Schütze, Birgit; Rakoczy, Katrin; Klieme, Eckhard; Besser, Michael; Leiss, Dominik

Training effects on teachers' feedback practice. The mediating function of feedback knowledge and the moderating role of self-efficacy

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
ZDM : the international journal on mathematics education 49 (2017) 3, S. 475-489

Training effects on teachers' feedback practice: The mediating function of feedback knowledge and the moderating role of self-efficacy

Birgit Schütze

University of Münster

Katrin Rakoczy and Eckhard Klieme

German Institute for International Educational Research

Michael Besser

University of Education Freiburg

Dominik Leiss

Leuphana University Lüneburg

Author Note

Birgit Schütze, Institute of Psychology in Education, University of Münster, Germany; Katrin Rakoczy, Department of Educational Quality and Evaluation, German Institute for International Educational Research, Germany; Eckhard Klieme, Department of Educational Quality and Evaluation, German Institute for International Educational Research, Germany; Michael Besser, Faculty of Mathematics, Science, and Technology, University of Education Freiburg, Germany; Dominik Leiss, Institute for Didactics of Mathematics, Leuphana University of Lüneburg, Germany.

Correspondence concerning this article should be addressed to Birgit Schütze, Institute of Psychology in Education, University of Münster, Fliednerstraße 21, D-48149 Münster, Germany.

Phone: +49 (0) 251 83-34167

E-mail: birgit.schuetze@uni-muenster.de

Abstract

Formative assessment has been identified as a promising intervention to support students' learning. How to successfully implement this means of assessment, however, is still an open issue. This study contributes to the implementation of formative assessment by analyzing the impact of a training measure on teachers' formative feedback practice, with a special focus on mediating and moderating variables. Research questions are as follows: (1) Is there an indirect training effect on teachers' instructional feedback practice via (a) teachers' declarative feedback knowledge and (b) the ability to generate feedback in a test situation? (2) Is this indirect effect moderated by teachers' self-efficacy?

A total of 67 secondary education mathematics teachers participated in the study, taking part in professional development either on formative assessment and feedback (*PD-FA*) or on mathematical modelling and problem solving (*PD-PM*). Training was provided in two sessions (T1 and T2; each lasting three days) with ten weeks in between T1 and T2. Teachers' self-efficacy regarding feedback was measured before T1 with a questionnaire. Declarative feedback knowledge and the ability to apply this knowledge were tested after T2. Teachers' instructional feedback practice was assessed with a student questionnaire (before T1 and 4-6 weeks after T2). Path analyses show that (1) there is no indirect training effect (*PD-FA vs. PD-PM*) on the development of teachers' feedback practices in mathematics instruction; but an indirect effect on the ability to generate feedback in a test situation via teachers' declarative feedback knowledge. Teachers participating in PD-FA show a higher level of declarative feedback knowledge than teachers in the PD-PM condition. Declarative feedback

knowledge in turn is positively related to the ability to generate feedback in a test situation. (2) This indirect effect is moderated by teachers' self-efficacy. Teachers with a high level of self-efficacy are better able to use their knowledge to generate feedback in a test situation than teachers with a low level of self-efficacy.
Keywords: professional development; formative assessment; teacher knowledge; mathematics instruction; self-efficacy

Training effects on teachers' feedback practice: The mediating function of feedback knowledge and the moderating role of self-efficacy

1 Introduction

Formative assessment has been identified as a promising intervention to support students' learning (e.g., Black and Wiliam 1998b; Hattie 2009; see also Dunn and Mulvenon 2009; Kingston and Nash 2011). Defined as "…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 and Wiliam 1998a, pp. 7-8; see also Black and Wiliam 2009), formative assessment has been gaining significance as "next best hope" (Cizek 2010, p. 2) and "powerful tool" (Wylie et al. 2012, p. 121) in educational practice, research, and policy.

In the domain of mathematics education, formative assessment has taken on an increasingly important role since the 1990s (Brown and Hodgen 2010). In 1994, Ruthven examined 50 contributions of the Second International Congress on Mathematical Instruction (ICMI) and identified "increasing the integration of processes of teaching, learning and assessment" (p. 433) as a central trend. Since then more and more researchers have dealt with the impact and design of formative assessment in mathematics instruction (e.g., Fuchs et al. 1999; Kim and Lehrer 2015;Koedinger et al. 2010; Lehmann and Seeber 2005; Phelan et al.

2011; Rakoczy et al. in press; Thompson and Kaur 2011; Toh et al. 2011; Wiliam et al. 2004; Ysseldyke and Bolt 2007; for an overview see, e.g., Harks 2013 and Suurtamm et al. 2016).

The implementation of formative assessment, still, is very demanding, time-consuming, and frequently rather unsystematic (e.g., Black and Wiliam 1998a; Yin et al. 2008). Evidence-based hints on how to implement formative assessment in everyday instruction are scarce (Shavelson 2008). In their review on assessment practices in mathematics, Brown and Hodgen (2010) conclude, "While formative assessment, in general, has been shown to be a particularly effective pedagogic strategy, it is poorly described in ways that can be implemented in practice" (p. 282).

One reason for the scarce information on how to implement formative assessment might be the broadness and complexity of the construct (see, e.g., the definition by Black and Wiliam 1998a, above) thus making it difficult to deduce specific instructions for implementation. To better understand how to transfer formative assessment into practice, it might be useful to focus on specific aspects or key strategies first, and systematically investigate their implementation.

1.1 Feedback as a key strategy of formative assessment

Wiliam and Thompson (2008) describe five key strategies of formative assessment (see also Bennett 2011): (a) clarifying, sharing, and understanding learning intentions and criteria for success; (b) engineering effective classroom discussions, questions, and tasks that elicit evidence of learning; (c) providing feedback that moves learners forward; (d) activating students as instructional resources for one another; and (e) activating students as the owners of their own learning. In particular, the provision of feedback has been emphasized as a central element in formative assessment (e.g., Hattie 2003; Sadler 1998) and has been shown to be a powerful intervention if provided carefully (e.g., Hattie 2009; Hattie and Timperley 2007). If it is to benefit learning, feedback should answer three questions–"Where am I going?", "How am I going?", and "Where to next?"–and focus on task performance, understanding processes, or regulatory metacognitive processes (e.g., Hattie and Timperley 2007). Feedback should furthermore establish unstable attributions in the case of failure (e.g., Dresel and Haugwitz 2008) and compare a student's performance to a predefined content-related criterion (e.g., Krampen 1987) or the individual prior performance (e.g., Mischo and Rheinberg 1995; Shih and Alexander 2000; Thompson 2007), but not to the performance of others (for an overview of the impact of different feedback characteristics see, e.g., Hattie and Timperley 2007; Kluger and DeNisi 1996; Mason and Bruning 2001; Shute 2008). In comparison with other domains, in mathematics instruction, a high percentage of class time is spent working on tasks and problems (Hiebert et al. 2003), the identification of errors is relatively unambiguous, and even complex solution processes frequently proceed comparatively schematically. Thus, from a theoretical point of view, criterial elaborated feedback with specific hints regarding what went right, what went wrong, and how to improve should prove comparatively beneficial in this domain.

The present study focuses on feedback as a key element of formative assessment and investigates processes and conditions underlying a successful implementation of formative feedback practice in everyday mathematics instruction.

