Comparing Pre-Service Teachers', In-Service Teachers' and Educational Researchers' Evidence-Informed Reasoning About Classroom Situations – Results of a Mixed Methods Investigation

Ein Vergleich des evidenzorientierten Denkens von Lehramtsstudierenden, Lehrkräften und Bildungswissenschaftler*innen über Unterrichtssituationen – Ergebnisse einer Mixed Methods-Untersuchung

Christina Wekerle[1], Katharina Kiemer[1], Kai Wagner[3], Kati Trempler[2], Theresa Wilkes[3], Martin Greisel[1], Robin Stark[3] & Ingo Kollar[1]

[1]University of Augsburg, Universitätsstraße 10, 86159 Augsburg, Germany
Christina Wekerle; +49 (0)821 598 5526; christina.wekerle@phil.uni-augsburg.de
Dr. Katharina Kiemer; katharina.kiemer@phil.uni-augsburg.de
Dr. Martin Greisel; +49 (0)821 598 4392; martin.greisel@phil.uni-augsburg.de
Prof. Dr. Ingo Kollar; +49 (0)821 598 2607; ingo.kollar@phil.uni-augsburg.de

[2]University of Wuppertal, Gaußstraße 20, 42119 Wuppertal, Germany
Dr. Kati Trempler; +49 (0)202 439-3941; trempler@uni-wuppertal.de

[3]Saarland University, Campus, 66123 Saarbrücken, Germany
Dr. Kai Wagner; 13kw@gmx.de
Theresa Wilkes; +49 (0)681 302 4163; theresa.wilkes@uni-saarland.de
Prof. Dr. Robin Stark; +49 (0)681 302 3765; r.stark@mx.uni-saarland.de

Korrespondierende Autorin:
Christina Wekerle
University of Augsburg
Universitätsstraße 10
86159 Augsburg
E-Mail: christina.wekerle@phil.uni-augsburg.de

This work was supported by the German Research Foundation [grant number 395758789].

RUNNING HEAD: EVIDENCE-INFORMED REASONING ABOUT CLASSROOM SITUATIONS

Comparing Pre-Service Teachers', In-Service Teachers' and Educational Researchers' Evidence-Informed Reasoning About Classroom Situations – Results from a Mixed Methods Investigation

Ein Vergleich des evidenzorientierten Denkens von Lehramtsstudierenden, Lehrkräften und Bildungswissenschaftler*innen über Unterrichtssituationen – Ergebnisse einer Mixed Methods-Untersuchung

Abstract

Pre-service teachers rarely engage in evidence-informed reasoning when they are confronted with problematic classroom situations. We argue that interventions that target pre-service teachers' acquisition of evidence-informed reasoning skills should be informed by research that compares pre-service teachers', in-service teachers', and educational researchers' evidence-informed reasoning. We asked $N = 55$ pre-service teachers, in-service teachers, and educational researchers to think aloud about a written classroom scenario and complete a retrospective interview on their evidence-informed reasoning. Results indicate that educational researchers describe problematic events more often than pre- and in-service teachers but do not seem to differ on a number of other reasoning activities. However, educational researchers more often refer to academic knowledge than pre- and in-service teachers. Pre- and in-service teachers do not seem to differ from each other, neither with respect to their reasoning activities nor concerning their use of academic knowledge. Additional qualitative analyses illustrate these findings.

Keywords

Evidence-based practice, pre-service teachers, in-service teachers, educational researchers, mixed methods research

Zusammenfassung

Lehramtsstudierende gehen selten evidenzorientiert vor, wenn sie mit problematischen Unterrichtssituationen konfrontiert sind. Wir argumentieren, dass Interventionen, die auf eine Förderung evidenzorientierter Denkfertigkeiten bei Lehramtsstudierenden abzielen, auf Ergebnissen vergleichender Forschung zu den Denkprozessen von Lehramtsstudierenden, Lehrkräften und Bildungswissenschaftler*innen basieren sollten. Wir baten $N = 55$ Lehramtsstudierende, Lehrkräfte und Bildungswissenschaftler*innen, über ein schriftliches Unterrichtsszenario laut nachzudenken und an einem retrospektiven Interview zur Rekonstruktion ihrer evidenzorientierten Denkprozesse teilzunehmen. Die Ergebnisse zeigen, dass Bildungswissenschaftler*innen Situationen häufiger als Lehramtsstudierende und Lehrkräfte beschreiben, sich von diesen aber nicht in anderen evidenzorientierten Denkprozessen unterscheiden. Jedoch beziehen sich Bildungswissenschaftler*innen in ihren Analysen häufiger auf bildungswissenschaftliches Wissen als Lehramtsstudierende und Lehrkräfte. Lehramtsstudierende und Lehrkräfte unterscheiden sich hinsichtlich ihrer Denkaktivitäten und Verwendung bildungswissenschaftlichen Wissens nicht voneinander. Die Ergebnisse werden durch weiterführende qualitative Analysen illustriert.

Stichworte

Evidenzbasierte Praxis, Lehramtsstudierende, Lehrkräfte, Bildungswissenschaftler, Mixed Methods Forschung

Background and aims

In addressing problematic classroom situations, teachers are increasingly being asked to not only ground their decisions and actions on their individual experience, but also on evidence from educational research (e.g., Slavin, 2008). This kind of teachers' reasoning has been called "evidence-based education" (Davies, 1999). The demand for evidence-based education is driven by at least two sources: First, it is fueled by political goals such as an increase of the economic competitiveness of and the social cohesiveness in society (European Commission, 2007, p. 11). Second, research has demonstrated evidence-based, pedagogical-psychological knowledge of teachers to be positively related to instructional quality and students' learning outcomes (e.g., König & Pflanzl, 2016). However, such calls for teachers' evidence-based reasoning are not undisputed (Dekker & Meeter, 2022). For example, Sharples (2013) argues that the term "evidence-based education" is not to be mistaken for a clear recipe for any kind of educational problem. Educational science as a scientific discipline is characterized by its probabilistic rather than mechanistic perspective. Thus, in this article, we use the term evidence-*informed* reasoning to indicate that scientific evidence should not be regarded as a recipe. Instead, it serves as a resource that teachers can refer to as an orientation to reason in educational situations that are uncertain or that repeatedly proof difficult (Nelson, Mehta, Sharples, & Davey, 2017). Furthermore, building upon Zimmerman (2000), we propose a distinction between two dimensions of evidence-informed reasoning: The *form dimension* refers to the kind of reasoning activities teachers employ when reflecting about classroom problems (such as hypothesizing or explaining; Fischer et al., 2014). The *content dimension* concerns teachers' retrieval and application of (scientific and other kinds of) knowledge that might be helpful for understanding and solving the problem at hand (Csanadi, Kollar & Fischer, 2021).

Research has shown that pre-service teachers have considerable difficulties on both the form and content dimensions of evidence-informed reasoning (Bergner, 2018; Menz, Spinath, & Seifried, 2021; Santagata & Angelici, 2010). Thus, teacher training programs play an

important role in providing learning opportunities to practice evidence-informed reasoning (Hetmanek et al., 2015). We argue that the development of such learning opportunities should be rooted (1) in normative frameworks of professional evidence-informed reasoning (e.g., Fischer et al., 2014) and (2) in research that compares pre-service teachers' reasoning to that of experienced in-service teachers and educational researchers. We consider this comparison important for at least two reasons: First, comparing pre-service teachers and in-service teachers might illustrate the effects of *professional experience* on teachers' evidence-informed reasoning. Thus, this comparison may provide information on how to design instructional interventions that help pre-service teachers to acquire the skills that in-service teachers have acquired through practical experience. Second, comparing pre-service teachers and educational researchers may help to better understand the effect of *academic expertise* on evidence-informed reasoning. Educational researchers' reasoning should be insightful to the design of instructional interventions for pre-service teachers, given that educational researchers are supposedly experts in the application of evidence from educational research. However, comparing in-service teachers possessing professional experience and educational researchers with academic expertise might also be promising, as results from such a comparison may indicate how these two groups could profit from each other's particular kinds of competencies, for example, when establishing research-practice collaborations.

