Learning and Instruction 97 (2025) 102084

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

# Analysing classroom videos in teacher education— How different instructional settings promote student teachers' professional vision of classroom management☆

Isabell Tucholka *, Bernadette Gold

*Institute of General Didactics and School Pedagogy, TU Dortmund University, Germany*

ARTICLE INFO

*Keywords:*
Classroom video
Professional vision
Student teachers
Classroom management
Instruction

ABSTRACT

*Background:* Working with video cases may enhance student teachers' professional vision, i.e. their ability to notice and reason classroom events and generate alternatives. To foster professional vision successfully, videos need to be embedded into adequate instructional environments that need to be examined regarding their effectiveness.
*Aims:* This study investigates the effect of different instructional settings regarding video-based online courses on the development of student teachers' professional vision (noticing, reasoning, generating alternatives).
*Sample:* Participants were 280 student teachers.
*Methods:* In a 2 × 2-pretest-posttest design, students attended a course assigned to one of four conditions (video analysis before or after conceptual input; video analysis with a comparative or non-comparative task). For the pre- and posttest, they analysed video clips using open writing and rating items to measure noticing, reasoning (subskill 1: interpretation depth, subskill 2: evaluation) and generating alternatives. We applied a multivariate growth model with time points (level 1) nested in individuals (level 2), regressing according random slope coefficients on conditions.
*Results:* Students improved across all course conditions. Students working with comparative tasks excelled in generating alternatives, while receiving conceptual input before working with the videos was slightly advantageous regarding video evaluation skills.
*Conclusions:* The effectiveness of instructional settings depends on the intended learning goal. Concerning noticing and interpretation depth, the specific instructional setting seems less decisive.

## 1. Introduction

Teaching on a high quality level requires a high competence of teachers. This concerns not only competences regarding the contents they want to convey but also pedagogical aspects like classroom management (Baumert & Kunter, 2013; Pianta & Hamre, 2009). Apart from mere knowledge, however, teachers also need professional vision, that is, the ability to identify and interpret relevant classroom events and decide whether and how to react (Sherin & van Es, 2009). To foster these skills effectively already during initial teacher education, video cases have become an increasingly popular tool. However, there is agreement that videos can only be effective if embedded in appropriate instructional settings in line with the intended learning goals (Kang & van Es, 2019). In this vein, one important aspect seems to be the order of conceptual input and casework (Beitzel & Derry, 2009; Likourezos & Kalyuga, 2017; Loibl et al., 2020), but also whether videos are analysed independently of each other or comparatively might influence the learning outcome (Alfieri et al., 2013; Nagarajan et al., 2004; Rittle-Johnson & Star, 2007).

Only few studies have systematically investigated the influence of different instructional video-course settings on student teachers' professional vision so far (e.g. Seidel et al., 2013). Therefore, we conducted a 2 × 2-pretest-posttest intervention study to investigate how the order of conceptual input and casework (*concepts-first, casework-first*) and the observation task (*non-comparative, comparative*) influence student teachers' development of professional vision during a self-study online

☆ This study's design, hypotheses and main analyses were preregistered; see https://aspredicted.org/N16_7FW.

* Corresponding author. Martin-Schmeißer-Weg 4-8, 44227 Dortmund, Germany.
*E-mail address:* isabell.tucholka@tu-dortmund.de (I. Tucholka).

https://doi.org/10.1016/j.learninstruc.2025.102084
Received 14 August 2023; Received in revised form 13 January 2025; Accepted 16 January 2025
Available online 4 February 2025

course. As the interventions' topic, we chose classroom management as it is essential for teaching quality and learning (Hattie, 2009; Pianta & Hamre, 2009).

### 1.1. Teachers' professional vision

A lot of research has examined teachers' professional competence, in both a generic, pedagogical sense and a subject-specific one (Shulman, 2013). Blömeke et al. (2022) regard competence as a continuous, multidimensional construct in which the relationship between individual dispositions (including cognition and affect-motivation) and observable performance is mediated by situation-specific skills, namely perception, interpretation and decision-making. In a teaching context, perception, or else *noticing*, refers to the ability to identify classroom events noteworthy for the learning process (Sherin, 2007) which appears particularly crucial as goings-on in classrooms are complex and highly unpredictable (Doyle, 2006). Teachers have to reason these events on the basis of their individual knowledge (Sherin, 2007). These knowledge-based *reasoning* skills are often conceptualised to include various subskills as describing, evaluating, interpreting, explaining or predicting (Sherin, 2007; Sherin & van Es, 2009; Seidel & Stürmer, 2014). In essence, teachers need to quickly assess the significance of an event, analyse what may have caused it and how and why it interferes with the pupils' learning. Only then can they *generate* alternative courses of action they could take (Stahnke & Blömeke, 2021a) and decide whether and how to (re)act.

These skills are also called *professional vision* (Sherin, 2007) and represent typical characteristics of expert teachers: they identify more crucial situations in teaching and learning processes and concentrate primarily on student-related events while novices' attention is more often on the teacher (Stahnke & Blömeke, 2021b). Additionally, due to their flexibly organised and integrated knowledge structures, experts rather interpret and explain relevant events while novices tend to describe events superficially (Sherin & van Es, 2009; Wolff et al., 2015), and they generate more alternatives than novices do (Stahnke & Blömeke, 2021a).

This triad of professional vision—*noticing, reasoning* as well as *generating alternatives*—is relevant for teaching performance (Blömeke et al., 2022; Krauss et al., 2020). Empirical studies revealed that those skills relate to aspects of instructional quality (Blömeke et al., 2022; Jamil et al., 2015) and student learning outcomes (Kersting et al., 2012; Roth et al., 2011). Professional vision is, therefore, regarded a prerequisite for professional teaching behaviour and should be addressed early in student teacher education. Being bound to specific situations, professional vision is often fostered using classroom videos (Gaudin & Chaliès, 2015).

### 1.2. Using classroom videos to foster professional vision

During the last decades, classroom videos have become increasingly popular in teacher education as they authentically convey classroom complexity (Baecher et al., 2018; Santagata et al., 2021). While classroom videos can illustrate concepts and theories and hence help attain relevant knowledge, they can also facilitate contextualising acquired knowledge and pondering on if and how it might be applied in the displayed situation (Blomberg et al., 2014). Studies confirmed a positive impact of working with videos on (student) teachers' professional vision (Gaudin & Chaliès, 2015; Marsh & Mitchell, 2014). However, there is consensus that videos are not effective per se but have to be embedded in adequate instructional environments to exploit their full potential (e.g. Kang & van Es, 2019). Apart from setting learning objectives and selecting suitable video material, this also includes specifying objective-related observation tasks and prompts and providing adequate scaffolding, for example in the form of conceptual input that supports analysing and reflecting on the observed situation (Kang & van Es, 2019).

In this vein, our study shall take a closer look at two instructional settings in designing an online self-study teacher education course based on video casework. We will investigate how the specific observation task (comparative, non-comparative) and the order of conceptual input and video casework (concepts-first, casework-first) influence the development of student teachers' professional vision during the course.

#### 1.2.1. The observation task

Various tasks might be set for analysing video sequences; for example, they might be analysed individually or in explicit comparison to each other. Comparing cases can enhance learning and understanding. In a meta-analysis, Alfieri et al. (2013) found an average effect size of $d = 0.5$ in favour of case comparison as opposed to sequential analyses, single case approaches or other forms of case analysis. Comparing contrasting cases, that is examples mostly alike but differing regarding one essential conceptual feature (Loibl et al., 2020), helps focussing the learners' attention upon distinctive features of interest (Kurtz & Gentner, 2013; Loibl et al., 2017). According to Schwartz and Bransford's (1998) knowledge-differentiation hypothesis, becoming aware of these distinctive features helps students develop a rather rich and differentiated knowledge base, which is considered a prerequisite of further competence acquisition (Kumschick et al., 2017). For instance, comparing a video showing a very smooth transition from one lesson phase to the other with a video showing a rather erratic transition, therefore, might shift attention towards this aspect and trigger reflection on what influences transitions in general. As videos cannot be observed simultaneously and learners are not likely to actively compare sequentially presented cases without specific prompting (Gentner et al., 2003; Rittle-Johnson & Star, 2007; Wilkes et al., 2022), the explicit task to find similarities and differences between the videos seems necessary. Hirstein et al. (2017) found hints that, given an according task, especially videos representing very strong and distinctive contrasts may be supportive for student teachers' professional vision development. It seems, therefore, that video comparison may be advantageous for students' professional vision development.

However, classroom videos are very rich in detail and highly complex. Even if constant and varying features are being controlled for, the high complexity may cause ambiguity, make it harder to detect the essential elements and thereby distract the observer from the germane features (Alfieri et al., 2013). Therefore, students require additional guidance to learn from contrasting cases effectively; working with contrasting videos without sufficient scaffolding might distract and overwhelm them and lead to rather superficial understanding (Nagarajan et al., 2004; Nagarajan & Hmelo-Silver, 2006). In this vein, Beitzel and Derry (2009) assume that the knowledge structures created when comparing video cases might be individually different and potentially inconsistent with subsequent input, which is why students should learn about important concepts first to build relevant knowledge structures and use video comparison afterwards to elaborate on the acquired schemas (schema-elaboration hypothesis). This goes to show that the choice of observation tasks is not a separate issue, but interlinked with the question of when to incorporate conceptual input, namely before or after video casework.