1.2 The implementation of formative feedback through teacher training

The implementation of feedback in classrooms can be guided in different ways (e.g., Pellegrino et al. 2001), teacher training being one promising variant (e.g., Desimone 2009). Research on the effectiveness of professional development suggests that training be structured into input-, field trial-, and reflection phases (Lipowsky and Rzejak 2012) and identifies (a) content focus (focus on subject matter content), (b) active learning (e.g., discussions and observations), (c) coherence (consistency of training contents with school, district, and state reforms), (d) duration (sufficient time, e.g., spread over one semester, at least 20 hours of contact time), and (e) collective participation (promoting interaction and discourse between teachers) as essential training characteristics (e.g., Desimone 2009). The sparse literature on training design in the area of formative assessment (Heritage 2007; Hill et al. 2010; Schneider and Randel 2010) further recommends that *input* be provided on (formative) assessment in general (e.g., on types of formative assessment) as well as on key strategies of formative assessment, namely diagnostics (e.g., on task development) and feedback (e.g., on feedback characteristics). Moreover, content-specific input is desirable (e.g., on subject-specific competencies). Knowledge should be acquired by informing teachers, in discussions and active learning phases (hands-on-practices, e.g., by analyzing videos, student solutions, or feedback examples). The acquired knowledge should be road-tested in everyday instruction (*field trial*). An exchange should be facilitated among teachers as well as between teachers and researchers, for example by means of group meetings, online-chats, or by attendance at school (*reflection phase;* e.g., James and Pedder 2006; Schneider and Randel 2010). Based on these suggestions, teacher training on formative assessment focusing on feedback (and diagnostics as a prerequisite for providing feedback) was developed and evaluated for this study.

1.3 Indirect effects of formative assessment training

The effectiveness of formative assessment interventions (including formative assessment training) has typically been tested by investigating its impact on students' performance (e.g., Schneider and Randel 2010). Performance, however, is a rather distal success criterion and little is known about the mechanisms that mediate training effects on students' performance. If training on formative assessment did not lead to improved student performance, previous research in most cases left unclear at which point the intervention failed. Desimone (2009) proposes a model for studying the impact of professional development that focuses on more proximal variables which might mediate the influence from teacher training on students' achievement. According to Desimone, (1) professional development positively affects teachers' knowledge and skills and influences their attitudes and beliefs. (2) A change in knowledge, skills, attitudes, and beliefs is in turn assumed to positively influence the quality of instruction. (3) Instructional improvement, finally, should lead to increased student performance. Each step in Desimone's (2009) model has been supported by empirical evidence: For example, professional development has been shown to positively affect teachers' professional knowledge (e.g., Besser et al. 2015, based on the same sample as the present study; Diamond et al. 2014). Previous research furthermore found that teacher knowledge is positively associated with instructional quality (e.g., Kunter et al. 2013; Voss et al. 2014) which in turn is related to students' achievement (e.g., Fauth et al. 2014). Many studies, however, have focused on single steps in the model (e.g., on the relation between teacher knowledge and instructional quality) and do not deal with teacher training in particular. To our knowledge, the indirect effect from teacher training on instructional quality (via teacher knowledge) or on students' achievement (via teacher knowledge and via instructional quality) has not been studied empirically, at least not in the field of formative assessment.

More work is furthermore needed to specifically investigate how the knowledge acquired in training enables a change in instructional practice (see Step 2 in the model of Desimone 2009). We know from learning psychology that people have difficulties in applying knowledge in everyday problem situations (Renkl 1996). This phenomenon is called "inert knowledge" (Whitehead 1929) and is also debated in the context of teacher training. Concrete mechanisms underlying this gap between knowing and action (Mandl and Gerstenmaier 2000) are speculative. When a teacher fails to implement training contents, he/she might have difficulties in applying training knowledge in general, regardless of the specific situation. On the other hand, the teacher might be able to apply the knowledge per se, but struggle to do so in complex everyday teaching situations. To gain a better understanding of such transfer processes, in our study, we distinguish between (a) declarative knowledge of feedback, (b) application of feedback knowledge (i.e., generating feedback) in a test situation, and (c) application of feedback knowledge in everyday instruction (i.e., instructional feedback practice). Based on this differentiation, we study the indirect effect of teacher training on instructional feedback practice (i.e., the sequence of effects by which teacher training exerts its effect on instructional feedback practice) and investigate declarative feedback knowledge and the generation of feedback in a test situation as intervening variables.

1.4 The moderation of training effects

There is a strong consensus that effects of teacher training are moderated by (i.e., vary as a function of) (a) student characteristics; (b) individual teacher characteristics; (c) contextual factors at the classroom, school, and district levels; and (d) policy conditions (for an overview see Desimone 2009). Mostly the same moderators are discussed in the literature on formative assessment (e.g., Black and Wiliam 1998a; Heritage 2007). Obviously, the success of training interventions should strongly depend on those who are supposed to put training contents into practice. Thus, teacher motivation, and specifically teachers' self-efficacy, has been stressed as a central moderating variable (e.g., Gegenfurtner 2011). Self-efficacy can be defined as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura 1986, p. 391). Self-efficacy is positively associated with effort and persistence (Multon et al. 1991; Schunk 1995) and positively related to the efficient use of learning strategies (Pintrich and DeGroot 1990; Zimmerman 2000), thus teachers with a high level of self-efficacy should learn more than others and might be more likely to put this knowledge into action. Empirical studies indeed show a positive relationship between trainees' self-efficacy and transfer of training content (e.g., Pugh and Bergin 2006; Simosi 2012; Stein and Wang 1988; for an overview see Gegenfurther 2011). Many of these studies, however, have not specifically focused on teacher training (for an exception see Stein and Wang 1988), and none of them has investigated the moderation of indirect training effects.

1.5 Present study: research questions and hypotheses

The aim of the present study is to contribute to a better understanding of the impact of teacher training on teachers' instructional feedback practice in mathematics instruction. In particular, we focus on the intervening role of (a) teachers' feedback knowledge and (b) the ability to generate feedback in a test situation as well as the moderating function of teachers' self-efficacy. Our research questions are as follows: (1) Is there an indirect training effect on teachers' instructional feedback practice via (a) teachers' declarative feedback knowledge and (b) the ability to generate feedback in a test situation? (2) Is this indirect effect moderated by teachers' self-efficacy? Regarding the first research question, we hypothesized that teachers participating in a training measure on formative assessment know more about feedback than teachers who take part in training that does not deal with assessment (*Hypothesis 1a,* Step 1 of the indirect effect). We further expected that feedback knowledge relates positively to the ability to generate feedback in a test situation (*Hypothesis 1b,* Step 2 of the indirect effect), which in turn should be positively associated with teachers' instructional feedback practice (*Hypothesis 1c,* Step 3 of the indirect effect). Correspondingly, we hypothesized that there would be an indirect effect on the generation of feedback in a test situation via feedback knowledge (*Hypothesis 1d*) and an indirect effect on instructional feedback practice via feedback knowledge and feedback generation in a test situation (*Hypothesis 1e*). With regard to the second research question, we assume the indirect effect to be moderated by teachers' self-efficacy (*Hypothesis 2;* see Figure 3).

2 Methods

The present study was based on data from the project "*Conditions and Consequences of Classroom Assessment*" (Co²CA; for an overview see Rakoczy et al., in press).

2.1 Participants

In this study, 67 mathematics teachers (65.67% female; 79.1% younger than 46 years) in Bavaria, Germany, took part. Each teacher participated with one of their mathematics classes (in total: N = 1.528 students). Participation in the study was voluntary.

2.2 Design

We conducted a quasi-experimental field trial with two conditions (professional development on formative assessment, PD-FA; professional development on problem solving and modeling competence, PD-PM) and three measurement points. Each teacher participated in one of the two professional developments. Depending on their own individual interest, teachers were assigned to the respective course: $n = 30$ teachers enrolled in the course on formative assessment, $n = 37$ teachers chose the course on problem solving and modeling. Teachers' instructional feedback practice was assessed as dependent variable, teachers' analyzed as intervening variables mediating between professional development and instructional feedback practice. Teachers' self-efficacy was studied as a moderating variable.

As teachers were not randomly assigned to the two courses, pretreatment differences between conditions regarding teacher variables were analyzed. Teachers in the two conditions did not differ with regard to their vocational education (second state examination, teaching mathematics outside of one's subject area, number of professional development courses attended in the last three years) or their teaching experience. They furthermore did not vary with regard to intervention-related variables assessed before the treatment (instructional feedback practice, formative assessment practice, formative assessment beliefs, interest in feedback), with the exception of self-efficacy regarding feedback (PD-FA: $M = 62.83$, $SD = 12.21$; PD-PM: $M = 69.16$, $SD = 11.95$; $t(64) = 2.12$; $p = .04$[1]; Cohen's $d = -.53$). Teachers' self-efficacy was controlled for in the model.

