Using authentic representations of practice in teacher education: Do direct instructional and problem-based approaches really produce different effects?

Abstract

This paper investigates effects of different learning approaches (problem-based vs. direct instructional) when using authentic representations of practice in teacher education. We assigned 638 student teachers from 21 equivalent professional development courses to one of the two conditions. Student teachers' analyses of practice were evaluated on selective attention, reflective thought, and theory-practice integrations in a pre-post-design. In line with inconsistent findings from prior research, we were able to produce evidence for equivalent effects of the instructional approaches on all depended variables using Bayesian data analyses. As called for in a review on the topic, we additionally explored the role of the instructors administering the field study interventions. Findings revealed that a positive attitude toward the learning approaches they administered was related to more theory-practice integrations in the students' analyses.

Keywords

teacher education, analysis of practice, authentic representations, problem-based learning, direct instruction, attitudes

1. Introduction

Over the past 15 years, there has been a steady increase in research interest on authentic representations of practice (particularly via video) in teacher education. Its substantial potential for professional development has been argued theoretically from the perspectives of professional vision (Sherin & van Es, 2009) or core practices (Grossman & Pupik Dean, 2019), as well as corroborated by empirical studies (Beisiegel, Mitchell & Hill, 2018). In their systematic review, Gaudin and Chaliès (2015) provide an overview of repeatedly focused learning objectives in teacher education: Specific categories of knowledge (e.g., content knowledge, pedagogical content knowledge), forms of knowledge (e.g., case knowledge, strategic knowledge) or the development of skills for interpretation and reflection (Shulman, 1986). There is broad consensus that the latter is a central component of professional development over a teacher's lifetime. In the current paper, we focus on reflection-on-action (Schön, 1983) in teacher education and related concepts. Due to the large amount of research, we first summarize the current state of research, referring to two narrative and two systematic reviews in the field.

2. Current research on authentic representations of practice in teacher education

In a narrative review, Marsh and Mitchell (2014) found learning via video representations to enhance reflection and reflection-related skills in teachers. Investigated concepts were diverse and extended from analysis (e.g., Harford & MacRuairc, 2008) to reflection-on-action (Barnett, 2006) and metacognitive reflection (Charteris & Smardon, 2013). Based on empirical findings (e.g., Mitchell et al., 2010), they argued that collaborative video-based learning is particularly effective due to its discursive nature, which scaffolds the process of reflection. Further, video-based representations facilitate collaborative reflection by allowing many learners to observe the same classroom situations in a comparable way. Compared to collaborative settings, Marsh

and Mitchell discovered less evidence for the effectiveness of video-based teacher education using individualized learning settings. Exceptions constitute mentor-mentee relations, where videos may serve as a stimulus for their interactions. Based on earlier work, Marsh, Mitchell, and Adamczyk (2010) further argued that video-based reflection may foster relations of theory and practice through contextualization and abstraction.

Marsh and Mitchell (2014) also found that video-based learning approaches develop noticing skills. Noticing (or selective attention), as defined by Sherin and van Es (2009), describes the ability to perceive and be aware of relevant, in-depth features of instruction. Novices tend to focus overly on the surface features of classroom interactions (Castro et al., 2005) and on themselves rather than student learning (Colestock & Sherin, 2009). There are numerous studies that show the effectiveness of video-based learning arrangements in supporting the development of noticing (e.g., Tripp & Rich, 2012). However, it appears challenging to determine the instructional approaches most promising for the development of noticing (or to describe their differential effects) due to a wide variety of instructional approaches and operationalizations of noticing implemented in these studies.

In a systematic review, Gaudin and Chaliès (2015) found similar results but focused more narrowly on the teachers' activities of selective attention and knowledge-based reasoning while viewing video. Parallel to noticing, selective attention describes the "ability to identify certain classroom events" (Gaudin & Chaliès, 2015, p. 46), for which they listed several terms and operationalizations from the research literature. Gaudin and Chaliès (2015) gathered empirical evidence for video-based learning arrangements to develop selective attention among pre-service and in-service teachers. Again, the lack of comparability (and detailed description) of instructional approaches in the original studies makes it challenging to estimate their relative

effectiveness. Concerning knowledge-based reasoning, defined as the ability to describe and interpret classroom situations, Gaudin and Chaliès (2015) identified a series of empirical studies that support the effectiveness of video-based learning arrangements. Notably, most of these studies are designed as problem-based learning approaches. Gaudin and Chaliès (2015) argued that problem-based learning plays a key role in the development of interpretation and reflection of practice.

In their narrative review, Orland-Barak and Maskit (2017) utilized a broader approach in which they described methodologies for the mediation of learning. One methodology included the use of "cases as recording experience" (p. 37), in which the authors described representations of practice as raw data for studying classroom practice. For video as well as text representations, they highlighted the potential of this methodology for relating theory and practice and further for enhancing reflection in teacher education programs. Similar to Marsh and Mitchell (2014), they pointed out that video may bring challenges with it that text representations escape, like a high level of complexity for novices or legal and ethical issues. In line with prior reviews, Orland-Barak and Maskit (2017) described another methodology they called "video as observing experience" (p. 51). They emphasized the potential of video to facilitate the relation of theory and practice and the ability to notice and analyze practice. Beside "traditional" observations of video such as video clubs (Sherin & Han, 2004), creating a video-based teaching episode (Li, 2007) appears to foster new perceptions and interpretations of their practice. What's more, Orland-Barak and Maskit (2017) pointed out the effectiveness of video representations as a basis for mentor-mentee interactions (see also Marsh & Mitchell, 2014).

In a systematic review, Baecher, Kung, Ward, and Kern (2018) gave an overview of current practices of facilitating video analysis. The review was descriptive in nature, giving insight into

"how facilitators carry out mediation tasks through video in teacher education" (p. 189). It was aimed at instructors designing professional development for teachers, as well as unveiling challenges and gaps in research. The authors concluded that researchers are "reinventing the wheel" (p. 206): There is a lack of sharing learning materials and profound descriptions of complex treatments, without which, practitioners cannot recreate the arrangements and researchers cannot replicate findings. Baecher et al. (2018) highlighted the role of instructors, which is rarely clarified in scholarly publications. Guiding discussions and structuring analyses of participants, the instructor plays an important part in the success of the learning process (Gröschner et al., 2014; Author, 2018 [details removed for peer review]). Consequently, there exists a large body of exploratory research on a variety of video-based learning arrangements in teacher education. Since the treatments are rarely compared and tested through experimental variation, we will continue to struggle in gathering knowledge about the differential effects of video-based approaches. We emphasize the importance of Baecher et al.'s call for more detailed reporting on the video-based learning arrangement (e.g., using their checklist) to solve the information shortage in scholarly publications. Additionally, to accumulate evidence on the effectiveness of video-based approaches, we need studies conducting experimental variation and comparison.