Differentiating the content and form dimensions of evidence-informed reasoning

In line with Zimmerman (2000), we propose a distinction between a content and a form dimension of evidence-informed reasoning. We do so for the following reasons: First, the dimensions constitute theoretically different constructs. The form dimension refers to a cross-domain approach, which focuses on the procedural knowledge of a person, which can be applied across multiple domains. In contrast, the content dimension refers to a domain-specific approach to scientific reasoning, wherein the emphasis is placed on the conceptual knowledge

of a person about phenomena within a given scientific field (Zimmerman, 2000). Second, the dimensions have been found to not only constitute theoretically, but also empirically different constructs. There is evidence that training measures may yield differential effects on the two distinct dimensions (e.g., Wekerle & Kollar, 2021). Thus, this differentiation might provide a more complete picture of the problems pre-service teachers experience as compared to in-service teachers and educational researchers. A line of research that provides a basis for conceptualizing the form and content dimension of evidence-informed reasoning is research on professional vision (e.g., van Es & Sherin, 2010), which concerns the abilities members of a professional group share in interpreting typical field-specific phenomena.

The form dimension of evidence-informed reasoning

This dimension refers to the set of reasoning activities pre-service teachers engage in when confronted with a problematic classroom situation and include procedural knowledge about different reasoning steps and how to perform them. Research on professional vision identified reasoning activities such as noticing, describing, and explaining relevant classroom situations as well as generating alternative strategies to handle such situations as central components in the successful analysis of classroom situations (Kersting et al., 2012; van Es & Sherin, 2008). Comparable reasoning activities (termed ‘epistemic activities’ or ‘diagnostic activities’) have also been determined in research on teachers’ scientific reasoning (e.g., Fischer et al., 2014; Kramer, Förtsch, Seidel, & Neuhaus, 2021). Both strands of research have in common that they do not further differentiate the activity of generating alternatives or drawing conclusions. Yet, research on self-management (Lenzen, Daniëls, van Bokhoven, van der Weijden, & Beurskens, 2017) shows that this activity should be further specified into deriving goals and developing courses of action. Thus, we suggest teachers’ professional evidence-informed reasoning about classroom situations to optimally include the following five activities:

(1) *Notice* significant instances: In order to develop a solution for a problematic classroom situation, teachers first identify the problem at hand. For example, a teacher might realize that their students are inattentive to the present task.

(1) *Describe* significant instances: In a next step, teachers reconstruct or categorize the problem by applying experiential or academic concepts and principles. The said teacher might understand their students' inattention as stemming from uncertainty about what exactly to do.

(2) *Explain* significant instances: Subsequently, teachers try to explain the problem based on well-reflected academic or experiential knowledge. This includes ordering, ranking, and weighing individual aspects of the situation and integrating them into a comprehensive explanatory model. In our example, the teacher might come up with the explanation that their students are inattentive because of a lack of motivation to learn. This state might be caused by a relatively low task value (see Wigfield & Eccles, 2000), as the task might be abstract and hold little personal relevance (see Krapp, 2000).

(3) *Derive* objectives for action: Once teachers have developed a suitable explanation for the classroom situation, they select objectives for subsequent actions. In our example, the teacher might reason that there is a need to increase the task value.

(4) *Develop* options for action: Finally, pre-service teachers select actions they deem effective for reaching their objectives. The teacher in our example might seek to increase the task value by interrupting the task, mentioning their observation to their students, explaining why the content is important and where it might reappear in their everyday lives.

Empirical research has shown that pre-service teachers rarely follow such a structured reasoning process when confronted with problematic classroom situations. For example, Santagata and Angelici (2010) asked pre-service teachers to reason about a short video-based

case vignette of a classroom situation. Results revealed that participants failed to provide explanations of the classroom event if not prompted to do so, but instead tended to only recite significant instances (see also Csanadi et al., 2021).

The content dimension of evidence-informed reasoning

The content dimension is orthogonal to the form dimension, as it reflects the type of mainly conceptual knowledge that (pre-service) teachers apply in each of these steps when analyzing a challenging classroom situation. In line with Menz et al. (2021), we suggest a distinction between academic and experiential knowledge. Experiential knowledge refers to knowledge that is reflected in subjective theories. Teachers use such knowledge to derive subjective hypotheses which they consider true (Richardson, 2003) based on subjective practical experience (see also "tacit knowledge"; Dimmock, 2016). Academic knowledge, in contrast, encompasses knowledge about academic theories and empirical evidence from educational research. This type of knowledge is typically laid out in external information sources (Chinn, Buckland & Samarapungavan, 2011) such as journal articles, research reports, or handbooks. Despite the potential of academic knowledge (e.g., its predictive value for instructional quality; König & Pflanzl, 2016), studies suggest that pre-service teachers rarely actually have such knowledge. For example, Menz et al. (2021) showed that pre-service teachers stated experience as the main source of their knowledge about teaching. In contrast, academic knowledge was mentioned significantly less often. Also, in an interview study, Bergner (2018) demonstrated that less than one third of pre-service teachers' explanations of a classroom problem contained any traces of educational theories or evidence. Moreover, only roughly 10% of their solutions referenced academic knowledge.

Overall, previous research provides evidence that pre-service teachers tend to only focus on some of the proposed activities and make rare use of academic knowledge. Consequently, both dimensions should be considered when designing a curricular intervention.

A comparative approach to analyzing pre-service teachers' evidence-informed reasoning

In line with an understanding of evidence-informed reasoning as being guided by professional experience and academic expertise (Sharples, 2013), we argue that curricular interventions for pre-service teachers should account for both kinds of expertise. A comparison between pre-service teachers, in-service teachers and educational researchers should allow for the assessment of pre-service teachers' prerequisites as well as for the definition of learning goals based on the competencies demonstrated by more experienced and knowledgeable persons (Auerbach et al., 2018).

Comparing pre-service teachers' and in-service teachers' evidence-informed reasoning

Research about pre- and in-service teachers' reasoning practices with respect to the form dimension has resulted in contradictory findings. In line with research on teacher expertise (e.g., Gegenfurtner, Lewalter, Lehtinen, Schmidt, & Gruber, 2020), in-service teachers' greater classroom experience should result in richer, well-developed reasoning schemata than those of pre-service teachers. For example, in a study by Wolff, Jarodzka and Boshuizen (2017), experienced in-service teachers offered more interpretations of problematic classroom situations than pre-service teachers. In addition, when being asked to analyze photographs of classroom situations, Gegenfurtner et al. (2020) found in-service teachers to describe relevant information more often than pre-service teachers. However, in several professional vision development studies, Sherin and van Es (Sherin & van Es, 2009; van Es & Sherin, 2008, 2010) showed that in-service teachers tend to mainly *recite* and/or *evaluate* (video-based) classroom situations, rather than actually *explaining* them. Also, Kim and Klassen (2018) were not able to discern any differences in the frequencies of in- and pre-service teachers' reasoning activities when being confronted with school-based scenarios.

Likewise, empirical research on pre- and in-service teachers' reasoning regarding the content dimension yielded ambiguous results. Experiential knowledge seems to be a much more valuable information resource for in-service teachers when deciding which teaching approaches

to employ, as compared to academic knowledge (Nelson et al., 2017). Further studies have shown that in-service teachers rarely make use of academic knowledge. For example, Cain (2015) provided in-service teachers with research reports about teaching gifted and talented students and supported them in using these findings in their teaching over a 12-month period. Analyses indicated that the teachers used the research reports only very occasionally. Hetmanek et al. (2015) further concluded that even when in-service teachers report that they apply academic knowledge, this knowledge is often superficial and not well-suited to justifying pedagogical decisions. Thus, based on these findings, pre- and in-service teachers might not extensively differ regarding their use of academic knowledge. However, results of the aforementioned study by Gegenfurtner et al. (2020) indicated that in-service teachers were more inclined to refer to pedagogical content knowledge than pre-service teachers. Furthermore, in video-based reasoning settings, in-service teachers reached significantly higher professional vision scores than pre-service teachers (e.g., Gold & Holodynski, 2017). Due to these ambiguous findings on the form and content dimension, a comparison of pre- and in-service teachers' reasoning based on rich open-ended data might yield new insights for the design of curricular interventions to promote pre-service teachers' evidence-informed reasoning skills.