2.3 Procedure

Data collection took place from February to May 2013 and from September to December 2013, respectively. Each professional development consisted of two three-day training sessions spanning on average 14 hours each. After the first training session, teachers were expected to implement central ideas of the training in one of their mathematics classes.

[1] The significance threshold is adjusted to p = .05/5 = .01 when applying a Bonferroni-correction.

Following ten weeks of implementation, a second training session took place. In both conditions, training sessions were conducted by the same two researchers. Teachers' pre-training self-efficacy was assessed with a teacher questionnaire and teachers' pre-training feedback practice with a student questionnaire before the first training. Teachers' knowledge about feedback and their ability to generate feedback in a test situation were tested after the second training. Finally, teachers' post-training feedback practice was measured with a student questionnaire four to six weeks after the second training session (see Figure 1).

2.4 Professional development

Following suggestions on how to design teacher training described in section 1.2, both courses comprised phases of (a) input, (b) field trial, and (c) reflection. Input was provided in the first and in the second training session. The *input phase* was characterized by direct instruction in combination with content-focused active hands-on learning phases in which teachers worked collaboratively on authentic material (such as videos or tasks from textbooks). To promote a comprehensive examination of training content, homework was given after each session (e.g., development of mathematics tasks). Teachers were furthermore expected to road-test the newly acquired content in the classroom after the first and the second session (*field-trial phase;* e.g., administering the self-developed tasks). During the field-trial, teachers exchanged experiences and materials via an online learning platform. They reflected the implementation of their interventions in the last training session (*reflection phase*).

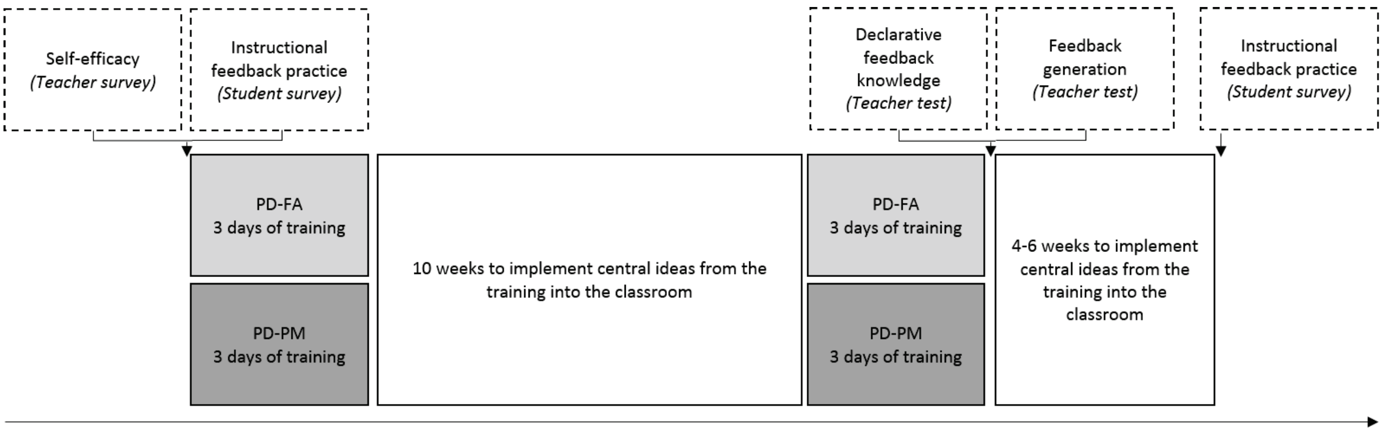


Figure 1. Design and procedure. PD-FA = professional development on formative assessment; PD-PM: professional development on problem solving and modelling competence.

2.4.1 Professional development on formative assessment (PD-FA).

The professional development on formative assessment aimed at helping teachers to implement formative assessment (esp. the diagnostic and the feedback component) in everyday mathematics instruction. To this end, pedagogical and psychological input was provided on central ideas of formative assessment (e.g., aims, types), performance diagnostics (e.g., psychometric quality of tests), and performance feedback (esp. beneficial feedback characteristics). Teachers additionally received content-specific, didactical input on how to implement formative assessment in their mathematics instruction. We specifically focused on instruction that deals with mathematical modeling. Input was provided on central ideas of mathematical modeling (e.g., theoretical considerations based on the German national standards, Kultusministerkonferenz 2003), how to use modeling tasks for diagnostic purposes (e.g., typical steps and mistakes in solving modeling tasks, the modeling cycle as a diagnostic aid), and how to provide feedback for such tasks. Furthermore, lesson plans and corresponding assessment materials were developed and discussed.

2.4.2 Professional development on problem solving and modeling competence (PD-PM).

The aim of the professional development on problem solving and modeling competence was to provide basic, general didactic information on how to implement competence oriented mathematics instruction (focusing on mathematical problem solving and modeling competence). For this purpose, central ideas on problem solving and modeling competence (e.g., theoretical considerations based on the German national standards, Kultusministerkonferenz 2003) were imparted. Lesson plans as well as problem solving and modeling tasks were analyzed, developed, and discussed. In contrast to PD-FA, the diagnostic aspect of the tasks was not explicitly addressed.

2.5 Measures

2.5.1 Declarative knowledge about feedback.

Declarative knowledge about feedback was assessed with 11 test items consisting of multiple-choice, short answer, and open answer items. Items were self-developed and measure declarative knowledge about central feedback characteristics (e.g., degree of elaboration, feedback level, feedback's frame of reference). An example of an item was: "Which three central questions should be answered when providing feedback?" Short answer and open answer items were double-coded by two trained graduate students using a standardized coding guideline. Average Cohen's kappa was $\kappa = .78$ across all items with a range from $\kappa = .56$ to $\kappa = 1.00$. Internal consistency of the scale was Cronbach's $\alpha = .84$.

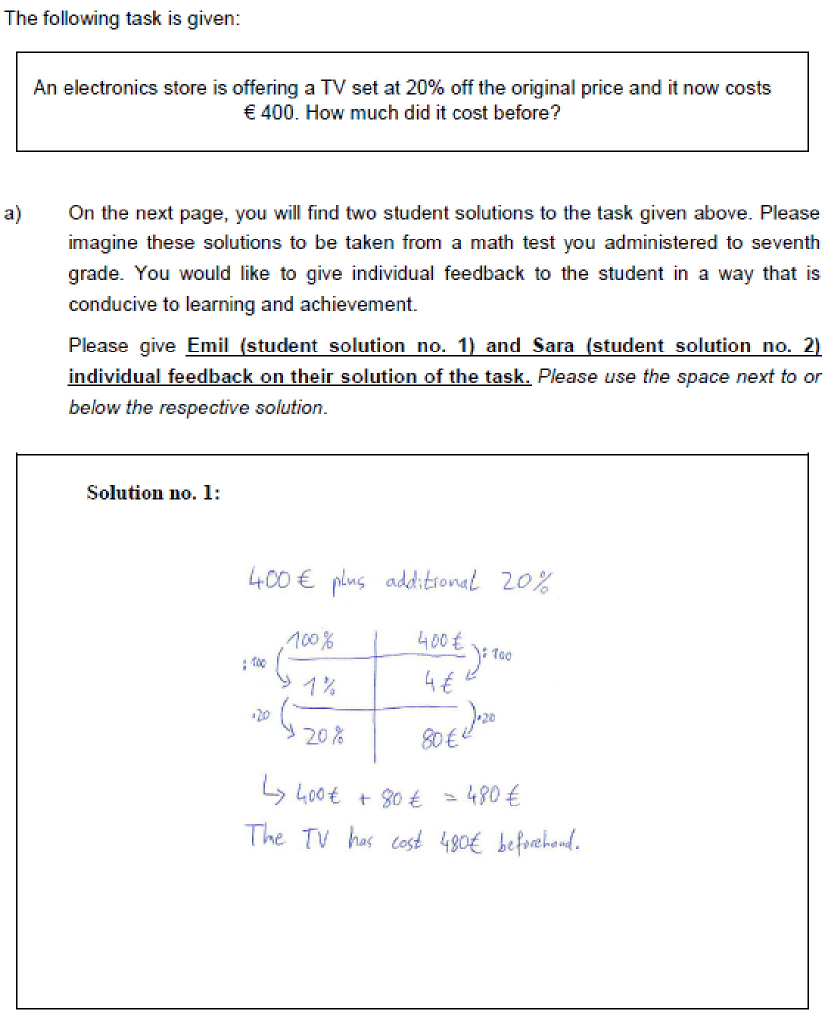
2.5.2 Application of feedback knowledge.