Summarizing the current state of research, we want to draw attention to six points. First, video-based learning arrangements appear to be a promising approach to develop selective attention and reflection (as well as related concepts). Second, the authors emphasize the benefit of video-based learning for relating theory and practice. Third, collaborative approaches seem to be particularly effective. Further, in the reported studies, the majority of learning approaches are designed as problem-based. Fourth, some authors point out challenges of video-based arrangements and highlight the effectiveness of text-based representations. Therefore, we will

use the term *authentic representations of practice* to include both video and text-based representations. Fifth, there is still little attention being paid to the central role of the instructor. Sixth, there is little robust evidence on the differential effects of instructional approaches using authentic representations of practice. More detailed reporting and experimental approaches are needed to accumulate evidence.

Based on the findings described above, we conducted an experimental field study comparing different instructional approaches and their effects on selective attention, reflective thought (which corresponds to knowledge-based reasoning), and integration of theoretical knowledge with authentic representations of practice. We also address the role of the instructor, focusing on their attitude toward the treatment.

2.1 Problem-Based and Direct Instruction

The instructional approach in which authentic representations are leveraged plays a pivotal role in the effectiveness of the learning outcome (Hatch & Grossman, 2009). These instructional approaches may have differential effects on how learners perceive and reason about practice (Borko et al., 2008). Even though most learning arrangements using authentic representations in teacher education follow a problem-based (PB) approach, it is still unclear which differential effects direct instruction (DI) will evoke, due to lack of empirical studies that systematically compare these two approaches.

The use of authentic representations is an inherent part of PB instruction. It opens a problem space that initiates the learning process. (Fogarty, 1997). Learners explore the problem and engage in problem-solving strategies, such as generating hypotheses and researching and

applying knowledge in a self-regulated manner (Duch, Groh & Allen, 2001). PB learning takes place in small groups, whose members divide tasks, share knowledge and engage in discursive reasoning to reach a conclusion (Savery, 2006). Accordingly, Hmelo-Silver (2004) argues that knowledge-based reasoning constitutes an integral part of PB learning and is therefore practiced in the learning process. In contrast, DI can be described as "providing information that fully explains the concepts and procedures that students are required to learn" (Kirschner et al., 2006, p. 75). Within DI, authentic representations are used to exemplify or practice a solution presented previously by the instructor (Atkinson et al., 2000). Learners may subsequently work on similar authentic representations to replicate the solution process, and thus, do not necessarily reason about different solutions. Thus, from a theoretical perspective, DI would be suitable to practice the selective attention of participants. Moreover, DI as a structured approach may be particularly helpful for novice learners, who may suffer from cognitive overload otherwise (Author, 2016a [details removed for peer review]).

2.2 Empirical Findings on Effects of Different Instructional Approaches

In our study, we compare the two instructional approaches and their effects on student teachers' selective attention, reflective thought, and integration of theoretical knowledge with representations of practice. To our knowledge, there are only two studies directly comparing different instructional strategies. Seidel, Blomberg, and Renkl (2013) compared two different instructional strategies that share characteristics with the DI and PB methods—rule-example versus example-rule. In the rule-example strategy, definitions about the topic (goal clarity, scaffolding, and learning climate) were given and subsequently exemplified in a classroom video. The students were then asked to practice the demonstrated observation on further video representations. The example-rule strategy used group discussions stimulated by observations on classroom videos. These discussions and a subsequent moderated collection of ideas were

intended to lead to an understanding of the topic. Aggregated scores on noticing and evaluating classroom situations showed that the rule-example group outperformed the example-rule group. Unfortunately, the paper did not report separate scores for noticing and evaluation; thus, it is challenging to draw more precise inferences from the effects. In a test on factual knowledge, the rule-example group outperformed the example-rule group. However, when asked to plan a lesson, the example-rule group more frequently elaborated on theoretical ideas that included situational references rather than being merely general.

Barth et al. (2019) compared two collaborative learning approaches using authentic representations of practice (vignettes) that included either direct instructional or traditional problem-based elements. Interventions were implemented in professional development courses for student teachers on the topic of classroom management. In the DI condition students received lecture-like instruction on theoretical aspects of classroom management, then analyzed several vignettes guided by a three-step worksheet that depicted selective attention, knowledge-based reasoning and theory-practice integrations. In the PB condition students first observed one problematic vignette, then independently read theoretical literature on classroom management in a self-study phase. Afterwards, they proceeded to analyze the given vignette using the same worksheet as the DI condition. The intervention therefore focused more on different approaches concerning the acquisition of theoretical knowledge and less on different approaches concerning the analysis process. Measures included selective attention and theory-practice integrations in written analyses on video-based classroom vignettes. Posttest comparison of the groups in a first study yielded no differences on measures of selective attention and a small to medium effect on theory-practice integration favoring the DI group. In a second study Barth et al. (2019) focused solely on the DI group revealing no differences in selective attention and a strong effect on theory-practice integrations from pretest to posttest.

2.3 Empirical Findings on the Effects of Instructors' Attitudes

As Baecher et al. (2018) argued in their systematic review, more attention needs to be directed toward instructors that facilitate the learning arrangements being studied in teacher education. One important aspect when implementing and studying the effects of teaching approaches in field studies are the attitudes of instructors administering the treatment conditions. Instructors of teacher education courses may differ in their attitudes about teaching styles and learning scenarios (Courtland & Leslie, 2010). These attitudes influence the way instructors practice their teaching. Even when facing obstacles (e.g., being told to teach in a different manner), instructors may try to maintain alignment between their attitudes and their practice (Hallett, 2010). Given that field studies involve instructors in the delivery of treatments, their attitudes toward the treatment can be an important source of information and predict outcomes. Several studies revealed that a positive attitude of instructors toward the treatment may have an effect on the subject's performance concerning critical measurements (e.g., seminal first findings by Rosenthal & Fode, 1963). Since there has been little research in this regard in teacher education field studies, this part of our study is exploratory.

2.4 Hypotheses

In this study, we investigate how different learning approaches (PB vs. DI) and the instructors' attitudes are related to student teachers' selective attention, reflective thought, and integration of theoretical knowledge with authentic representations of practice. Based on the lack of robust research in this area all hypotheses were labelled as exploratory. One of the hypotheses on reflective thinking (H_{2a}) constitutes an exception, as strong assumptions can be derived from theory.

1. Selective attention:

a. DI and PB learning approaches both foster selective attention (no difference between groups).
b. Positive attitudes of the instructors about the learning approach are positively related to selective attention.

2. Reflective thought:
   a. A PB learning approach leads to more elaborate reflective thought compared with DI. (confirmatory)
   b. Positive attitudes of the instructors about the learning approach are positively related to reflective thought.

3. Theory-practice integrations:
   a. DI and PB learning approaches both lead to similar amounts of theory-practice integrations (no difference between groups).
   b. Positive attitudes of the instructors about the learning approach are positively related to theory-practice integrations.

Consistent with our hypotheses (H_{1a} and H_{3a}), we use Bayesian inferential statistics in our data analysis. This allows us (in contrast to frequentist null hypothesis testing) to make statements about the equivalence of groups (Quintana & Williams, 2018).