#### 1.2.2. The order of conceptual input and casework

It is a common question concerning instructional settings whether to incorporate explicit conceptual input on relevant contents before or after student-driven activities (Likourezos & Kalyuga, 2017) as video casework, that is whether to use a concepts-first setting or a casework-first setting.

The issue is controversially discussed (Kirschner et al., 2006; Loibl et al., 2020). *Concepts-first* settings provide relevant conceptual input before engaging students in complex activities as video casework. This initial input offers students the opportunity to build prior knowledge necessary to process a subsequent complex task without being overwhelmed by it (Kalyuga & Singh, 2016; Kirschner et al., 2006;

Kumschick et al., 2017). Furthermore, it helps directing the students' focus to germane features of the target concepts (Likourezos & Kalyuga, 2017; Paas & van Gog, 2006), potentially facilitating subsequent video casework because students already know more precisely what to look out for. In a concept-first setting, classroom videos mainly serve the purpose of exemplifying previously provided concepts (Blomberg et al., 2014). *Casework-first* settings, on the other hand, aim to contextualise knowledge acquisition from the start. Observing an authentic teaching situation via video provides a context that subsequent conceptual input can refer to or even be derived from (Santagata & Angelici, 2010; Seidel et al., 2013). Casework-first settings offer the opportunity to activate students' prior knowledge for future learning to build on (Trninic et al., 2022) and to become aware of potential knowledge gaps (Loibl & Rummel, 2014) that might be purposefully resolved by subsequent conceptual input. Therefore, they may prepare students to grasp subsequent conceptual input more deeply (Trninic et al., 2022). Additionally, this offers the chance to raise students' motivation and curiosity (see also Lamnina & Chase, 2019) to gain knowledge on how the presented situation might be dealt with.

Empirical studies from mathematical or science contexts revealed a positive impact of casework-first settings on learning (Loibl et al., 2017; Sinha & Kapur, 2021). However, according topics (e.g. fractions) offer less scope for ambiguous and debatable solutions as is the case in analysing complex teaching situations. In teacher education, however, only few studies systematically examined the impact of a concepts-first and a casework-first setting on the development of student teachers' professional competence in general and their professional vision in particular. In Seidel et al.'s (2013) video-based intervention study, 56 preservice teachers were assigned to courses dealing with general aspects of teaching and learning, and students in the concepts-first condition showed a significantly stronger development in their professional vision. Similar studies underscored these findings (Barth et al., 2019; Blomberg et al., 2014; Kumschick et al., 2017). Participants of Seidel et al.'s (2013) casework-first group, on the other hand, identified significantly more potential lesson challenges, a rather action-related form of knowledge going beyond the observable video scene, similar to generating alternatives. Concerning our study, therefore, the order of conceptual input and casework seems relevant regarding the development of professional vision. However, while concepts-first settings seem to foster video observation skills (noticing, reasoning) particularly well, casework-first settings might be more advantageous regarding generating alternatives.

For these reasons, our study looks at the influence of the specific observation tasks (comparative, non-comparative) in combination with the order of conceptual input and video casework (concepts-first, casework-first) in video-based teacher education courses on the development of student teachers' professional vision.

## 2. Research questions

Working with classroom videos allows for different instructional settings. We investigated differential effects of selected aspects (order of conceptual input and casework; specific observation task) in a self-study online course on student teachers' professional vision as represented by the subskills *noticing, reasoning* and *generating alternatives*. We chose classroom management as the central content of the intervention as it is an essential element of teaching quality (Hattie, 2009; Pianta & Hamre, 2009) and relevant across all subjects and grade levels.

Based on the theoretical insights and empirical results described above, we addressed the following hypotheses.

**Hypothesis 1. Observation task (comparative, non-comparative):** We assumed students who explicitly compared classroom videos to make more progress regarding *noticing, reasoning* and *generating alternatives* than students who received a general analysis task across all clips as video case comparison proved effective given sufficient support (Nagarajan & Hmelo-Silver, 2006).

**Hypothesis 2. Order of casework and conceptual input[1]:** In analogy to Seidel et al. (2013), we expected students in concepts-first settings to make more progress regarding *noticing* and *reasoning* than students in casework-first settings (*H2a*). As *generating alternatives*, however, goes beyond analysing classroom situations and requires more action-oriented knowledge usage fostered especially in casework-first settings (Blomberg et al., 2014), we expected students in casework-first settings to outperform those in concepts-first settings regarding *generating alternatives* (*H2b*).

**Hypothesis 3. 2nd-order effects of order and task:** We expected ordinal interaction effects and therefore assumed the effects to be additive.

## 3. Methods

### 3.1. Design and sample

We conducted a quasi-experimental intervention study (non-equivalent groups) in a 2 × 2-pretest-posttest-design.[2] A total of 326 student teachers participated in one of multiple video-based self-study online courses on classroom management (CM) that were part of the regular teacher education schedule and mandatory for students in the according master's programme. We provided participants with identical contents and material, but designed the online learning environments to differ regarding two aspects of their instructional settings (order: concepts-first, casework-first; task: analyse videos in a comparative or a non-comparative task). This resulted in four course conditions: (1) *concepts-first/non-comparative, (2) concepts-first/comparative, (3) casework-first/comparative,* and (4) *casework-first/non-comparative.* For administrative reasons, we offered a total of eleven courses randomly assigned to these conditions, and while students could freely choose which course to attend based on time slots, they did not know which condition they signed up for. Four instructors, distributed equally across conditions, offered opportunities to discuss upcoming questions as needed and maintained the online learning environments, which were kept identical across courses of the same condition.

We excluded 46 students from analyses because they denied permission for data usage (37 students), did not attend the posttest (3 students) or skipped viewing at least one video during pre- or posttesting (6 students). Consequently, data analyses were based on the data of 280 students (81.1% female; $M_{age} = 23.55$, $SD_{age} = 2.48$) who, on average, studied in their eighth semester ($M_{semester} = 8.03$, $SD_{semester} = 0.52$) for primary (75.4%), secondary (10.7%) or special education (13.9%). This is a sufficient sample size according to an a priori power analysis.[3]

ANOVA and Chi$^2$-tests revealed no significant differences between the four groups regarding gender ($\chi^2(3) = 2.88$, $p = .411$), age ($F(3, 275) = 0.79$, $p = .499$) and number of semesters ($F(3, 276) = 1.53$, $p = .208$). Further, a MANOVA revealed that pretest results did not differ significantly between conditions, $F(12, 714.64) = 0.83$, $p = .622$, Wilk's $\Lambda = 0.964$. To control for potential instructor bias, we included according dummy-variables into our main analyses, showing no significant influence on the development of the dependent variables over time.

[1] While we did not include a directed hypothesis concerning the order of input and casework in pre-registration due to inconsistent empirical findings, further discussion in our research group after pre-registration led us to include a directed hypothesis in the paper.

[2] Participants were treated in accordance with the ethical guidelines of the German Psychological Society (DGPs).

[3] Originally, we intended to conduct mixed ANOVAs for data analysis, resulting in the following power analysis results: ANOVA repeated measures, between factors; 4 groups; f = 0.25, α = .05, power = 0.90, $r = 0.5$; required sample size = 176.

### 3.2. Intervention

The intervention courses were completely online-based and consisted of five 90-minute sessions (see Fig. 1): one reading session and four sessions of individual video analysis. The reading session provided scientific literature on all CM aspects the video sessions focussed on. In the video sessions, we asked the students to observe three to four short clips and prepare a written analysis regarding selected aspects of the observed CM. Literature and video material were identical across all course conditions and students had to hand in their analyses regularly.

#### 3.2.1. Video selection

For the video sessions, we used authentic videos of real primary school science teaching from the platform *ProVision.* Initially, a range of clips had been coded regarding whether they featured CM-relevant situations and which specific CM aspects could be observed. This includes, amongst others, a clear structure, high group focus and a constant monitoring of the classroom to detect and deal with occurring disruptions (Doyle, 2006; Kounin, 1970). In subsequent discussions, the number of clips was reduced to 30. Expert teachers ($N = 29$) who had been selected for the task by their headmasters based on their expertise rated these 30 clips again regarding whether those CM aspects could be observed and whether the clips represented rather positive or negative examples regarding the implementation of these aspects. For the video course of the present study, we chose three to four clips (duration: 1′16″ to 6′22″) for each video session representing contrasting cases regarding the selected CM aspect the session focussed on, that is, videos representing different levels of effectiveness in implementing this aspect as assessed by the expert teachers.

#### 3.2.2. Intervention conditions

*3.2.2.1. Observation tasks.* In the courses using *comparative* tasks (6 courses), the students observed each sessions' videos to find similarities and differences of the teachers' (re)actions. In the *non-comparative* tasks (5 courses), they had to describe and interpret events they considered relevant in the clips and find suitable alternative actions the teacher could have taken concerning each event. In all conditions, the students filled in accordingly designed documents.