Application of feedback knowledge was measured with three self-developed test items that required teachers to generate feedback on students' solutions of a mathematics task on percentage calculation (for an example see Figure 2).

Teachers' feedback was coded by two trained graduate students using a standardized coding guideline. Raters coded whether the teacher feedback exhibited characteristics proven to be beneficial for learning. It was coded whether the feedback answered the three central feedback questions ("How am I going?", "Where am I going?", and "Where to next?"), whether the feedback was provided on the process-, task- or self-regulation level, but not on the self-level, and whether an individual frame of reference, criterial frame of reference, but not a social frame of reference was used. Raters furthermore coded whether errors were marked or corrected and whether the language used was comprehensible and specific. Teacher feedback was double-coded (with overall 17 coding categories per person). Average Cohen's kappa was κ = .90 across all coding categories with a range from κ = .74 to κ = 1.00. Cronbach's α was .93.

2.5.3 Perceived feedback practice in instruction.

Students' perception of central aspects of teachers' feedback practice was reported on a five-item scale (Klimczak et al. 2012) ranging from 1 (*does not apply*) to 4 (*fully applies*). Students indicated to which extent their teacher provided elaborated, criterial feedback answering the three feedback questions (How am I going? Where am I going? Where to next?). An item example was "In mathematics instruction, I get feedback on how I can improve in areas I am not good in". Internal consistency of the scale was Cronbach's α .76 in the pre-questionnaire and .86 in the postquestionnaire. Individual student perceptions of feedback practice were aggregated on class level (for classroom level aggregation of climate variables see Marsh et al. 2012). The intraclass correlation ICC(1) (proportion of total variance explained by between-class differences) shows that perceived feedback practice varies between classrooms, $ICC(1)_{Pre-/Post-questionnaire} = .19$. The ICC(2) (estimated by applying the Spearman-Brown formula to the ICC(1), considering the average class size) indicates a sufficient degree of reliability of the class-mean ratings, $ICC(2)_{Prä} = .84$/$ICC(2)_{Post} = .77$ (Bliese 2000; Lüdtke et al. 2009; Snijders and Bosker 1999).


The following task is given:

An electronics store is offering a TV set at 20% off the original price and it now costs € 400. How much did it cost before?

a) On the next page, you will find two student solutions to the task given above. Please imagine these solutions to be taken from a math test you administered to seventh grade. You would like to give individual feedback to the student in a way that is conducive to learning and achievement.

Please give **Emil (student solution no. 1) and Sara (student solution no. 2) individual feedback on their solution of the task.** *Please use the space next to or below the respective solution.*

Solution no. 1:

400 € plus additional 20%

100 %	400 €
1 %	4 €
20 %	80 €

:100 ·20

↳ 400 € + 80 € = 480 €

The TV has cost 480€ beforehand.

Figure 2. Item from the feedback generation test. The introduction refers to Item 1 (i.e., student solution no. 1, depicted) and Item 2 (i.e., student solution no. 2, not depicted).

2.5.4 Self-efficacy.

Teachers' self-efficacy regarding feedback was assessed with a self-developed scale asking to what degree teachers felt able to provide students with feedback. An example item was "I can provide students with feedback on their strengths and weaknesses," with a certainty rating scale ranging from 0 (*I am not able at all*) to 100 (*I am able with high certainty*). The estimate of reliability of the five-item scale was Cronbach's $\alpha = .67$.

2.6 Data analyses

To address our research questions, the manifest path model depicted in Figure 3 (Model 1) was applied to the data. In Model 1, training was entered as a dummy-coded predictor variable (1 = PD-FA, 0 = PD-PM). Pre-training instructional feedback practice was analyzed as *z*-standardized covariate and post-training instructional feedback practice as *z*-standardized criterion variable. Declarative feedback knowledge and the ability to generate feedback in a test situation were entered as *z*-standardized intervening variables. *Z*-standardized self-efficacy and the corresponding interaction terms (self-efficacy × training, self-efficacy × declarative feedback knowledge, self-efficacy × feedback generation) were included to test moderation.

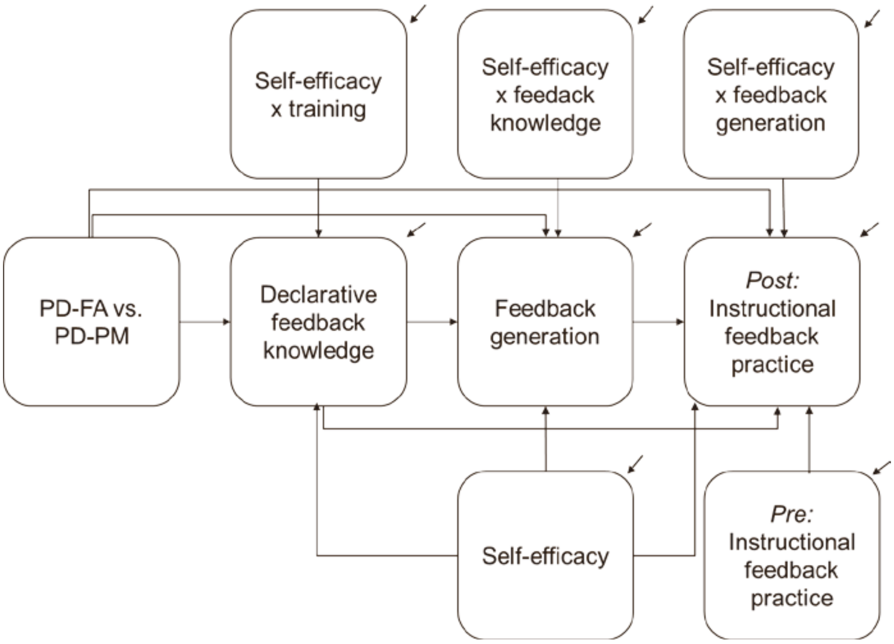


Figure 3. Path diagram for analyzing moderated indirect effects from training (PD-FA vs. PD-PM) on instructional feedback practice (Model 1).

To test our hypotheses, we examined the following effects:

- *Hypothesis 1a*: direct effect of training on declarative feedback knowledge
- *Hypothesis 1b:* direct effect of declarative feedback knowledge on the ability to generate feedback in a test situation
- *Hypothesis 1c*: direct effect from feedback generation on instructional feedback practice
- *Hypothesis 1d*: indirect training effect on feedback generation via declarative feedback knowledge (see, e.g., MacKinnon 2008)

- *Hypothesis 1e*: indirect training effect on instructional feedback practice via feedback knowledge and feedback generation
- *Hypothesis 2:* effects from interaction terms (e.g., self-efficacy × training) on intervening variables and the criterion variable (see, e.g., Edwards and Lambert 2007).