3. Method

3.1 Participants

638 student teachers participated in the study, recruited from 21 introductory courses for secondary teachers. The 21 courses covered the same content on teaching, learning, and instruction over 15 weekly 90-minute sessions. They all occurred in the same academic semester and in the same student teacher program, allowing for comparable conditions. The treatment was a regular part of the course, measurements were administered through an online

pretest and posttest survey. Participants were M_{age} = 21.01 (SD_{age} = 2.36) years old on average and predominantly female (65.0% female, 33.2% male, and 1.8% "other" or none). Students usually attend this course in their second of 10 semesters; hence, 72.3% of the participants were studying in their second semester.

3.2 Design

The study was conducted within our institute's teacher preparation program, allowing for field study conditions. In each of the 21 courses included in the study, one of the two interventions was implemented. For the interventions, we redesigned two of the courses' weekly 90-minute sessions (6^{th} and 7^{th} of 15 sessions) and an assignment between these two sessions using authentic representations of practice. The topic of these two sessions and the assignment was classroom management. The other 13 sessions before and after the treatment remained as planned by the instructors ("business as usual"). Courses were either realized as problem-based (n_{PB} = 11) or direct instructional (n_{DI} = 10). Thus, the interventions were implemented at the course level. Students were assigned to the courses via a university-administrated system we had no influence on. We allocated courses to the different interventions such that each intervention was uniformly distributed across days between Monday and Friday, time slots (8am to 8 pm) and the six instructors. These instructors were the same as those teaching throughout the semester, further strengthening field study conditions. We used the R package 'BayesFactor' (Morey & Rouder, 2018) to test whether conditions were equivalent concerning several potentially confounding context variables. The Bayes factor is a measure of relative evidence comparing two hypotheses, one of which can be specified as a null hypothesis. As opposed to (classical frequentist) null hypothesis significance testing, an analysis using Bayes factors allows us to gather relative evidence for a null hypothesis, and therefore, test for equivalence. In our Bayes factor analyses for two independent samples with standard priors,

we tested the conditions for equivalence (H_0) or differences (H_1) of the groups concerning gender (BF_{10} = .109), teaching experience (BF_{10} = .116), experience with video-based analyses (BF_{10} = .099), topic-related literature read (BF_{10} = .114), and prior knowledge on classroom management theories (BF_{10} = .101). All Bayes factors pointed toward evidence of equivalence between the groups prior to the treatment.

3.3 Treatments, Materials, and Procedure

The procedure of the treatment is shown in Table 1. To ensure that the instructors were well trained conducting the intervention sessions, they taught these sessions one semester before the study and received feedback from two researchers videotaping and observing the sessions. In multiple subsequent meetings with all instructors, we were able to further optimize treatment compliance. Choosing from a pool of 14 normal-practice classroom videos, we found four vignettes suitable as authentic representations of classroom management, each with a duration of approximately 5 minutes. These vignettes were deemed appropriate because the classroom management strategies were particularly visible in them (thus providing an authentic representation) and were appropriate for the student teachers' grade level (secondary). In the two sessions, students learned about the classroom management strategies of Kounin (1970), Evertson (2006), and Mayr (2009). To offer text-based vignettes, we transcribed the video vignettes and added nonverbal information. Text and video vignettes were evenly distributed between the two treatment groups, representing another factor (2 x 2 design) not considered as a predictor in this paper (for description and results see Author, 2016b [details removed for peer review]). One vignette was used in the first session, another as an assignment (to be completed until the second session), and two in the second session.

Table 1

297 Procedure for measurement and intervention.

	Problem-based learning (PB)	*Direct instruction (DI)*
Before intervention (40 minutes)	(Online) pretest (analysis of practice, demographics, and covariates)	
Session 1 (90 minutes)	Instructor lectures on classroom management theories and steps for analyzing classroom vignettes	
	- Students analyze video or text-based vignette in small groups, focusing on classroom management - Students participate in collaborative discussions about interesting situations	- Instructor analyzes a video or text-based vignette on classroom management, step by step - Students replicate the analysis with a new situation
Assignment (60 minutes)	Analysis of a video or text-based vignette at home	
Session 2 (90 minutes)	- Students analyze video or text-based vignette in small groups, focusing on classroom management - Students participate in collaborative discussion about interesting situations	- Instructor analyzes a video or text-based vignette on classroom management, step by step - Students replicate the analysis with a new situation
After intervention (40 minutes)	(Online) posttest (analysis of practice, covariates)	

298

299 The first session had two parts. In the first part, the instructor introduced students to the topic
300 of classroom management and its theoretical approaches (Evertson, 2006; Kounin, 1970; Mayr,
301 2009). Instructors gave definitions and clarified strategies concerning classroom management
302 in a short exercise. Subsequently, students were made familiar with the steps of how to analyze
303 the authentic representations of practice (i.e. vignettes): “Describe the problem/situation,
304 describe the teacher’s action, reason about alternative courses of action, anticipate reasons and
305 consequences of these actions, decide on one of the alternatives”. This first part was taught

equally over all courses of both conditions. In contrast, the second part varied between the two conditions.

3.3.1 Direct Instructional Treatment

In the second part, instructors introduced students to the lesson represented in the vignette, giving background information on the topic, the lesson structure, the class level, and the specific sequence of the lesson they were about to witness. The vignette was then shown to the entire course without interruption, so that students could visualize the classroom activities. In a second round, the instructor stopped at certain situations and demonstrated a step-by-step analysis that integrated theoretical aspects of classroom management. After the situation was sufficiently analyzed, the instructor continued by repeating the analysis with several more situations. After this, students analyzed some more situations chosen by the instructor. These analyses were performed as a dialogue with the instructor and other students contributing. In the second session, the same procedure of demonstration and exercise was repeated for two additional vignettes.

3.3.2 Problem-Based Treatment

Problem-based courses also started with the instructor giving background information on the subsequent vignette. After that, the vignettes were handed out either as text vignettes on paper or video vignettes viewed on laptops. Students came together in groups of four to five members. The students, not the instructor, selected which situations they analyzed. They observed several situations and discussed the analysis. To guide the analysis, students received key questions that targeted the analysis of practice steps. In a final discussion, the whole course talked about two or three of these situations. Students were asked to describe situations that stood out to them and perform analyses of them. The instructor acted as a moderator and tried to include

different students' suggestions about the situation and promote a discursive discussion. In the second session, the same procedure of small group discussion and final course discussion was repeated with two vignettes.

Treatments checks were administered by one of three trained raters who visited the treatment sessions. They judged the implementation of the treatment on eight items that measured the degree of DI of the instructor on a 6-point Likert scale (from "doesn't apply at all" to "applies completely"). A (reversed) example item is as follows: "Students chose which situations to have a closer look at while working with the vignettes." Raters also recorded the time students were effectively able to work on the vignettes. For both measures, the inter-rater reliability showed good intraclass correlation (ICC = .96–.99). Internal consistency concerning the treatment check scale showed good scores (Cronbach's α = .96 for both sessions). In both sessions, we found evidence of equivalence between the groups for the time students worked on the vignettes (1st session: $BF_{10} = .514$; 2nd session: $BF_{10} = .545$). In contrast, we were able to provide evidence of differences in the degree of direct instruction for the treatments (1st session: $BF_{10} = 3.76 \cdot 10^6$; 2nd session: $BF_{10} = 2.34 \cdot 10^5$) using Bayes factor analyses for two independent samples with standard priors. These findings corroborate that the instructors realized the different instructional methods as planned with comparable times on task.

