Check for updates

Preservice teachers' evaluation of evidential support in causal arguments about educational topics

Andreas Lederer[1]*, Eva Thomm[1,2] and Johannes Bauer[1,2]

[1]Educational Research and Methodology, Faculty of Education, University of Erfurt, Erfurt, Germany,
[2]Institute for Planetary Health Behaviour, Erfurt, Germany

Many questions about educational topics—such as the effectiveness of teaching methods—are of causal nature. Yet, reasoning about causality is prone to widespread fallacies, such as mistaking correlation for causation. This study examined preservice teachers' ability to evaluate how various types of evidence provide adequate support for causal claims, using psychology students as a comparison group. The experiment followed a 2 × 3 mixed design with the within-participant factor evidence type (i.e., anecdotal, correlational, experimental) and the between-participants factor study field (i.e., teacher education, psychology). Participants ($N = 135$) sequentially read short texts on three different educational topics, each presenting a claim and associated evidence. For each topic, participants indicated their claim agreement, and evaluated the convincingness of the argument and the strength of the evidential support. Results from mixed ANOVAs displayed main effects for evidence type on the convincingness of the argument and strength of evidential support, but not on individual claim agreement. Participants found experimental evidence to be more convincing and to provide stronger support for causal claims compared to anecdotal evidence. This pattern occurred similarly for both student groups and remained stable when controlling for cognitive and motivational covariates. Overall, preservice teachers seem to possess a basic understanding of different kinds of evidence and their differential strength in supporting causal arguments. Teacher education may build upon this foundational knowledge to enhance future teachers' competencies in critically appraising evidence from educational research and relating it to school-related claims and issues.

KEYWORDS

argumentation, causality, evidence-informed reasoning, fallacies, preservice teachers, teacher education

1 Introduction

Many—if not most—questions about educational topics are inherently causal (e.g., Kvernbekk, 2016; Shavelson and Towne, 2002). This concerns not only research questions, such as the effectiveness of teaching methods or educational interventions, but also frequently encountered concerns among educational practitioners. For instance, teachers may wonder how to best explain a difficult topic to enhance students' comprehension, boost engagement and motivation or support a struggling child. Though other types of questions, such as diagnostic ones, are certainly important (Shavelson and Towne, 2002), education and teaching,

as goal-directed endeavours, naturally involve analyses of whether specific actions causally contribute to achieving desired outcomes.

Unfortunately, human reasoning about causality poses considerable challenges and is notoriously susceptible to biases (Bleske-Rechek et al., 2015; Cunningham, 2021; Hernán and Robbins, 2023). The common adage "correlation does not imply causation" underscores the highly prevalent fallacy of mistaking mere coincidence or correlation between events as evidence of a cause-and-effect relationship (Bleske-Rechek et al., 2015). Because of such biases, reasoning about what practices might be effective in a given context may pose some challenges for practitioners. For instance, teachers wishing to learn about the effects of utilising digital media in instruction on students' learning may encounter a plethora of sources that present diverse evidence: colleagues' experiences, media coverage, educational guidebooks and tutorials but also literature based on educational research. Although both research and professional experiences hold value for practitioners (Rousseau and Gunia, 2016), studies consistently indicate that teachers might lack fundamental research knowledge (Rochnia et al., 2023; Schmidt et al., 2023). This is not surprising, given that teacher education curricula typically offer less systematic methodological training compared with disciplines more oriented towards empirical science (e.g., psychology). In line with this, teachers often favour anecdotal evidence, such as personal or colleagues' experiences, over research-based sources when forming conclusions about effective practices (Fischer, 2021; van Schaik et al., 2018). However, because of these biases, anecdotal evidence must be viewed as the least conclusive for supporting generalised causal claims. Moreover, even within research, studies involving randomised experiments are typically deemed more causally informative compared with observational study designs, such as surveys (Hernán and Robbins, 2023; Shavelson and Towne, 2002). Therefore, relying solely on experiences or anecdotes, as well as studies with lower internal validity, could foster ill-advised practices or pedagogical misconceptions (Asberger et al., 2021; Menz et al., 2021; Michal et al., 2021).

The present study aimed to investigate preservice teachers' assessments of different types of evidence (i.e., anecdotal, correlational, experimental) in supporting causal claims on educational topics. The study thus contributes to the expanding body of literature on preservice teachers' evidence-informed reasoning and engagement with research-based knowledge in teacher education (Kollar et al., 2023). Surprisingly, although research has examined preservice teachers' reasoning abilities (e.g., Csanadi et al., 2021) or evidence evaluation (e.g., Reuter and Leuchter, 2023), studies have hardly focused on causal argumentation. This lack of research is noteworthy considering the pivotal role that causal issues play in educational practice (Kvernbekk, 2016).

1.1 Causal reasoning about educational topics

Reasoning about causal claims essentially involves constructing arguments (Hahn et al., 2017; Kuhn and Dean, 2004). Argumentation follows a basic structure that involves presenting an assertion (Claim, C) supported by pertinent evidence (E) that is connected by a logical-theoretical link (Warrant, W) (Toulmin, 2003). These elements enable an assessment of the claim's validity (Toulmin, 2003; Moshman and

Tarricone, 2016). Analysing whether a causal claim is justified often requires examining the pertinence and conclusiveness of the evidence. This is pivotal because, as discussed above, not all evidence can equally substantiate causal claims (Hernán and Robbins, 2023; Kvernbekk, 2016; Shavelson and Towne, 2002).

Although there are various ways to identify causal relationships (Cunningham, 2021; Hernán and Robbins, 2023; Pearl, 2009), anecdotal evidence (alone) will rarely provide sufficient support for a robust causal argument (Kuhn, 1991). For instance, although anecdotal evidence from teaching practice can be valuable in recognising important co-occurring events and inspiring hypotheses about putative reasons, it falls short in excluding alternative explanations (Bleske-Rechek et al., 2015; Kuhn, 1991). In contrast, typical causal claims in education, such as those regarding the effectiveness of instructional methods, require evidence from controlled experimental settings that have high internal validity (Shavelson and Towne, 2002). Randomised experiments are generally considered the simplest and gold standard method to establish causality within the methodological literature (Holland, 1986) and, thus, hold a high position in evidence hierarchies for evidence-based practice (Kvernbekk, 2016). Conversely, observational study designs typically have lower internal validity although, under specific circumstances, they may allow for the identification of causal effects (Cunningham, 2021; Hernán and Robbins, 2023). This is mainly because observational studies often lack adequate control over extraneous factors that may influence the outcome of interest. As a result, relying on observational evidence to support a causal claim frequently leads to flawed arguments (Kuhn, 1991; Kuhn and Dean, 2004).

In essence, evaluating causal arguments necessitates recognising the relative quality of available evidence supporting a claim and aligning the type of evidence with that claim (Kuhn, 1991). This requires to coordinate one's theoretical understanding of the claim with the quality of the evidence to reach a valid conclusion (Kuhn and Dean, 2004; Kuhn, 2012). However, research indicates that aligning claims with evidence is a challenging task that is prone to fallacies (Bromme and Goldman, 2014; Kuhn, 2012; Kuhn and Modrek, 2022). For instance, people often find anecdotal evidence to be as compelling as experimental evidence when substantiating a causal theory (Kuhn, 1991; Hoeken and Hustinx, 2009). This tendency seems particularly prevalent in emotionally engaging situations (Freling et al., 2020) when information is presented in narrative form (Kuhn, 1991), aligns with preexisting beliefs (Schmidt et al., 2022) and has high plausibility (Michal et al., 2021). Even when the flawed nature of evidence is explicitly highlighted, biased reasoning can occur (Braasch et al., 2014; Steffens et al., 2014), such as drawing causal conclusions from observational data (Bleske-Rechek et al., 2015).

1.2 (Preservice) teachers' abilities in claim−evidence coordination

Given the aforementioned widespread prevalence of causal fallacies in society (Bleske-Rechek et al., 2015; Kuhn, 2012; Seifert et al., 2022), it seems plausible that preservice teachers may also have difficulties in effectively coordinating claims with evidence. This expectation is also backed up by curricular analysis indicating that teacher education programmes typically do not offer the systematic

methodological training that would provide a thorough foundation in scientific (causal) reasoning (Darling-Hammond, 2017; Engelmann et al., 2022; Rochnia et al., 2023; Pieschl et al., 2021). Moreover, although specific research on preservice teachers' claim–evidence coordination is scarce, the broader literature on preservice and in-service teachers' engagement with research highlights substantial motivational and skill-related barriers (e.g., Ferguson et al., 2023; Kiemer and Kollar, 2021; see for review van Schaik et al., 2018). We will elaborate on the discussed issues before delving into additional individual factors that might affect claim–evidence coordination.[1]

1.2.1 Lack of methodological learning opportunities in teacher education

Despite the increasing recognition of teaching as a research-based profession and related developments in teacher education (Bauer and Prenzel, 2012; Darling-Hammond, 2017), study curricula commonly lack dedicated learning opportunities in research methods and statistics (Rochnia et al., 2023). Initial teacher education primarily aims to equip future teachers with a scientifically founded knowledge base essential for achieving high instructional quality and advancing student learning (Darling-Hammond, 2017; Rochnia et al., 2023). Although many countries, including Germany where the present study was conducted, emphasise fostering a scientific mindset, often through practitioner-research projects (Bock et al., 2024; Böttcher-Oschmann et al., 2021; Westbroek et al., 2022), comprehensive training in research methods akin to disciplines like psychology or sociology remains notably absent (Pieschl et al., 2021; Thomm et al., 2021b). Teacher education also seems to offer less methodological training compared with other profession-oriented study programmes, such as medicine (Rochnia et al., 2023). Empirical studies among preservice teachers and in-service teachers have frequently found that they tend to exhibit limited methodological knowledge and skills in scientific reasoning and argumentation (Groß Ophoff et al., 2017; Schmidt et al., 2023; Williams and Coles, 2007). These competencies are foundational for understanding and critically engaging with research (e.g., Joram et al., 2020; Niemi, 2008), including coordinating claims and evidence within (causal) argumentation.

