



Fostering prospective teachers' explaining skills during university education—Evaluation of a training module

Stefanie Findeisen¹  · Viola Katharina Deutscher²  · Jürgen Seifried² 

Published online: 7 September 2020

© The Author(s) 2020

Abstract

Providing instructional explanations is a core component of effective instruction and an important teaching skill. Teaching skills are generally regarded as learnable, and teacher education programs aim to improve teachers' professional competences. In this study, we analyze to what extent explaining skills can be fostered during teacher education at university by means of a specific training module. We designed a training (university module) for prospective economics teachers at vocational schools (candidates in a Master's teaching program). By means of videotaped simulated interactions at two measurement points, we analyzed the development of teacher candidates' explaining skills. Teacher candidates were asked to explain the neoclassical supply and demand model (treatment group: $n = 48$; control group: $n = 30$) to an actor playing the role of a school student. The quality of the explanations was operationalized in respect of five aspects of successful explanations, which were derived from a literature review: (1) Content, (2) Student-teacher interaction, (3) Process structure, (4) Representation, and (5) Language. The results show that there was a treatment effect on the development of the *Process structure* aspect, while *Student-teacher interaction* appeared to develop "naturally" through experience, regardless of participation in the training. The quality aspects *Content*, *Representation*, and *Language* appeared stable over time. Hence, the findings show that some aspects of explaining skills are learnable even in a short training module. Learning effects are attributable partly to the instructional input received and partly to repeated practice. Both imply the importance of further opportunities to practice instructional explanation in teacher education.

✉ Stefanie Findeisen
stefanie.findeisen@uni-konstanz.de

¹ Business and Economics Education, University of Konstanz, Universitaetsstrasse 10, 78464 Konstanz, Germany

² Area of Economic and Business Education, University of Mannheim, L4,1, 68161 Mannheim, Germany

Keywords Explaining skills · Teacher education · Pedagogical content knowledge · Video-based analysis · Economics education

Objectives

Current research on teaching and instruction emphasizes the importance of teaching quality for learner achievement. Empirical findings show that teachers' competences have a crucial impact on students' learning processes and learning outcomes (Hattie 2009; Kunter et al. 2013). Hence, there has been growing interest in examining prospective teachers' professional competences (Ball et al. 2005; Blömeke and Delaney 2014; Hill et al. 2008; Kunter et al. 2013). One central competence of teachers is explaining subject matter to students (Brown 2006; Leinhardt 1987), and explaining is regarded as a core activity in a teacher's daily practice (Ball et al. 2005; Charalambous et al. 2011; Leinhardt 2010). Although the teacher is not necessarily the only one engaged in explaining content in the classroom (for evidence on the importance of self-explanations, or explanations by fellow students, see, e.g., Chi et al. 1989, 1994), there are still certain classroom events that call specifically for teacher explanations (e.g., reaction to student errors, demonstration of processes; Hargie 2011). Also, in line with the findings outlined above, evidence shows that teachers' explanations contribute to student learning (Eisenhart et al. 1993; Evans and Guyson 1978; Hines et al. 1985).

Against this background, one can state that a crucial part of teacher education programs is to foster relevant teaching skills. In this regard, empirical evidence demonstrates significant effects of university-based learning opportunities on teacher candidates' competences (Fritsch et al. 2015; Kleickmann and Anders 2013). At the same time, empirical evidence clearly documents that prospective teachers struggle when it comes to providing instructional explanations (Findeisen 2017; Borko et al. 1992; Guler and Celik 2016; Halim and Meerah 2002; Thanheiser 2009). Moreover, there are significant differences between the explaining skills of expert and novice teachers (Charalambous 2016; Leinhardt 1989; Leinhardt and Greeno 1986; Ma 2010). These results raise the question whether explaining skills can be fostered during university teacher education.

This paper aims to contribute to our understanding of this issue by examining the effect of a training course (teacher education module) on teacher candidates' explaining skills. The study focuses on prospective vocational school teachers in Germany (teacher candidates in university-based teacher education programs), in the domain of economics. To measure prospective teachers' explaining skills, we used an interactive simulation in order to achieve a performance-based assessment that also addresses the main feature of explanation processes—the interaction between teacher and student. Using an experimental design, we videotaped 78 teacher candidates (treatment group: 48 teacher candidates; control group: 30 teacher candidates). Each participant provided an explanation at two measurement points (at the beginning and the end of the treatment); thus, an empirical base of 156 videos was amassed. The treatment group participated in a university module (five sessions of 90 min each) devoted to the quality of explanations and the use of video-based self-reflections on one's performance. We applied video analyses to examine teacher candidates' explaining skills.

With this paper, we aim to answer the following research questions:

- 1 To what extent does the intervention improve teacher candidates' explaining skills?
- 2 Which aspects of teacher candidates' explaining skills are improvable by means of a short-term treatment; which aspects do not demonstrate development?

Theoretical framework

Quality of instructional explanations

Instructional explanations can be defined as “unique communicative forms that support the learning and understanding of others” (Leinhardt 1997, 231). Instructional explanations are characterized by three central features (Findeisen 2017): (1) they are provided in an interaction between the person who is explaining and the listener(s); (2) there is a difference in knowledge between speaker and recipients; and (3) the speaker intends to clarify something for the audience. To evaluate the quality of an instructional explanation, several aspects need to be taken into account. In previous work (Findeisen 2017), we systematically analyzed previous findings on the quality of explanation and developed an assessment framework consisting of the following five quality aspects: (1) Content, (2) Student-teacher interaction, (3) Process structure, (4) Representation, and (5) Language.

Undeniably, an explanation, first of all, has to present the *Content* in a suitable manner. The reviewed sources highlight the importance of the correctness and comprehensiveness of explanations (e.g., Brown and Daines 1981; Cabello Gonzalez 2013; Duffy et al. 1986), the logical structure of the content (e.g., Duffy et al. 1986; Gage et al. 1968; Sevian and Gonsalves 2008), and the definition of new terms (e.g., Brown 2006; Charalambous et al. 2011; Sevian and Gonsalves 2008). A teacher should also be able to provide multiple approaches to the same content and to change an explanatory approach flexibly, if a student does not understand the explanation (e.g., Geelan 2013).