3.4 Measures

Dependent variables and covariates were assessed via an online survey the students were asked to complete as part of the course, before (pretest) and after (posttest) the two treatment sessions. As stated in the research questions, we are looking into the dependent variables of selective attention, reflective thought, and integration of theoretical knowledge with authentic representations of practice. These three variables were assessed by means of written comments

that students gave on classroom vignettes presented in the survey. Students were instructed to comment on situations they perceived relevant in the topic of classroom management. Since students might have been familiar with the idea but not with the term classroom management, we asked them to observe the lesson planning, control of behavior, and shaping of relationship witnessed in the vignettes—the three dimensions of classroom management that were taught in the treatment sessions. They were instructed to write and save each analysis separately using a "save comment" button below the text box (Fig. 1). After saving an analysis, they were free to continue the observation and write further analyses.

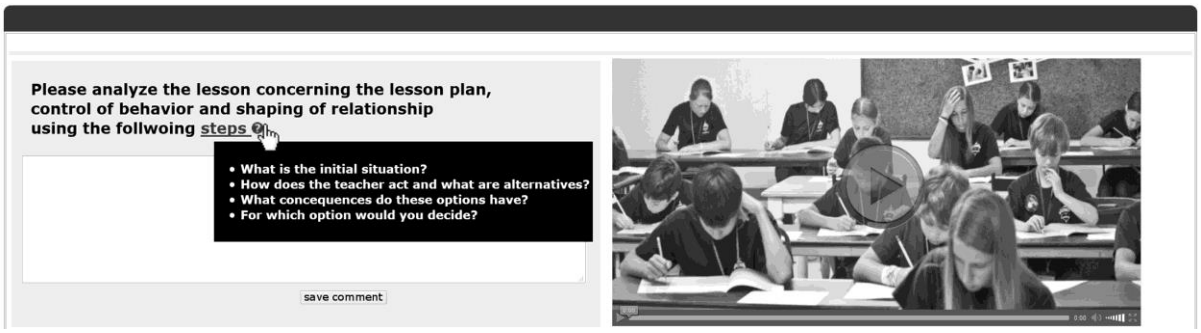


Fig. 1. Screenshot of the web-based survey: One of three vignettes to be analyzed by students.

The pretest and posttest each showed three vignettes of a classroom video that add up to ten minutes. To avoid a test effect, different vignettes from the same videotaped classroom lesson were used in the pretest and posttest. Vignettes were matched to either the pretest or posttest based on the ratings of three experts (in the fields of practice-oriented teacher education and classroom management) on 14 dimensions (e.g., complex, interesting, or classroom management) with a mean (standard deviation) agreement of r_{WG}= .79 (.31). We used these ratings to conduct cluster and graphical analyses to find three pairs of similar vignettes that were then separately allocated to either the pretest or posttest.

3.4.1 Quantity of Selected Situations as Selective Attention

Selective attention as defined above denotes the ability to be aware of relevant classroom features. In teacher education programs, what is defined as relevant may depend on the current learning goal of the course or learning arrangement (e.g., teacher guidance and support; Seidel & Stürmer, 2014). In our study, we focus on *classroom management*; the topic of the two sessions in the course our treatment took place.

Literature adopting a theoretical perspective on selective attention, it is generally placed in a larger context of reflection (e.g., van Es & Sherin, 2008). Based on this notion, we operationalize selective attention based on a reconstruction of reflection in the literature. The research literature showed a broad consensus that reflection on teaching plays a central role in the development of professional teachers (Day, 1993). However, there is a wide variety of definitions of reflective thinking (Loughran, 2002), many of which derive their ideas from Dewey's thoughts on an "analysis of a complete act of thought" (Dewey, 1910, p. 68). The thinking process in reflection consists of two main elements—the selection of situations and the inquiry about these situations (Dewey, 1910). Dewey explains the selection of situations based on its potential to respond to it in multiple different ways without one being the obvious way to go from the start (like standing in front of an unknown crossroad). Selected situations are characterized as having attracted one's attention and been deemed worthy of further inquiry (Zeichner, 1981). Thus, we operationalize selective attention as situations the participants select to discuss a relevant topic, where "relevant" means that the selected situations are instances of the focused topic of our study, "classroom management." We use the quantity of selected situations in the web-based survey students commented on (i.e., number of saved analyses). Participants were instructed to comment on every situation they perceived as being relevant concerning the classroom management in the vignette. Therefore, they were able to save as many analyses as they pleased. A trained coder rated whether the analyses focused on

classroom management or were off topic. Counts were summed up for each of the three vignettes in the pretest and posttest, thus constituting the number of selected (and analyzed) situations for each vignette.

3.4.2 Realized Inquiry Steps Representing Reflective Thought

From a theoretical stance on reflection, reasoning about different possible options to the selected situation (labeled as "inquiry") constitutes reflective thought (Dewey, 1910). Therefore, one needs to pass through several processes, which are as follows: describing the selected situation, exploring options for the situation, anticipating consequences, and making a decision based on these processes. Reflective thought may mean jumping back and forth between these processes. Thus, what matters is not the order in which they are carried out, but whether the processes are carried out. We focus on the processes of *describing the situation (e.g., the teacher's action)*, *anticipating alternatives*, and inquiry about *consequences* that constitute the heart of reflective thought (Dewey, 1910).

We operationalized reflective thought by students' ability to apply the inquiry steps to the selected situations from the vignettes in the online survey. On the survey, students were reminded about the inquiry steps in the item question (see Fig. 1). Thus, we measured whether students could *transfer and apply* these steps on the practice situations observed rather than whether they remembered the steps. The comments written and saved by the students were coded as whether they contained the individual inquiry steps (dichotomous). The scale from 0 to 3 categorizes whether we detected none, one, two, or all of these steps in each subject's written comments (Table 2).

Table 2

Levels of reflective thought based on coded inquiry steps.

Level	Steps realized	Example
3	- Description, alternatives, and consequences	The teacher collects the notes from the students. It becomes increasingly loud in the classroom as the teacher hangs notes on the blackboard. The teacher gives the students a warning. It gets quieter. Overall, it would be better to include the students more in hanging up the notes. That would enable a quieter learning situation.
2	- Description and alternatives	Students insult each other, but the teacher tries to ignore it by continuing with the lesson. She should really try to stop the insults before recommencing.
	- Description and consequences	While checking the results of the assignment, the teacher praises the students a lot. That'll motivate them.
	- Alternatives and consequences	The class is loud as some students begin to present their results. The teacher should not let them begin to present until the rest of the class is quiet. Thus, half the information wouldn't be lost.
1	- Description	There was quite a bit of noise in the classroom, then the teacher placed her finger on the lips.
	- Alternatives	The teacher should react to the student wandering around the room.
	- Consequences	The students don't seem to heed the teacher's actions.
0	none	Everything seems normal.