*3.2.2.2. Order of concepts and casework.* In the *concepts-first* settings (6 courses), the reading session was the introductory session. The following video sessions focussed on one CM aspect each; students first reread the relevant literature and, subsequently, observed and analysed the selected videos regarding this aspect. The tasks explicitly prompted them to refer to the presented literature. In the *casework-first* settings (5 courses), students started with the video sessions and received only a short definition of the relevant CM aspect beforehand to guide their attention and avoid confusion as to the complexity of classroom video stimuli (Nagarajan & Hmelo-Silver, 2006). Without further conceptual input, the students were asked to extract video events they considered relevant regarding the respective CM aspect and to label identified features with a suitable term. Only in the following session were they provided with the relevant conceptual reading and asked to link their intuitive labels to scientific terminology from the literature. The concluding reading session combined the previous conceptual input and served as a subsumption.

### 3.3. Instruments

In the pre- and posttest, students received brief definitions of the central CM aspects and subsequently analysed video clips to measure their skills in noticing, reasoning and generating alternatives, i.e. the triad of professional vision (see Fig. 2). As reasoning is a multi-facetted skill including how thoroughly students interpret relevant CM events but also how they evaluate them, we included both a written video analysis and a standardised video evaluation into our pre- and posttest to reach a comprehensive picture of students' professional vision development and how it is affected by the instructional settings.

#### 3.3.1. Written video analysis

The written video analysis was adapted from Gippert et al. (2022). The students watched a primary school video (2′16″) showing relevant events regarding all major CM aspects the intervention addressed. Using a provided chart, the students' task was to identify and interpret all events they considered relevant regarding CM and to generate appropriate alternatives for each of those events (time limit: 30′). Their answers were coded regarding the variables *noticing, interpreting* and *generating alternatives.* After intensive coder training, two independent raters reached sufficiently high intercoder agreement (Cohen's κ: noticing = .82; interpreting = .76; alternatives = .84) across 100 of 550 documents (18 %); major discrepancies were clarified in subsequent discussions.

*Noticing.* First, we coded how many relevant CM events the students identified. We did so based on a master analysis of six experts from both CM research and teaching practice, who identified nine events in the

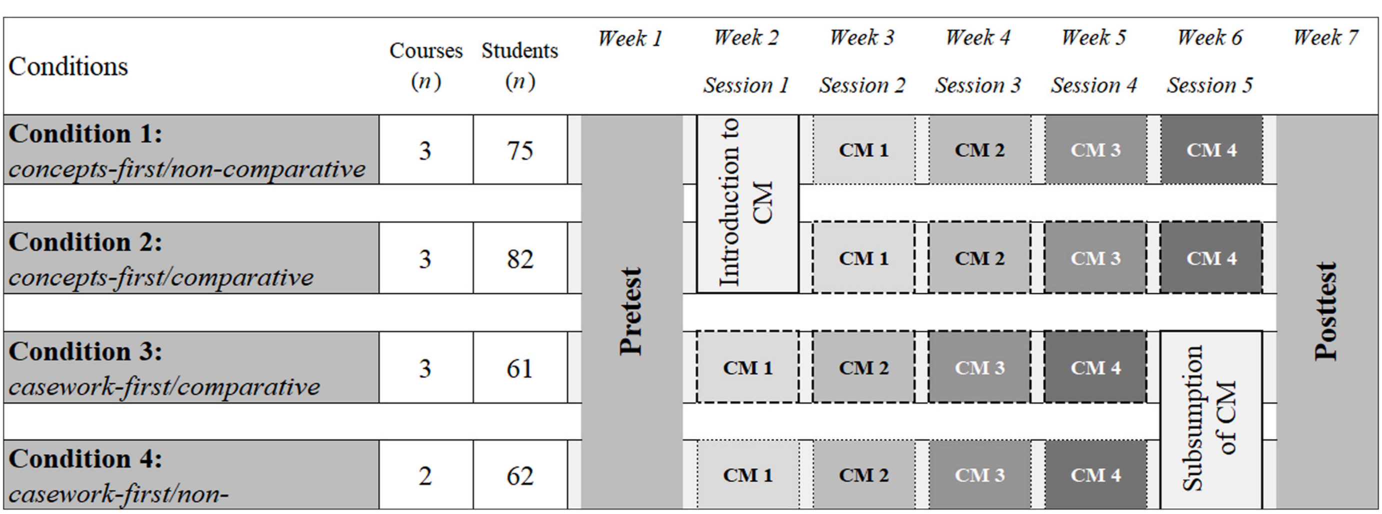


**Fig. 1.** Intervention Design

*Note.* $N = 280$. CM = Classroom management. Topics of video-based sessions: CM 1—monitoring, CM 2—structure and instruction, CM 3—group focus, CM 4—emotional support.

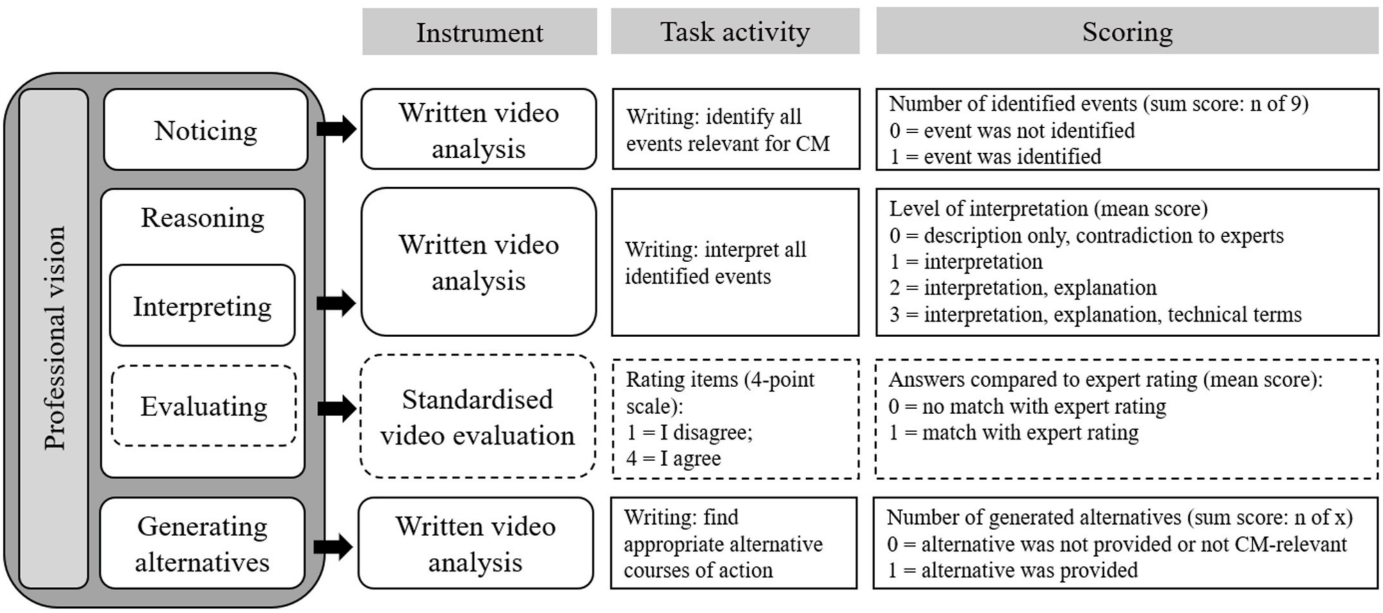


**Fig. 2.** Overview of Instruments and Scoring
*Note.* CM = Classroom management.

video as crucial, for example the teacher's repeatedly inappropriate reaction to disruptions. Students received one point for each of the nine events they identified, while any additional events coded as "irrelevant" were ignored. As the various observable events covered very different aspects of CM leading to only low correlations between items (Supplement 1; see also Stadler et al., 2021, on the discussion of scale consistency versus content validity) we provide a supplement giving detailed information on each event as well as event-specific analyses (Supplement 2).

*Interpreting* (reasoning component 1). For each of the identified events, we assigned up to three points depending on how thoroughly students had reasoned it, that is, whether they only described or actually interpreted the situation, explained their interpretation or used technical terms (see Fig. 2). We did not assign an interpreting score to events that went unnoticed and calculated a mean score based on those they had identified.

*Alternatives.* Moreover, we determined how many CM-specific alternatives participants generated.

#### 3.3.2. Standardised video evaluation

Naturally, the written analysis takes only those events into account that the students actually noticed. Therefore, we included a standardised video test (adapted from Gold & Holodynski, 2017) to measure how much students agreed with experts in their *evaluation* of various CM events while simultaneously shedding light on a wider range of CM aspects that, without the item prompt, the students might not have noticed in the first place.