Since we analyzed indirect effects (which are products of direct paths) and since the distribution of product terms is only asymptotically normal (e.g., MacKinnon et al. 2004; Shrout and Bolger 2002), we computed standard errors of indirect effects through a bootstrapping procedure. We used 5000 bootstrap samples and estimated bias-based 95% confidence intervals (CIs) for the effects (e.g., MacKinnon et al. 2004). The null hypothesis that no indirect effect is present was rejected if zero was not included in the CI. To gain a more comprehensive understanding of the training effect, we additionally exploratively analyzed direct and total effects (see Figure 3). Full information maximum likelihood estimation (FIML) was used. All analyses were conducted with MPlus 6.1 (Muthén and Muthén 1998–2007).

To confirm the findings from Model 1, a further model (Model 2) was estimated: Model 2 was analyzed due to the high percentage of missing data for instructional feedback practice (4.48% in the pre survey, 38.8% in the post survey, at a sample size of $N = 67$; for all other variables the percentage of missing data was lower than 1.5%.). The high percentage of missing data in the post survey resulted from a drop out of 26 classes. The teachers of these classes did not send the completed student questionnaires per mail. As this drop out is neither related to the training condition (50% of missing classes belong to the PD-PM-condition, p [Fischer] = .615, *Phi* = -.08) nor to teachers' pretraining feedback practice ($r = .15$, $p = .24$), we do not assume missingness to be related to posttraining feedback practice and thus expect the data to be missing at random (see e.g., Lüdtke et al. 2007, Schlomer et al. 2010).

FIML including extra, auxiliary variables (correlates of the missingness or correlates of the variables that have missing values, e.g., Collins et al. 2001) has been identified as a method applicable to dealing with (large amounts of) missing data (e.g., Graham 2003; Lüdtke et al. 2007; Schlomer et al. 2010). As the bootstrapping-option (applied in Model 1) cannot be combined with the inclusion of auxiliary variables in MPlus, Model 2 was added. Model 2 is based on Model 1 (without bootstrapping), supplemented by the following auxiliary variables: personal relevance of training (pre survey), formative assessment practice (pre and intermediate survey[2]), available time for dealing with training contents (intermediate survey), avoidance intention regarding training (intermediate survey). All of these variables show a positive correlation with instructional feedback practice or its missingness.

3 Results

Sample sizes, means, standard deviations, and correlations of all variables in Model 1 are set out in Table 1. Model 1 shows a good fit to the data (χ^2 = 8.12, *df* = 8, *p* = .42; root mean square error of approximation, RMSEA = .02, 90% CI [.00, .15], *p* = .54; comparative fit index, CFI = 1.00; standardized root mean square residual, SRMR = .04). The direct and interaction effects estimated in Model 1 are set out in Table 2. As hypothesized, the results reveal that PD-FA was associated with a significantly higher level of declarative feedback knowledge than PD-PM, direct effect: β = .52, 95% CI [.07, .95] (*Hypothesis 1a*). Declarative feedback knowledge was as expected significantly related to the ability to generate feedback in a test situation, direct effect: β = .49, 95% CI [.31, .65] (*Hypothesis 1b*). Contrary to expectations, there was no direct effect from the ability to apply feedback in a test situation to instructional feedback practice, direct effect: β = .13, 95% CI [-.18, .40] (*Hypothesis 1c*).

---

[2] The intermediate survey was conducted at the second training session. It is not depicted in Figure 1 or described in the section “Procedure” as the variables assessed at training session 2 are of no direct relevance for answering our research questions. However, as these variables were correlated with instructional feedback practice and its missingness, respectively, they were included as auxiliary variables.

In line with the results on the corresponding direct effects, Table 3 shows that following expectation, in comparison with PD-PM, PD-FA had a significant positive indirect effect on the ability to generate feedback in a test situation (*Hypothesis 1d*) via declarative feedback knowledge, β = .25, 95% CI [.05, .54]. Contrary to our hypothesis, however, there was no significant indirect effect on instructional feedback practice via declarative feedback knowledge and via the ability to generate feedback in a test situation, β = .03, 95% CI [-.02, .18] (*Hypothesis 1e*).

There was, in line with our expectation, a significant moderation of the indirect effect by teachers' self-efficacy (*Hypothesis 2*; see Table 2): The direct effect from declarative feedback knowledge on the ability to generate feedback was positively moderated by teachers' self-efficacy (self-efficacy × declarative feedback knowledge on feedback generation: β = .20, 95% CI [.03, .35]). Direct effects from training on declarative feedback knowledge (self-efficacy × training on declarative feedback knowledge: β = -.19, 95% CI [-.62, .24]) and from the ability to generate feedback on instructional feedback practice (self-efficacy × feedback generation on instructional feedback practice: β = -.06, 95% CI [-.29, .24]), by contrast, did not depend on teachers' level of self-efficacy.

In addition, Table 2 shows a negative direct effect from training on feedback generation, β = -.58, 95% CI [-.91, -.22], no direct effect from training on instructional feedback practice, β = .25, 95% CI [-.30, .79], and no direct effect from feedback knowledge on instructional feedback practice, β = -.18, 95% CI [-.40, .04]. Moreover, we found no total training effect on feedback generation (β = -.33, 95% CI [-.72, .10]) or on instructional feedback practice (β = .11, 95% CI [-.36, .50], see Table 3).

The findings from Model 2 (including auxiliary variables) correspond to the results of Model 1, with the exception that the indirect training effect on feedback generation (via declarative feedback knowledge, β = .25, SE = 0.13, p = .06) and the moderation effect (self-efficacy × declarative feedback knowledge on feedback generation: β = .20, *SE* = 0.11, *p* = .06) are marginally significant only (see Table A1-A2).

Table 1

Correlations, means, and standard deviations of all variables in the study

		Total sample									PD-FA			PD-PM		
Variable		*n*	1	2	3	4	5	6	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
1.	Training[a]	67	–						0.45	0.50	30	1.00	0.00	37	0.00	0.00
2.	Declarative feedback knowledge	67	.30*	–					5.01	3.01	30	6.00	3.07	37	4.20	2.74
3.	Ability to generate feedback in a test situation	67	-.12	.38**	–				6.18	5.12	30	5.53	5.12	37	6.70	5.12
4.	Pre-training instructional feedback practice	64	-.06	.12	.04	–			2.94	0.28	30	2.92	0.34	34	2.95	0.22
5.	Post-training instructional feedback practice	41	.03	-.11	.03	.76**	–		2.69	0.31	17	2.70	0.37	24	2.69	0.28
6.	Self-efficacy	66	-.26*	-.15	.00	.05	.00	–	66.28	12.39	30	62.83	12.21	36	69.16	11.95

Note. Means and standard deviations are given for the total sample and separately for both training conditions.

[a] Dummy-coded (PD-FA = 1, PD-PM = 0).

* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table 2

Direct and interaction effects in Model 1

Variable	Declarative feedback knowledge			Feedback generation			Instructional feedback practice (post)		
		95% CI			95% CI			95% CI	
	β	*LL*	*UL*	β	*LL*	*UL*	β	*LL*	*UL*
Predictor									
Training (PD-FA vs. PD-PM)[a]	.52*	.07	.95	-.58*	-.91	-.22	.25	-.30	.79
Intervening variables									
Declarative feedback knowledge				.49*	.31	.65	-.18	-.40	.04
Feedback generation							.13	-.18	.40
Moderator									
Self-efficacy	-.05	-.38	.23	-.03	-.21	.18	-.05	-.23	.14
Training × Self-efficacy	-.19	-.62	.24						
Declarative feedback knowledge × Self-efficacy				.20*	.03	.35			
Feedback generation × Self-efficacy							-.06	-.29	.24
Covariate									
Instructional feedback practice (pre)							.82*	.63	1.03

Note. All variables are *z*-standardized except for Training. Confidence intervals are bias-based. CI = confidence interval; *LL* = lower limit, *UL* = upper limit.

[a] Dummy-coded (PD-FA = 1, PD-PM = 0).

* Significant regression coefficient (zero was not included in the 95% CI).

Table 3

Summary of specific indirect and total training effects on feedback generation and feedback practice in Model 1.