1.2.2 Abilities for engagement with evidence and claim–evidence coordination

The expanding literature investigating teachers' evidence-informed reasoning and engagement with educational research (e.g., Kollar et al., 2023; Thomm et al., 2021c) documents barriers linked to limited abilities (e.g., Thomm et al., 2021b; Williams and Coles, 2007) and dysfunctional motivational orientations, attitudes, and beliefs (e.g., Bråten and Ferguson, 2015; Merk et al., 2017; Voss, 2022). Research reception requires sufficient skills in finding, reading, evaluating, and applying relevant research knowledge and evidence (Thomm et al., 2021c). Previous research suggests that preservice and in-service teachers often show only insufficient skills to draw on and

reason along with research findings or report low confidence in their abilities to do so (e.g., Duke and Ward, 2009; Ferguson et al., 2023; van Schaik et al., 2018; Wenglein, 2018). At the same time, teachers frequently tend to devaluate the relevance and applicability of educational research to inform their professional actions and decisions (e.g., Farley-Ripple et al., 2018; Thomm et al., 2021b; Voss, 2022). Hence, compared with students in other profession-oriented disciplines like medicine, preservice teachers seem to develop a research-oriented mindset to a lesser extent (Rochnia et al., 2023). Notably, preservice teachers exhibit a strong and persistent preference for anecdotal evidence sources, such as personal experiences or reports from colleagues (e.g., Ferguson et al., 2023; Kiemer and Kollar, 2021). As a result, active teachers rarely draw upon research-based knowledge to inform professional action and decision-making (van Schaik et al., 2018).

Concerning argumentation, preservice teachers often encounter difficulties in constructing evidence-based arguments unless they receive specific training (Iordanou and Constantinou, 2014; Uçar and Cevik, 2020). For instance, Wenglein et al. (2015) and Wenglein (2018) found that preservice teachers struggled to integrate scientific evidence they had been presented with to construct evidence-based arguments on educational topics, unless they received dedicated training. Instead, a substantial proportion of participants constructed weak arguments, relying only on anecdotal evidence or no evidence at all. Regarding evidence evaluation and coordination, some studies suggest that preservice teachers possess basic abilities to also differentiate different types of evidence (Reuter and Leuchter, 2023); however, this ability may primarily pertain to distinguishing between anecdotal and scientific evidence rather than discerning evidence derived from different scientific study types, such as observational versus experimental designs (see List et al., 2022). Although such research sheds light on the relevant abilities for causal argumentation, further investigation into how (preservice) teachers discern different evidence types and evaluate their respective strengths to support causal arguments is warranted.

1.2.3 Additionally relevant factors

Beyond the abilities described above, reasoning and arguing about educational issues can depend on perceptions of the specific topic at hand (Asberger et al., 2021). Specifically, teachers' prior knowledge and interest in a topic can substantially shape the way they interpret the available evidence and use it for argumentation (Schmidt et al., 2022; Yang et al., 2015). People frequently assess the plausibility of a claim or its consistency with evidence based on their preexisting beliefs about the topic (Abendroth and Richter, 2023; Futterleib et al., 2022; Michal et al., 2021; Thomm et al., 2021a; Wolfe et al., 2009).

Furthermore, research suggests that engaging in epistemic activities, such as evaluating evidence, can depend on personal epistemic orientations that define an individual's subjective understanding of what makes a valid argument and constitutes evidence (Fischer et al., 2014; Fives et al., 2017; Garrett and Weeks, 2017). For instance, individuals vary in their perceived need for supporting claims with valid evidence and in whether they consider intuition adequate support for establishing a claim's truth, as opposed to seeking factual evidence (Chinn et al., 2014; Garrett and Weeks, 2017).

Given the potential role of these factors in preservice teachers' abilities to differentiate between various types of evidence and

---

1   Because claim-evidence coordination primarily involves individual cognitive processes, we refrain from elaborating on contextual factors that might influence teachers' engagement with research, such as constraints in availability or time or dysfunctional social pressures (see, e.g., Gold et al., 2023; Greisel et al., 2023; Thomm et al., 2021b; van Schaik et al., 2018).

coordinate their judgement with causal claims, we have included them as additional (exploratory) covariates in our study.

2 The present study

To address the abovementioned research gaps, the present study investigated preservice teachers' ability to discern different types of evidence and its strength in supporting causal arguments about educational topics. The particular interest in substantiating causal arguments aligns with the causal nature of many—if not most—issues relevant to teachers' work and schooling, such as the effectiveness of teaching methods (e.g., Kvernbekk, 2016; Shavelson and Towne, 2002). Given that anecdotal or correlational evidence provides insufficient support for causal arguments but is frequently preferred by preservice teachers (e.g., Ferguson et al., 2023), gaining a deeper understanding of their ability to judge evidential support seems crucial (see List et al., 2022). For this purpose, we conducted a repeated measures experiment to scrutinise whether and how preservice teachers can discern the evidential support provided by various types of evidence (i.e., anecdotal, correlational, and experimental).

To better gauge preservice teachers' ability of judging evidential support, we compared them to psychology students as a benchmark. We chose the latter as a reference group, first, because both teacher education and psychology are pertinent to educational topics, such as issues of teaching and learning (cf. Asberger et al., 2020, 2021). This common ground aided a meaningful comparison. However, second, study programmes in teacher education and psychology are strikingly different regarding methodological learning opportunities. Psychology programmes contain comprehensive training in research methods and statistics, as required by established curricular standards [e.g., American Psychological Association (APA), 2023; Deutsche Gesellschaft für Psychologie (DGPs), 2014]. In contrast, as detailed above (see Section 1.2), systematic methodological training is rarely part of teacher education programmes. Moreover, psychology traditionally places a strong emphasis on the use of experimental methods for causal inference (e.g., Shadish et al., 2002). This specific methodological training can be expected to facilitate psychology students in recognising different types of evidence, such as correlational or experimental (Morling, 2014; Mueller and Coon, 2013; Seifert et al., 2022), and may foster skepticism towards unsupported beliefs and anecdotal evidence (Green and Hood, 2013; Leshowitz et al., 2002).

Before conducting the experiment, we preregistered the following hypotheses based on the theoretical reasoning and prior research outlined above.[2]

First, we expected that preservice teachers would agree more with a causal claim (H1a; *claim agreement*) and perceive it as more convincing (H1b; *convincingness*) when supported by anecdotal evidence rather than by correlational or experimental evidence, respectively. Additionally, we hypothesised that preservice teachers would attribute greater support strength to anecdotal evidence compared with correlational or experimental evidence, respectively

(H1c; *strength of evidential support*). Second, we expected that psychology students would agree more with a causal claim (H2a) and find it more convincing (H2b) when supported by experimental evidence than by correlational or anecdotal evidence, respectively. Furthermore, we assumed that psychology students would attribute higher strength of evidential support to experimental evidence compared with correlational or anecdotal evidence, respectively (H2c).

In an additional exploratory analysis, we controlled for potential effects of the factors previously discussed: (a) familiarity with research methods and statistics, (b) topic-related prior knowledge and interest and (c) faith in intuition and the need for evidence, which represent crucial aspects of epistemic orientations. Although not central to the internal validity of testing our primary hypotheses, we considered investigating these covariates as promising for gaining a more nuanced understanding of the personal factors contributing to claim–evidence coordination.[3]

3 Methods

3.1 Data and materials availability

Open data, as well as all materials, are available on the Open Science Framework (OSF) under https://osf.io/xw64c.

3.2 Design and participants

The current study followed a 2×3 mixed design with *evidence type* (anecdotal vs. correlational vs. experimental evidence) as a within-participant factor and *study field* (preservice teachers vs. psychology students) as a between-participant factor. The participants evaluated three causal claims about different educational topics: the testing effect (Rowland, 2014), the effectiveness of advance organisers in teaching (Stone, 1983) and the benefits of self-regulated learning (Dent and Koenka, 2016). Each claim was supported by one of the three evidence types, with the assignment of evidence type being balanced across subjects and presented in randomised order. For details, see Section 3.4.