Students observed three short video clips of primary teaching and rated the observed CM implementation via a total of 41 items (e.g. 'The teacher notices all up-coming disruptions') on a four-point Likert scale (1 = *I disagree*; 4 = *I agree*). The answers were compared to an expert rating (Gold & Holodynski, 2017) and earned one point for a match and zero points for a mismatch. We validated this adapted test version with 439 student teachers (80.4 % female, 18.9 % male, 0.7 % diverse, $M_{age}$ = 23.63, $SD_{age}$ = 2.77); a confirmatory bi-factor analysis revealed good fit values for a model with all 41 items loading on a one-dimensional factor for CM evaluation and one of three video factors representing the items' affiliation to one of the three clips, $\chi^2(738) = 916.50$, $p < .001$, RMSEA = 0.023, 90% CI [0.018, 0.028], CFI = 0.976. The scale's internal consistency was high (Cronbach's α; pretest: .88; posttest: .86) and we created a mean score representing the dependent variable *evaluating* (reasoning component 2).

### 3.4. Analyses

Using M*plus* 8, we set up a multivariate growth model with the two measurement points (level 1) nested within individuals (level 2). Noticing, interpreting, evaluating and generating alternatives were entered into the model as dependent variables. First, we fit several unconditional growth models, starting with an intercept-only model (model 0) assuming that there was no change in the outcomes over time. Level-2 mean levels for each outcome and variance at both levels were estimated. Intraclass correlations coefficients (ICC) were substantial for most outcomes; only concerning interpreting, the value was marginal (see Table 2). In model 1, we included fixed slope effects, regressing each outcome on measurement occasion (0 = pretest, 1 = posttest) to calculate the mean growth rate from pretest to posttest across all participants. Variables were allowed to covary. In model 2, we added random slope effects instead to account for individual differences in growth. As the online course aimed to foster all aspects of professional vision and especially as noticing and generating alternatives were highly interrelated (see Supplement 3), we allowed for covariance between the random slopes.

Finally, we set up two conditional models. In model 3, we added task (0 = non-comparative, 1 = comparative), order (0 = concepts-first, 1 = casework-first) and their interaction term as between-level variables on level 2. The representation of this model, which is the main model of this study, is visualised in the Appendix. In model 4, to control for potential instructor bias, we included three dummy-coded variables at level 2.

For all analyses, we applied maximum likelihood estimation with robust standard errors and an alpha level of .05. We checked for outliers ($z > 2.56$) and ran analyses with and without them; as results did not differ significantly, they were not excluded.

We applied Satorra-Bentler scaled chi-square tests based on loglikelihood values (Satorra & Bentler, 2001; see also Muthén & Muthén, 2017) to compare model fit sequentially to establish whether the next level of model complexity fitted our data better than the previous level. Additionally, AIC and BIC fit indices were provided.

## 4. Results

Mean scores and standard deviations regarding all dependent variables and conditions are presented in Table 1 and visualised in Fig. 3 (see Supplement 3 for correlational data).

The analyses revealed that there was significant growth over time concerning all dependent variables. First, Satorra-Bentler scaled chi-square tests showed that a model with fixed slopes fitted the data significantly better than the intercepts-only model and that the integration of random slopes increased model fit even more (see Table 2). Second, model 2 showed highly significant average growth rates for all outcomes. The variance of these growth rates was significantly different from zero for noticing ($\widehat{\sigma}^2 = 0.45$; 95% CI [0.01; 0.88]) and interpreting ($\widehat{\sigma}^2 = 0.25$; 95% CI [0.19; 0.32]), showing differences between students' development of these skills over time. Also, variance in generating alternatives was notable, although the difference from zero was not significant ($\widehat{\sigma}^2 = 0.38$; 95% CI [-0.04; 0.80]) while the growth rate of evaluation seemed rather similar between participants.

Regarding our hypotheses, model 3 tested whether and how various instructional settings influenced these growth rates (see Table 3). The conditional model 3 fitted our data significantly better than the unconditional models 0 to 2, although intervention groups predicted only few slope scores. However, as these slope scores (evaluating and generating alternatives) did not show significant interindividual variance in the first place, these results have to be interpreted very cautiously. Model 4 revealed no significant influence of different instructors on the results; model fit did not improve significantly compared to model 3.

*Observation task (H1)*. The observation task had a significant effect on how many alternatives the participants generated. In a *non-comparative* task format, the average number of suggested alternatives remained rather stable over both measurement points, while student teachers who had worked with *comparative* tasks generated nearly one alternative more after the intervention than they did before, resulting in a significantly steeper slope (see Table 3).

*Order of conceptual input and casework (H2)*. We found no significant effect of order on the slopes regarding the variables noticing, interpreting and generating alternatives. However, there was a significant effect regarding video evaluation, revealing a small advantage for those students who had worked in a concepts-first setting, that is who received conceptual input before engaging in video casework (see Table 3).

*2$^{nd}$-order effects (H3)*. There were no significant effects of order × task concerning any dependent variable, that is there was no combination of order and observation task that fostered the students' professional vision particularly well.

Having found a significant effect of order on the development of evaluation skills and of task on generating alternatives, analyses showed that no condition seemed to foster noticing and the thoroughness of interpretation as measured in the written video analysis particularly well. However, we found significant slope scores, revealing that, on average, participants perceived more relevant CM events after the intervention than before and interpreted them more thoroughly, apparently notwithstanding the course design.

## 5. Discussion

We aimed to investigate whether different instructional settings in working with video cases have differential effects on the development of student teachers' professional vision regarding CM. In a 2 × 2-intervention study framed by a pre- and posttest, student teachers were enrolled in a video-based self-study online course on CM that was assigned to one of four instructional settings: (1) concepts-first/non-comparative, (2) concepts-first/comparative, (3) casework-first/comparative and (4) casework-first/non-comparative. Before and after the course, participants engaged in video analysis tasks to measure their professional vision via the dependent variables noticing, reasoning (subskill 1: interpretation depth; subskill 2: evaluation) and generating alternatives.

On average, participants across all groups identified more relevant events after the intervention than before, interpreted these events more elaborately, evaluated them in higher agreement with experts and generated more alternative courses of action. For lack of an untreated control group, we cannot rule out the possibility of retest effects. Previous studies, however, proved videos' effectiveness to foster student teachers' professional vision (Gold et al., 2021; Santagata & Angelici, 2010), leading us to assume that at least part of these learning effects can be ascribed to the intervention.

Zooming into the hypothesised interaction effects of time with different course conditions, however, we are confronted with quite diverse results for the different aspects of professional vision.

### 5.1. Effects of the observation task

As expected, we did find the assumed advantageous effect of video comparison on generating alternatives (*H1*) as student teachers in the courses with comparative tasks generated more alternatives than student teachers in the courses with non-comparative tasks. However, dissident to our assumptions, the comparative observation task provided no additional value for students' noticing and reasoning skills, be it interpretation or evaluation, but was similarly effective as the non-comparative task. It is possible that the existing evidence on contrasting cases cannot be transferred to such complex cases as classroom videos. Although videos were chosen to represent contrasting cases regarding CM implementation, classroom interactions are far too complex to control all aspects that might vary between clips and potentially influence student teachers' analyses. Even when explicitly instructed to compare videos regarding selected aspects, the high complexity apparently makes it harder to detect the relevant contrasting elements needed to discover deep conceptual features *without* preceding conceptual input (*casework-first/comparative*). *With* preceding input (*concepts-first/comparative*), on the other hand, the emphasis is strongly on the relevant CM concepts from the beginning, ensuring a very pre-defined focus during video observation and thereby potentially make the additional support of a comparative task (Alfieri et al., 2013) redundant. Possibly, due to the complexity of video comparison, such tasks also require even more specific scaffolding measures (Nagarajan & Hmelo-Silver, 2006) like purposeful prompting (Martin et al., 2022) or a certain level of practical experience and qualified feedback (Weber et al., 2018) to unfold their potential.

As we did find the assumed advantageous effect of video comparison on generating alternatives, however, there seems to be a notable difference between the application of knowledge to a given, *observable* video scene and to imaginary alternative actions *beyond* it, potentially requiring more creative than analytical processes. Bringing in personal experiences and enriching these by observing videos during the intervention may have resulted in a pool of actions the students could draw from. In other words, they may have stored the different teacher actions as case-based knowledge and retrieved these cases from memory when confronted with a novel situation to either revise or retain the originally learned case (Jonassen, 2006). The comparative task actively guided the students' attention to similarities and differences and included the deliberate juxtaposition of different teacher actions in similar situations, which might have initiated a more effective memorising and application in a novel video situation and therefore turned out particularly effective for fostering generating of alternatives.