Comparing pre-service teachers' and educational researchers' evidence-informed reasoning

Educational researchers should have supposedly acquired extensive academic knowledge due to their ongoing experience in this domain and be prone to look at teaching challenges through a scientific lens. Results based on self-report data corroborate these assumptions: The greater the research experience possessed by teacher educators, the more they report possessing practical knowledge for the implementation of evidence-informed teaching practices, experiencing a sense of self-efficacy in relation to evidence-informed actions, and relying on evidence in their own teaching (Georgiou, Mok, Fischer, Vermunt, & Seidel, 2020). Yet, educational researchers' reasoning might differ from how evidence-informed reasoning

should look like based on theoretical and normative considerations. However, pre-service teachers' and researchers' actual evidence-informed reasoning have rarely been compared systematically in the past. When investigated at all, research has mostly been conducted in well-structured STEM domains with experts and novices without a teacher education background. Here, a well-known finding points to different problem-solving strategies by researchers and undergraduates in the field of physics: Researchers tend to use a forward-working strategy in which they begin problem-solving from what is given in a problem statement and elaborate on these givens. In contrast, undergraduates tend to work backwards by focusing on the solution and identifying possible strategies that might lead to that result, often in a trial and error-like fashion (Kohl & Finkelstein, 2008). Similarly, Chi, Feltovich, and Glaser (1981) found physics researchers to represent problems and their solutions in terms of deeper principles not stated in the problem (so called "second-order features"), while undergraduates rely on general, surface characteristics mentioned in the problem statement when describing problems. Based on these findings, we tentatively assume that educational researchers will engage in different reasoning activities and apply different kinds of knowledge than pre-service teachers. However, research by Feist (1994) indicates that educational researchers might use different reasoning strategies when reasoning about classroom situations compared to when they engage in research. Also, a study by Randles and Overton (2015) illustrates that (chemistry) undergraduates' and researchers' primary strategies for problem-solving might not differ much. Consequently, differences between pre-service teachers' and educational researchers' reasoning might be smaller than expected. Due to the conflicting results and lacking results outside of the STEM domain, we suggest comparing educational researchers' and pre-service teachers' form- and content-related reasoning capabilities to help teacher curriculum designers derive implications for effective interventions.

Aims and research questions

The ambiguous findings on the form and content dimensions based on comparisons of pre-service teachers with in-service teachers and educational researchers suggest that investigating these groups using open-ended data could inform curricular interventions to enhance the evidence-informed reasoning in pre-service teachers.

Therefore, we aimed to compare the evidence-informed reasoning of (a) pre-service teachers to those of (b) in-service teachers, and (c) educational researchers with a mixed methods approach based on rich, open-ended data in the form of thinking aloud and interview protocols.

Our exploratory research questions were: What are the differences between pre-service teachers, experienced in-service teachers, and educational researchers regarding their use of evidence-informed reasoning in addressing classroom problems in terms of the form dimension (RQ1) and the content dimensions (RQ2)?

Method

Participants and design

Our sample consisted of $N = 55$ participants. Nineteen participants were pre-service teachers at the beginning of their studies ($M_{age} = 21.79$, $SD_{age} = 2.80$, $\%_{female} = 94.7$; $M_{semester} = 4.37$, $SD_{semester} = 4.59$). One pre-service teacher was enrolled in a primary, 2 in a lower secondary, and 16 in an upper secondary school teacher education undergraduate program at two German state universities. Twelve pre-service teachers who were enrolled in upper secondary programs studied a language subject, 11 a social science subject, 3 a STEM subject, and 2 an arts subject. Note that pre-service teachers study two subjects in upper secondary school teacher education programs in Germany. Three pre-service teachers in the primary school and lower secondary school teacher education program studied a wide range of subjects.

Eighteen participants were experienced in-service teachers (M_{Age} = 39.83, SD_{age} = 8.46, $\%_{female}$ = 66.7) with more than 5 years of teaching experience (M_{years} = 15.10, SD_{years} = 4.82). Three of the in-service teachers were primary school teachers, 4 lower secondary and 10 upper secondary school teachers (1 teacher of a different school type). Eight of the upper secondary school teachers taught a language subject, 2 a STEM subject, 7 a social science subject and none an arts subject. Again, 7 primary school and lower secondary school teachers taught a wide range of subjects.

Finally, 18 participants were educational researchers (M_{age} = 37.72, SD_{age} = 6.62; $\%_{female}$ = 50.0) at three German state universities with at minimum a Ph.D. in educational science or psychology. Eight of these educational researchers reported to mainly work in the field of educational science, 9 researchers in the field of psychology/educational psychology and 1 in both fields. Participants were required to at least state a minimum relationship to teacher education in their research, which applied to 17 researchers (no relationship = 0, minimum relationship = 5, significant relationship = 8, solely research on teacher education = 4), or to have already taught courses in teacher education programs, which also applied to 17 researchers (number of courses: M = 23.75, SD = 23.16, Min = 6). One researcher had completed teacher training. No researcher had previously worked as a school teacher.

Pre-service teachers were recruited in mandatory educational psychology courses at two German state universities and received course credit for their participation. In-service teachers were approached via teacher networks of two German state universities. Educational researchers were approached via Germany-based research networks of the authors. In-service teachers and educational researchers received no compensation for their participation.

Procedure

Each participant was invited to an individual meeting with an experimenter. Each meeting consisted of four parts. First, participants were asked to answer a paper-pencil questionnaire on demographic variables and several control variables. Then, they were

introduced to the think aloud method. Afterwards, they were provided with a written, fictitious description of a problematic classroom lesson developed by the authors. Even though the case was fictitious, it was inspired by authentic classroom situations reported in informal exchanges with teachers. The written case focused on different student motivational issues in the context of a learning circle and how a teacher trainee handled the situation. The description of these motivational issues was based on well-known concepts and theories in the field of motivational research such as self-determination theory (Deci & Ryan, 1985), attribution theory (Weiner, 1985), theory of interest (Krapp, 2000), self-concept (Marsh, 1986), achievement motivations (Pekrun, Elliot, & Maier, 2006), control-value theory (Pekrun, 2006), and expectancy-value theory (Wigfield & Eccles, 2000). The written case vignette was 650 words long and can be found in the appendix. Participants were asked to thoroughly read the description of the classroom lesson, underline text segments, were allowed to take notes, and were asked to express every thought that came to their mind while reflecting on the case vignette. In line with Ericsson and Simon's (1993) suggestions, every time participants remained silent for some time, the experimenter asked them to continue speaking out loud what was going through their mind. On average, the think-aloud phase took 9.35 (*SD* = 3.51) minutes.

After that, participants were interviewed by one of four trained experimenters. The interviews consisted of questions addressing participants' retrospective perceived engagement in the cognitive activities (e.g., "Did you reason about actions that you could have taken in place of Ms. Sander to achieve the goals you mentioned?", "Which actions did you consider?"), how they proceeded when performing each activity (e.g., "How did you proceed in order to select teacher actions?"), and what kind of knowledge they applied in each step (e.g., "How do you know that these actions might be successful?"). Questions were asked one by one. If participants did not give an answer that addressed the question, the question was repeated or reframed by the experimenter. The interviews took 20.11 (*SD* = 5.95) minutes on average.

Finally, participants were asked to verbally rate their knowledge of different motivational concepts and theories.

Variables

Form dimension

To analyze the form dimension, we first transcribed the think-aloud data from each participant. We then segmented the data into dialogic units involving complete lines of reasoning, which mostly consisted of several sentences. The segments were introduced by changes in speaker or when a participant started a new line of reasoning after a break of at least five seconds. The segments were then coded with a coding scheme that distinguished between participants' engagement in the five reasoning activities:

Noticing significant instances. A segment was coded as "noticing significant instances" if participants revealed that they detected a problem in the case description (e.g., "Ms. Sander focuses too much on the two students").

Describing significant instances. This code was assigned when a problem in the case description was categorized using terms or concepts not mentioned in the case description (e.g., "There are expectancy problems and there are value problems").

Explaining significant instances. A segment was coded as an explanation when participants hypothesized about cause-effect relationships regarding a problem based on information in the case or their own prior knowledge (e.g., "They believe that they won't get it right as the other teacher has probably always inculcated them with the belief 'We don't have any German language skills', which might be why they may have started to believe that they don't have any German language skills").

Deriving objectives. This code was applied when participants pointed out which aspect of the situation should be addressed by the teacher and what might be desirable goals to reach to solve the situation (e.g., "When designing the lesson and introducing the topic of the lesson, Ms. Sander should make sure to draw on the students' motivation.").

Developing options for actions. A segment was coded in this category when participants made suggestions about how to introduce concrete changes to achieve certain goals (e.g., "She could have also said: 'You have been highly engaged in today's work, so no homework for today'.").

Each segment received a dichotomous code of *did not occur* (0) or *did occur* (1) on all five categories. Thus, the codes were not mutually exclusive, as more than one cognitive activity could occur in each segment. Interrater agreement based on two independent trained undergraduate research assistants who coded data from 22% of participants was above 80% for all variables (noticing: 81%, describing: 85%, explaining: 82%, deriving objectives: 86%, developing options for action: 89%). Due to the low prevalence of some categories, we used prevalence-adjusted bias-adjusted kappa (PABAK; see Byrt, Bishop, & Carlin, 1993), which ranged between $\kappa = .62$ and $\kappa = .77$ (noticing: .62, describing: .70, explaining: .64, deriving objectives: .73, developing options for action: .77). Disagreements between the two coders were resolved through discussion. The remaining data were split between the two coders. The mean for each category across all segments for each participant was used as a variable in further analyses.