Inter-rater reliability was satisfactory (κ = .64–.77). Each vignette was tested for one-dimensionality using confirmatory factor analysis (CFA) with individual comments used as indicators and the robust maximum likelihood estimator (MLR) obtained by full information maximum likelihood (FIML; Praetorius et al., 2012). The data showed a good fit regarding all six vignettes, $p[\chi^2]$ = .13–.99, CFI = .95–1, RMSEA ≤ .001–.03, p[RMSEA < .05] = .86–1. Furthermore, model comparisons indicated that we were able to assume strict measurement invariance for all vignettes between treatment groups. In addition, reliability between the comments revealed good internal consistency for all vignettes, McDonald's ω = .70–.80 (McDonald, 1999). Thus, we computed mean scores for each vignette in the pretest and posttest, reflecting the average quality of reflective thought.

3.4.3 Theory–Practice Integration

We assessed students' theory–practice integration regarding classroom management principles by coding their written analyses for terms and principles from the classroom management literature. When detected, coders evaluated whether the theoretical principle fit the situation to which the comment referred. We did not decide to use a sample solution (predetermined list of theoretical principles associated with specific situations) because it would not do justice to the complexity of classroom situations. There can be no certainty about what students were truly referring to within these situations when only coding for predetermined theoretical principles while excluding further information from the participants' view. Written analyses that referred to theoretical principles of classroom management that fit the situation described were coded dichotomously with 1 if they met these criteria (e.g., "the teacher does not manage to show withitness by maintaining eye contact and keeping students under control or quiet") or 0 if they did not fulfill these criteria.

Inter-rater-reliability was satisfactory ($\kappa = .74$). We computed mean scores for each vignette in the pretest and posttest, reflecting the percentage of comments that included theoretical principles of classroom management.

3.4.4 Theoretical Literature on Classroom Management

We allege that the students were unfamiliar with the topic of classroom management since the teacher education program's curriculum did not cover it up to this point (2nd semester). Concerning the two sessions on classroom management, three recommended readings (by Kounin, Evertson, and Mayr) were provided for download on the university's content management system used for this course. As theory–practice integration was part of the treatment and a central dependent variable, we needed to control for different preconditions of

theoretical knowledge caused by sources outside the treatment sessions. Students were thus asked which of the articles they had read. After the two treatment sessions, 41% of the students read none, and 28% read all three texts (median = 1). The theoretical literature read was used as a control variable.

3.4.5 Attendance

As the treatment was part of two sessions of a regular course, attendance of students may have varied and influenced the efficacy of the treatment. The more sessions they attended, the more the treatment could influence their cognition and behavior. Therefore, we asked the subjects how many of the two treatment sessions they attended (none, one, or two). The maximum of two sessions were attended by 79%, whereas 20% attended one session, and 1% attended neither of the two sessions. Attendance was used as a control variable.

3.4.6 Instructors' Positive Attitude Toward the Treatment

During the yearlong training instructors received for the intervention, we noticed their divergent attitude toward the two learning approaches. To assess their attitudes, we separately asked for both treatments they conducted: "Think about the concept of the PB [DI] course: I like the way learning with classroom vignettes is handled here." Neither of the two treatments seemed to be tendentially more popular among all the instructors: On the 6-point Likert scale (1 = disagree strongly to 6 = agree strongly), instructors' attitudes ranged from 2 to 6 for the problem-based treatment and 1 to 6 concerning the direct instructional treatment. This indicates that there is no consistent bias with one or the other treatment, thus, we used it as a covariate, rather than a mediator, of the treatment.

3.5 Statistical Analysis

The collected data contained 18% missing values overall. Therefore, in the first step, we decided to impute the data via chained equation with the R package 'mice' (van Buuren & Groothuis-Oudshoorn, 2011). Second, with the complete datasets, we computed separate full structural equation models for each of the three dependent variables (selective attention, reflective thought, and theory–practice integration) with the predictors' treatment (0: DI, 1: PB), attendance, theoretical literature on classroom management, and instructors' attitudes about the treatment (Fig. 2). Note that the three vignettes we used to measure the dependent variables in the pretest and posttest were matched with great effort, yet they were not exactly the same. As a result, comparisons between the pretest and posttest on absolute scores must be interpreted carefully for all three dependent variables. Hence, we preferred to use the pretest scores as predictors of the posttest, accounting for differences before the treatment. Other predictors can be interpreted as increasing or decreasing the posttest's selective attention score under the control of pretest scores. Reflecting our study design, we obtained clustered data. Given our research interest, we are not interested in modeling these clustered data, but consider them a nuisance (Snijders & Bosker, 2012). Accordingly, we used robust standard errors and adjusted χ^2 that take the clustered structure into account.

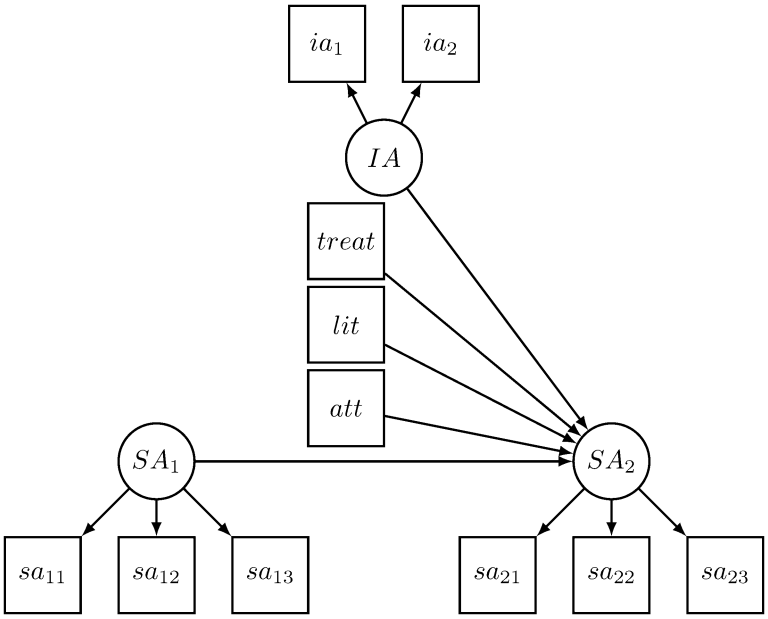


Fig. 2. Computed structure model for one dependent variable (i.e., selective attention, models on reflective thought and theory–practice integration structured accordingly). *SA*: Latent variable "selective attention" in pretest (SA_1) as control variable and posttest (SA_2) as dependent variable; *sa*: Manifest variables of "selective attention" representing the three vignettes within each measurement point; *att*: Attendance; *lit*: Theoretical literature on classroom management; *treat*: Treatment; *IA*: Latent variable of instructors' attitudes about the treatment; *ia*: Manifest variables of instructors' attitudes.

Fit of the three models were good: Taking the $N = 638$ students into account, it is not surprising that the χ^2-test shows a significant result ($\chi^2_{(40)} = 56.743–67.988$, $p = .008–.053$). Furthermore, the CFI = .962–.981, TLI = .950–975, and RMSEA = .035–.033 with $CI_{95\%}$ = [.014–.042] indicated good fit (lowest and highest value respectively).