### 5.2. Effects of the order of conceptual input and casework

In line with our original assumptions, students in concepts-first settings, who received conceptual input *before* engaging in video casework (schema-elaboration hypothesis, Beitzel & Derry, 2009), could increase

**Table 1**
Mean scores (*M*) and standard deviations (*SD*) by measure, time and treatment condition.

| Variable | — | Concepts-first/Non-Comparative | | Concepts-first/Comparative | | Casework-first/Comparative | | Casework-first/Non-comparative | | Across Groups | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| *Noticing* | t1 | 3.92 | 1.21 | 3.85 | 1.33 | 4.02 | 1.49 | 3.82 | 1.17 | 3.90 | 1.30 |
| | t2 | 4.58 | 1.46 | 4.76 | 1.41 | 4.88 | 1.39 | 4.17 | 1.45 | 4.61 | 1.44 |
| *Interpreting* | t1 | 0.77 | 0.43 | 0.81 | 0.28 | 0.79 | 0.36 | 0.71 | 0.34 | 0.77 | 0.37 |
| | t2 | 0.95 | 0.66 | 1.00 | 0.59 | 0.89 | 0.62 | 1.08 | 0.62 | 0.98 | 0.62 |
| *Evaluating* | t1 | 0.40 | 0.20 | 0.40 | 0.18 | 0.42 | 0.17 | 0.44 | 0.22 | 0.41 | 0.19 |
| | t2 | 0.56 | 0.19 | 0.52 | 0.19 | 0.53 | 0.18 | 0.52 | 0.19 | 0.53 | 0.19 |
| *Generating Alternatives* | t1 | 3.38 | 1.38 | 2.98 | 1.41 | 3.24 | 1.10 | 3.25 | 1.39 | 3.20 | 1.34 |
| | t2 | 3.47 | 1.55 | 3.99 | 1.42 | 3.92 | 1.44 | 3.13 | 1.33 | 3.65 | 1.47 |

**Table 2**
Model parameters and goodness of fit for unconditional linear growth models.

| Parameter | Noticing | | Interpreting | | Evaluation | | Alternatives | |
|---|---|---|---|---|---|---|---|---|
| | Est. | *SE* | Est. | *SE* | Est. | *SE* | Est. | *SE* |
| Model 0: Intercept-only model | | | | | | | | |
| Fixed effects | | | | | | | | |
| Intercept | 4.21*** | 0.07 | 0.88*** | 0.02 | 0.47*** | 0.01 | 3.39*** | 0.07 |
| Random effects | | | | | | | | |
| Level 2 variance (person) | 0.49*** | 0.13 | 0.01 | 0.02 | 0.02*** | 0.00 | 0.62*** | 0.13 |
| Level 1 variance (measures) | 1.66*** | 0.16 | 0.27*** | 0.03 | 0.02*** | 0.00 | 1.52*** | 0.13 |
| ICC | 0.23 | | 0.04 | | 0.49 | | 0.28 | |
| Model fit statistics | | | | | | | | |
| 2 log likelihood | 4550.27 | | | | | | | |
| scaling correction factor for MLR | 1.0100 | | | | | | | |
| AIC | 4574.27 | | | | | | | |
| BIC | 4626.20 | | | | | | | |
| Model 1: Linear growth model with random intercepts and fixed slopes | | | | | | | | |
| *Fixed effects* | | | | | | | | |
| Mean initial score | 3.85*** | 0.08 | 0.78*** | 0.02 | 0.41*** | 0.01 | 3.16*** | 0.08 |
| Mean growth rate | 0.73*** | 0.10 | 0.21*** | 0.04 | 0.12*** | 0.01 | 0.46*** | 0.10 |
| Random effects | | | | | | | | |
| Initial score variance | 0.32*** | 0.09 | 0.01 | 0.02 | 0.02*** | 0.00 | 0.25** | 0.09 |
| Level 1 residual variance | 1.66*** | 0.15 | 0.26*** | 0.03 | 0.01*** | 0.00 | 1.77*** | 0.14 |
| Model fit statistics | | | | | | | | |
| 2 log likelihood | 4118.19 | | | | | | | |
| scaling correction factor for MLR | 1.0109 | | | | | | | |
| AIC | 4162.19 | | | | | | | |
| BIC | 4257.41 | | | | | | | |
| Chi$^2$-difference test[a] | $\chi2(10)$[b] = 426.96*** | | | | | | | |
| Model 2: Linear growth model with random intercepts and random slopes | | | | | | | | |
| *Fixed effects* | | | | | | | | |
| Mean initial score | 3.85*** | 0.08 | 0.78*** | 0.02 | 0.41*** | 0.01 | 3.16*** | 0.08 |
| Mean growth rate | 0.73*** | 0.10 | 0.21*** | 0.04 | 0.12*** | 0.01 | 0.46*** | 0.10 |
| Random effects | | | | | | | | |
| Initial score variance | 0.32*** | 0.09 | 0.00 | 0.01 | 0.02*** | 0.00 | 0.25** | 0.09 |
| Growth rate variance | 0.45* | 0.22 | 0.25*** | 0.04 | 0.00 | 0.00 | 0.38 | 0.22 |
| Level 1 residual variance | 1.43*** | 0.17 | 0.13*** | 0.03 | 0.01*** | 0.00 | 1.58*** | 0.16 |
| Model fit statistics | | | | | | | | |
| 2 log likelihood | 4038.55 | | | | | | | |
| scaling correction factor for MLR | 1.0669 | | | | | | | |
| AIC | 4102.55 | | | | | | | |
| BIC | 4241.05 | | | | | | | |
| Chi$^2$-difference test[a] | $\chi2(10)$[c] = 66.92*** | | | | | | | |

Note.
*$p < .05$. **$p < .01$. ***$p < .001$.
[a] As analyses used MLR estimation, Satorra-Bentler scaled chi-square tests (Satorra & Bentler, 2001) were applied based on loglikelihood values to compare model fits.
[b] Model 1 adds four growth values and six covariances between the outcome variables.
[c] Model 2 adds four values of growth rate variance and six covariances between the slopes because courses aimed to foster all professional vision skills simultaneously.

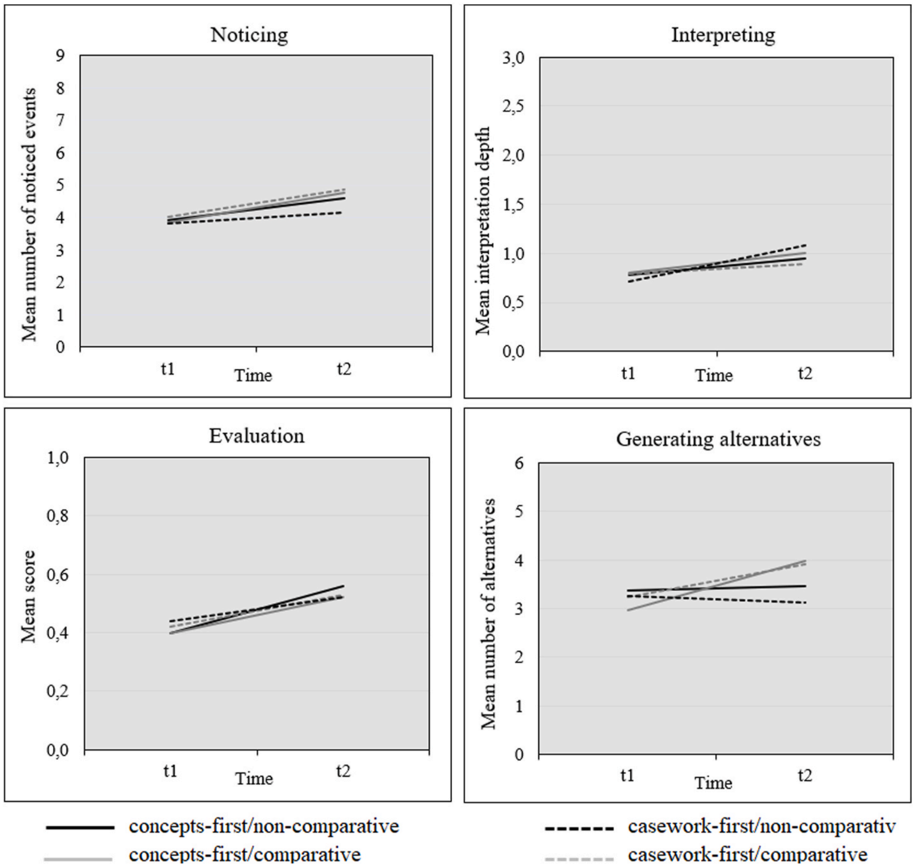


**Fig. 3.** Development of student teachers' professional vision in different instructional settings.

their skills in CM evaluation to a slightly greater extent than students in casework-first settings (*H2a*; see Seidel et al., 2013), although results should be interpreted very carefully due to the small β-value and residual variance. Students in the more advantageous concepts-first conditions first familiarised with given CM concepts and, afterwards, identified exemplifying video events. This rather guided knowledge acquisition (Blomberg et al., 2014; Schworm & Renkl, 2007) zeroed in on the relevant concepts from the very beginning, providing a clear focus without cognitively overstraining students (Kirschner et al., 2006; Kumschick et al., 2017). Additionally, the thematised concepts were largely in line with those covered by the standardised video evaluation test and therefore presumably facilitated the retrieval of according knowledge structures when triggered by the pre-formulated items.