As an explanation is always provided in an *Interaction* with the audience, the teacher should have the students in mind. This means to consider students’ prior knowledge and characteristics when designing the explanation (e.g., Kennedy 1996; Wittwer and Renkl 2008), as well as involving students actively in the explanatory process (e.g., Brown 2006; Kulgemeyer and Schecker 2013; Sevian and Gonsalves 2008) and adapting explanations in response to cues from students (e.g., Duffy et al. 1986; Sevian and Gonsalves 2008).

The term *Process structure* refers to a logical sequencing of the explanatory process: i.e., following different steps that make the explanation more comprehensible to learners. The steps involve clarifying the aim of an explanation and outlining its structure (e.g., Charalambous et al. 2011), evaluating prior knowledge (e.g., Kennedy 1996; Sevian and Gonsalves 2008), repeating key features (e.g., Brown 2006; Miltz 1972), and assessing understanding (e.g., Duffy et al. 1986; Kulgemeyer and Schecker 2013).

Another important aspect of explaining is the *Representation* of the explanatory content. Here, the literature highlights the importance of using examples (e.g., Brown and Daines 1981; Duffy et al. 1986; Sevian and Gonsalves 2008), visualizations (e.g., Brown and Daines 1981; Geelan 2013; Sevian and Gonsalves 2008), or analogies (e.g., Duffy et al. 1986; Geelan 2013; Sevian and Gonsalves 2008). Teachers should also be able to combine different representational forms (e.g., Kulgemeyer and Schecker 2013; Leinhardt 1987; Sevian and Gonsalves 2008) and to point out similarities and differences between representations and the explanatory content (e.g., Cabello Gonzalez 2013; Leinhardt 1987; Sevian and Gonsalves 2008).

Finally, the *Language* used by the instructor is important for an explanation’s success. Language aspects mentioned in previous work on the quality of explanations comprehend the appropriateness of the level of speech (e.g., Charalambous et al. 2011; Gage et al. 1968; Sevian and Gonsalves 2008), avoiding vagueness (e.g., Brown 2006; Duffy et al. 1986; Miltz 1972), and the supportive use of gestures and movement (e.g., Brown 2006; Cabello Gonzalez 2013).

The explanation quality framework (Findeisen 2017) offers several advantages compared with existing lists of quality features. A first advantage is that the framework is based on a literature review of the various aspects of explanation quality found both in empirical studies and in didactical conceptualizations. During the review, we examined 24 sources and extracted those quality criteria that were mentioned in at least three different sources. This resulted in identifying 23 quality criteria (e.g., logical structure, involving students actively, providing examples). Hence, our framework includes the most important quality aspects. In the following step, the 23 quality criteria were categorized into five quality aspects: (1) Content, (2) Student-teacher interaction, (3) Process structure, (4) Representation, and (5) Language. Therefore, the single aspects of explanation quality operationalize the five quality criteria. Another advantage of the framework stems from the fact that the framework is very general. Thus, it is applicable across different disciplines and can easily be adapted to domain-specific particularities. Finally, the framework deconstructs explanation quality in five different important dimensions, giving researchers the option to analyze those features separately, and teacher educators the possibility of providing detailed and specific feedback on different aspects of explaining.

For the present study, we define teachers' explaining skills as the ability to provide a high-quality explanation; meaning, an explanation that is correct and coherent that actively involves the audience and is presented in a logical, sequenced set of steps that use suitable representations and appropriate language. With this definition, it becomes apparent that in order to explain subject matter to students, a teacher needs to have solid content knowledge, as well as an understanding of how students learn (e.g., learning processes, typical misconceptions, pedagogical content knowledge; Brown 2006).

Empirical evidence on teacher candidates' explaining skills

As mentioned above, previous findings demonstrate deficits in prospective teachers' explaining skills. Those deficits occur across the different aspects of explanation quality. For instance, teacher candidates' explanations have been shown to be often error-prone or incoherent (Borko et al. 1992; Guler and Celik 2016; Halim and Meerah 2002; Leinhardt 1989; Thanheiser 2009). Furthermore, pre-service teachers struggle when it comes to evaluating and activating prior knowledge (Sánchez et al. 1999), tailoring explanations to students' needs (Halim 1998), or reacting flexibly to unexpected events or to students' questions (Borko and Livingston 1989; Leinhardt 1989). Also, they experience difficulties when designing suitable representations (Borko et al. 1992; Inoue 2009; Kinach 2002a; Wheeldon 2012) or examples (Borko et al. 1992; Inoue 2009; Wheeldon 2012). Deficits in prospective teachers' explaining skills are often attributed to a lack of content knowledge (Borko et al. 1992; Eisenhart et al. 1993; Halim and Meerah 2002; Thanheiser 2009). Kulgemeyer and Riese (2018) show that both content knowledge (CK) and pedagogical content knowledge (PCK) correlate with explaining skills (CK: $r = .38$, $p < .01$; PCK: $r = .38$, $p < .001$). However, the effect of CK on explaining skills is mediated by PCK.

Another set of studies analyzes the possibility of fostering teacher candidates' explaining skills. Several studies on this research question were conducted between the 1970s and the early 1990s. The research topic became popular again with the rising interest in teachers' professional competences in the 2000s. In some of the earlier studies, instructors were asked to rate the learnability of different features of explaining (Brown and Daines 1981; Miltz 1972).

The following aspects were ranked amongst the most learnable features: clarity, simplicity, selection of appropriate content, logical organization, focusing of attention on important points, summarizing, use of visualizations and examples, appropriate vocabulary, and avoidance of vagueness. At the same time, the experts rated aspects like flexibility, enthusiasm, verbal fluency, and explaining links as the most difficult to learn. Moreover, studies have examined the impact of specific interventions on teacher candidates' explaining skills as well as the development of explaining skills during teacher education. For instance, Miltz (1972) implemented a training for prospective teachers and found that participants' explanations improved significantly compared with a control group. Further intervention studies also provide positive evidence of the possibility of fostering the explaining skills of teacher candidates (Cabello Gonzalez 2013; Charalambous 2008; Charalambous et al. 2011; Kinach 2002b; Zembal-Saul et al. 2000). There is also evidence that, after receiving feedback from their mentor during teaching placements, prospective teachers' explanations improved in later lessons (Borko et al. 1992; Eisenhart et al. 1993). In respect of field placements, Kulgemeyer et al. (2020) showed that professional knowledge prior to the teaching placement (pre-test) influenced explaining skills after the teaching placement (post-test). In fact, only teacher candidates entering the school internship with high professional knowledge showed significant development in explaining skills after a 5-month placement. There was no development in explaining skills for teacher candidates with low professional knowledge at the start of the internship.