Based on an *a priori* power analysis (Faul et al., 2007), we aimed at a sample size of $N = 142$ participants to have sufficient statistical power (95%) for detecting medium-sized effects ($f^2 = 0.25$) in a mixed ANOVA with a significance level of $\alpha = 0.05$. Undergraduate students were recruited online from mail distribution lists and university lecturers across multiple German universities. Participation was voluntary. The participants could either enter a lottery of vouchers or receive course credit as an incentive. We followed APA ethical

---

2　https://doi.org/10.17605/OSF.IO/R2TVW

3　For additional exploratory purposes unrelated to the primary research questions of the present study, we gathered supplementary data concerning the ascribed trustworthiness and expertise of the evidence sources. These variables were measured at the end of each respective measurement point and, therefore, could not have influenced the primary outcomes collected before. Because these variables are beyond the scope of this paper, we do not include them herein. Aside from this omission, we confirm the reporting of all experimental conditions and dependent variables.

standards, and the study was approved by the ethics committee of the University of Erfurt.

A total of $N = 230$ participants responded to the questionnaire and provided informed consent. In line with our preregistration, data of those participants who indicated they had not responded sincerely ($n = 3$), withdrew their consent at the end of the study ($n = 5$), were not enrolled in initial teacher training or psychology ($n = 10$) or missed responses on complete evidence conditions were excluded ($n = 77$). The final sample consisted of $N = 135$ university students (83.7% female, $M = 22.48$ years, $SD = 3.88$). Of this total, $n = 61$ were preservice teachers (78.7% female, $M = 23.70$ years, $SD = 3.63$), with 63.9% enrolled in a bachelor's degree programme ($M = 5.05$ semesters, $SD = 1.11$) and 32.8% in a master's degree programme ($M = 2.94$ semesters, $SD = 1.11$). Participants aimed to become teachers for either elementary school (63.9%), general and vocational secondary schools (19.6%), and special needs education (16.4%). The larger percentage of primary education is due to the specialization of the addressed universities in teacher education. Further, $n = 74$ participants studied psychology (87.8% female, $M = 21.47$ years, $SD = 3.80$). Most of them were enrolled in a bachelor's degree programme (93.2%, $M = 4.00$ semesters, $SD = 2.00$), while some completed a master's degree (6.8%, $M = 3.00$ semesters, $SD = 1.58$).

To ensure data quality, in a preliminary analysis, we checked for extreme outliers. As a criterion, we used z-values ($z < 3.29$; Field, 2012) for univariate outliers and Mahalanobis distance (MAH > 50; Tabachnick and Fidell, 2013) for multivariate outliers. This resulted in the identification of four outlying cases. Following our preregistration, we present the complete sample analysis. An additional analysis excluding outliers did not lead to different substantive conclusions compared with the complete sample. The results, excluding outliers, are available in Supplementary material S1.

3.3 Procedure

The study was conducted online. After introductions and providing informed consent, the participants were instructed that they were to read texts on the three educational topics mentioned above and answer related questions. For each of the three texts, the assessments followed the same sequence. First, the participants were told the topic and rated their prior knowledge and interest in it. Subsequently, they read the argument and assessed (a) their personal agreement with the claim (*claim agreement*), (b) how convincing they perceived the argument to be (argument *convincingness*), and (c) the strength of the evidential support provided in the text (strength of evidential *support*). Subsequently, they were asked for a brief written justification for their assessment of the evidential support. Having evaluated the arguments, we measured the participants' familiarity with research methods/statistics, captured their need for evidence and faith in intuition and asked for demographic information (i.e., study field, study degree, number of semesters studied, gender, and age).

After completing the study, the participants could withdraw their participation, responded to a seriousness check and received a debriefing entailing information about the study's goal, the experimental manipulation and additional scientific information about the presented topics.

3.4 Experimental manipulation

The participants read three short texts, each presenting an argument about the respective educational topic (see Table 1 for an example). We chose the following topics for their sound evidence base in educational-psychological research: the testing effect (e.g., Rowland, 2014), the effectiveness of advance organisers (e.g., Stone, 1983) and the benefits of self-regulated learning (e.g., Dent and Koenka, 2016).

TABLE 1 Example of argument text ("testing effect").

	Evidence types		
	Anecdotal	Correlational	Experimental
Introduction	Tests and learning assessments are an integral part of teaching and schooling. Tests are mainly used to assess learning performance. However, when used correctly, tests can also be used specifically to support learning after an initial period of teaching and learning. This is known as formative testing.		
Causal claim	*[Experience from school practice/Scientific studies]* show that learning processes are more effective if, after an initial learning phase, the content is actively recalled from memory through tests that accompany the learning process.		
Evidential support	*In an interview, a teacher describes her many years of experience with formative testing.* The teacher stated that learning assessments are easy to prepare and can be integrated flexibly and beneficially into the classroom. Regular or repeated use is also possible. The teacher reports that the learning assessment gives the students an assessment of their level of knowledge, reveals gaps and helps them remember the content through repetition.	*A recent study investigated this using a correlative study design.* Students from several schools answered a questionnaire about how regularly tests were used in class to accompany learning. In addition, their current level of learning was assessed using established achievement tests. The data from each school were then combined and analysed. The results show highly positive correlations between the regular use of tests that accompany learning and student achievement.	*A recent study investigated this using an experimental study design.* Students in two randomly assigned learning groups received identical instruction over several weeks. In addition, one group took regular tests to monitor their learning, while the other group repeated the material at their own pace over the same period. At the end of the experiment, the students' progress was assessed using established achievement tests. The group with regular tests performed significantly better than the repetition group.
Conclusion	Thus, the use of tests that accompany learning results in better learning outcomes.		

The example illustrates the manipulation of anecdotal, correlational, and experimental evidence. The original text was in German. Arguments were developed in German and pretested using cognitive interviews ($n = 10$) regarding clarity and comprehensibility.

All texts followed a parallel structure. Each began with a sentence introducing the topic. Next, the text presented the causal claim of interest. The phrasing of the claim already pointed to the evidential support that followed. Evidential support was systematically varied: it briefly described either *anecdotal evidence* of an experienced in-service teacher, the design and result of a correlational study (*scientific correlational evidence*) or the design and result of an experimental study (*scientific experimental evidence*). The combination of evidential support and topic was balanced, and the order of the presented evidence was randomised to prevent sequence effects. Each text led to the same causal conclusion about the presented learning or teaching method.

The texts with nine possible combinations of topic and evidence were of a similar length ($M = 147.33$ words, $SD = 2.87$, range = 143 words to 154 words) and difficulty (Flesh Reading Index: $M = 26.11$, $SD = 4.56$). Before the experiment, we conducted cognitive interviews with 10 preservice teachers to examine the comprehensibility of the text materials and modified them by following the participants' comments.

3.5 Measures

3.5.1 Dependent variables

All dependent variables (i.e., *claim agreement, argument convincingness, and strength of evidential support*) were measured using single items on a 7-point rating scale with higher numbers indicating higher prevalence. Item texts were as follows: "How much do you agree with the statement that [topic claim (e.g., learning processes are more effective when the content is actively recalled from memory after an initial learning phase through tests that accompany learning)]?", (1 = do not agree at all; 7 = fully agree) for *claim agreement*. "How convincing do you find the argument given in the text that [topic claim]?", (1 = not convincing at all; 7 = fully convincing) for *argument convincingness*. "In the text segment, the statement [Topic claim] is supported by [Evidence type]. How good does this evidence support the statement in your opinion?", (1 = does not provide support at all; 7 = provides very good support) for the strength of evidential support.

3.5.2 Covariates

For exploratory purposes (see Section 2), we collected data on the covariates listed below. Unless stated otherwise, all used a 7-point rating scale similar to the dependent variables. For item texts, please see Supplementary material S2.

Prior knowledge (i.e., "I have comprehensive prior knowledge of the topic") and *topic interest* (i.e., "I am very interested in this topic") were assessed by single items. *Familiarity with research methods/ statistics* was measured by an established scale asking the participants to rate their understanding of methodological concepts (e.g., "quasi-experiment", "correlation") on a 4-point scale ranging from 1 (= do not know the concept) to 4 (= understand the concept and could explain it to someone else) (7 items, α = 0.81; Mang et al., 2018; Thomm et al., 2021b). Finally, we collected data on two aspects of *epistemic orientations*, adopting scales from Garrett and Weeks (2017): the *need for evidence* (e.g., "Evidence is more important than whether something feels true", 4 items, α = 0.72) and *faith in intuition* (e.g., "I trust my gut to tell me what's true and what's not", 4 items, α = 0.77).

3.6 Analyses

To examine H1 to H3, we used mixed ANOVAs and follow-up *t*-tests applying Bonferroni correction. The significance level was set to α < 0.05. We judged the effect sizes according to Cohen's criteria (Cohen, 1988).