Content dimension

The content dimension was investigated by analyzing the interview data, as its dialogue-based nature made it easier to determine the kind of knowledge participants used to reason about the case. This approach assumed that individuals do not consistently state their sources of knowledge, especially when they possess encapsulated, not easily retrievable knowledge (see Boshuizen & Schmidt, 1992) and when they are not challenged to do so. Again, we first segmented the data into dialogic units. We then applied a coding scheme to each segment that measured participants' references of academic knowledge. An expression within a segment was identified as a reference to academic knowledge when participants verbalized educational constructs, theories, empirical findings, or academically derived teaching actions (e.g., "In this

case, the exam causes extrinsic motivation."). This category was coded by assigning points to each segment assessing the extent of academic references (max = 6 points). Up to three points were awarded for mentioning educational constructs, theories, or empirical findings ranging from *no references to educational constructs, theories, or empirical findings* (0) to *distinct references to educational constructs, theories, or empirical findings* (3). Up to three points were awarded for mentioning academic strategic knowledge ranging from *no references to academically derived teaching actions* (0) to *distinct references to academically derived teaching actions* (3).

Interrater agreement was determined based on two independent trained coders who coded the data of 22% of participants. Both coders were trained undergraduate research assistants. Interrater agreement was good, ICC(2,1) = .76. The remaining data were coded by one coder. The mean across all segments for each participant was used as a variable in the analyses.

Results

Preliminary analyses

First, to confirm the expected differences between educational researchers' and teachers' (subjective) knowledge prerequisites, we asked participants to verbally rate their knowledge of the following motivational concepts and theories on a scale ranging from (1) *no knowledge* to (10) *profound knowledge*: (1) self-determination theory (Deci & Ryan, 1985), (2) attribution theory (Weiner, 1985), (3) theory of interest (Krapp, 2000), (4) self-concept (Marsh, 1986), (5) achievement motivations (Pekrun et al., 2006), (6) control-value theory (Pekrun, 2006), and (7) expectancy-value theory (Wigfield & Eccles, 2000). Internal consistency was very good (Cronbach's α = .92). Differences between pre-service teachers' (M = 3.31, SD = 1.80), in-service teachers' (M = 2.92, SD = 1.62), and educational researchers' (M = 6.80, SD = 1.32) knowledge of motivational concepts and theories were statistically significant, resulting

in a large effect, $F(2, 52) = 32.53$, $p < .001$, part. $\eta^2 = .56$. Post-hoc Bonferroni-corrected tests showed that educational researchers reported to have significantly more knowledge than pre-service teachers, $p < .001$, and in-service teachers, $p < .001$. The differences between pre- and in-service teachers were not significant, $p = 1.00$.

Second, we checked for possible differences in the number of segments produced by the different groups during thinking aloud and interview. The descriptives point to comparable numbers for pre-service teachers (thinking aloud: $M = 11.68$, $SD = 5.82$; interview: $M = 7.32$, $SD = 1.83$), in-service teachers (thinking aloud: $M = 12.5$, $SD = 5.68$; interview: $M = 8.06$, $SD = 1.92$) and educational researchers (thinking aloud: $M = 11.11$, $SD = 5.33$; interview: $M = 7.94$, $SD = 2.31$). No significant differences were found; thinking aloud: $F(2, 52) = 0.28$, $p > .05$, interview: $F(2, 52) = 0.72$, $p > .05$.

Third, we analyzed the bivariate correlations between form, content dimension variables and subjective knowledge. While the manifest correlations (see Table 1) illustrate that the different noticing and reasoning activities (form dimension) were rather closely associated with each other, they were barely associated with the content dimension, except for one significant positive correlation between references to academic knowledge and describing significant instances. Academic knowledge was also positively associated with subjective knowledge.

Table 1

Descriptive statistics and manifest correlations of the form dimension, content dimension and subjective knowledge for pre-service teachers, in-service teachers and educational researchers.

	M	*SD*	Min	Max	Skew	1	2	3	4	5	6
Form dimension											
(1) Noticing	0.52	0.25	0.00	1.00	0.04						

(2) Describing	0.21	0.27	0.00	1.00	1.75	.69**					
(3) Explaining	0.25	0.20	0.00	1.00	1.67	.81**	.75**				
(4) Deriving objectives	0.17	0.24	0.00	1.00	2.36	.55**	.81**	.65**			
(5) Developing options for action	0.28	0.30	0.00	1.00	1.24	.45**	.50**	.61**	.63**		
(6) Content dimension	0.76	0.76	0.00	3.00	1.24	.11	.33*	.06	.23	–.09	
(7) Subjective knowledge	4.32	2.35	1.00	9.14	0.26	–.11	.17	–.13	.08	–.28*	.57**

Note. $N = 55$. * $p < .05$; ** $p < .01$

Research question 1: form dimension

Table 2 presents the extent to which pre- and in-service teachers as well as educational researchers engaged in the different noticing and reasoning activities. While all groups most often engaged in noticing significant instances, educational researchers engaged in this activity slightly more often than pre- and in-service teachers. Large differences were found for describing significant instances: educational researchers engaged in this activity more often than pre- or in-service teachers. Explaining significant instances was used by all groups to a comparable amount. Educational researchers engaged in deriving objectives most frequently, followed by in-service teachers and then pre-service teachers. Finally, pre- and in-service teachers developed options for actions slightly more frequently than educational researchers.

A MANCOVA with group as the independent variable, the five activities as dependent variables and average number of words per segment as covariate revealed a significant, large effect, $F(10, 94) = 3.00$, $p < .01$, part. $\eta^2 = .24$, Wilk's $\Lambda = 0.58$. Post-hoc univariate ANOVAs showed a significant, large difference between groups in describing significant instances, $F(2, 51) = 5.91$, $p < .01$, part. $\eta^2 = .19$. Pair-wise Bonferroni-corrected comparisons revealed that educational researchers described significant instances significantly more often than pre-service

teachers, p = .02, and in-service teachers, p = .01. No other significant group differences regarding the four remaining activities were found, all $F(2, 51) < 1.81$, $p > .05$.

Table 2

Means for use of activities among pre-service teachers, in-service teachers and educational researchers.

	All		Pre-service teachers		In-service teachers		Educational researchers	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Noticing	0.52	0.25	0.52	0.28	0.47	0.21	0.57	0.25
Describing	0.21	0.27	0.13	0.18	0.14	0.23	0.37	0.33
Explaining	0.25	0.20	0.26	0.18	0.21	0.17	0.27	0.25
Deriving objectives	0.17	0.24	0.08	0.10	0.17	0.24	0.25	0.32
Developing options for action	0.28	0.30	0.28	0.24	0.33	0.28	0.23	0.37
Content dimension	0.76	0.78	0.33	0.30	0.44	0.47	1.53	0.81

Research Question 2: content dimension

Descriptive statistics regarding the degree to which pre-service teachers, in-service teachers, and educational researchers referred to academic knowledge when explaining their reasoning are presented in Table 2. While pre- and in-service teachers only rarely referred to academic knowledge, educational researchers tended to use academic knowledge in almost every line of argumentation.

An ANCOVA with group as the independent variable, references to academic knowledge as dependent variables, and average number of words per segment as covariate revealed a significant, large effect, $F(2, 51) = 24.59$, $p < .001$, part. $\eta^2 = .49$. Pair-wise Bonferroni-corrected comparisons found significantly more references to academic knowledge by educational researchers compared to both pre-service teachers, $p < .001$, and in-service

teachers, $p < .001$. The difference between pre- and in-service teachers was not significant, $p = 1.00$.