	Feedback generation			Feedback practice		
		95% CI			95% CI	
Training effect	β	*LL*	*UL*	β	*LL*	*UL*
Total	-.33	-.72	.10	.11	-.36	.50
Specific indirect via						
Declarative feedback knowledge	.25*	.05	.54	-.09*	-.34	-.003
Feedback generation				-.08	-.32	.08
Declarative feedback knowledge via feedback generation				.03	-.02	.18

Note. Training is dummy-coded (PD-FA = 1, PD-FA = 0). All other variables are *z*-standardized. Confidence intervals are bias-based. CI = confidence interval; *LL* = lower limit, *UL* = upper limit.

* Significant regression coefficient (zero was not included in the 95% CI).

4 Discussion

In the present study, we investigated the impact of teacher training on teachers' instructional feedback practice in mathematics instruction. We specifically focused on the role of teachers' feedback knowledge and the ability to generate feedback as intervening mechanisms and the function of teachers' self-efficacy as a moderating variable.[3]

4.1 Summary and interpretation of results

As hypothesized, results reveal that teachers participating in a training on formative assessment know more about feedback than teachers who took part in a training on mathematical modeling and problem solving (*Hypothesis 1a*). This finding is consistent with the model of Desimone (2009) and findings from previous studies (e.g., Diamond et al 2014), indicating a positive impact of teacher training on teacher knowledge.

Our study furthermore showed that in line with our assumptions, knowledge about feedback has a positive effect on the ability to generate feedback in a test situation (*Hypothesis 1b*). Accordingly, we found that our six-day teacher training yielded a positive indirect impact on the ability to provide feedback in a test situation mediated via the acquisition of feedback knowledge (*Hypothesis 1d).* Teachers thus were able not only to recall, but also to apply the newly acquired knowledge in a feedback generation task.

This ability, however, was not positively associated with teachers' feedback practice in mathematics instruction (*Hypothesis 1c*). In line with this result, no indirect training effect on instructional feedback practice via feedback knowledge and feedback generation (*Hypothesis 1e*) was shown. These findings were not expected and do not correspond to (a) Desimone's (2009) theoretical model suggesting an indirect impact from teacher training via teacher

[3] Our findings from Model 1 mostly correspond to the findings from Model 2, except for small differences in the significance level which might be due to the use or non-use of the bootstrapping procedure.

knowledge on instruction and (b) previous findings indicating a positive relationship between teacher knowledge and instructional quality (e.g., Kunter et al. 2013; Voss et al. 2014). It should be noted, however, that neither these studies nor Desimone's model focus on formative assessment. Kunter, Voss and colleagues (e.g., Kunter et al. 2013; Voss et al. 2014) did not even particularly deal with teacher training, and the indirect effect suggested in Desimone's model has so far not been tested empirically.

Our finding that teachers do not transfer the newly acquired feedback knowledge into everyday instructional practice rather supports the well-documented phenomenon of "inert knowledge" (Mandl and Gerstenmaier 2000; Renkl 1996; Whitehead 1929). Our analysis of the indirect training effect helps to better understand this gap between knowledge and action; it shows that teachers indeed are well able to use the newly acquired knowledge per se–at least in isolated test situations. However, the transfer to everyday instruction is not successful. One explanation might be the comparatively high complexity and lack of structuredness of instructional situations (compared to test situations), which might make it difficult to identify and use relevant information to provide adequate individual feedback. Another possible explanation is that teachers–despite being able to do so–do not provide individual feedback because they find it too time-consuming. A third explanation could be that teachers indeed changed their feedback practice, but students have not realized this change yet. The questionnaire scale on teachers' feedback practices did potentially not assess every aspect of change in teachers' feedback behavior. It focuses on central feedback characteristics, but does not cover all aspects of feedback practice that were dealt with in the training.

Corresponding to our hypothesis, the indirect effect on feedback generation was moderated by teachers' self-efficacy (*Hypothesis 2*). Teachers who have a high level of self-efficacy benefit more from training than teachers with a low level of self-efficacy. This finding supports previous studies showing a positive relationship between trainees' self-efficacy and transfer of training contents (e.g., Pugh and Bergin 2006; Simosi 2012; Stein and

Wang 1988; for an overview see Gegenfurther 2011). Most of these studies, however, have not focused on teacher training (for an exception see Stein and Wang 1988), and none has investigated the moderation of indirect training effects. In our study, in contrast, we specifically investigated at which point the impact of training–(1) effect from training on feedback knowledge; (2) effect from feedback knowledge on feedback generation; (3) effect from feedback generation on instructional feedback practice–is influenced by teachers' self-efficacy. Our results show the following: (1) The impact of training on teachers' feedback knowledge is not affected by self-efficacy, whereas (2) knowledge effects are; teachers with a high level of self-efficacy are better able to use their knowledge to generate feedback in a test situation than teachers with a low level of self-efficacy. (3) The relationship between feedback generation and instructional feedback practice, in contrast, does not depend on the teacher's belief in his/her capability to provide feedback.

Self-efficacy is positively associated with effort and persistence (Multon et al. 1991; Schunk 1995) and positively related to increased cognitive engagement and the efficient use of learning strategies (Pintrich and DeGroot 1990; Zimmerman 2000). Thus, teachers with a higher degree of self-efficacy might make stronger efforts or be better able to put the newly acquired feedback knowledge into action (Step 2 of the indirect effect: effect from feedback knowledge on feedback generation). Gegenfurtner (2011) showed that the relationship between self-efficacy and transfer varies as a function of knowledge type, with a stronger relation between self-efficacy and procedural knowledge than between self-efficacy and declarative knowledge. This might explain the missing moderation for Step 1 of the indirect effect (effect from training on declarative feedback knowledge). Regarding Step 3 of the indirect effect (effect from feedback generation on instructional feedback practice), Stajkovic and Luthans (1998) demonstrated that the relationship between self-efficacy and performance depends on the complexity of the task with the magnitude of relationship decreasing for high levels of task complexity. Possibly, the transfer of training content to real-life instruction

(Step 3) was too difficult and complex, even for highly self-efficient teachers.

4.2 Limitations, implications for practice and suggestions for future research

When interpreting the results, some methodological limitations of the study should be considered. *First*, we have a high percentage of missing data for instructional feedback practice. Although we used FIML under consideration of auxiliary variables, effects on instructional feedback practice should be interpreted with caution. *Second*, when interpreting the moderation effect, one has to bear in mind that self-efficacy was assessed in the pre survey only but might have changed in the course of training. *Third*, whereas instructional feedback practice was assessed before and after the training, declarative feedback knowledge and feedback generation were tested only once. Thus, the study provides no information on the development of these variables. *Fourth*, teachers were not randomly assigned to conditions, but enrolled in a professional development course depending on their personal interests. Thus, no premature causal conclusions should be drawn (see, e.g., Muthén and Asparouhov 2015). There is a group difference for self-efficacy regarding feedback, which, however, was controlled for in the model. Furthermore, the two treatment groups do not systematically differ with regard to their vocational education (second state examination, teaching mathematics outside of one's subject area, number of professional development courses attended in the last three years) or their teaching experience. They furthermore do not vary with regard to pretreatment instructional feedback practice, formative assessment practice, formative assessment beliefs, and interest in feedback. When taking these variables (instructional feedback practice, formative assessment practice, formative assessment beliefs, interest in feedback) as covariates into account, the results of the path analysis remain the same. Still, results should be interpreted carefully as unmeasured variables may confound the association between treatment, intervening, and outcome variables. *Fifth, o*ur path model had 19 free parameters. As our sample size was $N = 67$, the requirement for sufficient sample sizes (i.e., a cases/parameter ratio higher than 5:1, Kline 2005) was not fulfilled satisfactorily and parameters should thus be interpreted carefully (see, e.g., Gelman and Carlin 2014).

Despite these limitations, we believe that the present study makes a significant contribution to research on formative assessment practice in mathematics instruction. The teacher training on formative assessment was developed and evaluated, drawing on current training and formative assessment research. Furthermore, the mechanisms that mediate and moderate the effects of this training were studied. We identified teachers' pre-training self-efficacy as a significant motivational moderator helping teachers to make use of training contents and found that the application of feedback knowledge in everyday instruction is a critical point at which the training intervention might fail.