However, the order of conceptual input and video casework did neither have an effect on *noticing* and *interpreting* of observed CM events (*H2a*) nor on *generating alternatives* (*H2b*). Apparently, the advantageous effects of this initial focus on germane CM concepts in concepts-first settings (see also Likourezos & Kalyuga, 2017; Paas & van Gog, 2006) did not surface equally regarding the variables bound to the written video analysis, underscoring that these tasks require different cognitive skills and are only partially comparable (Müller & Gold, 2022). In an open writing task, students have to filter the highly complex classroom situation for themselves, identify those events they consider important and autonomously conjure possible interpretations based on their prior knowledge. Generating alternatives, even, reaches beyond mere video analysis for it does not connect to events that are observable in a video but to imaginary events that might have taken place instead of an observable action. It seems that both concepts-first and casework-first settings can be equally successful in promoting these skills. The standardised evaluation items, conversely, channel students' attention towards specific concepts (Weyers et al., 2023), namely the very concepts they focussed on from the beginning in the concepts-first courses, thereby triggering according recall. This might explain why a stronger guidance in instruction fosters improvement of evaluative skills.

The variability in our results also indicates that analysing videos is subject to many different influences, not only instructional settings. Apart from student teachers' knowledge, their beliefs on CM or practical teaching experience might to be taken into account as possible moderators in future studies, but also affective dispositions as students' motivation for video work (Gaudin & Chaliès, 2015) or how they perceive videos emotionally (Tucholka et al., 2025).

### 5.3. General discussion

Our results highlight that the effects of different instructional settings and observation tasks depend on the concrete aspect of professional vision. Teacher educators should make instructional decisions carefully with regard to the specific skill they aim to promote, underscoring the need to consider different cognitive processes of professional vision individually when investigating the effects of different course conditions. These dependencies may illuminate why we only found punctual effects of selected conditional aspects and accordingly explain the absence of any additive higher-order effects. Using videos in teacher education, apparently, does not only include working with highly complex stimuli but represents a highly complex endeavour itself (see also Baecher et al., 2018), including the choice between manifold observation tasks or working modes and the potential influence of general circumstances and individual student dispositions, not all of which could be considered in this study. While we made sure that the groups did not differ significantly in their pretest results, different levels

**Table 3**
Model parameters and goodness of fit for conditional linear growth models.

| Parameter | Noticing | | Interpreting | | Evaluation | | Alternatives | |
|---|---|---|---|---|---|---|---|---|
| | Est. | *SE* | Est. | *SE* | Est. | *SE* | Est. | *SE* |
| Model 3: Linear growth model with random intercepts, random slopes and level-2 predictors | | | | | | | | |
| Fixed effects | | | | | | | | |
| Mean initial score | 3.85*** | 0.08 | 0.78*** | 0.02 | 0.41*** | 0.01 | 3.16*** | 0.08 |
| Mean growth rate | 0.68*** | 0.18 | 0.17* | 0.08 | 0.15*** | 0.02 | 0.23 | 0.19 |
| Predictors Level 2 | | | | | | | | |
| Task (0 = non-comparative) | 0.21 | 0.23 | 0.05 | 0.10 | −0.03 | 0.02 | 0.64** | 0.23 |
| Order (0 = concepts-first) | −0.40 | 0.26 | 0.14 | 0.11 | −0.06* | 0.02 | −0.34 | 0.24 |
| Task × Order | 0.54 | 0.35 | −0.23 | 0.15 | 0.06 | 0.04 | 0.23 | 0.34 |
| Random effects | | | | | | | | |
| Initial score variance | 0.31** | 0.09 | 0.01 | 0.01 | 0.02*** | 0.00 | 0.30** | 0.09 |
| Growth rate residual variance | 0.37 | 0.22 | 0.25*** | 0.04 | 0.00 | 0.00 | 0.27 | 0.21 |
| Level 1 residual variance | 1.44*** | 0.18 | 0.13*** | 0.03 | 0.01*** | 0.00 | 1.50*** | 0.16 |
| Model fit statistics | | | | | | | | |
| 2 log likelihood | 4003.87 | | | | | | | |
| scaling correction factor for MLR | 1.0427 | | | | | | | |
| AIC | 4091.87 | | | | | | | |
| BIC | 4282.30 | | | | | | | |
| Chi$^2$-difference test[a] | $\chi^2(12) = 35.45$*** | | | | | | | |
| Model 4: Linear growth model with random intercepts, random slopes and level-2 predictors | | | | | | | | |
| Fixed effects | | | | | | | | |
| Mean initial score | 3.85*** | 0.08 | 0.78*** | 0.02 | 0.41*** | 0.01 | 3.16*** | 0.08 |
| Mean growth rate | 0.54* | 0.22 | 0.12 | 0.10 | 0.13*** | 0.02 | 0.16 | 0.24 |
| Predictors Level 2 | | | | | | | | |
| Task (0 = non-comparative) | 0.34 | 0.23 | 0.03 | 0.10 | −0.03 | 0.03 | 0.71** | 0.24 |
| Order (0 = concepts-first) | −0.26 | 0.32 | 0.08 | 0.13 | −0.07 | 0.03* | −0.28 | 0.32 |
| Task × Order | 0.05 | 0.40 | −0.18 | 0.16 | 0.04 | 0.04 | 0.10 | 0.39 |
| Instructor (dummy 1)[a] | 0.32 | 0.24 | 0.10 | 0.11 | 0.02 | 0.03 | 0.07 | 0.24 |
| Instructor (dummy 2)[a] | 0.04 | 0.32 | 0.03 | 0.13 | 0.04 | 0.03 | 0.16 | 0.32 |
| Instructor (dummy 3)[a] | −0.40 | 0.30 | 0.15 | 0.13 | 0.02 | 0.03 | −0.05 | 0.31 |
| Random effects | | | | | | | | |
| Initial score variance | 0.27** | 0.09 | 0.01 | 0.01 | 0.02*** | 0.00 | 0.31** | 0.09 |
| Growth rate residual variance | 0.31 | 0.21 | 0.25*** | 0.04 | 0.00 | 0.00 | 0.26 | 0.21 |
| Level 1 residual variance | 1.48*** | 0.18 | 0.13*** | 0.03 | 0.01*** | 0.00 | 1.50*** | 0.15 |
| Model fit statistics | | | | | | | | |
| 2 log likelihood | 3991.38 | | | | | | | |
| scaling correction factor for MLR | 1.0293 | | | | | | | |
| AIC | 4103.39 | | | | | | | |
| BIC | 4345.75 | | | | | | | |
| Chi$^2$-difference test[b] | $\chi^2(12) = 12.74$ | | | | | | | |

Note.
*$p < .05$. **$p < .01$. ***$p < .001$.
[a] As courses were supervised by four different instructors, we included three dummy-coded instructor variables.
[b] As analyses used MLR estimation, Satorra-Bentler scaled chi-square tests (Satorra & Bentler, 2001) were applied based on loglikelihood values were applied to compare model fits.

of knowledge on CM or practical experience in teaching might moderate the effect of different instructional settings on students' professional vision development.[4]

That, on average, students improved their noticing and interpreting skills similarly well in all instructional settings seemingly represents a distinct opposition to the overall consensus that classroom videos have to be embedded in an adequate instructional setting to be effective (Kang & van Es, 2019). At second glance, though, an instructional setting was in fact provided in every course condition but apparently *all* conditions were equally suitable to promote these competencies. Across all instructional settings, student teachers dealt with the same concepts, received the same readings and analysed the same video clips, which obviously was a sufficient base for all groups to notably improve their noticing and interpreting skills (see also Trninic et al., 2022 on effects of preparatory activities). Regarding our data, however, it seems that, to improve noticing and interpreting through video analysis, it is important to *somehow* engage in concept acquisition and connect those concepts to specific video sequences, *regardless* of which specific instructional setting is applied. For future studies, it might be interesting to review these results regarding face-to-face courses allowing for more discussion and interaction than online and self-study settings (Janeczko et al., 2024; see also Loibl & Rummel, 2015). In the current study, we gave a higher priority to standardisation across conditions and thus to controlling confounding factors.

Despite the general growth regardless of course conditions, however, we still found significant variance in the respective growth rates. Therefore, apart from the conditional factors we investigated, other factors seem at play influencing students' professional vision development. Professional vision requires knowledge and the application of that knowledge in a specific situation (Sherin, 2007). As previous knowledge facilitates learning new information (Witherby & Carpenter, 2022), students with more previous CM knowledge were potentially better able to build and extent according knowledge structures in all course conditions, while students without previous knowledge made slower progress. Additionally, students with previous experience in teaching or in video analysis might have been better prepared for the complexity these situations involve, while students with less experience might have been overwhelmed by it (Mayer & Fiorella, 2014), aggravating the transfer of knowledge to the presented scene. While professional vision development has mostly been considered focussing on the competence as such, a more individualised approach might be called for (Müller & Gold, 2025).