As mentioned in the hypothesis section, a Bayesian approach is needed to test the formulated hypotheses. Accordingly, with the results from the models, we applied a Bayesian informative

hypothesis approach using the R package ‘bain’ (Hoijtink et al., 2019). As opposed to the commonly used frequentist null hypothesis significance testing, this allows us to quantify and compare relative evidence of hypotheses (including a null hypothesis). Hypotheses to be compared were derived from the formulated hypotheses above. They will be reported directly in the respective results part to reduce complexity (all analyses and results can be examined in the supplemental material).

4. Results

4.1 Selective Attention

The quantity of selected situations slightly declined from pretest to posttest in both groups (Tab. 3). This might be an unexpected result; however, please be reminded that a direct comparison of pretest and posttest should be interpreted with caution.

Table 3

Means (and standard deviations) of the three dependent variables.

	Selective attention		*Reflective thought*		*Theory-practice integration*	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
DI	2.44 (1.71)	2.19 (1.41)	2.18 (.58)	2.16 (.59)	.05 (.14)	.23 (.33)
PB	2.42 (1.81)	2.15 (1.44)	2.17 (.56)	2.22 (.60)	.04 (.11)	.27 (.34)

Note: Selective attention: count of analyzed situations per test; reflective thought: realized inquiry steps in the analysis process [0–3]; theory–practice integration: relative frequency of analyses including theoretical target aspects

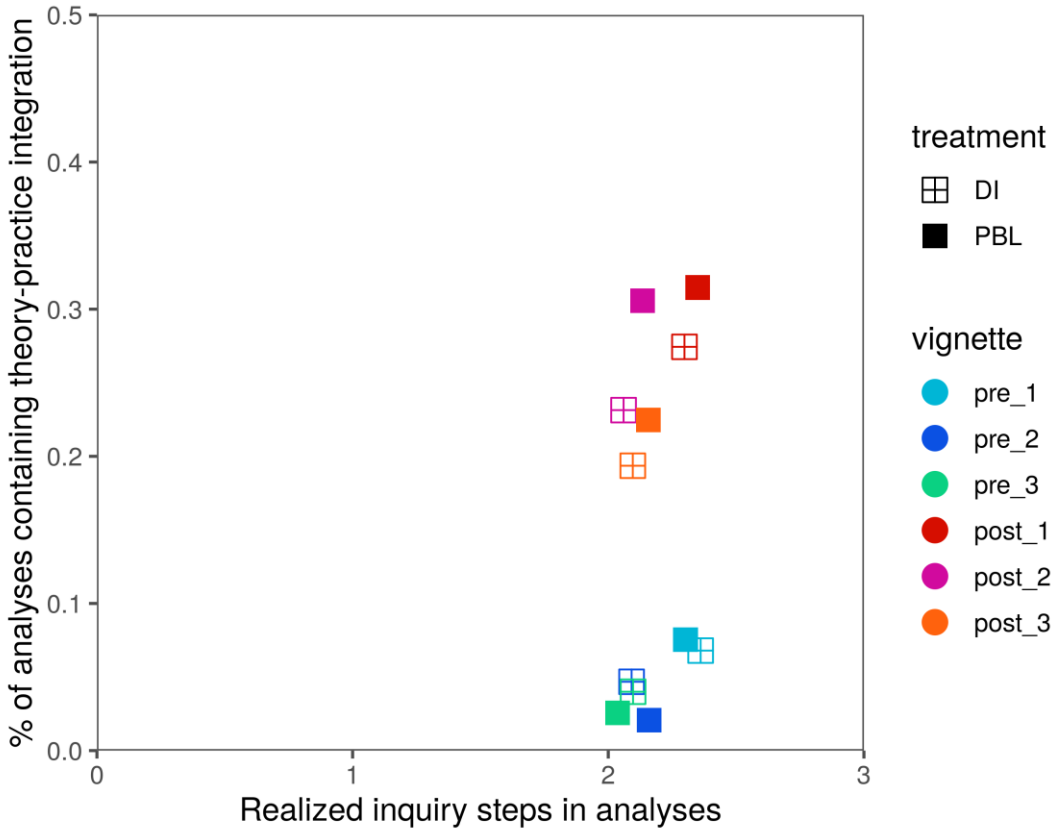
We expected the DI and PB learning approaches to foster selective attention (no difference between groups) and for positive attitudes of the instructors about the learning approach to be positively related to selective attention. Therefore, the statistical hypothesis to test can be

formulated as H_1: $\beta_{treat} = 0$ & $\beta_{IA} > 0$, with the dichotomous treatment variable coded as DI = 0 and PB = 1 (see Fig. 2). This hypothesis will be tested against H_2: $\beta_{treat} = 0$ & $\beta_{IA} < 0$ as one may also assume that instructors convinced of the treatment they are teaching make students analyze situations in greater detail. Students will in consequence select fewer situations to analyze. These hypotheses are further compared with a null hypothesis, H_0: $\beta_{treat} = 0$ & $\beta_{IA} = 0$, and an unrestricted hypothesis that will have the greatest probability in case all our formulated hypotheses are implausible, H_u: β_{treat}; β_{IA}.

To describe the results, we indicate which hypothesis received the greatest posterior probability, then report the Bayes factor against its complement (opposite of the hypothesis) and the Bayes factors of the hypothesis against the other hypotheses that were also tested (see supplement for detailed results). Evidence pointed toward the exploratory hypothesis H_2, as well as the null hypothesis H_3. We found solid evidence of these hypotheses against their complement ($BF_{2c} = 27.32$; $BF_{0c} = 25.59$) and H_1 ($BF_{21} = 50.97$; $BF_{01} = 47.73$). Comparing the two hypotheses H_2 and H_3 against each other yielded no clear result ($BF_{20} = 1.07$). We conclude that there is strong evidence that the two learning approaches have an equivalent effect on selective attention. In addition, instructors' positive attitudes toward the treatment had either no relation or a negative relation to the quantity of selected situations. However, regarding the instructors' attitude, we cannot make a concluding statement.

4.2 Reflective Thought

Students' reflective thought (as measured by realized inquiry steps in the analyses) was already well developed before they entered the treatment sessions and showed only little change afterwards (Fig. 3).



568

569 **Fig. 3.** Students' scores in reflective thought and theory–practice integration differentiated by

570 treatment and vignettes of the pretest and posttest.

571

572 We expected the PB learning approach to foster reflective thought (compared with DI) and

573 instructors' positive attitude to show a positive relation, H_1: $\beta_{treat} > 0$ & $\beta_{IA} > 0$. Two

574 exploratory hypotheses tested whether only one of the effects holds, H_2: $\beta_{treat} = 0$ & $\beta_{IA} > 0$ and

575 H_3: $\beta_{treat} > 0$ & $\beta_{IA} = 0$. As before, we included a null (H_0) and an unrestricted hypothesis (H_u).

576

577 As one may expect from the descriptive results, we found substantial evidence for the null

578 hypothesis against the other hypotheses and its complement (BF_{0u}= 82.00; BF_{0c} = 82.00;

579 BF_{01} = 28.22; BF_{02} = 4.09; BF_{03} = 6.26). Therefore, we conclude that the effect on students'

580 reflective thought is equivalent between learning approaches and the instructor's attitude is not

581 related to students' reflective thought.