Prior to the analyses, we checked for potential topic differences. Preliminary ANOVA with follow-up *t*-tests did not reveal significant differences in the outcome measures across the three topics: *claim agreement, $F(2, 135) = 2.848$, $p = 0.060$; argument convincingness, $F(2, 135) = 2.683$, $p = 0.070$; and strength of evidential support, $F(2, 135) = 0.204$, $p = 0.815$*. Therefore, we considered them comparable and averaged across topics. Furthermore, we examined whether the data met the assumptions for conducting ANOVAs. Inspection of P–P plots, skewness and kurtosis pointed to a violation of the normal distribution assumption in all dependent variables. Following the recommendations of Tabachnick and Fidell (2013), we used square root transformations for moderate negative skewness to correct for these violations and reran all analyses. The results from these analyses led to identical conclusions, as with the untransformed data. Conforming with our preregistration, we will report the results from the analyses with untransformed data. The results for transformed data are available in Supplementary material S3.

In the exploratory analyses, we examined the potential effects of controlling for the mentioned covariates. To this end, we employed multilevel models (MLM) with repeated measures of the dependent variables at level one and individual-level covariates at level two. MLM were conducted following the procedures laid out by Field et al. (2012), using the *nlme* package (3.1–164) in R (4.2.3). We expanded the models stepwise[4]: Model 1 (M1) served as a baseline for further analyses and was set up analogously to the mixed ANOVA tests of H1 to H3. Hence, M1 can also be considered a robustness check for the ANOVA results (Field et al., 2012). Model 2 (M2) added the individual level covariates related to knowledge and topical content relation (i.e., prior knowledge, topic interest, and participants' familiarity with research methods/statistics). Model 3 (M3) added individual faith in intuition and need for evidence as further individual level predictors.

4 Results

4.1 Examining the effects of evidence type and field of study

Regarding the hypotheses on *claim agreement* (H1a/H2a), the results from the mixed ANOVA indicated no statistically significant effects [evidence type, $F(2,135) = 0.30$, $p = 0.739$, *part.* $\eta^2 = 0.00$; study field, $F(1,135) = 0.01$, $p = 0.938$, *part.* $\eta^2 = 0.00$; interaction evidence

---

4   Compared with the interaction or random-effects models, the fixed effects model consistently demonstrated superior fit across all dependent measures. M1 formula: $Y_{ij} = \beta0_{(Intercept)} + \beta1_{(Evidence)ij} + \beta2_{(Study Field)ij} + u0_{(Random Intercept)j} + e_{ij}$. M2 formula: $Y_{ij} = \beta0_{(Intercept)} + \beta1_{(Evidence)ij} + \beta2_{(Study Field)ij} + \beta3_{(Topic Prior Knowledge)ij} + \beta4_{(Topic Interest)ij} + \beta5_{(Knowledge Methods/Statistics)ij} + u0_{(Random Intercept)j} + e_{ij}$. M3 formula: $Y_{ji} = \beta0_{(Intercept)} + \beta1_{(Evidence)ij} + \beta2_{(Study Field)ij} + \beta3_{(Topic Prior Knowledge)ij} + \beta4_{(Topic Interest)ij} + \beta5_{(Knowledge Methods/Statistics)ij} + \beta6_{(Faith Intuition)ij} + \beta7_{(Need Evidence)ij} + u0_{(Random Intercept)j} + e_{ij}$.

type x study field, $F(2,135) = 1.86$, $p = 0.157$, *part.* $\eta^2 = 0.01$]. That is, in contrast to H1a and H2a, both preservice teachers and psychology students agreed equally high with the presented claim, regardless of the type of evidential support (see Figure 1).

Concerning *argument convincingness* (H1b/H2b), the mixed ANOVA revealed a significant main effect of evidence type, $F(1.89,135) = 4.32$, $p = 0.016$, *part.* $\eta^2 = 0.03$. In contrast, the main effect of study type, $F(1,135) = 0.06$, $p = 0.815$, *part.* $\eta^2 = 0.00$, and the interaction evidence type x study field, $F(1.89, 135) = 1.54$, $p = 0.218$, *part.* $\eta^2 = 0.01$, failed to reach significance (see Figure 2). Follow-up *t*-tests on the main effect of evidence type showed that, overall, the participants judged arguments supported by experimental evidence more convincing than those drawing on anecdotal evidence, $t(134) = 2.56$, $p = 0.011$, $d = 0.22$. However, experimental evidence was not perceived as more convincing than correlational evidence, $t(134) = 0.61$, $p = 0.541$, $d = 0.05$. Finally, the participants considered correlational evidence as more convincing than anecdotal, $t(134) = 2.54$, $p = 0.012$, $d = 0.22$. Regarding the interaction, there was a descriptive group difference in the expected direction, indicating that psychology students judged anecdotal evidence as less convincing than experimental evidence, but it was not statistically significant. In summary, these results provide support for H2b, even though H1b cannot be maintained.
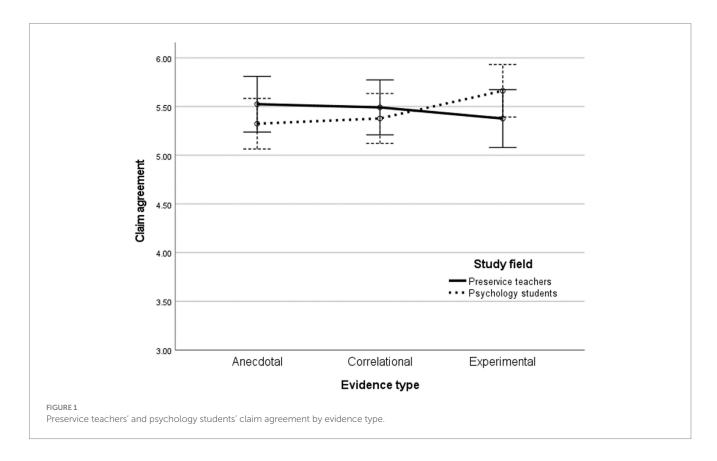
Regarding *support strength* (H3a/H3b), the mixed ANOVA showed a significant main effect of evidence type, $F(1.92, 135) = 18.48$, $p < 0.001$, *part.* $\eta^2 = 0.12$, but again, no further significant effects for study field, $F(1,135) = 0.10$, $p = 0.753$, *part.* $\eta^2 = 0.00$, or for their interaction, $F(1.92,135) = 1.22$, $p = 0.295$, *part.* $\eta^2 = 0.01$, were found (see Figure 3). Consistent with the findings on convincingness, follow-up *t*-tests for the main effect of evidence type indicated that participants assessed experimental evidence to provide stronger
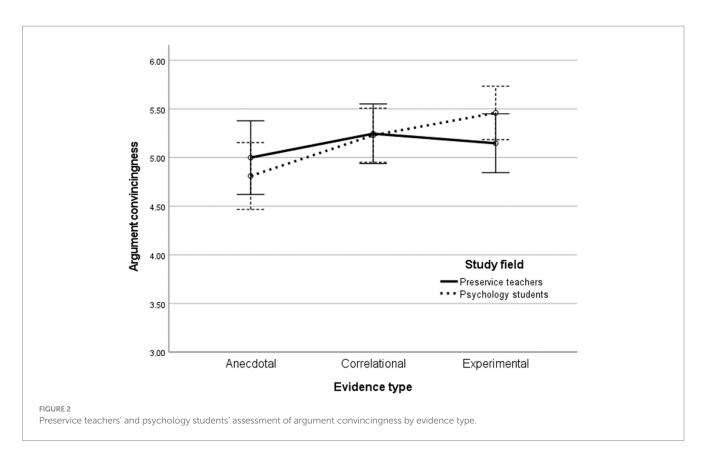
evidential support than anecdotal evidence, $t(134) = 6.00$, $p < 0.001$, $d = 0.52$, or correlational evidence, $t(133) = 3.36$, $p = 0.001$, $d = 0.29$. Moreover, they assessed correlational evidence to provide stronger support than anecdotal evidence, $t(133) = 3.12$, $p = 0.002$, $d = 0.27$. These results are not in line with H1c but corroborate H2c.
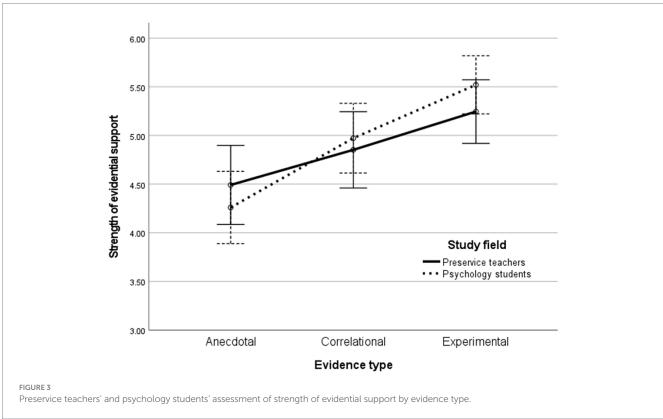
4.2 Exploring the individual differences in participants' assessments

Regarding the model fit of the MLM, the results showed that, for *claim agreement,* both M2 [$\chi^2(3) = 9.09$, $p = 0.028$] and M3 [$\chi^2(5) = 14.74$, $p = 0.012$] significantly improved the fit over the baseline M1, while differences between M2 and M3 were not statistically significant [$\chi^2(2) = 5.64$, $p = 0.060$]. Concerning *argument convincingness,* M1 and M2 had similar fit [$\chi^2(3) = 7.03$, $p = 0.071$]. However, M3 yielded significant improvements over both M1 [$\chi^2(5) = 17.79$, $p = 0.003$] and M2 [$\chi^2(2) = 10.77$, $p = 0.005$]. Finally, regarding the *strength of evidential support*, there was no significant difference between M1 and M2 [$\chi^2(3) = 2.75$, $p = 0.431$]. However, M3 provided a significantly better fit than both M1 [$\chi^2(5) = 22.36$, $p < 0.001$] and M2 [$\chi^2(2) = 19.60$, $p < 0.001$].