Concerning different aspects of explanation quality, findings show that teacher candidates, for instance, improve with respect to content structure (Miltz 1972), clarity (Cabello Gonzalez 2013), and the use of examples (Miltz 1972). Following a course of training, explanations were more appropriate for students (Cabello Gonzalez 2013; Charalambous 2008) and more comprehensible (Charalambous et al. 2011). In addition, teacher candidates evaluated prior knowledge more often, involved students more frequently, and showed higher flexibility when it came to adjusting an explanation during the explanatory process (Zembal-Saul et al. 2000).

To sum up, the empirical evidence suggests that providing high-quality explanations can be learned. However, certain limitations of the studies mentioned above should be taken into account. First, and most importantly, the reported studies—with the exception of the work by Miltz (1972)—do not include control groups. This must be viewed critically, as it seems possible therefore that the basis for teacher candidates improving is not their participation in a teacher education course, but rather their experience of practice in instruction. Drawing on expertise research, in which learning is mostly premised on practice (Berliner 1994; Ericsson 1996), this possibility seems even likely. Second, the studies are often based on individual case analysis (Borko et al. 1992; Eisenhart et al. 1993) or on small sample sizes (Cabello Gonzalez 2013: $n = 20$; Charalambous et al. 2011: $n = 4$; Clermont et al. 1993: $n = 8$; Zembal-Saul et al. 2000: $n = 2$). An exception is, again, the study of Miltz (1972) with 60 participants (30 in the treatment group, 30 in the control group). Third, explaining skills are either measured during teaching placements, where conditions cannot be controlled by the researchers (e.g., Borko et al. 1992; Kinach 2002b), or in laboratory conditions (e.g., Charalambous et al. 2011; Miltz 1972) where—with the exception of Cabello Gonzalez (2013), who assessed explaining skills in microteaching episodes—a crucial feature of instructional explanations is neglected: the interaction with the addressee. If student-teacher interactions are to be part of the analysis, video analysis would appear to be the right method to best capture the different aspects of the interaction and to allow for an analysis of all elements of the explanation (e.g., generating visualizations).

The present study

The present study aims to contribute to the research on teachers' professional competences with regard to explaining skills. Aiming to examine the effects of a targeted training module in explaining skills, we designed a module for students of a Master's teacher education program in Germany (see the section "[Treatment design and curricular embedding](#)"). Our approach had three particularities. First, we used a treatment-control-group design that allowed for the segregation of a treatment effect from a mere time or repetition effect, in relation to explaining skills. Second, by using interactive simulations (see the section "[Assessment of explaining skills](#)"), controlled test conditions were realized that allowed a performance-based assessment of explaining skills, while accounting for the interactive nature of the explanatory process. Third, deconstructing the explanation quality into five different aspects allowed us to analyze how different quality criteria changed in the course of the training.

In light of previous findings, reported above, we assume that the quality of teacher candidates' explanations benefits from training. Hence, we hypothesize that prospective teachers in the experimental group will improve their performance. However, it is possible that the control group will also show some improvement, even without attending the module, due to a repetition effect (explaining the same content 5 weeks later). Any such improvement in the control group would, however, be expected to prove smaller than that in the experimental group. Hence, we formulate the following hypotheses for our five quality aspects:

- h_1 : The improvement in the quality aspect Content from explaining situation 1 (t_0) to explaining situation 2 (t_1) is greater for the experimental group than for the control group.
- h_2 : The improvement in the quality aspect Student-teacher interaction from explaining situation 1 (t_0) to explaining situation 2 (t_1) is greater for the experimental group than for the control group.
- h_3 : The improvement in the quality aspect Process structure from explaining situation 1 (t_0) to explaining situation 2 (t_1) is greater for the experimental group than for the control group.
- h_4 : The improvement in the quality aspect Representation from explaining situation 1 (t_0) to explaining situation 2 (t_1) is greater for the experimental group than for the control group.
- h_5 : The improvement in the quality aspect Language from explaining situation 1 (t_0) to explaining situation 2 (t_1) is greater for the experimental group than for the control group.

In our analysis, we are interested in differences between improvements across the five aspects of explanation quality. We assume that not all five aspects are equally improvable in a short-term treatment of 5 weeks. Based on the findings in previous studies (see the section "[Empirical evidence on teacher candidates' explaining skills](#)"), we expect that the aspect *Content* (e.g., correctness, completeness) would have potential for improvement, meaning that teacher candidates' explanations on the neoclassical supply and demand model would improve in regard to correctness, completeness, and structure. Improvements have also been documented for *Student-teacher interaction* (e.g., actively involving students), *Process structure* (e.g., evaluating prior knowledge, summarizing), and *Representation* (e.g., use of visualizations and examples). Thus, we hypothesize that in the experimental group, the quality of explanations will improve with respect to these aspects. In respect of *Language*, previous studies also show certain improvements (e.g., on using appropriate vocabulary, avoiding

vagueness). Nonetheless, we assume that the language used in explanations might emerge to be less subject to change over time than other features. Hence, compared with the other four quality aspects, we expect less improvement on *Language*.

Research design

Treatment design and curricular embedding

The treatment group participated in a university module intended to foster teacher candidates' explaining skills. In its conception, this module was theoretically rooted in educational action research (Elliott 1981, 1991), where teachers evaluate their instructional performance as researchers in order to reflect on and improve their educational practice (in the present case, their explaining skills). To secure close interaction with the lecturer and to encourage discussions in the group, we divided the treatment group ($n = 48$ teacher candidates) into three separate instruction groups of 16 participants each.