Following these findings, future training on formative assessment and corresponding studies should specifically attempt to help teachers apply their newly acquired knowledge in complex everyday instruction. To that end, interventions could be intensified through longer-term and more continuous teacher support measures (e.g., through sitting in on classes). In addition, implementation of training content might be improved by highlighting the application orientation of training (e.g., through the analysis and discussion of videos showing the teacher's own assessment practices). Furthermore, teachers could be provided with more specific guidelines, prepared task material, and feedback sheets to increase confidence and reduce time burden when providing feedback in classrooms. To gain a better insight into what actually happens in classrooms, future studies might not only base the assessment of teachers' feedback practice on students' judgements but also on teacher questionnaires or video recordings.

The present study not only provides hints on how to design feedback trainings, it also contributes to a more differentiated understanding of the frequently described phenomenon of inert knowledge (e.g., Mandl and Gerstenmaier 2000; Renkl 1996) by analyzing the relationship between knowledge acquisition and knowledge application, and distinguishing between knowledge application in simple test situations and complex everyday situations. It furthermore helps researchers better to comprehend the impact of teacher training in general by empirically testing parts of Desimone's model (2009). Following studies might additionally assess students' performance as a criterion variable to analyze the entire mediation process proposed by Desimone and test whether the primary goal of formative assessment–improving students' learning (e.g., Black and Wiliam 1998a)–has been reached. Taking these aspects into account, future studies might contribute to enhancing feedback culture in secondary mathematics instruction.

5 Acknowledgments

The present study is embedded in the project "Conditions and Consequences of Classroom Assessment" (Co^2CA), which is collaboratively conducted by researchers at the German Institute for International Educational Research, University of Kassel, and Leuphana University of Lüneburg. The preparation of this paper was supported by grants from the German Research Foundation (DFG, KL1057/10-3, BL2751/17-3, LE2619/1-3) in the Priority Program "Models of Competencies for Assessment of Individual Learning Outcomes and the Evaluation of Educational Processes" (SPP 1293).

6 References

Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive view.* Englewood Cliffs, NY: Prentice-Hall.

Bennett, R. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25. doi:10.1080/0969594X.2010.513678.

Besser, M., Leiss, D., & Klieme, E. (2015). Wirkung von Lehrerfortbildungen auf Expertise von Lehrkräften zu formativem Assessment im kompetenzorientierten Mathematikunterricht [Impact of professional development on teachers' expertise on formative assessment in competence oriented mathematics instruction]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 47*(2), 110–122. doi:10.1026/0049-8637/a000128.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assesment in Education, 5*(1), 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom. *Phi Delta Kappan, 80*(2), 139–147.

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

Bliese, P. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. Klein & S. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Brown, M., & Hodgen, J. (2010). Assessment in schools–mathematics. In E. Baker, B. McGaw, & P. Peterson (Eds.), *International encyclopedia of education* (3 ed., pp. 279–284). Oxford: Elsevier.

Cizek, G. (2010). An introduction to formative assessment: History, characteristics and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3–17). New York, NY: Routledge.

Collins, L., Schafer, J., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330–351. doi: 10.1037/1082-989X.6.4.330.

Desimone, L. (2009). Improving impact studies of teachers’ professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. doi: 10.3102/0013189X08331140.

Diamond, B. S., Maerten-Rivera, J., Rohrer, R. E., & Lee, O. (2014). Effectiveness of a curricular and professional development intervention at improving elementary teachers' science content knowledge and student achievement outcomes: Year 1 results. *Journal of Research in Science Teaching, 51*(5), 635–658.

Dresel, M., & Haugwitz, M. (2008). A computer-based approach to fostering motivation and self-regulated learning. *The Journal of Experimental Education, 77*(1), 3–18. doi:10.3200/JEXE.77.1.3-20.

Dunn, K. E., & Mulvenon, S. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research and Evaluation, 14*(7), 1–11.

Edwards, J., & Lambert, L. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12*, 1–22. doi:10.1037/1082-989X.12.1.1.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9. doi: 10.1016/j.learninstruc.2013.07.001.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L. & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal, 36* (3), 609-646. doi:10.3102/00028312036003609

Gegenfurtner, A. (2011). Motivation and transfer in professional training: A meta-analysis of the moderating effects of knowledge type, instruction, and assessment conditions. *Educational Research Review, 6*(3), 153–168.

Gelman, A. , & Carlin, J (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science, 9*(6), 641–651. doi: 10.1177/1745691614551642

Graham, J. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10*, 80–100. doi: 10.1207/S15328007SEM1001_4.

Harks, B. (2013). *Kompetenzdiagnostik und Rückmeldung–zwei Komponenten formativen Assessments* [Competence diagnostics and feedback–two components of formative assessment]. Unpublished dissertation. Hattie, J. (2003). *Formative and summative interpretations of assessment information.* http://web.auckland.ac.nz/uoa/fms/default/education/staff/Prof.%20John%20Hattie/Documents/John%20Hattie%20Papers/assessment/Formative%20and%20Summative%20Assessment%20(2003).pdf.
Accessed 1 May 2013.

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Routledge.

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

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89*, 140–145.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study.* NCES (2003-013). Washington DC: U.S. Department of Education/National Center for Education Statistics.

Hill, M., Cowie, B., Gilmore, A., & Smith, L. (2010). Preparing assessment-capable teachers: What should preservice teachers know and be able to do? *Assessment Matters, 2*, 43–64.

James, M., & Pedder, D. (2006). Professional learning as a condition for assessment for learning. In J. Gardner (Ed.), *Assessment and learning* (pp. 25–27). London: Sage.

Kim, M.-J., & Lehrer, R. (2015). Using learning progressions to design instructional trajectories. In C. Suurtamm & A. Roth McDuffie (Eds.), *Annual perspectives in mathematics education:Assessment to enhance teaching and learning* (pp. 27–38). Reston, VA: National Council of Teachers of Mathematics.

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

Klimczak, M., Kampa, M., Bürgermeister, A., Harks, B., Rakoczy, K., Besser, M., . . . Leiss, D. (2012). *Dokumentation der Befragungsinstrumente des Feldexperimentes im Projekt "Conditions and Consequences of Classroom Assessment" (Co²CA)* [Documentation of the questionnaires of the laboratory experiment in the project 'Conditions and Consequences of Classroom Assessment' (Co²CA)].

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Kluger, A., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284. doi:10.1037//0033-2909.119.2.254.

Koedinger, K. R., McLaughlin, E., & Heffernan, N. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research, 43* (1), 489-510. doi:org/10.2190/EC.43.4.d

Krampen, G. (1987). Differential effects of teacher comments. *Journal of Educational Psychology, 79*, 137–146. doi:10.1037//0022-0663.79.2.137.

Kultusministerkonferenz. (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss* [National standards in mathematics for secondary school]. München: Luchterhand.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820. doi: 10.1037/a0032583.

Lehmann, R. H., & Seeber, S. (2005). *Accelerated mathematics in grades 4-6. Summary of a quasi-experimental study in Northrhine-Westphalia, Germany*. Madison, WI: Renaissance Learning, Inc. Retrieved from http://research.renlearn.com/research/pdfs/192.pdf

Lipowsky, F., & Rzejak, D. (2012). Lehrerinnen und Lehrer als Lerner–Wann gelingt der Rollentausch [Teachers as learners–when does the role reversal work]? *Schulpädagogik Heute, 5*(3), 1–17.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung [Dealing with missing data in psychological research]. *Psychologische Rundschau, 58*(2), 103–117.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*(2), 120–131. doi: 10.1016/j.cedpsych.2008.12.001.

MacKinnon, D. (2008). *Introduction to statistical mediation analysis.* New York, NY: Erlbaum.

MacKinnon, D., Lockwood, C., & William, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99–128. doi:10.1037/1082-989X.7.1.83.