Inspecting the results of the MLM (Table 2), including the covariates in Models 2 and 3, did not lead to substantial changes regarding the pattern of the experimental treatment effects for any of the dependent variables. Notably, the estimated parameters for the type of evidence remained numerically almost identical when comparing M1 with M2, and M1 with M3, respectively. Moreover, the MLM results were consistent with the ones from the ANOVAs reported above regarding the significant effects of correlational

FIGURE 1
Preservice teachers' and psychology students' claim agreement by evidence type.

FIGURE 2
Preservice teachers' and psychology students' assessment of argument convincingness by evidence type.

FIGURE 3
Preservice teachers' and psychology students' assessment of strength of evidential support by evidence type.

evidence and experimental evidence on argument convincingness and the strength of the evidential support.

Furthermore, we observed several statistically significant effects of the covariates that, however, occurred mostly differently for the respective dependent variables. Of the knowledge and topic-related variables, prior knowledge had a small negative effect on argument convincingness, and topic interest had a small positive effect on claim agreement. Both effects were stable across M2 and M3. Familiarity

TABLE 2 Multilevel models of claim agreement, argument convincingness, and strength of evidential support.

	Claim agreement					Argument convincingness					Strength of evidential support				
	Coef	*Est*	*SE*	*T*	*p* value	*Coef*	*Est*	*SE*	*T*	*p* value	*Coef*	*Est*	*SE*	*T*	*p* value
Model 1															
Inter.	β_{00}	**5.43**	**0.21**	**25.62**	**<0.001**	β_{00}	**4.84**	**0.26**	**18.63**	**<0.001**	β_{00}	**4.28**	**0.29**	**14.53**	**<0.001**
ECor	β_{01}	0.01	0.13	0.11	0.912	β_{01}	**0.34**	**0.14**	**2.35**	**0.020**	β_{01}	**0.56**	**0.17**	**3.40**	**0.001**
EExp	β_{01}	0.12	0.13	0.88	0.377	β_{01}	**0.42**	**0.14**	**2.91**	**0.004**	β_{01}	**1.04**	**0.16**	**6.34**	**<0.001**
Stud.	β_{02}	−0.01	0.12	−0.08	0.938	β_{02}	0.04	0.15	0.23	0.815	β_{02}	0.05	0.17	0.29	0.770
Model 2															
Inter.	β_{00}	**5.61**	**0.23**	**24.57**	**<0.001**	β_{00}	4.86	0.28	17.16	<0.001	β_{00}	4.08	0.32	12.69	<0.001
ECor	β_{01}	0.01	0.13	0.08	0.932	β_{01}	**0.32**	**0.15**	**2.17**	**0.031**	β_{01}	**0.57**	**0.17**	**3.39**	**0.001**
EExp	β_{01}	0.09	0.14	0.68	0.500	β_{01}	**0.43**	**0.15**	**2.87**	**0.004**	β_{01}	**1.03**	**0.17**	**6.03**	**<0.001**
Stud.	β_{02}	−0.12	0.13	−0.89	0.375	β_{02}	0.03	0.17	0.17	0.868	β_{02}	0.18	0.19	0.94	0.351
PriK.	β_{03}	−0.07	0.04	−1.72	0.087	β_{03}	**−0.11**	**0.04**	**−2.53**	**0.012**	β_{03}	−0.02	0.05	−0.42	0.678
Int.	β_{04}	**0.10**	**0.05**	**2.09**	**0.038**	β_{04}	0.09	0.05	1.73	0.086	β_{04}	0.04	0.06	0.66	0.507
Stat.	β_{05}	**0.16**	**0.07**	**2.12**	**0.036**	β_{05}	−0.00	0.09	−0.05	0.958	β_{05}	−0.16	0.11	−1.49	0.140
Model 3															
Inter.	β_{00}	**5.51**	**0.23**	**24.11**	**<0.001**	β_{00}	**4.70**	**0.28**	**16.87**	**<0.001**	β_{00}	**3.86**	**0.31**	**12.60**	**<0.001**
ECor	β_{01}	0.01	0.14	0.09	0.929	β_{01}	**0.32**	**0.15**	**2.16**	**0.032**	β_{01}	**0.56**	**0.17**	**3.35**	**0.001**
EExp	β_{01}	0.09	0.14	0.68	0.496	β_{01}	**0.43**	**0.15**	**2.88**	**0.004**	β_{01}	**1.03**	**0.17**	**6.03**	**<0.001**
Stud.	β_{02}	−0.06	0.13	−0.42	0.674	β_{02}	0.13	0.17	0.77	0.443	β_{02}	0.32	0.18	1.74	0.085
PriK.	β_{03}	−0.06	0.04	−1.60	0.111	β_{03}	**−0.11**	**0.04**	**−2.42**	**0.016**	β_{03}	−0.01	0.05	−0.24	0.814
Int.	β_{04}	**0.09**	**0.05**	**1.98**	**0.049**	β_{04}	0.08	0.05	1.59	0.111	β_{04}	0.03	0.06	0.51	0.608
Stat.	β_{05}	0.13	0.08	1.53	0.128	β_{05}	−0.03	0.10	−0.29	0.776	β_{05}	−0.18	0.11	−1.60	0.112
FI	β_{06}	**0.12**	**0.05**	**2.31**	**0.023**	β_{06}	**0.22**	**0.07**	**3.31**	**0.001**	β_{06}	**0.33**	**0.07**	**4.52**	**<0.001**
NFE	β_{07}	0.04	0.06	0.72	0.468	β_{07}	0.02	0.07	0.22	0.827	β_{07}	0.01	0.08	0.13	0.899

N = 135. ECor = "evidence correlational," EExp = "evidence experimental," Stud = studyfield, PriK = "prior knowledge," Int = "topic interest," Stat = "familiarity with research methods/statistics," FI = "faith in intuition," NFE = "need for evidence," Anecdotal evidence = reference condition. Bold face indicates statistical significance.

with research methods and statistics had a small positive effect on claim agreement; however, this was only the case for Model 2. Of the epistemic orientations added in M3, faith in intuition had small to medium positive effects on all dependent variables, whereas the need for evidence was statistically non-significant throughout.

5 Discussion

Experimental evidence, where available, is the most appropriate support for substantiating causal claims (Holland, 1986). This is also true for causal questions frequently encountered in educational topics (Kvernbekk, 2016; Shavelson and Towne, 2002). Unfortunately, general cognitive biases such as mistaking coincidence and correlation for causation (Bleske-Rechek et al., 2015; Seifert et al., 2022) alongside strong preferences for anecdotal evidence (e.g., Bråten and Ferguson, 2015; Merk et al., 2017; van Schaik et al., 2018) can lead (preservice) teachers to adopt unreliable knowledge and one-sided advice, which may subsequently influence their professional judgements, decisions and actions. Therefore, promoting evidence-informed practice in the teaching profession necessitates that future teachers be able to discriminate the quality of evidence for appropriately supporting claims (Reuter and Leuchter, 2023; Wenglein et al., 2015). The present

study expands prior research by investigating this basic ability in the context of evaluating causal arguments. Specifically, we examined preservice teachers' evaluations of anecdotal, correlational and experimental evidence in supporting causal claims about educational topics and compared their evaluations with those of psychology students as a benchmark group.

Our study revealed a notable and somewhat surprising pattern of results. Contrary to our expectations, we found no significant differences in the participants' evaluations based on their study field. Despite indications from teacher education curricula and prior evidence suggesting that preservice teachers are less trained in the evaluation of evidence, they demonstrated comparable abilities to judge causal support as psychology students. Both groups could discriminate between the quality of the investigated evidence types in terms of argument convincingness and strength of support. Specifically, the participants perceived experimental evidence as more convincing than anecdotal evidence and as providing the strongest support for causal claims. However, these correct assessments did not translate into their personal claim agreement, which remained unaffected by the type of evidence they encountered. Notably, this pattern of results persisted even when controlling for relevant covariates, such as topic interest, prior knowledge, and epistemic orientations. These findings contrast with previous studies that have

raised concerns about the skills of in-service and preservice teachers to engage adequately with research evidence and argumentation (Lytzerinou and Iordanou, 2020; Rochnia et al., 2023; van Schaik et al., 2018; Zimmermann and Mayweg-Paus, 2021).