Furthermore, there seems to be a difference between the application of knowledge to an *observable* video scene (as required for noticing, interpreting and evaluating CM events) and to imaginary possible actions *beyond* it (as required for generating alternative courses of actions). At the same time, generating alternatives seems closer to the observers' own teaching behaviour than analysing somebody else's teaching. Having been neglected for a long time, especially more recent studies consider alternative generation a core aspect of professional vision (Gippert et al., 2022; Stahnke & Blömeke, 2021a), not least because it represents an important link to real teaching performance in the classroom (Blömeke et al., 2022) and is connected to pupils' learning gains (Kersting et al., 2012). Therefore, further research should put a stronger focus on how to support future teachers in generating situation-appropriate alternatives.

### 5.4. Limitations

We are aware that our study comes with some limitations. First, our intervention included four experimental groups but lacked an untreated control group as it was the study's main aim to compare differential impacts of instructional settings rather than proving the general impact of video analysis on the development of professional vision (Gold et al., 2021; Seidel et al., 2013). Furthermore, the seminars were part of the student teachers' regular university schedule, which did not allow for leaving one group untreated. While a total intervention duration of seven weeks allows for possible influences from outside concerning internal validity, this design makes our study highly ecologically valid.

Both pre- and posttest were identical, included a writing assignment and used the same video clips to keep the tests comparable. Although video observation is considered a motivating activity (Gaudin & Chaliès, 2015), this might have reduced student teachers' motivation. Moreover, professional vision is seen as a focus-specific competence (Gippert et al., 2022) and our findings only apply to the domain of CM. Finally, more than just the four specific instructional settings we included might be considered, for instance different amounts and ways of scaffolding and guidance (Nagarajan & Hmelo-Silver, 2006; Santagata & Angelici, 2010) provided during video analysis. Further studies on video-based courses should take different evidence-based aspects of instructional settings into account (e.g. video segmentation: Martin et al., 2022; expert feedback: Prilop et al., 2021).

## 6. Conclusion

Specific instructional settings in video courses influence the development of student teachers' professional vision of CM to different extents and in different qualities. In a nutshell, our results lead to two main conclusions. First, the instructional settings applied in video-based CM courses strongly depend on the intended learning goal (Kang & van Es, 2019). For example, it seems advisable to choose *comparative tasks* to foster the generation of multiple alternative courses of action. Secondly, however, we also found that—while it is important *that* video situations are embedded into the course design (Kang & van Es, 2019)—it seems less decisive *how* this is done regarding noticing and interpreting skills while it seems more decisive regarding inventive activities as generating alternatives.

## CRediT authorship contribution statement

**Isabell Tucholka:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Bernadette Gold:** Writing – review & editing, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2025.102084.

[4] A knowledge test on classroom management, as included in pre-registration, did not show sufficient reliability values and was excluded from data analysis.

## Appendix

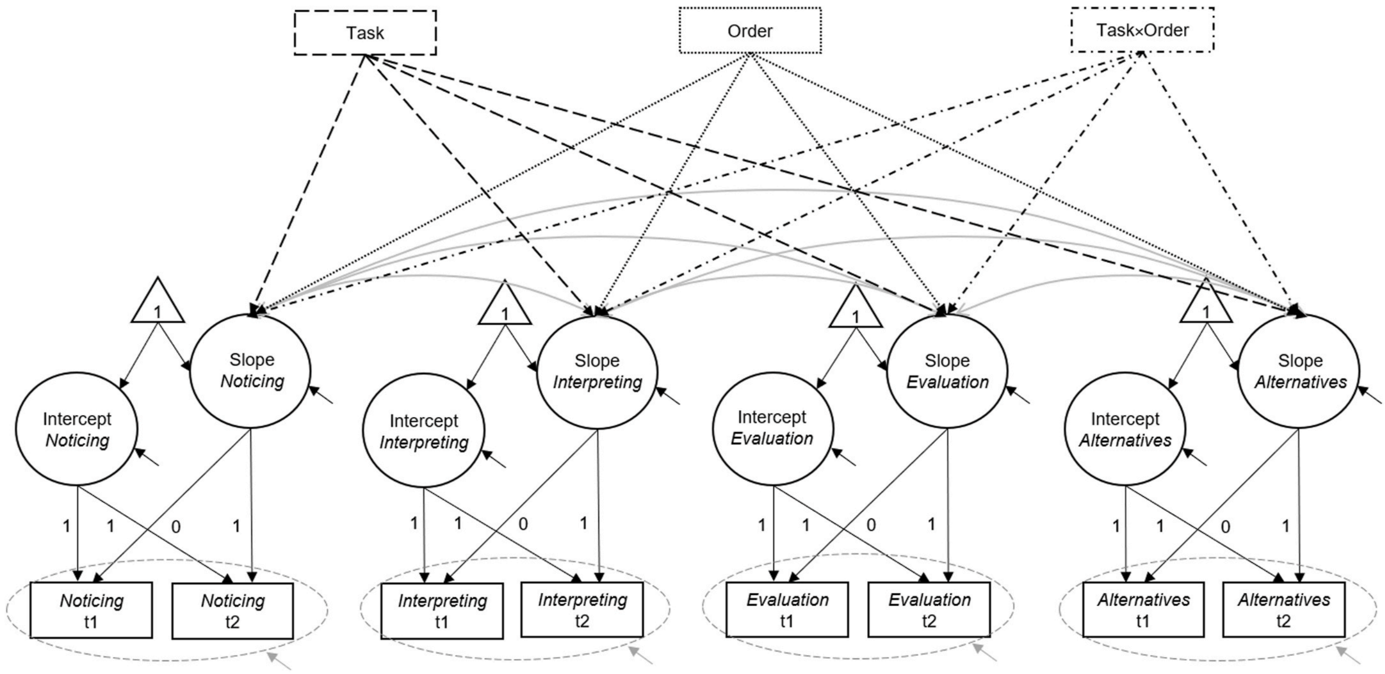


**Fig. A1.** Graphical representation of linear growth model with random intercepts, random slopes and level-2 predictors (Model 3)
*Note.* The figure illustrates the structure of our multivariate latent growth model (model 3). The lower rectangles represent the four outcome variables at two measurement points (t1, t2), with combined residual variances at level 1 (short arrows). Covariances between the outcome variables were allowed but are omitted from the diagram for clarity. The manifest variables are linked to the latent variables (circles) through fixed factor loadings: 1 for the random intercepts and 0 and 1 for the random slopes, reflecting the growth trajectories over time. The triangles represent a constant (1) used to calculate the means of the latent intercepts and slopes. Variances of the latent variables are represented by short arrows while covariances between the slopes are shown as grey double-headed arrows.
To transition from Model 2 to Model 3, dichotomous predictors (task, order, and their interaction) were added to predict the random slope scores, as indicated by dotted/dashed lines. Model 4 further included additional predictors (instructors) as dummy variables; these were modelled similarly but are not shown in the figure.

## Data availability

Data will be made available on request.

## References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712
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.
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
Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. 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. 25–48). Springer. https://doi.org/10.1007/978-1-4614-5149-5_2.
Beitzel, B. D., & Derry, S. J. (2009). When the book is better than the movie: How contrasting video cases influence text learning. *Journal of Educational Computing Research, 40*(3), 337–355. https://doi.org/10.2190/EC.40.3.e
Blomberg, G., Sherin, M. G., Renkl, A., Glogger, I., & Seidel, T. (2014). Understanding video as a tool for teacher education: Investigating instructional strategies to promote reflection. *Instructional Science, 42*(3), 443–463. https://doi.org/10.1007/s11251-013-9281-6
Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction, 79*, Article 101600. https://doi.org/10.1016/j.learninstruc.2022.101600
Doyle, W. (2006). Ecological management and classroom management. In C. M. Evertson, & C. S. Weinstein (Eds.), *Handbook of classroom management* (pp. 97–126). Erlbaum.
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
Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393–408. https://doi.org/10.1037/0022-0663.95.2.393
Gippert, C., Hörter, P., Junker, R., & Holodynski, M. (2022). Professional vision of teaching as a focus-specific or focus-integrated skill: Conceptual considerations and video-based assessment. *Teaching and Teacher Education, 117*, Article 103797. https://doi.org/10.1016/j.tate.2022.103797
Gold, B., & Holodynski, M. (2017). Using digital video to measure the professional vision of elementary classroom management: Test validation and methodological challenges. *Computers & Education, 107*, 13–30. https://doi.org/10.1016/j.compedu.2016.12.012
Gold, B., Pfirrmann, C., & Holodynski, M. (2021). Promoting professional vision of classroom management through different analytic perspectives in video-based learning environments. *Journal of Teacher Education, 72*(4), 431–447. https://doi.org/10.1177/0022487120963681
Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. https://doi.org/10.4324/9780203887332
Hirstein, A., Denn, A., Jurkowski, S., & Lipowsky, F. (2017). Entwicklung der professionellen Wahrnehmungs- und Beurteilungsfähigkeit von Lehramtsstudierenden durch das Lernen mit kontrastierenden Videofällen—Anlage und erste Ergebnisse des Projekts KONTRAST. *Beiträge zur Lehrerinnen- und Lehrerbildung, 35*(3), 472–486. https://doi.org/10.25656/01:16995
Jamil, F. M., Sabol, T. J., Hamre, B. K., & Pianta, R. C. (2015). Assessing teachers' skills in detecting and identifying effective interactions in the classroom: Theory and measurement. *The Elementary School Journal, 115*(3), 407–432.
Janeczko, J. M., Junker, R., Hörter, P., & Holodynski, M. (2024). Förderung der professionellen Wahrnehmung und Selbstwirksamkeitsüberzeugungen über Klassenführung mithilfe eines digitalen, videobasierten Mastery-Learning-Moduls. *Unterrichtswissenschaft*, 1–28. https://doi.org/10.1007/s42010-024-00197-2
Jonassen, D. H. (2006). Typology of case-based learning: The content, form, and function of cases. *Educational Technology, 46*(4), 11–15.
Kalyuga, S., & Singh, A. M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*, 831–852.