Qualitative analysis of typical evidence-informed reasoning processes among pre-service teachers, in-service teachers and educational researchers

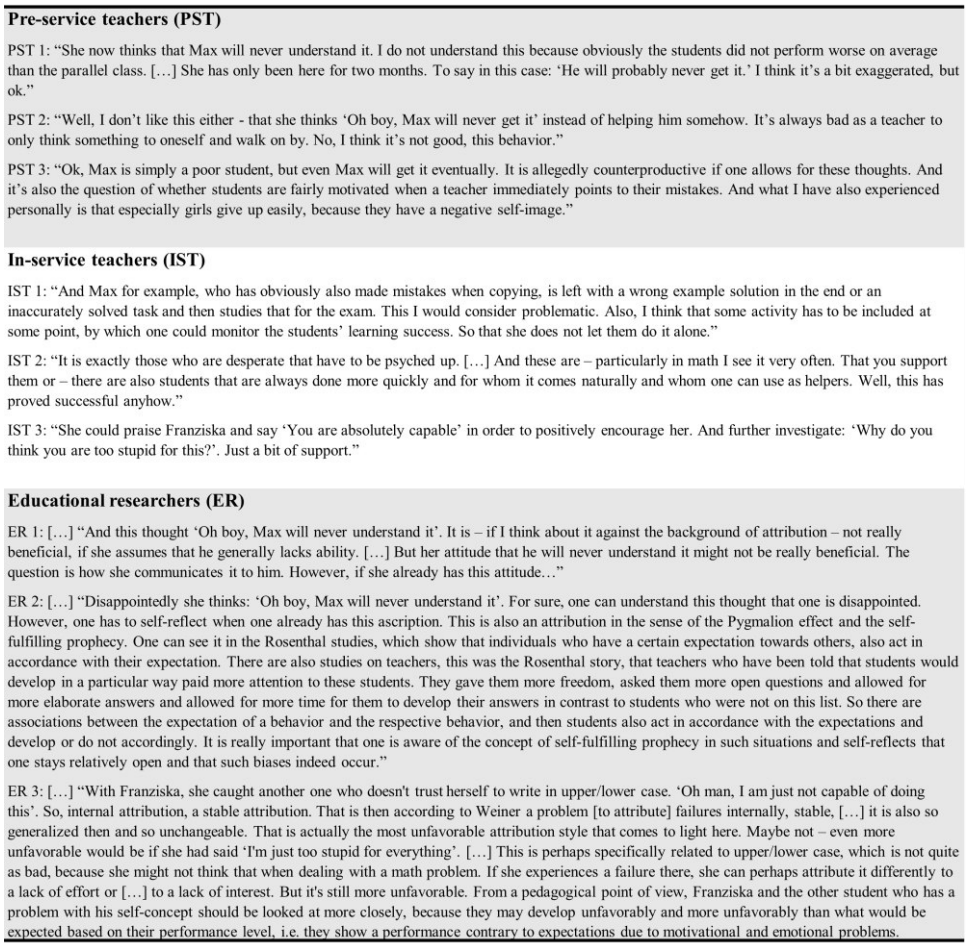
In order to better illustrate the quantitative research findings and to derive conclusions for the specific design of teacher education curricula, we elaborate on typical think-aloud statements by three pre-service teachers ($M_{Years} = 20.33$, $SD_{Years} = 0.58$; $\%_{female} = 100$), three in-service teachers ($M_{Years} = 36.67$, $SD_{Years} = 3.79$; $\%_{female} = 100$) and three educational researchers ($M_{Years} = 41.33$, $SD_{Years} = 11.50$; $\%_{female} = 66.7$). All statements refer to the following section of the case vignette:

"As Ms. Sander looks over Max's shoulder, she instantly spots several mistakes in the comma placement task. Disappointedly she thinks: 'Oh boy! Max will never get this…'. As she tries to point out to Franziska some mistakes regarding her use of capital and small initial letters, Franziska answers in desperation: 'Oh man, I am just not capable of doing this.'"

We have chosen this particular situation for its richness, as it simultaneously focuses on the learner's and teacher's perspectives, providing diverse starting points for analysis. To identify typical participants' statements, we first selected all segments in which participants referred to the outlined section of the case vignette by using the keywords 'Max', 'disappointed', 'Franziska', 'in desperation', 'never get this', and 'just not capable of doing this'. We then identified the segments in which at least one of the two most prominent reasoning activities for each group were demonstrated, based on the quantitative result pattern. For educational researchers, we further considered segments in which they demonstrated the use of scientific knowledge. Finally, we opted to choose three segments of the remaining segments for the respective groups. Based on the concept of "information power" for sample sizes by Malterud, Siersma and Guassora (2016), we decided on three segments per group, as one segment was too few for a cross-case analysis of unprompted statements. Yet, with our sampling based on specific characteristics and our specific objective of illustration, we

considered three segments as sufficient to capture the diversity of the respective target groups. Thus, the presented statements represented in Figure 1 roughly capture the average reasoning patterns in each respective group.

Figure 1



Pre-service teachers (PST)

PST 1: "She now thinks that Max will never understand it. I do not understand this because obviously the students did not perform worse on average than the parallel class. […] She has only been here for two months. To say in this case: 'He will probably never get it.' I think it's a bit exaggerated, but ok."

PST 2: "Well, I don't like this either - that she thinks 'Oh boy, Max will never get it' instead of helping him somehow. It's always bad as a teacher to only think something to oneself and walk on by. No, I think it's not good, this behavior."

PST 3: "Ok, Max is simply a poor student, but even Max will get it eventually. It is allegedly counterproductive if one allows for these thoughts. And it's also the question of whether students are fairly motivated when a teacher immediately points to their mistakes. And what I have also experienced personally is that especially girls give up easily, because they have a negative self-image."

In-service teachers (IST)

IST 1: "And Max for example, who has obviously also made mistakes when copying, is left with a wrong example solution in the end or an inaccurately solved task and then studies that for the exam. This I would consider problematic. Also, I think that some activity has to be included at some point, by which one could monitor the students' learning success. So that she does not let them do it alone."

IST 2: "It is exactly those who are desperate that have to be psyched up. […] And these are – particularly in math I see it very often. That you support them or – there are also students that are always done more quickly and for whom it comes naturally and whom one can use as helpers. Well, this has proved successful anyhow."

IST 3: "She could praise Franziska and say 'You are absolutely capable' in order to positively encourage her. And further investigate: 'Why do you think you are too stupid for this?'. Just a bit of support."

Educational researchers (ER)

ER 1: […] "And this thought 'Oh boy, Max will never understand it'. It is – if I think about it against the background of attribution – not really beneficial, if she assumes that he generally lacks ability. […] But her attitude that he will never understand it might not be really beneficial. The question is how she communicates it to him. However, if she already has this attitude…"

ER 2: […] "Disappointedly she thinks: 'Oh boy, Max will never understand it'. For sure, one can understand this thought that one is disappointed. However, one has to self-reflect when one already has this ascription. This is also an attribution in the sense of the Pygmalion effect and the self-fulfilling prophecy. One can see it in the Rosenthal studies, which show that individuals who have a certain expectation towards others, also act in accordance with their expectation. There are also studies on teachers, this was the Rosenthal story, that teachers who have been told that students would develop in a particular way paid more attention to these students. They gave them more freedom, asked them more open questions and allowed for more elaborate answers and allowed for more time for them to develop their answers in contrast to students who were not on this list. So there are associations between the expectation of a behavior and the respective behavior, and then students also act in accordance with the expectations and develop or do not accordingly. It is really important that one is aware of the concept of self-fulfilling prophecy in such situations and self-reflects that one stays relatively open and that such biases indeed occur."

ER 3: […] "With Franziska, she caught another one who doesn't trust herself to write in upper/lower case. 'Oh man, I am just not capable of doing this'. So, internal attribution, a stable attribution. That is then according to Weiner a problem [to attribute] failures internally, stable, […] it is also so generalized then and so unchangeable. That is actually the most unfavorable attribution style that comes to light here. Maybe not – even more unfavorable would be if she had said 'I'm just too stupid for everything'. […] This is perhaps specifically related to upper/lower case, which is not quite as bad, because she might not think that when dealing with a math problem. If she experiences a failure there, she can perhaps attribute it differently to a lack of effort or […] to a lack of interest. But it's still more unfavorable. From a pedagogical point of view, Franziska and the other student who has a problem with his self-concept should be looked at more closely, because they may develop unfavorably and more unfavorably than what would be expected based on their performance level, i.e. they show a performance contrary to expectations due to motivational and emotional problems.

Qualitative results: form dimension

Regarding the *form dimension*, the quantitative results had shown that pre-service teachers demonstrated the activity of noticing more frequently than the other activities. This activity is illustrated by pre-service teachers (PST) 1, 2 and 3, who identified Ms. Sander's thoughts about Max's learning process as problematic (PST 1: "exaggerated", PST 2: "I don't like this either", PST 3: "counterproductive"). However, rather than reasoning about the meaning and consequences of these problematic thoughts, PST 3 directly moved on to reason about Ms. Sander's interaction with Franziska, while PST 1 mentioned reasons why Ms.

Sander's behavior was not justified in their point of view ("students did not perform worse on average than the parallel class", "only been here for two months"). PST 2 at least pointed to an alternative, albeit quite general action the teacher could take ("helping him somehow"), which was also one of the more common reasoning activities engaged in by pre-service teachers in general. Moreover, in line with pre-service teachers' rather frequent use of explanations, PST 3 reasoned about possible consequences of Ms. Sander's corrective behavior ("give up easily"), albeit in a very generalized manner ("because they have a negative self-image").