4.3 Theory–Practice Integration

Students' theory–practice integrations when analyzing classroom situations greatly changed from the pretest to posttest (ΔM_{DI}= 18%; ΔM_{PB}= 23%). Even though the pretest and posttest vignettes are not identical, the difference in scores is certainly noteworthy. As opposed to reflective thought, students showed considerable room for potential with their theory–practice integrations in the pretest (Fig. 3). A mere 2–8% of analyses on the vignettes in the pretest contained theory–practice integrations, although the theoretical literature on classroom management was already provided before the test. To obtain the effect of the treatment independently of the amount of literature read by the students, we measured and controlled for this variable (see Fig. 1).

We expected the PB learning approach to generate a similar amount of theory–practice integration as DI and the instructors' attitudes to be positively related, H_1: $\beta_{treat} = 0$ & $\beta_{IA} > 0$. Like before, we explored in further hypotheses whether PBL shows a positive effect H_2: $\beta_{treat} > 0$ & $\beta_{IA} > 0$ or the attitudes make no difference H_3: $\beta_{treat} > 0$ & $\beta_{IA} = 0$, and included a null (H_0) along with an unrestricted hypothesis (H_u). Evidence indicates strong support for the hypothesis H_1 (BF_{1u} = 34.06; BF_{1c} = 34.06; BF_{12} = 12.24; BF_{13} = 90.95; BF_{10} = 41.09). From these results, we infer that both learning approaches lead to an equivalent effect on students' theory–practice integrations. What is more, instructors' positive attitudes toward the treatment are related to an increase in theory–practice integrations.

5. Discussion

5.1 Interpretation of Results

The goal of our field study was to compare different learning scenarios using authentic representations of practice and their differential effects on how students analyze classroom situations. More specifically, we aimed to consider selective attention, reflective thought, and theory–practice integration. Our data did not reveal differences between the DI and PB approaches on the selective attention of students. At first glance, the treatment's low impact does not necessarily contradict Seidel et al.'s (2013) results because students in the DI courses might have been able to notice more critical situations, but if given a choice, did not make use of that skill due to negative attitude or a lack of motivation. To test for this possible explanation, we included the students' willingness for effort and their attitude on readiness for reflection in the structure model. The variables had no significant relation with selective attention, and thus, cannot resolve this issue. Concerning the relation of the instructors' attitude toward the learning approaches with the selective attention of students, we were not able to make a final conclusion, as results from the data were inconclusive.

We were able to reveal evidence for PB and DI to have equivalent effects on the reflective thought of students. This contradicts our assumptions formulated before data collection. As the strongest predictor in the model was the pretest score on reflective thought and our intervention was rather short (two 90-minute sessions), this might support assumptions that reflective thought is rather a stable skill, and thus, more challenging to influence. Further inquiry shows a negative relationship with the number of selected situations and a substantial correlation with theory–practice integration. The motto for high performers in inquiry seems to be less quantity, more quality. Overall, the means imply consistent high scores on the 0–3 scale from pre- to posttest for both groups. Moreover, instructors' attitudes toward the treatment did not play a role in students' reflective thought.

Students' theory–practice integrations when analyzing classroom situations greatly increased from pretest to posttest. Both learning approaches appear to have equivalent positive effects, as attendance in either represents a significant predictor. Thus, these results do not support Dochy et al.'s (2003) and Seidel et al.'s (2013) findings. Interestingly, the strongest predictor (stronger than reading literature on the topic) represents the instructors' attitude toward the learning approaches. What is more, their attitudes were rather heterogeneous: The instructors favored different learning approaches with no approach being universally preferred.

5.2 Limitations and Further Research

In our measurement tool, we operationalized selective attention as the quantity of comments students gave on the classroom vignettes. This conceptualization makes it difficult to compare the data with further studies in the field, such as those on professional vision (Sherin, 2007). With our performance-based operationalization, differences in the quantity of selected situations may be interpreted as differences in the ability to notice critical situations or as differences in motivation or attitude. We tried to address this limitation by including related variables (willingness for effort, readiness for reflection), but this did not improve the model or influence correlations between latent variables.

Our results underscore the significance of Baecher et al.'s (2018) claim that more attention should be paid to the role of instructors. However, it remains unclear how the instructors' attitudes about learning scenarios affect the students' performance in applying the analysis of practice. Further insight and research is needed on the instructors' and students' sides to clarify the path of effects and interactions: How does the instructors' attitude influence their teaching performance and how does the teaching performance influence the students' beliefs and performances? Lastly, the treatment was short compared with, for example, video clubs (van

Es & Sherin, 2008); this was due to the field study character of the teacher education program. An artificially prolonged treatment that exceeded the courses' two sessions on classroom management could have had a different effect, but this would have reduced the external validity in our case.

5.3 Conclusion

With the limitations in mind, we draw two major conclusions from our data. First, we refer to the question posed in the title: Do direct instructional and problem-based approaches really produce different effects? Based on our data, the answer is *no*. In our study we produced evidence, that using either DI or PB learning approaches to short-term interventions will yield equivalent effects on students' selective attention, reflective thought, and theory–practice integrations. Particularly, students' reflective thought proved to be a stable skill, and therefore, would need to be addressed with longer interventions (e.g., over one semester).

Second, encouragingly, both learning approaches can foster theory–practice integrations of students, with the instructors playing a key role. These results contribute to further uncover approaches to increase theory–practice integration, which was already labeled a "highly relevant endeavor" (Brouwer & Korthagen, 2005, p. 216). They also underscore the importance of examining the role of instructors within future field-based research. Based on our findings, we might not necessarily recommend forcing instructors to use certain (allegedly effective) learning methods, but rather, to draw on those about which they have positive views. We consider this result as a vital insight because it should be relevant for related field studies or the interpretation of results in laboratory studies (e.g., where researchers function as instructors). Coming from a field study design perspective, these results are applicable for teacher education practice, and thus, highly relevant.

References

Author, 2016a [details removed for peer review]

Author, 2016b [details removed for peer review]

Author, 2018 [details removed for peer review]

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of Educational Research*, *70*(2), 181–214. https://doi.org/10.3102/00346543070002181

Baecher, L., Kung, S.-C., Ward, S. L., & Kern, K. (2018). Facilitating Video Analysis for Teacher Development: A Systematic Review of the Research. *Journal of Technology and Teacher Education*, *26*(2), 185–216.

Barnett, M. (2006). Using a Web-based Professional Development System to Support Preservice Teachers in Examining Authentic Classroom Practice. *Journal of Technology and Teacher Education*, *14*(4), 701–729.

Barth, V. L., Piwowar, V., Kumschick, I. R., Ophardt, D., & Thiel, F. (2019). The impact of direct instruction in a problem-based learning setting. Effects of a video-based training program to foster preservice teachers' professional vision of critical incidents in the classroom. *International Journal of Educational Research*, *95*, 1–12. https://doi.org/10.1016/j.ijer.2019.03.002

Beisiegel, M., Mitchell, R., & Hill, H. C. (2018). The Design of Video-Based Professional Development: An Exploratory Experiment Intended to Identify Effective Features. *Journal of Teacher Education*, *69*(1), 69–89. https://doi.org/10.1177/0022487117705096

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, *24*(2), 417–436. https://doi.org/10.1016/j.tate.2006.11.012

Brouwer, N., & Korthagen, F. A. J. (2005). Can Teacher Education Make a Difference? *American Educational Research Journal*, *42*(1), 153–224. https://doi.org/10.3102/00028312042001153

Castro, A., Clark, K., Jacobs, J., & Givvin, K. B. (2005). Response to theory & practice question: Using video to support teacher learning. *AMTE Connections*, *14*(3), 8–12.