Regarding claim agreement (H1a/H2a), we expected preservice teachers to indicate stronger agreement with arguments entailing anecdotal rather than correlational or experimental evidence and assumed the reverse response of psychology students. However, there was neither an effect of evidence type nor of study field on claim agreement. Although these findings go against our assumptions, they appear reasonable in hindsight. The research on inert knowledge has shown that people often fail to utilise their existing knowledge in relevant task situations (Cavagnetto and Kurtz, 2016; Renkl et al., 1996). Accordingly, it is possible that the participants acknowledged differences in evidential support but did not draw appropriate conclusions to determine their personal claim agreement. An alternative explanation could be that, even though it seems normatively reasonable to base one's judgement on the evidence type, people may form their judgement based not only on this single factor. Previous research has observed similar discrepancies between what people may personally think about a scientific claim and what they consider scientifically more adequate or credible information (e.g., Scharrer et al., 2017; Thomm and Bromme, 2012, 2016). For example, Thomm and Bromme (2012) found that presenting information about scientific topics in a scientific text style (e.g., including citations and method information) compared with a factual one enhanced perceived credibility but had no effect on claim agreement. Hence, additional factors, such as prior beliefs and plausibility assessments, may affect claim agreement as well (Barzilai et al., 2020; Futterleib et al., 2022; Richter and Maier, 2017). Indeed, the results from our exploratory analysis indicated that the participants' topic interest and faith in intuition played a role in their claim agreement. This may suggest that participants weighed the quality of the encountered evidence with their personal plausibility judgements. These interpretations should be treated with due caution, however, because the mentioned covariate effects were exploratory and of a small size. This notwithstanding, future research should investigate the interplay of such factors more closely.

For the convincingness of the arguments (H1b/H2b), we expected to find higher assessments of anecdotal evidence by preservice teachers and, in contrast, higher assessments of experimental evidence by psychology students when compared with the other types of evidence. Although the results regarding psychology students were widely in line with these assumptions, both groups found arguments entailing anecdotal evidence to be less convincing than those entailing either correlational or experimental evidence. However, the participants did not differentiate between these two types of scientific evidence. Thus, the contrast between arguments entailing anecdotal and scientific evidence might have been more salient than the additional, more nuanced difference between scientific correlational and experimental evidence. This pattern of results ties in with List's (2024) finding that university students failed to discern the quality of correlational and causal evidence about a scientific claim, whereas they judged both to be of a higher quality than mere anecdotal evidence (see also List et al., 2022). This categorical difference might be easier to recognise, even for individuals with limited methodological knowledge.

For the strength of evidential support (H1c/H2c), we anticipated a similar pattern of assessments as observed for the convincingness of the argument. However, both preservice teachers and psychology students judged experimental evidence to be the strongest support for

a causal claim. Together with the findings on convincingness, this suggests that the participants were able to distinguish not only between anecdotal and scientific evidence but also to make the more subtle distinction between the two types of scientific evidence. This may seem surprising given the frequently raised doubts about such abilities, even among university students (e.g., List, 2024). However, this apparent discrepancy may not be as pronounced as it seems at first glance. In the present study, the participants read simple, well-structured arguments designed to capture a fundamental understanding and coordination of evidence types for supporting causal claims. This setting might have facilitated discerning different types of evidence, and participants might have struggled with more complex or controversial arguments (see List, 2024; Menz et al., 2020; Münchow et al., 2023). Moreover, Reuter and Leuchter (2023) caution that, even if preservice teachers can recognise differences between more and less robust research evidence, it remains unclear whether they have a consistent understanding of the features that constitute evidence strength. That is, the existing studies offer no insights into which characteristics of the presented evidence participants were focussing on to form their judgements. Ideally, these judgements would be based on their knowledge of methodological principles and an understanding of why scientific evidence is preferable to anecdotal evidence, as well as why experimental evidence outweighs correlational evidence in supporting a causal claim. However, participants might also have referred to more superficial characteristics, such as the appearance of "scientificness" (Thomm and Bromme, 2012) in some texts, deeming them more trustworthy. Future research might delve deeper into whether participants indeed possess an adequate understanding of the principles that render different types of evidence more or less conclusive for causal hypotheses. So far, our findings demonstrate that preservice teachers possess at least a basic understanding of different kinds of evidence and their differential strength in supporting causal arguments.

Thus, taken together, our findings speak against simplistic and mainly deficit-oriented assumptions about preservice teachers' competences in engaging with research. Although the skills investigated in the present study are certainly basic, they also address the very core of understanding research and using claim–evidence coordination in argumentation. This is remarkable given the frequent lack of methodological training in teacher education. However, the less clear distinction between the two investigated types of scientific evidence might indicate that both preservice teachers and psychology students still face challenges in understanding the crucial distinction in correlational and experimental research. As a result, there is a need to support the development of these skills to prevent correlation–causation fallacies when reasoning about causal questions in education. More generally, it seems valuable to support preservice teachers in better understanding the nature of diverse types of evidence and their qualities regarding claims about educational topics.

Limitations. Several limitations of the present study deserve attention. First, as mentioned, we presented participants with clearly structured text materials and used topics for which there is a sound research base. This allowed us to capture basic skills of claim–evidence coordination in a controlled way, without the danger of introducing extraneous variance that might have resulted from differential understanding and engagement with longer and more difficult texts. More authentic text materials that more closely resemble real research would certainly need to be more complex. For example, future studies

could present participants with multiple documents that provide different types of evidence or contain inconsistent results (see Thomm et al., 2021c). Additionally, features of the addressed educational topic—for instance, to what degree it is politically controversial and emotionally laden—might play a role in what type of evidence people perceive as compelling (Aguilar et al., 2019; Darner, 2019; Sinatra et al., 2014). Hence, follow-up studies could systematically vary such topic-related characteristics.

Second, the within-participant design of the study offers several advantages, but it also introduces certain limitations. One notable advantage lies in the sequential presentation of texts featuring all three types of evidence, which enhances their contrast (cf. Birnbaum, 1999) and, thus, augments the systematic experimental variance. However, given this design, it is crucial to interpret participants' differential judgements of evidential support as relative comparisons across texts rather than absolute evaluations. Therefore, our findings may be more applicable to situations where preservice teachers evaluate different evidence sources against each other as opposed to assessing a single piece of evidence. Nonetheless, encountering multiple information sources on a given topic is a realistic scenario, as noted earlier (Bromme and Goldman, 2014; Sinatra and Lombardi, 2020; Thomm et al., 2021c).

Third, although within-participant designs are advantageous for controlling individual-level factors, potential order effects pose a threat to internal validity. Even though the fully balanced sequence of conditions effectively eliminates position effects, controlling differential carryover effects in within-participant designs is challenging. These effects occur when the (potential) carryover effect of treatment condition A (e.g., *anecdotal evidence*) on condition B (e.g., *experimental evidence*) differs from the carryover effect of condition B on condition A (Maxwell et al., 2018). Although, in our view, there is no substantive reason to consider differential carryover as particularly likely for our three types of evidence, its potential occurrence is a concern, especially in a within-design with a small washout period between treatments, as in the present study. Evaluating the stability and strength of our results would require reconceptualising the study by employing a between-participants design.

Finally, it is an open issue to what degree our findings generalize to in-service teachers. Teachers with longer work tenure might differ in evidence evaluation based on their professional experience and larger distance to academia as compared to preservice teachers (Hillmayr et al., 2024; van Schaik et al., 2018). Moreover, in future research on claim-evidence coordination it could be interesting to include comparison groups from other fields that are unrelated to educational issues (e.g., STEM disciplines, economy; cf. Asberger et al., 2020, 2021). Doing so would aid our understanding what role of prior knowledge of the topic domain plays in causal reasoning about educational issues.

Notwithstanding these limitations, we believe that our findings provide important insights into the ability of preservice teachers to evaluate different types of evidence that they may encounter when faced with causal questions in education, such as the effectiveness of instructional methods. Building on the foundational skills that we observed would require additional learning opportunities in teacher education that promote their understanding of different qualities of evidence, also in relation to the claim it shall support. This is important because, like many other people, they are susceptible to the fallacy of misinterpreting coincidence and correlation as sufficient support for causal claims.

Author's note

This study is preregistered with OSF preregistrations: doi: 10.17605/OSF.IO/R2TVW.

Data availability statement

The datasets presented in this study can be found online at https://osf.io/xw64c.

Ethics statement

The studies involving humans were approved by Ethics Committee of the University of Erfurt. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

AL: Writing – original draft, Writing – review & editing. ET: Writing – review & editing. JB: Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1379222/full#supplementary-material

References

Abendroth, J., and Richter, T. (2023). Reading perspectives moderate text-belief consistency effects in eye movements and comprehension. *Discourse Process.* 60, 119–140. doi: 10.1080/0163853X.2023.2172300

Aguilar, S. J., Polikoff, M. S., and Sinatra, G. M. (2019). Refutation texts: a new approach to changing public misconceptions about education policy. *Educ. Res.* 48, 263–272. doi: 10.3102/0013189X1984

American Psychological Association (APA). (2023). Guidelines for the undergraduate psychology major. Empowering People to Make a Difference in Their Lives and Communities Version 3.0. Available at: https://www.apa.org/about/policy/undergraduate-psychology-major.pdf (Accessed April 12, 2024).