In-service teachers and pre-service teachers mostly engaged in comparable reasoning activities, only (descriptively) differing regarding in-service teachers' stronger emphasis on objectives and actions. For example, with respect to Max, in-service teacher (IST) 1 identified the need to "monitor the students' learning success" and to "not let them do it alone" as objectives. Similar patterns can be seen for IST 2 and IST 3: In their statements, both in-service teachers mostly mentioned objectives to be reached (IST 2: "have to be psyched up", "that you support them", IST3: "positively encourage", "further investigate"), and subsequent practical measures to be taken (IST 2: "for whom it comes naturally and whom one can use as helpers", IST 3: "You are absolutely capable", "Why do you think…?").

The quantitative results had illustrated that educational researchers' use of reasoning activities differed more strongly from that of pre-service teachers. They engaged to a particularly high extent in describing significant instances, but also (at least descriptively) in deriving objectives. In line with these results, all three educational researchers in this example categorized Ms. Sander's thought about Max and interaction with Franziska as instances of inappropriate attributions, which led them to further categorize the problem situation (ER 1: "ability", "her attitude"), assess the severity of the problem situation (ER3: "even more unfavorable would be…"), or refer to relevant consequences and objectives (ER 2: "Pygmalion effect", "associations between the expectation of a behavior and the respective behavior",

"aware of the concept of self-fulfilling prophecy"; ER3: "develop unfavorably than what would be expected based on their performance", "should be looked at more closely").

Qualitative results: content dimension

With regard to the *content dimension*, the quantitative results had demonstrated that pre-service teachers were rather unlikely to refer to academic knowledge when reasoning about the classroom case. An example is PST 3, who explicitly referred to his own experience by stating "and what I have also experienced personally…". However, the pre-service teachers also used two academic expressions ("motivated", "negative self-image"), indicating that some kind of academic knowledge may be represented as well, even if somewhat masked by personal experience.

Here again, the pattern among in-service teachers was comparable to that among pre-service teachers. For example, IST 2 introduced experiential knowledge with the statement "particularly in math, I see it very often" and "this has proved successful". Furthermore, IST 2 and IST 3 used expressions such as "support", "helper" and "positive encouragement", which seem to refer to subjective rather than academically informed concepts, even though they might have an academic source.

Finally, the quantitative results had shown that educational researchers referred to academic knowledge to a higher degree compared to the other two teacher groups. To illustrate this, ER 2's and ER3's use of academic knowledge was particularly prominent, as these participants not only mentioned academic concepts (ER2: "attribution", "Pygmalion effect", "self-fulfilling prophecy"; ER3: "internal attribution", "stable attribution", "self-concept"), but also named authors or elaborated in more detail on academic findings (ER2: "Rosenthal story"; ER3: "Weiner").

Discussion

In this exploratory mixed methods study, we adopted a comparative perspective to investigate the reasoning approaches applied by pre-service teachers, in-service teachers, and educational researchers when reflecting on a prototypical classroom case. The results illustrated only slight differences between educational researchers, pre-service teachers, and in-service teachers regarding the form dimension, but large differences regarding the content dimension.

Differences among pre-service teachers, in-service teachers and educational researchers regarding their engagement in the five activities (RQ1) were mostly negligible, apart from educational researchers who unlike pre- and in-service teachers engaged more often in descriptive activities. Our data do not suggest that in-service teachers focus more on describing and explaining classroom events than pre-service teachers, as was the case in the studies by Wolff et al. (2017) and Gegenfurtner et al. (2020). Nevertheless, our research findings align well with a study by Kim and Klassen (2018) in which pre- and in-service teachers showed a comparable degree of reasoning activities. A reason for the lacking effect in the study of Kim and Klassen (2018) and our study might be that the samples consisted of in-service teachers who were approached based on their teaching *experience* but not on their teaching *excellence* as in the studies of Wolff et al. (2017) and Gegenfurtner et al. (2020). Thus, the reasoning of our in-service teacher sample might rather illustrate a "typical" competent than an expert-like trajectory of pre-service teachers' reasoning pattern based on Dreyfus' (2014) model of skill acquisition. Therefore, professional experience alone might not affect (pre-)service teachers' reasoning to a significant degree, and we cannot draw any specific conclusions for the promotion of pre-service teachers.

Our finding that educational researchers engaged in conceptual descriptions of problematic situations more frequently than pre-service teachers seems to confirm what Chi et al. (1981) detected in the field of physics: Researchers represent problems by means of second-order features abstracted from the problem statement, which might in turn strongly influence

their subsequent solution procedure. This guidance by concepts was also a feature of the example statements by educational researchers, as they directly linked academic concepts to explanations or objectives. However, to better understand the role of or the association between describing and the other reasoning activities, a more detailed and sequentially structured analysis would be necessary in future research. Still, it is surprising that no other differences were found between pre-service teachers' and educational researchers' reasoning in our sample from a perspective of academic expertise. There might be two reasons for this: First, as demonstrated by Feist (1994), educational researchers' reasoning might depend on the domain at hand. This means that educational researchers might engage in more explanatory activities when reasoning about their research than when reasoning about a "practical" teaching problem in a classroom context. Second, rather unspecific, generalized lines of reasoning suggest that the differences between pre-service teachers' and educational researchers' reasoning activities might only become visible at a more fine-grained level of analysis. It thus might be worthwhile to develop further coding criteria to determine the quality of reasoning activities on the form dimension by inductively analyzing educational researchers' reasoning activities. Nevertheless, based on the difference regarding pre-service teachers' and educational researchers' frequency of descriptions, support measures should scaffold pre-service teachers to more often describe significant instances when reasoning about classroom situations. As in-service teachers' frequencies of descriptions also differed from those of educational researchers, in-service teachers might also profit from prompting to describe significant instances when reasoning about uncertain classroom teaching (Nelson et al., 2017). As we could not detect any reasoning activities that were more prominent in in-service teachers' compared to educational researchers' reasoning about the case vignette, we may not draw any particular conclusions for how educational researchers could benefit from in-service teachers' professional experience when collaborating.

Regarding the content dimension of evidence-informed reasoning (RQ2), pre- and in-service teachers did not differ in the extent to which they referred to academic knowledge. Thus, our findings do not replicate findings based on research on professional vision (see Gegenfurtner et al., 2020; Gold & Holodynski, 2017; Meschede et al., 2017) which demonstrated that in-service teachers refer more to scientific pedagogical content knowledge than pre-service teachers. Yet, they do confirm earlier results on teachers' scientific reasoning, which illustrated that both pre- and in-service teachers only partly use academic knowledge (Bergner, 2018; Nelson et al., 2017). These contradictory results might be due to the same argument we outlined for the form dimension: our sample may have included experienced competent, but not expert teachers. Consequently, the negligible use of academic knowledge by experienced in-service teachers indicates that "typical" pre-service teachers might show comparable patterns in their future careers as they do during teacher training. That said, it is difficult to infer particular instructional measures to support pre-service teachers in their use of academic knowledge. However, from a perspective of professional experience and based on our qualitative analyses, we do not know to which degree in-service teachers might possess academic knowledge that is encapsulated in their experiential knowledge (Boshuizen & Schmidt, 1992). Nevertheless, considering the content of our case vignette, the objectively as well as the subjectively assessed data suggest pre- and in-service teachers lack academic knowledge on motivational theories.

In contrast, the fact that our sample of educational researchers used academic knowledge more extensively compared to pre-service teachers confirms insights gained in STEM-domains (e.g., Chi et al., 1981) about educational researchers' approach to problem situations. Based on the subjectively assessed knowledge in our study, pre- and in-service teachers seem to lack knowledge of academic concepts such as motivational theories and therefore might not have been able to apply it. The correlation between subjective knowledge and the use of academic knowledge further indicates that persons who stated to possess

conceptual knowledge were also more likely to apply academic knowledge to a teaching setting. Consequently, we suggest helping pre-service teachers as well as in-service teachers to acquire conceptual knowledge as a prerequisite to deal with problematic motivation-related classroom situations in an evidence-informed way. In this regard, higher-order constructive (such as self-explaining evidence) and interactive learning activities (explaining evidence/questioning learning partners) might be promising (Engelmann, Hetmanek, Neuhaus, & Fischer, 2022). However, in a next step, it might be worthwhile to further investigate pre-service teachers', in-service teachers', and educational researchers' use of knowledge in more detail by analyzing to what degree they come to equivalent actions despite using different kinds and sources of knowledge. This might be particularly valuable to better understand the potential of experiential knowledge when reasoning about classroom situations.