Mandl, H., & Gerstenmaier, J. (2000). *Die Kluft zwischen Wissen und Handeln. Empirische und theoretische Lösungsansätze* [The gap between knowledge and action]. Göttingen, Germany: Hogrefe.

Marsh, H., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A., Abduljabbar, A., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*(2), 106–124. doi: 10.1080/00461520.2012.670488.

Mason, B., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us*. http://dwb.unl.edu/Edit/MB/MasonBruning.html. Accessed 1 May 2013.

Mischo, C., & Rheinberg, F. (1995). Erziehungsziele von Lehrern und individuelle Bezugsnormen der Leistungsbewertung [Educational goals of teachers and individual frames of reference of performance assessment]. *Zeitschrift für Pädagogische Psychologie, 9*(3/4), 139–151.

Multon, K., Brown, S., & Lent, R. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30–38. doi: 10.1037/0022-0167.38.1.30.

Muthén, B., & Asparouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling, 22*(1), 12–23. doi: 10.1080/10705511.2014.935843

Muthén, L., & Muthén, B. (1998-2007). Mplus (Version 6.1) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment.* Washington, DC: National Academic Press.

Phelan, J., Choi, K., Vendlinski, T. Baker, E. & Herman, J. (2011). Differential improvement in student understanding of mathematical principles following formative assessment intervention. *The Journal of Educational Research, 104*, 330-339. doi.org/10.1080/00220671.2010.484030

Pintrich, P., & De Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40. doi: 10.1037/0022-0663.82.1.33.

Pugh, K., & Bergin, D. (2006). Motivational influences on transfer. *Educational Psychologist, 41*(3), 147–160. doi: 10.1207/s15326985ep4103_2.

Rakoczy, K., Klieme, E., Leiß, D., & Blum, W. (in press). Formative assessment in mathematics instruction: Theoretical considerations and empirical results of the Co²CA-project. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.). *Competence assessment in education: Research, models, and instruments.* Berlin: Springer.

Renkl, A. (1996). Träges Wissen: Wenn Erlerntes nicht genutzt wird [Inert knowledge: When knowledge is not used]. *Psychologische Rundschau*, *47*, 78–92.

Ruthven, K. (1994). Better judgement: Rethinking assessment in mathematics education. *Educational Studies in Mathematics, 27*, 433-450. doi:org/10.1007/BF01273382

Sadler, D. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice, 5*(1), 77–84. doi:10.1080/0969595980050104.

Schlomer, G., Bauman, S., & Card, N. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology, 57*(1), 1–10. doi: 10.1037/a0018082.

Schneider, M., & Randel, B. (2010). Research on characteristics of effective professional development programs for enhancing educator' skills in formative assessment. In H. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp. 251–276). New York, NY: Routledge.

Schunk, D. (1995). Self-efficacy and education and instruction. In J. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment theory, research,and application* (pp. 281–303). New York, NY: Springer.

Shavelson, R. (2008). Guest editor's introduction. *Applied Measurement in Education, 21*(4), 293–294. doi: 10.1080/08957340802347613.

Shih, S., & Alexander, J. (2000). Interacting effects of goal setting and self- or other-referenced feedback on children's development of self-efficacy and cognitive skill within the Taiwanese classroom. *Journal of Educational Psychology, 92*, 536–544. doi:10.1037//0022-0663.92.3.536.

Shrout, P., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422–445. doi:0.1037/1082-989X.7.4.422.

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

Simosi, M. (2012). The moderating role of self-efficacy in the organizational culture–training transfer relationship. *International Journal of Training and Development, 16*(2), 92–106. doi: 10.1111/j.1468-2419.2011.00396.x.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Stajkovic, A., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261. doi: 10.1037/0033-2909.124.2.240.

Stein, M., & Wang, M. (1988). Teacher development and school improvement: The process of teacher change. *Teaching and Teacher Education, 4*(2), 171–187. doi: 10.1016/0742-051X(88)90016-9.

Suurtamm, C., Thompson, D.R., Kim, R.Y., Moreno, L.D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S., Vos, P. (2016). *Assessment in mathematics education. Large-Scale assessment and classroom assessment.* ICME 13 Topical Surveys. Springer Open.

Thompson, D. B. J. (2007). *Effects of evaluative feedback on math self-efficacy, grade self-efficacy, and math achievement of ninth grade algebra students: A longitudinal approach.* Unpublished doctoral dissertation, University of Louisville, KY. http://gradworks.umi.com/32/68/3268832.html.

Accessed 1 May 2013.

Thompson, D. R., & Kaur, B. (2011). Using a multi-dimensional approach to understanding to assess students' mathematical knowledge. In B. Kaur & K. Y. Wong (Eds.), *Assessment in the mathematics classroom: 2011 Association of Mathematics Educators Yearbook* (pp. 17-32). Singapore: World Scientific Publishing.

Toh, T. L., Quek, K. S., Leong, Y. H., Dindyal, J., & Tay, E. G. (2011). Assessing problem solving in the mathematics curriculum: A new approach. In B. Kaur & K. Y. Wong (Eds.), *Assessment in the mathematics classroom: 2011 Association of Mathematics Educators Yearbook* (pp. 33–66). Singapore: World Scientific Publishing.

Voss, T., Kunter, M., Seiz, J., Hoehne, V., & Baumert, J. (2014). Die Bedeutung des pädagogisch-psychologischen Wissens von angehenden Lehrkräften für die Unterrichtsqualität [The relevance of educational psychological knowledge of preservice teachers for instructional quality]. *Zeitschrift für Pädagogik, 60*(2), 184–201.

Whitehead, A. (1929). *The aims of education.* New York, NY: Mac Millan.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy and Practice*, *11* (1), 49-65. doi:10.1080/0969594042000208994

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). New York, NY: Lawrence Erlbaum.

Wylie, E., Gullickson, A., Cummings, K., Egelson, P., Noakes, L., Norman, K., & et al. (2012). *Improving formative assessment practice to empower student learning.* Thousand Oakes, CA: Corwin.

Yin, Y., Shavelson, R., Ayala, C., Ruiz-Primo, M., Brandon , P., Furtak, E., & et. al. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education, 21*(4), 335–359. doi:10.1080/08957340802347845.

Ysseldyke, J. E., & Bolt, D. M. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review*, *36*, 453-467.

Zimmerman, B. (2000). Attaining self-regulation. A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39).

7 Appendix

Table A1

Direct and interaction effects in Model 2

Variable	Declarative feedback knowledge β	*SE*	Feedback generation β	*SE*	Instructional feedback practice (post) β	*SE*
Predictor						
Training (PD-FA vs. PD-PM)[a]	.52*	.24	-.58*	.23	.27	.24
Intervening variables						
Declarative feedback knowledge			.49**	.11	-.18	.13
Feedback generation					.15	.13
Moderator						
Self-efficacy	-.05	.17	-.03	.11	-.09	.10
Training × Self-efficacy	-.19	.25				
Declarative feedback knowledge × Self-efficacy			.20[†]	.11		
Feedback generation × Self-efficacy					-.08	.12
Covariate						
Instructional feedback practice (pre)					.75**	.12

Note. All variables are *z*-standardized except for Training.

[a] Dummy-coded (PD-FA = 1, PD-PM = 0).

[†] $p < .10$. * $p < .05$. ** $p < .01$.

Table A2

Summary of specific indirect and total training effects on feedback generation and feedback practice in Model 2.

	Feedback generation		Feedback practice	
Training effect	β	*SE*	β	*SE*
Total	-.33	.25	.12	.22
Specific indirect via				
Declarative feedback knowledge	.25†	.13	-.09	.08
Feedback generation			-.09	.08
Declarative feedback knowledge via feedback generation			.04	.04

Note. Training is dummy-coded (PD-FA = 1, PD-FA = 0). All other variables are *z*-standardized.

† $p < .10$. * $p < .05$. ** $p < .01$.