org/10.1016/j.learninstruc.2018.01.004

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y., Jorgensen, T. D., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., & Du, H. (2022). *Latent variable analysis: Package "lavaan"*. https://lavaan.ugent.be/. Accessed 28 Apr 2023.

Roy, A., Guay, F., & Valois, P. (2015). The big-fish–little-pond effect on academic self-concept: The mod-erating role of differentiated instruction and individual achievement. *Learning and Individual Differ-ences, 42*, 110–116. https://doi.org/10.1016/j.lindif.2015.07.009

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

Ryan, R. M., & Deci, E. L. (2002). An overview of self-determination theory: An organismic-dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3–33). The University of Rochester Press.

Schütze, B., Rakoczy, K., Klieme, E., Besser, M., & Leiss, D. (2017). Training effects on teachers' feedback practice: the mediating function of feedback knowledge and the moderating role of self-efficacy. *ZDM Mathematics Education, 49*(3), 475–489. https://doi.org/10.25656/01:17042

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

Souvignier, E., Förster, N., Hebbecker, K., & Schütze, B. (2021). Using digital data to support teaching practice – quop: An effective web-based approach to monitor student learning progress in reading and mathematics in entire classrooms. In A. Wilmers & S. Jornitz (Eds.), *International perspectives on school settings, education policy and digital strategies. A transatlantic discourse in education research* (pp. 283–298). Barbara Budrich. https://doi.org/10.2307/j.ctv1gbrzf4.20

Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determi-nation theory: Another look at the quality of academic motivation. *Educational Psychologist, 41*(1), 19–31. https://doi.org/10.1207/s15326985ep4101_4

Visscher, A. J. (2021). On the value of data-based decision making in education: The evidence from six intervention studies. *Studies in Educational Evaluation, 69*, 100899. https://doi.org/10.1016/j.stueduc.2020.100899

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning (pp. 53–82)*. Erlbaum. https://doi.org/10.4324/9781315086545-3

Wisniewski, B., Zierer, K., & Hattie, J. (2019). The power of feedback revisited: A meta-analysis of educa-tional feedback research. *Frontiers in Psychology, 10*, 3087. https://doi.org/10.3389/fpsyg.2019.03087

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., Tomita, M. K., & Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and

conceptual change. *Applied Measurement in Education, 21*(4), 335–359. https://doi.org/10.1080/08957 340802347845

**Larissa Aust.** University of Münster, Fliednerstraße 21, 48149 Münster, Germany. Email: larissa.aust@ uni-muenster.de

*Current themes of research:*

Effects of formative assessment on student outcomes. Implementation of formative assessment into school practice.

**Birgit Schütze (Harks).** University of Münster, Münster, Germany. Email: birgit.schuetze@uni-muenster. de
*Current themes of research:*

Formative assessment. Feedback.

*Most relevant publications in the field of Psychology of Education:*

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction,60*, 154-165. https://doi.org/10.1016/j.learninstruc.2018.01.004

Schütze, B., Rakoczy, K., Klieme, E., Besser, M. & Leiss, D. (2017). Training effects on teachers' feedback practice: the mediating function of feedback knowledge and the moderating role of self-efficacy. *ZDM Mathematics Education, 49*(3), 475–489. https://doi.org/10.25656/01:17042

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest, and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology, 34*(4), 269–290. https://doi.org/10.1080/01443410.2013.785384

Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction, 27*, 63–73. https://doi.org/10.1016/j.learninstruc.2013.03.002

**Jan Hochweber.** St. Gallen University of Teacher Education, St. Gallen, Switzerland. Email: Jan.Hochweber@phsg.ch.

*Current themes of research:*

Instructional validity of tests. Classroom assessment. Teaching quality. Multilevel modeling.

*Most relevant publications in the field of Psychology of Education:*

Kolovou, D., Naumann, A., Hochweber, J., & Praetorius, A.-K. (2021). Content-specificity of teachers' judgment accuracy regarding students' academic achievement. *Teaching and Teacher Education, 100*, 103298. doi.org/10.1016/j.tate.2021.103298

Hübner, N., Wagner, W., Hochweber, J., Neumann, M., & Nagengast, B. (2020). Comparing apples and oranges: Curricular intensification reforms can change the meaning of students' grades. *Journal of Educational Psychology, 112*, 204-220. doi.org/10.1037/edu0000351

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction, 60*, 154-165. doi.org/10.1016/j.learninstruc.2018.01.004

Hochweber, J., & Vieluf, S. (2018). Gender differences in reading achievement and enjoyment of reading: The role of perceived teaching quality. *The Journal of Educational Research, 111*, 268-283. doi.org/ 10.1080/00220671.2016.1253536

**Elmar Souvignier.** University of Münster, Münster, Germany. Email: elmar.souvignier@uni-muenster.de.

*Current themes of research:*

Learning progress assessment. Implementation research in school context. Promotion of reading and math skills

*Most relevant publications in the field of Psychology of Education:*

Hebbecker, K., Förster, N., Forthmann, B., & Souvignier, E. (2022). Data-based decision-making in schools: Examining the process and effects of teacher support. *Journal of Educational Psychology, 114*(7), 1695-1721. https://doi.org/10.1037/edu0000530

Peters, M. T., Hebbecker, K. & Souvignier, E. (2021). Effects of providing teachers with tools for implementing assessment-based differentiated reading instruction in second grade. *Assessment for Effective Intervention,* doi.org/10.1177/15345084211014926

Hebbecker, K., Förster, N. & Souvignier, E. (2019). Reciprocal effects between reading achievement and intrinsic and extrinsic reading motivation. *Scientific Studies of Reading, 23*, 419-436. https://doi.org/10.1080/10888438.2019.1598413

Förster, N., Kawohl, E. & Souvignier, E. (2018). Short- and long-term effects of assessment-based differentiated reading instruction in general education on reading fluency and reading comprehension. *Learning and Instruction, 56*, 98-109. https://doi.org/10.1016/j.learninstruc.2018.04.009