Limitations

Despite strengths there are also certain shortcomings in our study that need to be addressed. First, we only used a single case vignette to determine differences in pre-service teachers', in-service teachers', and educational researchers' reasoning. Thus, it remains open whether our results generalize to other classroom situations addressing different problems such as classroom management issues or challenges related to students' diversity. Second, we used different sources of data to investigate the form and content dimensions (think aloud vs. interview protocol data), because we considered think aloud data as a valid source for observing unprompted reasoning activities but did not expect participants to consistently elaborate on their sources of knowledge. While we do feel that this methodological decision was well warranted, the downside is that this approach does not allow for a comparison of results gained from two sources of data. Third, as illustrated in the presentation of the qualitative characteristics of IST 3's statements, the identification of academic knowledge was not always easy. Even though we examined rather large segments as our idea units and asked participants to name the source of

the knowledge used, our results may underestimate participants' use of academic knowledge in cases in which they fused academic knowledge with experiential knowledge (particularly among experienced in-service teachers; Boshuizen & Schmidt, 1992). Finally, while in our study in-service teachers can be compared on a continuum as experts relative to novice pre-service teachers, this might not be the case for educational researchers as they differ from pre-service teachers both in terms of their educational background and their research career. Thus, the implementation of support measures for pre-service teachers based on their comparison with educational researchers needs to carefully consider pre-service teachers' prerequisites.

Conclusions

Based on our study results, we suggest practitioners such as curriculum designers and higher education staff to attend to students' acquisition and transfer of academic concepts more closely. It might be particularly promising to help students understand specific (problematic) classroom situations as examples of more abstract theoretical problem categories. A well-established method in research on professional vision (e.g., Kersting et al., 2012; Sherin & van Es, 2009) to build on is case-based learning. Pre-service teachers might be scaffolded by means of prompts and/or worked examples (Wekerle & Kollar, 2021; Krause-Wichmann et al., 2023) to systematically analyze authentic cases like the classroom situations they experienced in their internships with the help of educational evidence as well as school data (Brown, Schildkamp, & Hubers, 2017). A prerequisite for successful case-based learning is the acquisition of conceptual academic knowledge next to an understanding of the types and structure of student and school context data. Therefore, higher education staff (in particular those with a limited research background) should be familiar with recent research evidence, how to evaluate research findings and school data, and should direct pre-service teachers to adequate resources. Support in doing so is provided by clearing house initiatives (e.g., Clearing House Teaching; Hetmanek, Diery, Knogler, Schneeweiss, & Seidel, 2023). However, to ensure that higher

education staff is equipped with the necessary capabilities and resources to address these issues effectively, it appears imperative that policymakers allocate more time for teaching preparation, invest in the continued professional development, and foster collaborative partnerships between teachers and academic researchers. These factors seem to prevent higher education staff from providing high-quality learning opportunities for pre-service teachers (Diery, Vogel, Knogler, & Seidel, 2020). In addition to these actions in higher education, measures such as the creation of high accountability by inspections and high-stake assessments appear to be effective in promoting evidence-informed reasoning by teachers (Malin et al., 2020).

Despite these conclusions for practitioners and policymakers, there are still future research paths to be taken to gain more insights into pre-service teachers', in-service teachers', and educational researchers' evidence-informed reasoning: It would be worthwhile to develop alternative measurement approaches that enable a better understanding of the degree to which pre- and in-service teachers possess encapsulated knowledge. Here, research in medicine might function as a template (Rikers, Schmidt, & Boshuizen, 2000). In addition to the established comparative model-based coding approach, it might be of interest to inductively work out features that are specific to each of the three groups. Finally, including a further subsample possessing academic as well as professional expertise should be helpful to analyze the integration of academic and professional expertise. These characteristics might be pertinent in a subsample of teacher educators who both have worked in schools and have been active in research (see Diery et al., 2020).

References

Auerbach, A. J., Higgins, P., Brickman, P. & Andrews, T. C. (2018). Teacher knowledge for active-learning instruction: Expert-novice comparison reveals differences. *CBE – Life Sciences Education, 17*(1), https://doi.org/10.1187/cbe.17-07-0149

Bergner, M. (2018). *Analyse und Förderung wissenschaftlichen Denkens bei Lehramtsstudierenden* [Analysis and facilitation of pre-service teachers‘ scientific reasoning; Dissertation]. Universität Oldenburg, Oldenburg. Retrieved from https://oops.uni-oldenburg.de/id/eprint/3517

Boshuizen, H., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, *16*(2), 153–184. https://doi.org/10.1016/0364-0213(92)90022-M

Brown, C., Schildkamp, K., & Hubers, M. D. (2017). Combining the best of two worlds: a conceptual proposal for evidence-informed school improvement. *Educational Research*, *59*(2), 154–172. https://doi.org/10.1080/00131881.2017.1304327.

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, *46*(5), 423–429. https://doi.org/10.1016/0895-4356(93)90018-V

Cain, T. (2015). Teachers’ engagement with research texts: Beyond instrumental, conceptual or strategic use. *Journal of Education for Teaching*, *41*(5), 478–492. https://doi.org/10.1080/02607476.2015.1105536

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Chinn, C. A., Buckland, L. A., Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist, 46*(3), 141–167. https://doi.org/10.1080/00461520.2011.587722

Csanadi, A., Kollar, I., & Fischer, F. (2021). Pre-service teachers' evidence-based reasoning during pedagogical problem-solving: better together? *European Journal of Psychology of Education, 36*(1)*,* 147-168. https://doi.org/10.1007/s10212-020-00467-4

Davies, P. (1999). What is evidence-based education? *British Journal of Education Studies, 47*(2), 108–121. https://doi.org/10.1111/1467-8527.00106

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

Dekker, I. & Meeter, M. (2022). Evidence-based education: Objections and future directions. *Frontiers in Education.* https://doi.org/10.3389/feduc.2022.941410

Diery, A., Vogel, F., Knogler, M., & Seidel, T. (2020). Evidence-based practice in higher education: Teacher educators' attitudes, challenges, and uses. *Frontiers in Education, 5*, 183. https://doi.org/10.3389/feduc.2020.00062

Dreyfus, H. L. (2014). Intuitive, Deliberative, and Calculative Models of Expert Performance. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (2nd ed., pp. 17–28). https://doi.org/10.4324/9781315806129

Dimmock, C. (2014). Conceptualising the research–practice–professional development nexus: Mobilising schools as 'research-engaged' professional learning communities. *Professional Development in Education, 42*(1), 36–53, https://doi.org/10.1080/19415257.2014.963884

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: The MIT Press.

European Commission (2007). *Towards more knowledge-based policy and practice in education and training, SEC (2007)1098*. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/962e3b89-c546-4680-ac84-777f8f10c590

Feist, G. J. (1994). Personality and working style predictors of integrative complexity: A study of scientists' thinking about research and teaching. *Journal of Personality and Social Psychology, 67*(3), 474–484. https://doi.org/10.1037/0022-3514.67.3.474

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., Strijbos, J.-W., Heene, M., & Eberle, J. (2014). Scientific

reasoning and argumentation: advancing an interdisciplinary research agenda in education. *Frontline Learning Research, 2(3), 28-45.* https://doi.org/10.14786/flr.v2i2.96

Gegenfurtner, A., Lewalter, D., Lehtinen, E., Schmidt, M., & Gruber, H. (2020). Teacher expertise and professional vision: Examining knowledge-based reasoning of pre-service teachers, in-service teachers, and school principals. *Frontiers in Education*, *5*(59). https://doi.org/10.3389/feduc.2020.00059

Georgiou, D., Mok, S. Y., Fischer, F., Vermunt, J. D., & Seidel, T. (2020). Evidence-based practice in teacher education: The mediating role of self-efficacy beliefs and practical knowledge. *Frontiers in Education, 5*, 256. https://doi.org/10.3389/ feduc.2020.559192

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

Hetmanek, A., Diery, A., Knogler, M., Schneeweiss, A., Seidel, T. (2023). Das Clearing House Unterricht. Eine Plattform zur Stärkung der Evidenzbasierung in der Aus- und Weiterbildung von Lehrkräften. In K. S. Besa, D. Demski, D., J. Gesang, J. H. Hinzke (Eds) *Evidenz- und Forschungsorientierung in Lehrer*innenbildung, Schule, Bildungspolitik und -administration. Educational Governance, vol 55.* Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-38377-0_6

Hetmanek, A., Wecker, C., Kiesewetter, J., Trempler, K., Fischer, M. R., Gräsel, M., & Fischer, F. (2015). Wozu nutzen Lehrkräfte welche Ressourcen? Eine Interviewstudie zur Schnittstelle zwischen bildungswissenschaftlicher Forschung und professionellem Handeln im Bildungsbereich. *Unterrichtswissenschaft, 43*(3), 194-210. https://doi.org/09201503193