With respect to its content, the treatment consisted of theoretical input on the quality of instructional explanations (five 90-min sessions), on professional reflection techniques (individual reflection prompts after a short introduction to professional reflection) and on video analysis (four 180-min sessions). The module followed the following structural sequence: At the beginning of the module, teacher candidates were given one week to prepare an explanation in the field of economics (the neoclassical supply and demand model) and were instructed to design the explanation for a student at a vocational school. The teacher candidates were allowed to prepare for this situation as they saw fit, deciding how much time to allocate and which resources to use for preparation. During their preparation, they were only asked to consider that they would have about 10 minutes to provide the explanation to the student. Over the following weeks, they were introduced (via slides, discussions, and accompanying literature) to theory of instructional explanations, as well as methods of video analysis. A large proportion of instructional time was spent on the instructional quality assessment framework (Findeisen 2017) where students were guided conceptually through the five dimensions of the framework. Each dimension of the framework was covered in equal depth during the training module. In another session, practical implications and relations between the different aspects of the quality model were discussed in a plenary session.

After this theoretical input, they were prompted to reflect on their instructional performance, on the basis of the videotape of their first explaining attempt as well as on the basis of their newly acquired knowledge regarding instructional quality. After four weeks, the teacher candidates were again given a week's preparation time, to prepare a second explanation on the same topic. Another week later (5-week treatment time in total), we videotaped the participants for a second time. Afterwards, they turned to analyzing their own videos, working in the video laboratory to code their video material and to write a report reflecting on their performance during the explanation exercise, as well as on their learning outcomes. During this final phase, in order to support them in their data analysis, the lecturers offered individual coaching sessions (one 45-min session per teacher candidate).

The training for the treatment group took place during the teaching of this module, which is graded and compulsory for teacher candidates in this particular Master's program. Meanwhile, data from a control group were gathered during a parallel period of time amongst teacher candidates of the same Master's program who had characteristics similar to the participants in

the treatment group. In doing so, we made use of the fact that as part of the curriculum, teacher candidates at the particular university where this study took place are required to participate as test subjects in empirical investigations, in order to gain insights into different research methods from a participant's perspective. Opportunities to gain those mandatory hourly credits as test subjects are offered regularly during the study program, and teacher candidates can choose which study to participate in. Participation is not graded. The control group faced the same assessment of explaining skills, under the same conditions (two prepared explanations in an interactive simulation setting, before and after a 5-week period); however, they did not participate in any part of the training module. Therefore, the control group neither received theoretical input regarding high-quality explanations nor reflected on their first videotaped explaining exercise. They did however know that they would have to perform the same explanation again after five weeks. All participants in both groups continued their university semester with modules in Business and Economics Education.

Sample

The sample consisted of 78 pre-service teachers in Business and Economics Education (teacher candidates in a university-based teacher education program). The treatment group comprised 48 teacher candidates (31 female, 17 male), the control group 30 teacher candidates (22 female, 8 male). All participants were in the second or third semester of the Master's program. Data were collected in the fall of 2016.

Assessment of explaining skills

In order to allow for a performance-based assessment of the prospective teachers' skills, while being able to provide comparable, controlled conditions across different participants, we used simulated interactions. Hence, teacher candidates were teamed up with a trained student (standardized individual) to whom they provided an explanation in a simulated student-teacher interaction. If they wanted to visualize certain aspects of the explanation, they could make use of a whiteboard or paper and pencils provided in the room (see Fig. 1). A similar approach has been used before in teacher education (see, e.g., Dotger 2009, 2013; Dotger et al. 2014).

Two female research assistants (equal in age), who received intensive training and who acted according to a written script, portrayed the standardized students. They were instructed to, for instance, request additional information during the explanation process (e.g., how the supply curve would shift after an external shock). The interactions were videotaped.

Video analysis

To evaluate the quality of teacher candidates' explanations, we conducted a video analysis. In order to assess each of the five quality aspects from the assessment framework of Findeisen (2017), we used rating scales that had been used in our previous work (Findeisen 2017). The quality aspects were *Content*, *Student-teacher interaction*, *Process structure*, *Representation*, and *Language*, rated on a four-point Likert scale ranging from 1



Fig. 1 Experimental set-up

(*candidate does not comply with the quality requirements*) to 4 (*candidate fully complies with the quality requirements*). After an intensive coder training, three independent raters coded the material (156 videos in total; about 10 minutes per video). Around 50% of the material (82 videos) was rated by two different coders, in order to ensure these high-inference ratings were as objective as possible. We use Spearman's rank correlation as measures of interrater reliability, which is recommended for ratings (Wirtz and Caspar 2002). The results showed satisfying to sufficient agreement for each aspect (*Content*: .79; *Student-teacher interaction*: .65; *Process structure*: .62; *Representation*: .60; *Language*: .69; see Field 2011, p. 170). We used the mean of all five quality aspects as a measure of teacher candidates' explaining skills (overall quality of explanation).

Data analysis

In order to test our hypotheses, we used a mixed ANOVA that combined analysis of dependent within-subjects factors (repeated-measured variables: in this case, teacher candidates' explaining skills at two points in time) with independent between-subjects factors (in this case, treatment group [module participants] vs. control group). This design allowed us to differentiate between group effects (differences between treatment and control group), time effects (development over time), and interaction effects (time \times group). An interaction effect would be evident if the development of the teacher candidates' explaining skills differed between the treatment and control groups. We tested for the assumption of homogeneity of variance using Levene's test. The results showed equal variances for nine out of ten variables. For *Student-teacher interaction* at the second measurement point, the variances were significantly different in the two groups ($F_{(1, 76)} = 17.71, p = .000$). As the sample sizes did not differ greatly, we expected the ANOVA to be robust. However, in order to make sure not to overestimate effects, we additionally applied non-parametric tests (Friedman's test, Mann-Whitney U test) to assess time effects and group effects on the variable *Student-teacher interaction*.

Results

Table 1 shows the means and standard deviations of the measured five aspects at two points in time: before (t_0) and after (t_1) the treatment, for the treatment and control groups.

The treatment group started out with a slightly higher level of skill on all quality aspects. The differences were significant regarding the overall quality of explanations ($M_T=2.90$, $M_C=2.51$, $t=2.95$, $p=.004$) as well as *Process structure* ($M_T=2.38$, $M_C=1.73$, $t=4.45$, $p=.000$) and *Representation* ($M_T=3.02$, $M_C=2.53$, $t=2.78$, $p=.008$). This has to be taken into account when interpreting the results. Differences between treatment and control group in t_0 are explicable by the more committed nature of a graded university module compared with voluntary participation. This also explains the higher variance of explaining skills in the control group compared with the treatment group. It is noteworthy that the treatment group reached higher scores at the second measurement point regarding the quality of *Student-teacher interaction* and *Process structure*, while the aspects *Content*, *Representation*, and *Language* remained essentially constant over time. The control group only seemed to improve with respect to *Student-teacher interaction*. Figure 2 summarizes the means across time according to the experimental condition. The solid line represents the treatment group, the dotted line the control group.