Asberger, J., Thomm, E., and Bauer, J. (2020). Empirische Arbeit: Zur Erfassung fragwürdiger Überzeugungen zu Bildungsthemen: Entwicklung und erste Überprüfung des Questionable Beliefs in Education-Inventars (QUEBEC). *Psychol. Erzieh. Unterr.* 67, 178–193. doi: 10.2378/peu2019.art25d

Asberger, J., Thomm, E., and Bauer, J. (2021). On predictors of misconceptions about educational topics: a case of topic specificity. *PLoS One* 16:e0259878. doi: 10.1371/journal.pone.0259878

Barzilai, S., Thomm, E., and Shlomi-Elooz, T. (2020). Dealing with disagreement: the roles of topic familiarity and disagreement explanation in evaluation of conflicting expert claims and sources. *Learn. Instr.* 69:101367. doi: 10.1016/j.learninstruc.2020.101367

Bauer, J., and Prenzel, M. (2012). European teacher training reforms. *Science* 336, 1642–1643. doi: 10.1126/science.1218387

Birnbaum, M. H. (1999). How to show that 9>221: collect judgments in a between-subjects design. *Psychol. Methods* 4, 243–249. doi: 10.1037/1082-989X.4.3.243

Bleske-Rechek, A., Morrison, K. M., and Heidtke, L. D. (2015). Causal inference from descriptions of experimental and non-experimental research: public understanding of correlation-versus-causation. *J. Gen. Psychol.* 142, 48–70. doi: 10.1080/00221309.2014.977216

Bock, T., Thomm, E., Bauer, J., and Gold, B. (2024). Fostering student teachers' research-based knowledge of effective feedback. *Eur. J. Teach. Educ.* 47, 389–407. doi: 10.1080/02619768.2024.2338841

Böttcher-Oschmann, F., Groß, O. J., and Thiel, F. (2021). Preparing teacher training students for evidence-based practice promoting students' research competencies in research-learning projects. *Front. Educ.* 6:642107. doi: 10.3389/feduc.2021.642107

Braasch, J. L., Bråten, I., Britt, M. A., Steffens, B., and Strømsø, H. I. (2014). "Sensitivity to inaccurate argumentation in health news articles: potential contributions of readers' topic and epistemic beliefs" in Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences. eds. D. Rapp and J. Braasch (Cambridge, MA: MIT Press), 117–137.

Bråten, I., and Ferguson, L. E. (2015). Beliefs about sources of knowledge predict motivation for learning in teacher education. *Teach. Teach. Educ.* 50, 13–23. doi: 10.1016/j.tate.2015.04.003

Bromme, R., and Goldman, S. R. (2014). The public's bounded understanding of science. *Educ. Psychol.* 49, 59–69. doi: 10.1080/00461520.2014.921572

Cavagnetto, A. R., and Kurtz, K. J. (2016). Promoting students' attention to argumentative reasoning patterns. *Sci. Educ.* 100, 625–644. doi: 10.1002/sce.21220

Chinn, C. A., Rinehart, R. W., and Buckland, L. A. (2014). "Epistemic cognition and evaluating information: applying the AIR model of epistemic cognition" in Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences. eds. D. Rapp and J. Braasch (Cambridge, MA: MIT Press), 425–453.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *2nd* Edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

Csanadi, A., Kollar, I., and Fischer, F. (2021). Pre-service teachers' evidence-based reasoning during pedagogical problem-solving: better together? *Eur. J. Psychol. Educ.* 36, 147–168. doi: 10.1007/s10212-020-00467-4

Cunningham, S. (2021). Causal inference: The mixtape. New Haven and London: Yale University Press.

Darling-Hammond, L. (2017). Teacher education around the world: what can we learn from international practice? *Eur. J. Teach. Educ.* 40, 291–309. doi: 10.1080/02619768.2017.1315399

Darner, R. (2019). How can educators confront science denial? *Educ. Res.* 48, 229–238. doi: 10.3102/0013189X19849415

Dent, A. L., and Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: a meta-analysis. *Educ. Psychol. Rev.* 28, 425–474. doi: 10.1007/s10648-015-9320-8

Deutsche Gesellschaft für Psychologie (DGPs). (2014). Empfehlungen des DGPs-Vorstands zu Bachelor-und Masterstudiengängen in Psychologie. Available at: https://www.dgps.de/fileadmin/user_upload/PDF/Empfehlungen/Empfehlungen_des_Vorstands_Bachelor_und_Master_15_12_14.pdf (Accessed April 19, 2024).

Duke, T. S., and Ward, J. D. (2009). Preparing information literate teachers: a meta synthesis. *Libr. Inf. Sci. Res.* 31, 247–256. doi: 10.1016/j.lisr.2009.04.003

Engelmann, K., Hetmanek, A., Neuhaus, B. J., and Fischer, F. (2022). Testing an intervention of different learning activities to support students' critical appraisal of scientific literature. *Front. Educ.* 7:977788. doi: 10.3389/feduc.2022.977788

Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., and McDonough, K. (2018). Rethinking connections between research and practice in education: a conceptual framework. *Educ. Res.* 47, 235–245. doi: 10.3102/0013189X18761042

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Ferguson, L. E., Bråten, I., Skibsted Jensen, M., and Andreassen, U. R. (2023). A longitudinal mixed methods study of Norwegian preservice teachers' beliefs about sources of teaching knowledge and motivation to learn from theory and practice. *J. Teach. Educ.* 74, 55–68. doi: 10.1177/00224871221105813

Field, A., Miles, J., and Field, Z. (2012). Discovering statistics using R. London: Sage Publications.

Fischer, F. (2021). Some reasons why evidence from educational research is not particularly popular among (pre-service) teachers: a discussion. *German J. Educ. Psychol.* 35, 209–214. doi: 10.1024/1010-0652/a000311

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. *Frontline Learn. Res.* 2, 28–45. doi: 10.14786/flr.v2i2.96

Fives, H., Barnes, N., Buehl, M. M., Mascadri, J., and Ziegler, N. (2017). Teachers' epistemic cognition in classroom assessment. *Educ. Psychol.* 52, 270–283. doi: 10.1080/00461520.2017.1323218

Freling, T. H., Yang, Z., Saini, R., Itani, O. S., and Abualsamh, R. R. (2020). When poignant stories outweigh cold hard facts: a meta-analysis of the anecdotal bias. *Organ. Behav. Hum. Decis. Process.* 160, 51–67. doi: 10.1016/j.obhdp.2020.01.006

Futterleib, H., Thomm, E., and Bauer, J. (2022). The scientific impotence excuse in education – disentangling potency and pertinence assessments of educational research. *Front. Educ.* 7:1006766. doi: 10.3389/feduc.2022.1006766

Garrett, R. K., and Weeks, B. E. (2017). Epistemic beliefs' role in promoting misperceptions and conspiracist ideation. *PLoS One* 12:e0184733. doi: 10.1371/journal.pone.0184733

Gold, B., Thomm, E., and Bauer, J. (2023). Using the theory of planned behaviour to predict pre-service teachers' preferences for scientific sources. *Br. J. Educ. Psychol.* 94, 216–230. doi: 10.1111/bjep.12643

Green, H. J., and Hood, M. (2013). Significance of epistemological beliefs for teaching and learning psychology: a review. *Psychol. Learn. Teach.* 12, 168–178. doi: 10.2304/plat.2013.12.2.168

Greisel, M., Wekerle, C., Wilkes, T., Stark, R., and Kollar, I. (2023). Pre-service teachers' evidence-informed reasoning: do attitudes, subjective norms, and self-efficacy facilitate the use of scientific theories to analyze teaching problems? *Psychol. Learn. Teach.* 22, 20–38. doi: 10.1177/14757257221113942

Groß Ophoff, J., Wolf, R., Schladitz, S., and Wirtz, M. (2017). Assessment of educational research literacy in higher education: construct validation of the factorial structure of an assessment instrument comparing different treatments of omitted responses. *J. Educ. Res. Online* 9, 37–68. doi: 10.25656/01:14896

Hahn, U., Bluhm, R., and Zenker, F. (2017). "Causal argument" in The Oxford handbook of causal reasoning. ed. M. R. Waldmann (New York: Oxford University Press), 475–493.

Hernán, M. A., and Robbins, J. M. (2023). Causal inference: What if. Boca Raton: CRC Press.

Hillmayr, D., Reinhold, F., Holzberger, D., and Reiss, K. (2024). STEM teachers' beliefs about the relevance and use of evidence-based information in practice: a case study using thematic analysis. *Front. Educ.* 8:1261086. doi: 10.3389/feduc.2023.1261086

Hoeken, H., and Hustinx, L. (2009). When is statistical evidence superior to anecdotal evidence in supporting probability claims? The role of argument type. *Hum. Commun. Res.* 35, 491–510. doi: 10.1111/j.1468-2958.2009.01360.x

Holland, P. W. (1986). Statistics and causal inference. *J. Am. Stat. Assoc.* 81, 945–960. doi: 10.1080/01621459.1986.10478354

Iordanou, K., and Constantinou, C. P. (2014). Developing pre-service teachers' evidence-based argumentation skills on socio-scientific issues. *Learn. Instr.* 34, 42–57. doi: 10.1016/j.learninstruc.2014.07.004

Joram, E., Gabriele, A. J., and Walton, K. (2020). What influences teachers' "buy-in" of research? Teachers' beliefs about the applicability of educational research to their practice. *Teach. Teach. Educ.* 88:102980. doi: 10.1016/j.tate.2019.102980

Kiemer, K., and Kollar, I. (2021). Source selection and source use as a basis for evidence-informed teaching. *German J. Educ. Psychol.* 35, 127–141. doi: 10.1024/1010-0652/a000302

Kollar, I., Greisel, M., Krause-Wichmann, T., and Stark, R. (2023). Editoral: evidence-informed reasoning of pre-and in-service teachers. *Front. Educ.* 8:1188022. doi: 10.3389/feduc.2023.1188022

Kuhn, D. (1991). The skills of argument. Cambridge: Cambridge University Press.