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos

predict teaching quality and student learning. *American Educational Research Journal*, *49*(3), 568–589. https://doi.org/10.3102/0002831212437853

Kim, L. E., & Klassen, R. M. (2018). Teachers' cognitive processing of complex school-based scenarios: Differences across experience levels. *Teaching and Teacher Education*, *73*, 215–226. https://doi.org/10.1016/j.tate.2018.04.006

König, J., & Pflanzl, B. (2016). Is teacher knowledge associated with performance? On the relationship between teachers' general pedagogical knowledge and instructional quality, *European Journal of Teacher Education*, *39*(4), 419–436, https://doi.org/10.1080/02619768.2016.1214128

Kohl, P. B. & Finkelstein , N. D. (2008). Patterns of multiple representation use by experts and novices during physics problem solving. *Physics Education Research, 4*(1). https://doi.org/10.1103/PhysRevSTPER.4.010111

Kramer, M., Förtsch, C., Seidel, T., & Neuhaus, B. J. (2021). Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Studies in Educational Evaluation*, *68*, 100973. https://doi.org/10.1016/j.stueduc.2020.100973

Krapp, A. (2000). Interest and human development during adolescence: An educational-psychological approach. In J. Heckhausen (Ed.), *Advances in psychology. Motivational psychology of human development: Developing motivation and motivating development* (Vol. 131, pp. 109–128). Elsevier. https://doi.org/10.1016/S0166-4115(00)80008-4

Krause-Wichmann, T., Greisel, M., Wekerle, C., Kollar, I. & Stark, R. (2023). Promoting future teachers‘ evidence-informed reasoning scripts: effects of different forms of instruction after problem-solving. *Frontiers in Education, 8.* https://doi.org/10.3389/feduc.2023.1001523

Lenzen S. A., Daniëls, R., van Bokhoven, M. A., van der Weijden, T. & Beurskens, A. (2017). Disentangling self-management goal setting and action planning: A scoping review. *PLoS One, 12*(11). https:// doi.org/10.1371/journal.pone.0188822

Malin, J. R., Brown, C., Ion, G., van Ackeren, I., Bremm, N., Luzmore, R., Flood, J. & Rind, G. M. (2020). World-wide barriers and enablers to achieving evidence-informed practice in education: what can be learnt from Spain, England, the United States, and Germany?. *Humanities and Social Sciences Communications*, *7*(1). https://doi.org/10.1057/s41599-020-00587-8

Malterud, K., Siersma V. D. & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research, 26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444

Marsh, H. W. (1986). Global self-esteem: Its relation to specific facets of self-concept and their importance. *Journal of Personality and Social Psychology, 51*(6), 1224–1236. https://doi.org/10.1037/0022-3514.51.6.1224

Menz, C., Spinath, B., & Seifried, E. (2021). Where do pre-service teachers' educational psychological misconceptions come from? The roles of anecdotal versus scientific evidence. *Zeitschrift Für Pädagogische Psychologie*, 35(2-3), 143–156. https://doi.org/10.1024/1010-0652/a000299

Nelson, J., Mehta, P., Sharples, J., & Davey, C. (2017). *Measuring teachers' research engagement: Findings from a pilot study*. London, UK: The Education Endowment Foundation & National Foundation for Educational Research.

Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology, 98*(3), 583–597. https://doi.org/10.1037/0022-0663.98.3.583

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341. https://doi.org/10.1007/s10648-006-9029-9

Randles, C. A. & Overton, T. L. (2015). Expert vs. novice: Approaches used by chemists when solving open-ended problems. *Chemistry Education Research and Practice, 16*(4), 811–823. https://doi.org/10.1039/C5RP00114E

Richardson, V. (2003). Preservice teachers' beliefs. In J. Raths & A. C. McAninch (Eds.), *Teacher Beliefs and Classroom Performance: The Impact of Teacher Education* (pp. 1-22). Greenwich, CT: Information Age Publishing Inc.

Rikers, R. M. J. P., Schmidt, H. G. & Boshuizen, H. P. A. (2000). Knowledge encapsulation and the intermediate effect. *Contemporary Educational Psychology, 2,* 150-166. https://doi.org/10.1006/ceps.1998.1000

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

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

Sharples, J. (2013). *Evidence for the frontline: A report for the alliance for useful evidence.* Alliance for Useful Evidence. Retrieved from https://apo.org.au/sites/default/files/resource-files/2013-06/apo-nid34800.pdf

Slavin, R. E. (2008). Evidence-based reform in Education: What will it take? *European Educational Research Journal, 7*(1), 124–128. https://doi.org/10.3102/0013189X08314117

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

van Es, E. A., & Sherin, M. G. (2010). The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education*, *13*(2), 155–176. https://doi.org/10.1007/s10857-009-9130-3

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548–573. https://doi.org/10.1037/0033-295X.92.4.548

Wekerle, C. & Kollar, I. (2021). Fostering pre-service teachers' situation-specific TPACK – Does learning by mapping and learning from worked examples help? *Computers in Human Behavior, 115,* 106617. https://doi.org/10.1016/j.chb.2020.106617

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

Wolff, C. E., Jarodzka, H., & Boshuizen, H. P. (2017). See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events. *Teaching and Teacher Education*, *66*, 295-308. https://doi.org/10.1016/j.tate.2017.04.015

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*(1), 99–149. https://doi.org/10.1006/drev.1999.0497

Appendix

Case vignette

Ms. Sander has been independently teaching a German class for two months as part of her training program. Sometimes she has the feeling that her lessons are not really well received by the students. She knows from her colleagues that this class had a very strict teacher in the previous school year. This teacher was very dissatisfied with the performance of the class and thought that many of the students would be better off at another school. Accordingly, there were many bad grades in the last report card.

Ms. Sander has observed that the students in this class often do not seem to have a particularly deep interest in German and often do not pay attention in class. However, they surprisingly did not perform worse than the parallel class in the last unannounced test.

Currently, Ms. Sander is covering the topic of "Spelling and Punctuation." Since a class test on this topic is due next week, Ms. Sander decides to hold another practice lesson. She insists that all students really participate. Therefore, she has prepared a learning circle with four stations, which each student should go through. At each station there is a worksheet with tasks of varying difficulty.

During the lesson, Ms. Sander gives the following task: "Today you will practice spelling and punctuation again. To do this, you will work on different exercises at four different stations. There is one station for upper- and lower-case spelling, one for the spelling of 'das' and 'dass', one for direct speech, and one for comma placement. There is a worksheet at each station. Always work through the tasks one by one! Under each task you will find the correct result so that you can check your solutions yourself. Since you only have one lesson to complete all the stations, you have 10 minutes for each station. You must do the tasks you cannot complete as homework. I will always give a signal when it is time to move on. We will now count from 1 to 4 so that you know at which station you should start." The students count from 1 to 4. Ms. Sander says: "Then let’s start now." Shortly after the start of the learning circle, Ms. Sander hears Luis mutter: "Oh no, I'm not going to get that right again...". Other voices murmur: "Once again, totally boring" or "What's the point of this?"

Ms. Sander even expected such feelings from the students because she knows that they are not yet aware of how meaningful the tasks are for them. Therefore, she has decided to keep all the students busy so that they do not get bored. So, she has set more tasks at each station than the students can complete in the given time. This should also help them to practice working under time pressure, which they will also have to do in tests.

Every 10 minutes, Ms. Sander claps her hands and calls out: "Please move on to the next station now." Ms. Sander observes how the students work attentively. The classroom is very quiet, and even during the station change, there is hardly any commotion. Accordingly, she repeatedly praises individual students by saying, for example, "Well done!", "Keep it up!" and "Great, it seems to come easily to you!" As Ms. Sander looks over Max’s shoulder, she instantly spots several mistakes in the comma placement task. Disappointedly she thinks: "Oh boy! Max will never get this…". As she tries to point out to Franziska some mistakes regarding her use of capital and small initial letters, Franziska answers in desperation: "Oh man, I am just not capable of doing this." As far as the other students are concerned, everything seems to be going well. Overall, Ms. Sander is more or less satisfied with how the station work is going. She believes that most of the students enjoy practicing. After all, they are practicing for the class test.

At the end of the lesson, she hears a conversation between two students as they leave: "Wow, there were so many tasks ... I'm glad it's break time now, because this afternoon we won't be able to do anything since we have the whole rest as homework!" "Yeah, and it wasn't even fun ... Shall we play soccer?"