Table 2 shows the results of the ANOVA. As p values alone are confounded indexes that reflect effect size as well as sample size (Lang et al. 1998), we additionally report effect sizes for the single direct effects, as well as for the interaction between the group and the time variables.

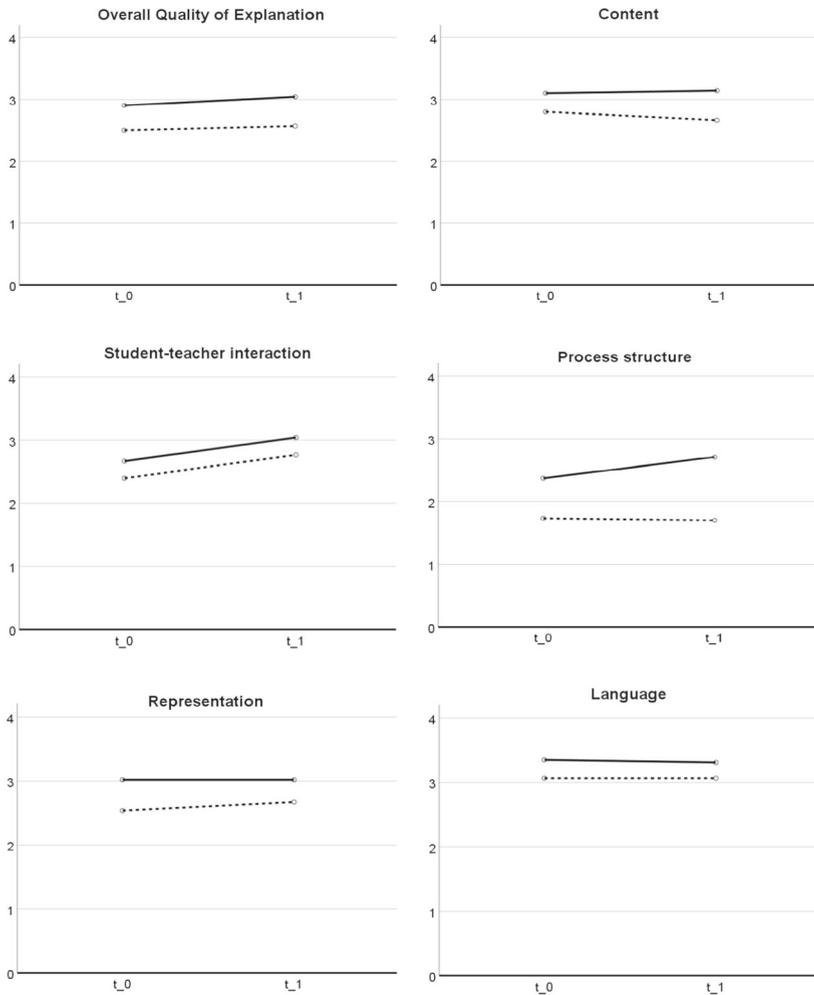
For overall quality of explanation, a medium-sized time effect ($F_{(1, 76)}=6.027$, $r=.28$) and a medium-sized group effect ($F_{(1, 76)}=12.775$, $r=.41$) were found. This means that the treatment group significantly outperformed the control group, but both groups improved over time.

Regarding the five specific quality aspects, we found small- to medium-sized group differences with respect to *Content*, *Representation*, and *Language*; in respect of *Process structure*, there was a large difference between the two groups. Although the treatment

Table 1 Description of variables for the treatment and control groups

Quality aspect	Measurement point	Treatment group		Control group	
		Mean (t)	SD (t)	Mean (c)	SD (c)
Overall quality	t_0	2.90	.533	2.50	.65
	t_1	3.06	.478	2.57	.598
Content	t_0	3.10	.881	2.80	.925
	t_1	3.15	.825	2.67	.922
Student-teacher interaction	t_0	2.67	.883	2.40	1.003
	t_1	3.04	.617	2.77	.971
Process structure	t_0	2.37	.606	1.73	.640
	t_1	2.71	.683	1.70	.651
Representation	t_0	3.02	.635	2.53	.819
	t_1	3.02	.668	2.67	.758
Language	t_0	3.35	.729	3.07	.640
	t_1	3.31	.589	3.07	.691

Repeat-measured, four-point Likert scale from 1 (*candidate does not comply with the quality requirements*) to 4 (*candidate fully complies with the quality requirements*)



Note. The two measurement points (t_0 and t_1) are displayed on horizontal axis, the vertical axis portrays the mean quality of the respective aspect on a four-point Likert scale from 1 (*candidate does not comply with the quality requirements*) to 4 (*candidate fully complies with the quality requirements*).

Fig. 2 Development of treatment group (solid line) and control group (dotted line) for different quality aspects between t_0 and t_1

group improved slightly with respect to *Content*, while the control group deteriorated slightly in the second explanation, the interaction effect was not significant. Hence, there was no generalizable treatment effect on the quality of *Content* in the explanations, and we reject h_1 .