Kang, H., & van Es, E. A. (2019). Articulating design principles for productive use of video in preservice education. *Journal of Teacher Education, 70*(3), 237–250. https://doi.org/10.1177/0022487118778549

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring useable knowledge. *American Educational Research Journal, 49*(3), 568–589. https://doi.org/10.3102/0002831212437853

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. https://doi.org/10.1207/s15326985ep4102_1

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

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N., & Blum, W. (2020). Competence as a continuum in the COACTIV study: The "cascade model". *ZDM, 52*(2), 311–327. https://doi.org/10.1007/s11858-020-01151-z

Kumschick, I. R., Piwowar, V., Ophardt, D., Barth, V., Krysmanski, K., & Thiel, F. (2017). Optimierung einer videobasierten Lerngelegenheit im Problem Based Learning Format durch Cognitive Tools. Eine Interventionsstudie mit Lehramtsstudierenden. *Zeitschrift für Erziehungswissenschaft, 20*(1), 93–113. https://doi.org/10.1007/s11618-017-0728-4

Kurtz, K. J., & Gentner, D. (2013). Detecting anomalous features in complex stimuli: The role of structured comparison. *Journal of Experimental Psychology: Applied, 19*(3), 219–232. https://doi.org/10.1037/a0034395

Lamnina, M., & Chase, C. C. (2019). Developing a thirst for knowledge: How uncertainty in the classroom influences curiosity, affect, learning, and transfer. *Contemporary Educational Psychology, 59*(1), 1–13. https://doi.org/10.1016/j.cedpsych.2019.101785

Likourezos, V., & Kalyuga, S. (2017). Instruction-first and problem-solving-first approaches: Alternative pathways to learning complex tasks. *Instructional Science, 45*(1), 195–219. https://doi.org/10.1007/s11251-016-9399-4

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*(4), 693–715. https://doi.org/10.1007/s10648-016-9379-x

Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction, 34*(1), 74–85. https://doi.org/10.1016/j.learninstruc.2014.08.004

Loibl, K., & Rummel, N. (2015). Discussing student solutions is germane for learning when providing or delaying instruction. In Y. Cho, I. Caleon, & M. Kapur (Eds.), *Authentic problem solving and learning in the 21st century* (pp. 229–248). Springer. https://doi.org/10.1007/978-981-287-521-1_13.

Loibl, K., Tillema, M., Rummel, N., & van Gog, T. (2020). The effect of contrasting cases during problem solving prior to and after instruction. *Instructional Science, 48*(2), 115–136. https://doi.org/10.1007/s11251-020-09504-7

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

Martin, M., Farrell, M., Seidel, T., Rieß, W., Könings, K. D., van Merriënboer, J. J., & Renkl, A. (2022). Focused self-explanation prompts and segmenting foster pre-service teachers' professional vision-but only during training. *International Journal of Educational Technology in Higher Education, 19*(1), 1–29. https://doi.org/10.1186/s41239-022-00331-z

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279–315). Cambridge University Press.

Müller, M. M., & Gold, B. (2022). Videobasierte Erfassung wissensbasierten Verarbeitens als Teilprozess der professionellen Unterrichtswahrnehmung–Analyse eines geschlossenen und offenen Verfahrens. *Zeitschrift für Erziehungswissenschaft, 26*(1), 7–29. https://doi.org/10.1007/s11618-022-01128-6

Müller, M. M., & Gold, B. (2025). *Exploring profiles of pre-service teachers' situation-specific skills of classroom management*. Submitted for publication].

Muthén, B. O., & Muthén, L. K. (2017). *Mplus User's guide* (8th ed.). Muthén & Muthén.

Nagarajan, A., & Hmelo-Silver, C. (2006, June). Scaffolding learning from contrasting video cases. *In Proceedings of the 7th international conference on learning sciences: Indiana University. Volume 1 (pp. 495–501). International Society of the Learning Sciences.*. https://doi.org/10.22318/icls2006.495

Nagarajan, A., Hmelo-Silver, C. E., & Chernobilsky, E. (2004). The benefits and challenges of learning from contrasting video cases. In Y. B. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of 6th international conference of the learning sciences* (p. 624). Erlbaum. https://doi.org/10.22318/icls2006.495.

Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction, 16*(2), 87–91. https://doi.org/10.1016/j.learninstruc.2006.02.004

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119. https://doi.org/10.3102/0013189X09332374

Prilop, C. N., Weber, K. E., & Kleinknecht, M. (2021). The role of expert feedback in the development of pre-service teachers' professional vision of classroom management in an online blended learning environment. *Teaching and Teacher Education, 99*, Article 103276. https://doi.org/10.1016/j.tate.2020.103276

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. Z. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching, 48*(2), 117–148. https://doi.org/10.1002/tea.20408

Santagata, R., & Angelici, G. (2010). Studying the impact of the lesson analysis framework on preservice teachers' abilities to reflect on videos of classroom teaching. *Journal of Teacher Education, 61*(4), 339–349. https://doi.org/10.1177/0022487110369555

Santagata, R., König, J., Scheiner, T., Nguyen, H., Adleff, A. K., Yang, X., & Kaiser, G. (2021). Mathematics teacher learning to notice: A systematic review of studies of video-based programs. *ZDM–Mathematics Education, 53*(1), 119–134. https://doi.org/10.1007/s11858-020-01216-z

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507–514. https://doi.org/10.1007/BF02296192

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522. https://doi.org/10.1207/s1532690xci1604_4

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99*(2), 285–296. https://doi.org/10.1037/0022-0663.99.2.285

Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education, 34*, 56–65. https://doi.org/10.1016/j.tate.2013.03.004

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, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 383–395). Erlbaum.

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. (2013). Those who understand: Knowledge growth in teaching. *Journal of Education, 193*(3), 1–11. https://doi.org/10.1177/002205741319300302

Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research, 91*(5), 761–798. https://doi.org/10.3102/00346543211019105

Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology, 60*, Article 100832. https://doi.org/10.1016/j.newideapsych.2020.100832

Stahnke, R., & Blömeke, S. (2021a). Novice and expert teachers' situation-specific skills regarding classroom management: What do they perceive, interpret and suggest? *Teaching and Teacher Education, 98*, Article 103243. https://doi.org/10.1016/j.tate.2020.103243

Stahnke, R., & Blömeke, S. (2021b). Novice and expert teachers' noticing of classroom management in whole-group and partner work activities: Evidence from teachers' gaze and identification of events. *Learning and Instruction, 74*, Article 101464. https://doi.org/10.1016/j.learninstruc.2021.101464

Trninic, D., Sinha, T., & Kapur, M. (2022). Comparing the effectiveness of preparatory activities that help undergraduate students learn from instruction. *Learning and Instruction, 82, Article 101688*. https://doi.org/10.1016/j.learninstruc.2022.101688

Tucholka, I., Busse, A., Bosse, D., & Gold, B. (2025). Emotions in classroom video evaluation—The interplay of student teachers' joy and anger and the quality of classroom management in video clips. *International Journal of Educational Research, 130*, Article 102539. https://doi.org/10.1016/j.ijer.2025.102539

Weber, K. E., Gold, B., Prilop, C. N., & Kleinknecht, M. (2018). Promoting pre-service teachers' professional vision of classroom management during practical school training: Effects of a structured online-and video-based self-reflection and feedback intervention. *Teaching and Teacher Education, 76*, 39–49. https://doi.org/10.1016/j.tate.2018.08.008

Weyers, J., König, J., Santagata, R., Scheiner, T., & Kaiser, G. (2023). Measuring teacher noticing: A scoping review of standardized instruments. *Teaching and Teacher Education, 122*, Article 103970. https://doi.org/10.1016/j.tate.2022.103970

Wilkes, T., Stark, L., Trempler, K., & Stark, R. (2022). Contrastive video examples in teacher education: A matter of sequence and prompts. *Frontiers in Education, 7*, Article 869664. https://doi.org/10.3389/feduc.2022.869664

Witherby, A. E., & Carpenter, S. K. (2022). The rich-get-richer effect: Prior knowledge predicts new learning of domain-relevant information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(4), 483–498. https://doi.org/10.1037/xlm0000996

Wolff, C. E., van den Bogert, N., Jarodzka, H., & Boshuizen, H. P. A. (2015). Keeping an eye on learning. *Journal of Teacher Education, 66*(1), 68–85. https://doi.org/10.1177/0022487114549810