Charteris, J., & Smardon, D. (2013). Second look – second think: A fresh look at video to support dialogic feedback in peer coaching. *Professional Development in Education*, *39*(2), 168–185. https://doi.org/10.1080/19415257.2012.753931

Colestock, A., & Sherin, M. G. (2009). Teachers' Sense-Making Strategies While Watching Video of Mathematics Instruction. *Journal of Technology and Teacher Education*, *17*(1), 7–29.

Courtland, M. C., & Leslie, L. (2010). Beliefs and Practices of Three Literacy Instructors in Elementary Teacher Education. *Alberta Journal of Educational Research*, *56*(1), 19–30.

Day, C. (1993). Reflection: A necessary but not sufficient condition for professional development. *British Educational Research Journal*, *19*(1), 83–93.

Dewey, J. (1910). *How we think*. Dover Publications.

Dochy, F., Segers, M., van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A metaanalysis. *Learning and Instruction*, *13*, 533–568.

Duch, B. J., Groh, S. E., & Allen, D. E. (2001). Why Problem-Based Learning? A Case Study of Institutional Chance in Undergraduate Education. In *The power of problem-based*

learning: A practical "how to" for teaching undergraduate courses in any discipline (pp. 3–11). Sterling, Va: Stylus Pub.

Evertson, C. M. (Hrsg.). (2006). *Handbook of classroom management: Research practice and contemporary issues*. Lawrence Erlbaum Associates.

Fogarty, R. (1997). *Problem-based learning and other curriculum models for the multiple intelligences classroom*. IRI/Skylight Training and Publishing.

Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, *16*, 41–67. https://doi.org/10.1016/j.edurev.2015.06.001

Gröschner, A., Seidel, T., Pehmer, A.-K., & Kiemer, K. (2014). Facilitating collaborative teacher learning: The role of "mindfulness" in video-based teacher professional development programs. *Gruppendynamik Und Organisationsberatung*, *45*(3), 273–290. https://doi.org/10.1007/s11612-014-0248-0

Grossman, P., & Pupik Dean, C. G. (2019). Negotiating a common language and shared understanding about core practices: The case of discussion. *Teaching and Teacher Education*, *80*, 157–166. https://doi.org/10.1016/j.tate.2019.01.009

Hallett, F. (2010). Do we practice what we preach? An examination of the pedagogical beliefs of teacher educators. *Teaching in Higher Education*, *15*(4), 435–448.

Harford, J., & MacRuairc, G. (2008). Engaging student teachers in meaningful reflective practice. *Teaching and Teacher Education*, *24*(7), 1884–1892. https://doi.org/10.1016/j.tate.2008.02.010

Hatch, T., & Grossman, P. (2009). Learning to Look beyond the Boundaries of Representation: Using Technology to Examine Teaching (Overview for a Digital Exhibition: Learning from the Practice of Teaching). *Journal of Teacher Education*, *60*(1), 70–85.

Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review*, *16*(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hoijtink, H., Gu, X., Mulder, J., & Rosseel, Y. (2019). Computing Bayes factors from data with missing values. *Psychological Methods*, *24*(2), 253–268. https://doi.org/10.1037/met0000187

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86.

Kounin, J. S. (1970). *Discipline and group management in classrooms*. Holt, Rinehart and Winston.

Li, Y. L. (2007). Teachers Talking about Effective Practice: Understanding the Knowledge and Practice of Teachers. *Journal of Early Childhood Teacher Education*, *28*(3), 301–310.

Loughran, J. J. (2002). Effective Reflective Practice: In Search of Meaning in Learning about Teaching. *Journal of Teacher Education*, *53*(1), 33–43. https://doi.org/10.1177/0022487102053001004

Marsh, B., & Mitchell, N. (2014). The role of video in teacher professional development. *Teacher Development*, *18*(3), 403–417. https://doi.org/10.1080/13664530.2014.938106

Marsh, B., Mitchell, N., & Adamczyk, P. (2010). Interactive video technology: Enhancing professional learning in initial teacher education. *Computers & Education*, *54*(3), 742–748. https://doi.org/10.1016/j.compedu.2009.09.011

Mayr, J. (2009). Klassen stimmig führen. *Pädagogik*, *61*(2), 34–37.

McDonald, R. P. (1999). *Test theory: A unified treatment*. L. Erlbaum Associates.

Mitchell, N., Marsh, B., Hobson, A. J., & Sorensen, P. (2010). 'Bringing theory to life': Findings from an evaluation of the use of interactive video within an initial teacher preparation programme. *Teacher Development*, *14*(1), 15–27. https://doi.org/10.1080/13664531003696543

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs*. https://CRAN.R-project.org/package=BayesFactor

Orland-Barak, L., & Maskit, D. (2017). *Methodologies of Mediation in Professional Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-319-49906-2

Praetorius, A.-K., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction*, *22*(6), 387–400.

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry*, *18*(1), 178. https://doi.org/10.1186/s12888-018-1761-4

Rosenthal, R., & Fode, K. L. (1963). The Effect of Experimenter Bias on the Performance of the Albino Rat. *Behavioral Science*, *8*(3), 183.

Savery, J. R. (2006). Overview of Problem-Based Learning: Definitions and Distinctions. *Interdisciplinary Journal of Problem-Based Learning*, *1*(1), 9–20. https://doi.org/10.7771/1541-5015.1002

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education*, *34*, 56–65.

Seidel, T., & Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal*, *51*(4), 739–771.

Sherin, M. G. (2007). The development of teachers' professional vision in video clubs. In R. Goldman (Hrsg.), *Video research in the learning sciences* (S. 383–395). Lawrence Erlbaum Associates.

Sherin, M. G., & Han, S. Y. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education*, *20*(2), 163–183. https://doi.org/10.1016/j.tate.2003.08.001

Sherin, M. G., & van Es, E. A. (2009). Effects of Video Club Participation on Teachers' Professional Vision. *Journal of Teacher Education*, *60*(1), 20–37. https://doi.org/10.1177/0022487108328155

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4–14.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd edition). Los Angeles; London; New Delhi; Singapore; Washington DC: Sage.

Tripp, T. R., & Rich, P. J. (2012). The influence of video analysis on the process of teacher change. *Teaching and Teacher Education*, *28*(5), 728–739.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3). https://doi.org/10.18637/jss.v045.i03

van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education*, *24*(2), 244–276. https://doi.org/10.1016/j.tate.2006.11.005

Zeichner, K. M. (1981). Reflective teaching and field-based experience in teacher education. *Interchange*, *12*(4), 1–22. https://doi.org/10.1007/BF01807805