Moreover, we found an effect of time on the variables *Student-teacher interaction* ($F_{(1, 76)} = 19.191$, $r = .45$) and *Process structure* ($F_{(1, 76)} = 3.992$, $r = .22$), indicating that all students performed better in their second performance in respect of these quality aspects. However, while there was a statistically significant interaction effect between the group and the time variables regarding *Process structure* ($F_{(1, 76)} = 5.963$, $r = .27$), meaning that the experimental

Table 2 Treatment and time effects

Quality aspect	Variable	Mean square	<i>F</i>	<i>p</i> value	Effect size <i>r</i>
Overall explaining skills	Time	.401	6.027	.016	.28
	Group	6.987	12.775	.001	.41
	Time × group	.052	.781	.380	.10
Content	Time	.078	.285	.595	.06
	Group	5.664	4.427	.039	.23
	Time × group	.283	1.038	.311	.12
Student-teacher interaction	Time	5.078	19.191	.000	.45
	Group	2.708	2.261	.137	.13
	Time × group	.001	.002	.961	.00
Process structure	Time	.831	3.992	.049	.22
	Group	25.131	40.195	.000	.59
	Time × group	1.241	5.963	.017	.27
Representation	Time	.164	.632	.429	.09
	Group	6.539	8.818	.004	.32
	Time × group	.164	.632	.429	.09
Language	Time	.016	.102	.750	.04
	Group	2.626	3.624	.061	.21
	Time × group	.016	.102	.750	.04

According to Cohen (1988), the effect size measure *r* calculating the strength of the relationship between the relative movements of two variables can be interpreted as follows: $r \geq .1$: small effect; $.3 \leq r < .5$: medium effect; $r \geq .5$: large effect

$df = 1$

group improved significantly more than the control group as a result of the treatment (supporting h_3), we found no interaction on *Student-teacher interaction*.¹ Since the control group also improved significantly with respect to *Student-teacher interaction*, the improvement cannot be attributed to our treatment (rejecting h_2). *Representation* and *Language* showed no effect that can be attributed to the treatment or to repetition, leading us to reject h_4 and h_5 .

Discussion

This study investigated the extent to which a specific treatment improved teacher candidates' explaining skills. Using a treatment and control group design, we analyzed two different research questions. First, we examined whether prospective teachers' explaining skills improved at all, on the basis of an intervention. Second, we differentiated between five different quality aspects to analyze which aspects of teacher candidates' explaining skills were found to improve in a short-term treatment and which remained relatively static. Concerning the first research question, we showed that explaining skills improved between the two measurement points. However, although there were some significant differences between the treatment and

¹ As the assumption of homogeneity of variance was violated for *Student-teacher interaction* at the second measurement point, we additionally applied non-parametric tests. These tests support the findings of the mixed ANOVA. A Friedman test (non-parametric repeated-measures ANOVA) also revealed significant time effects ($\chi^2(1) = 16.0, p = .000$). The Mann-Whitney *U* test showed no significant group differences at t_0 ($U = 614.0, z = -1.147, p = .252, r = -.13$) as well as at t_1 ($U = 596.0, z = -1.372, p = .170, r = -.16$). Hence, the results of the ANOVA were robust.

control groups, the improvement cannot be attributed to the treatment, as both groups improved roughly the same, with only a small and non-significant advantage for the treatment group. At the same time, it becomes apparent that there were different effects for different quality aspects (research question 2). Teacher candidates' explanations improved over time only in respect of *Student-teacher interaction* and *Process structure*. Moreover, *Process structure* appears to be the only quality aspect that was affected by the treatment: there was a significant interaction effect on this variable. Finally, there were neither time effects nor treatment effects for the quality aspects of *Content*, *Representation*, and *Language*.

It is notable that teacher candidates improved particularly in respect of those quality aspects where they had demonstrated greater deficits in their first performance (*Student-teacher interaction* and *Process structure*). For the variable *Student-teacher interaction*, the improvement appeared to be due to a repetition effect. Hence, participants in both the treatment and control groups learnt from their experience in the first simulation. However, when it comes to *Process structure*, we found a significant treatment effect. This finding suggests that after the treatment, teacher candidates were better able to clarify the aim of an explanation and more aware of the importance of assessing students' prior knowledge and understanding. This suggests that not all aspects of an explanation improve naturally through practice. In respect of the remaining three quality aspects, there are several possible explanations for the absence of time and treatment effects. It seems plausible that with respect to some aspects (especially *Language*), it is harder to reach improvements in a short period of time, as they require a greater investment of time and effort. This is in line with the results of Brown and Daines (1981). When it comes to the *Content* of the explanation, it is not too surprising that there was no significant improvement over time, as participants could use the same materials for both explanations, and explicit discussion of the explanatory content for this particular topic was not part of the module. This we believe is a significant opportunity for improvement of the module in the future. However, it was unexpected that the treatment did not affect the quality of *Representations*. As we discussed the importance of representations in the module and talked, for example, about how to tailor examples to different types of learners, we would have expected significant effects in the treatment group. One possible explanation is that the standardized simulation was not difficult enough. As the individuals had already reached high-quality ratings on the aspects *Content*, *Representation*, and *Language*, ceiling effects at individual level could conceivably conceal potential improvements at a group level. Finally, one could also question the quality of the treatment. However, the module was well received by the students and achieved good student evaluations (average: 1.8; scale from 1: very good to 5: insufficient).

The study was subject to several limitations that should be recognized and that may inform future research endeavors. In relation to the method of assessing typical teaching tasks, the simulation exercise that was used in this study proved to be a valuable tool to assess explaining skills as it (1) provided a standardized setting for all participants and (2) allowed concentrating on explaining skills while isolating other aspects (e.g., handling classroom disruptions). However, it has to be noted that the simulation is a simplification of reality, since in the classroom a teacher would have to interact with more than one student simultaneously. Hence, it is yet to be determined whether someone who performed well in the simulation would do so in a real classroom situation. Nonetheless, only with such an approach is the treatment group strictly comparable with the control group. This is clearly an advantage of this study, compared with prior research and a possible cause of the partly deviant results. In addition, in this research, we compared general quality ratings on a four-point Likert scale for each of the five

aspects as a whole. It is possible that participants improved with respect to single features of a given quality aspect (e.g., reduction of errors [*content*]; use of examples [*representation*]). More detailed analyses of the video data will be carried out to further examine more subtle effects of the treatment. This approach would not however negate the finding of little noticeable improvement overall from t_0 to t_1 , in our compressed measurement approach. Finally, conditions for the treatment and control group were not entirely equal. Since teacher candidates in the treatment group were part of a graded university module, they might have taken the explanation task more seriously. This could explain the group differences concerning quality of explanation. It does not, however, negate the findings of insignificant or absent improvement from t_0 to t_1 .

Our findings have implications for both research and teacher education. We see one paramount finding of this study as the possibility of the overestimation of treatment effects where no control group is used, as in such circumstances it is unclear whether the development occurs as a result of university teacher education, through targeted theoretical and reflective inputs, or whether such development would have occurred regardless of such inputs. Moreover, quite aside from the general absence of studies using a control group (with the exception of Miltz 1972), most studies in this field have focused on individual cases or on very small sample sizes—mostly in non-standardized settings.

Our results, gained through application of a controlled research design, show that only some aspects of explaining skills are learnable in a short training module. Learning effects can be partly attributed to the instructional treatment received, as the improvement for the *process structure* of the explanation shows. However, we also found a repetition effect for *student-teacher interaction*. This finding, following the idea of expertise research (Berliner 1994; Ericsson 1996), supports the view of teaching skill development as a process premised on practice. Both causes (treatment and repetitive exercise) call for more opportunities to practice explanations in teacher education. Teacher education programs should therefore implement opportunities to actively practice providing instructional explanations, guided by theoretical and reflective inputs in order to optimize learning progress. Here, interactive simulations are a valuable tool, as they allow not only for systematic reflection by teacher candidates but also for controlled conditions, allowing educators to specifically design learning situations as well as to compare and evaluate results.

Funding Information Open Access funding provided by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, 29(1), 14–17, 20–22, 43–46.
- Berliner, D. C. (1994). Expertise: the wonder of exemplary performances. In J. N. Mangieri & C. C. Block (Eds.), *Creating powerful thinking in teachers and students* (pp. 161–186). Fort Worth: Holt, Rinehart & Winston.
- Blömeke, S., & Delaney, S. (2014). Assessment of teacher knowledge across countries: a review of the state of research. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & W. H. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn: TEDS-M results* (pp. 541–585). Dordrecht: Springer.
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26(4), 473–498. <https://doi.org/10.3102/00028312026004473>.
- Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, 23(3), 194–222.
- Brown, G. A. (2006). Explaining. In O. Hargie (Ed.), *The handbook of communication skills* (3rd ed., pp. 195–228). London: Routledge.
- Brown, G. A., & Daines, J. M. (1981). Can explaining be learnt? Some lecturers' views. *Higher Education*, 10, 573–580.
- Cabello Gonzalez, V. M. (2013). *Developing skills to explain scientific concepts during initial teacher education: the role of peer assessment (doctoral dissertation)*. University of Dundee.
- Charalambous, C. Y. (2008). *Preservice teachers' mathematical knowledge for teaching and their performance in selected teaching practices: Exploring a complex relationship (unpublished doctoral dissertation)*. University of Michigan.
- Charalambous, C. Y. (2016). Investigating the knowledge needed for teaching mathematics. *Journal of Teacher Education*, 67(3), 220–237. <https://doi.org/10.1177/00224871166634168>.
- Charalambous, C. Y., Hill, H. C., & Ball, D. L. (2011). Prospective teachers' learning to provide instructional explanations: how does it look and what might it take? *Journal of Mathematics Teacher Education*, 14(6), 441–463.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science*, 13(2), 145–182. [https://doi.org/10.1016/0364-0213\(89\)90002-5](https://doi.org/10.1016/0364-0213(89)90002-5).
- Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3.
- Clermont, C. P., Krajcik, J. S., & Borko, H. (1993). The influence of an intensive in-service workshop on pedagogical content knowledge growth among novice chemical demonstrators. *Journal of Research in Science Teaching*, 30(1), 21–43.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hoboken: Taylor & Francis.
- Dotger, B. H. (2009). From a medicinal to educational context: Implementing a signature pedagogy for enhanced parent-teacher communication. *Journal of Education for Teaching*, 35(1), 93–94. <https://doi.org/10.1080/02607470802587186>.
- Dotger, B. H. (2013). *"I had no idea!": clinical simulations for teacher development*. Charlotte: Information Age Publishing.
- Dotger, B. H., Masingila, J. O., Bearkland, M., & Dotger, S. C. (2014). Exploring iconic interpretation and mathematics teacher development through clinical simulations. *Journal of Mathematics Teacher Education*, 18(6), 577–601. <https://doi.org/10.1007/s10857-014-9290-7>.
- Duffy, G. G., Roehler, L. R., Meloth, M. S., & Vavrus, L. G. (1986). Conceptualizing instructional explanation. *Teaching and Teacher Education*, 2(3), 197–214.
- Eisenhart, M., Borko, H., Underhill, R. G., Brown, C. A., Jones, D., & Agard, P. C. (1993). Conceptual knowledge falls through the cracks: complexities of learning to teach mathematics for understanding. *Journal for Research in Mathematics Education*, 24(1), 8–40. <https://doi.org/10.2307/749384>.
- Elliott, J. (1981). *Action Research: a framework for self-evaluation in schools*. Cambridge: Cambridge Institute of Education.
- Elliott, J. (1991). *Action research for educational change*. Buckingham: Open University Press.
- Ericsson, K. A. (1996). The acquisition of expert performance: an introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts, sciences, sports, and games* (pp. 1–50). Hillsdale: Lawrence Erlbaum Associates.
- Evans, W. E., & Guyson, R. E. (1978). *Clarity of explanation: a powerful indicator of teacher effectiveness*. Toronto: Paper presented at the Annual Meeting of the American Educational Research Association.

- Field, A. (2011). *Discovering statistics using SPSS (and sex and drugs and rock 'n' roll)* (3rd ed.). Los Angeles: Sage.
- Findeisen, S. (2017). *Fachdidaktische Kompetenzen angehender Lehrpersonen. Eine Untersuchung zum Erklären im Rechnungswesen [Professional Competences of Prospective Teachers. An Analysis of Instructional Explanations in Accounting Education]*. Wiesbaden: Springer.
- Fritsch, S., Berger, S., Seiffried, J., Bouley, F., Wuttke, E., Schnick-Vollmer, K., & Schmitz, B. (2015). The impact of university teacher training on prospective teachers' CK and PCK – a comparison between Austria and Germany. *Empirical Research in Vocational Education and Training*, 7(1). <https://doi.org/10.1186/s40461-015-0014-8>.
- Gage, N. L., Belgrad, M., Dell, D., Hiller, J. E., Rosenshine, B., & Unruh, W. R. (1968). *Explorations of the teacher's effectiveness in explaining (technical report no. 4)*. Stanford University: Stanford Center for Research and Development in Teaching.
- Geelan, D. (2013). Teacher explanation of physics concepts: a video study. *Research in Science Education*, 43(5), 1751–1762. <https://doi.org/10.1007/s11165-012-9336-8>.
- Guler, M., & Celik, D. (2016). A research on future mathematics teachers' instructional explanations: the case of algebra. *Educational Research and Reviews*, 11(16), 1500–1508. <https://doi.org/10.5897/ERR2016.2823>.
- Halim, L. (1998). Improving science education in schools from the perspective of teacher training. *Journal of Science and Mathematics Education in Southeast Asia*, 21(2), 19–28.
- Halim, L., & Meerah, S. M. (2002). Science trainee teachers' pedagogical content knowledge and its influence on physics teaching. *Research in Science & Technological Education*, 20(2), 215–225.
- Hargie, O. (2011). *Skilled interpersonal communication: research, theory, and practice* (5th ed.). London, New York: Routledge.
- Hattie, J. A. C. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: an exploratory study. *Cognition and Instruction*, 26(4), 430–511. <https://doi.org/10.1080/07370000802177235>.
- Hines, C. V., Cruickshank, D. R., & Kennedy, J. J. (1985). Teacher clarity and its relationship to student achievement and satisfaction. *American Educational Research Journal*, 22(1), 87–99. <https://doi.org/10.3102/00028312022001087>.
- Inoue, N. (2009). Rehearsing to teach: content-specific deconstruction of instructional explanations in pre-service teacher training. *Journal of Education for Teaching: International Research and Pedagogy*, 35(1), 47–60. <https://doi.org/10.1080/02607470802587137>.
- Kennedy, J. (1996). Classroom explanatory discourse: a look at how teachers explain things to their students. *Language Awareness*, 5(1), 26–39. <https://doi.org/10.1080/09658416.1996.9959889>.
- Kinach, B. M. (2002a). Understanding and learning-to-explain by representing mathematics: epistemological dilemmas facing teacher educators in the secondary mathematics “methods” course. *Journal of Mathematics Teacher Education*, 5(2), 153–186. <https://doi.org/10.1023/A:1015822104536>.
- Kinach, B. M. (2002b). A cognitive strategy for developing pedagogical content knowledge in the secondary mathematics methods course: toward a model of effective practice. *Teaching and Teacher Education*, 18(1), 51–71. [https://doi.org/10.1016/S0742-051X\(01\)00050-6](https://doi.org/10.1016/S0742-051X(01)00050-6).
- Kleickmann, T., & Anders, Y. (2013). Learning at university. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: results from the COACTIV project* (pp. 321–332). New York: Springer.
- Kulgemeyer, C., & Riese, J. (2018). From professional knowledge to professional performance: the impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, 30(14), 1393–1418. <https://doi.org/10.1002/tea.21457>.
- Kulgemeyer, C., & Schecker, H. (2013). Students explaining science – assessment of science communication competence. *Research in Science Education*, 43(6), 2235–2256.
- Kulgemeyer, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Reinhold, P., Riese, J., Schecker, H., Schröder, J., & Vogelsang, C. (2020). Professional knowledge affects action-related skills: the development of preservice physics teachers' explaining skills during a field experience. *Journal of Research in Science Teaching*. Advance online publication. <https://doi.org/10.1002/tea.21632>.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers: results from the COACTIV project*. New York: Springer.
- Lang, J. M., Rothman, K. J., & Cann, C. I. (1998). That confounded p-value. *Epidemiology*, 9(1), 7–8.
- Leinhardt, G. (1987). Development of an expert explanation: an analysis of a sequence of subtraction lessons. *Cognition and Instruction*, 4(4), 225–282.

- Leinhardt, G. (1989). Math lessons: a contrast of novice and expert competence. *Journal of Research in Mathematics Education*, 20(1), 52–75.
- Leinhardt, G. (1997). Instructional explanations in history. *International Journal of Educational Research*, 27(3), 221–232. [https://doi.org/10.1016/S0883-0355\(97\)89730-3](https://doi.org/10.1016/S0883-0355(97)89730-3).
- Leinhardt, G. (2010). Introduction: explaining instructional explanations. In M. K. Stein & L. Kucan (Eds.), *Instructional explanations in the disciplines* (pp. 1–5). New York, Dordrecht, Heidelberg, London: Springer.
- Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78(2), 75–95. <https://doi.org/10.1037/0022-0663.78.2.75>.
- Ma, L. (2010). *Knowing and teaching elementary mathematics: teachers' understanding of fundamental mathematics in China and the United States* (2nd ed.). Hoboken: Taylor & Francis.
- Miltz, R. (1972). *Development and evaluation of a manual for improving teachers' explanations (technical report no. 24)*. Stanford University: Stanford Center for Research and Development in Teaching.
- Sánchez, E., Rosales, J., & Cañedo, I. (1999). Understanding and communication in expository discourse: an analysis of the strategies used by expert and preservice teachers. *Teaching and Teacher Education*, 15(1), 37–58. [https://doi.org/10.1016/S0742-051X\(98\)00033-X](https://doi.org/10.1016/S0742-051X(98)00033-X).
- Sevian, H., & Gonsalves, L. (2008). Analysing how scientists explain their research: a rubric for measuring the effectiveness of scientific explanations. *International Journal of Science Education*, 30(11), 1441–1467.
- Thanheiser, E. (2009). Preservice elementary school teachers' conception of multidigit whole numbers. *Journal for Research in Mathematics Education*, 40, 251–281.
- Wheeldon, R. (2012). Examining pre-service teachers' use of atomic models in explaining subsequent ionisation energy values. *Journal of Science Education and Technology*, 21(3), 403–422. <https://doi.org/10.1007/s10956-011-9333-0>.
- Wirtz, M., & Caspar, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen [Inter-rater agreement and inter-rater reliability: Methods for determining and improving the reliability of assessments using category systems and rating scales]. Göttingen: Hogrefe.
- Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: a framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, 43(1), 49–64. <https://doi.org/10.1080/00461520701756420>.
- Zemba-Saul, C., Blumenfeld, P., & Krajcik, J. S. (2000). Influence of guided cycles of planning, teaching, and reflection on prospective elementary teachers' science content representations. *Journal of Research in Science Teaching*, 37(4), 318–339. [https://doi.org/10.1002/\(SICI\)1098-2736\(200004\)37:4<318::AID-TEA3>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1098-2736(200004)37:4<318::AID-TEA3>3.0.CO;2-W).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.