Kuhn, D. (2012). The development of causal reasoning. *Wiley Interdiscip. Rev. Cogn. Sci.* 3, 327–335. doi: 10.1002/wcs.1160

Kuhn, D., and Dean, D. Jr. (2004). Connecting scientific reasoning and causal inference. *J. Cogn. Dev.* 5, 261–288. doi: 10.1207/s15327647jcd0502_5

Kuhn, D., and Modrek, A. S. (2022). Choose your evidence: scientific thinking where it may most count. *Sci. Educ.* 31, 21–31. doi: 10.1007/s11191-021-00209-y

Kvernbekk, T. (2016). Evidence-based practice in education: Functions of evidence and causal presuppositions. London and New York: Routledge.

Leshowitz, B., DiCerbo, K. E., and Okun, M. A. (2002). Effects of instruction in methodological reasoning on information evaluation. *Teach. Psychol.* 29, 5–10. doi: 10.1207/S15328023TOP2901_02

List, A. (2024). The limits of reasoning: students' evaluations of anecdotal, descriptive, correlational, and causal evidence. *J. Exp. Educ.* 92, 1–31. doi: 10.1080/00220973.2023.2174487

List, A., Du, H., and Lyu, B. (2022). Examining undergraduates' text-based evidence identification, evaluation, and use. *Read. Writ.* 35, 1059–1089. doi: 10.1007/s11145-021-10219-5

Lytzerinou, E., and Iordanou, K. (2020). Teachers' ability to construct arguments, but not their perceived self-efficacy of teaching, predicts their ability to evaluate arguments. *Int. J. Sci. Educ.* 42, 617–634. doi: 10.1080/09500693.2020.1722864

Mang, J., Ustjanzew, N., Schiepe-Tiska, A., Prenzel, M., Sälzer, C., Müller, K., et al. (2018). Dokumentation PISA 2012 Skalenhandbuch der Erhebungsinstrumente [PISA 2012 documentation of measurement instruments]. Münster: Waxmann.

Maxwell, S. E., Delaney, H. D., and Kelley, K. (2018). Designing experiments and analyzing data. New York: Routledge.

Menz, C., Spinath, B., and Seifried, E. (2020). Misconceptions die hard: prevalence and reduction of wrong beliefs in topics from educational psychology among preservice teachers. *Eur. J. Psychol. Educ.* 36, 477–494. doi: 10.1007/s10212-020-00474-5

Menz, C., Spinath, B., and Seifried, E. (2021). Where do pre-service teachers' educational psychological misconceptions come from? *German J. Educ. Psychol.* 35, 143–156. doi: 10.1024/1010-0652/a000299

Merk, S., Rosman, T., Rueß, J., Syring, M., and Schneider, J. (2017). Pre-service teachers' perceived value of general pedagogical knowledge for practice: relations with epistemic beliefs and source beliefs. *PLoS One* 12:e0184971. doi: 10.1371/journal.pone.0184971

Michal, A. L., Zhong, Y., and Shah, P. (2021). When and why do people act on flawed science? Effects of anecdotes and prior beliefs on evidence-based decision-making. *Cogn. Res. Princ. Implic.* 6:28. doi: 10.1186/s41235-021-00293-2

Morling, B. (2014). Research methods in psychology: Evaluating a world of information. New York: W.W. Norton and Company.

Moshman, D., and Tarricone, P. (2016). "Logical and causal reasoning" in Handbook of epistemic cognition. eds. J. Greene, W. Sandoval and I. Bråten (New York: Routledge), 54–67.

Mueller, J. F., and Coon, H. M. (2013). Undergraduates' ability to recognize correlational and causal language before and after explicit instruction. *Teach. Psychol.* 40, 288–293. doi: 10.1177/0098628313501038

Münchow, H., Tiffin-Richards, S. P., Fleischmann, L., Pieschl, S., and Richter, T. (2023). Promoting students' argument comprehension and evaluation skills: implementation of two training interventions in higher education. *Z. Erzieh.* 26, 703–725. doi: 10.1007/s11618-023-01147-x

Niemi, H. (2008). Research-based teacher education for teachers' lifelong learning. *Lifelong Learn. Europe* 13, 61–69.

Pearl, J. (2009). Causality: models, reasoning, and inference. Cambridge: Cambridge University Press.

Pieschl, S., Budd, J., Thomm, E., and Archer, J. (2021). Effects of raising student teachers' metacognitive awareness of their educational psychological misconceptions on their misconception correction. *Psychol. Learn. Teach.* 20, 214–235. doi: 10.1177/1475725721996223

Renkl, A., Mandl, H., and Gruber, H. (1996). Inert knowledge: analyses and remedies. *Educ. Psychol.* 31, 115–121. doi: 10.1207/s15326985ep3102_3

Reuter, T., and Leuchter, M. (2023). Pre-service teachers' latent profile transitions in the evaluation of evidence. *Teach. Teach. Educ.* 132:104248. doi: 10.1016/j.tate.2023.104248

Richter, T., and Maier, J. (2017). Comprehension of multiple documents with conflicting information: a two-step model of validation. *Educ. Psychol.* 52, 148–166. doi: 10.1080/00461520.2017.1322968

Rochnia, M., Trempler, K., and Schellenbach-Zell, J. (2023). Two sides of the same coin? A comparison of research and practice orientation for teachers and doctors. *Soc. Sci. Humanit. Open* 7:100502. doi: 10.1016/j.ssaho.2023.100502

Rousseau, D. M., and Gunia, B. C. (2016). Evidence-based practice: the psychology of EBP implementation. *Annu. Rev. Psychol.* 67, 667–692. doi: 10.1146/annurev-psych-122414-033336

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol. Bull.* 140, 1432–1463. doi: 10.1037/a0037559

Scharrer, L., Rupieper, Y., Stadtler, M., and Bromme, R. (2017). When science becomes too easy: science popularization inclines laypeople to underrate their dependence on experts. *Public Underst. Sci.* 26, 1003–1018. doi: 10.1177/0963662516680311

Schmidt, K., Edelsbrunner, P. A., Rosman, T., Cramer, C., and Merk, S. (2023). When perceived informativity is not enough. How teachers perceive and interpret statistical results of educational research. *Teach. Teach. Educ.* 130:104134. doi: 10.1016/j.tate.2023.104134

Schmidt, K., Rosman, T., Cramer, C., Besa, K.-S., and Merk, S. (2022). Teachers trust educational science – especially if it confirms their beliefs. *Front. Educ.* 7:976556. doi: 10.3389/feduc.2022.976556

Seifert, C. M., Harrington, M., Michal, A. L., and Shah, P. (2022). Causal theory error in college students' understanding of science studies. *Cogn. Res. Princ. Implic.* 7:4. doi: 10.1186/s41235-021-00347-5

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston and New York: Houghton Mifflin.

Shavelson, R. J., and Towne, L. (Eds.) (2002). Scientific research in education. Washington DC: National Academy Press.

Sinatra, G. M., Kienhues, D., and Hofer, B. K. (2014). Addressing challenges to public understanding of science: epistemic cognition, motivated reasoning, and conceptual change. *Educ. Psychol.* 49, 123–138. doi: 10.1080/00461520.2014.916216

Sinatra, G. M., and Lombardi, D. (2020). Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments. *Educ. Psychol.* 55, 120–131. doi: 10.1080/00461520.2020.1730181

Steffens, B., Britt, M. A., Braasch, J. L., Strømsø, H., and Bråten, I. (2014). Memory for scientific arguments and their sources: claim–evidence consistency matters. *Discourse Process.* 51, 117–142. doi: 10.1080/0163853X.2013.855868

Stone, C. L. (1983). A meta-analysis of advance organizer studies. *J. Exp. Educ.* 51, 194–199. doi: 10.1080/00220973.1983.11011862

Tabachnick, B. G., and Fidell, L. S. (2013). Using multivariate statistics (vol. 6). Essex: Pearson, 497–516.

Thomm, E., and Bromme, R. (2012). "It should at least seem scientific!" textual features of "scientificness" and their impact on lay assessments of online information. *Sci. Educ.* 96, 187–211. doi: 10.1002/sce.20480

Thomm, E., and Bromme, R. (2016). How source information shapes lay interpretations of science conflicts: interplay between sourcing, conflict explanation, source evaluation, and claim evaluation. *Read. Writ.* 29, 1629–1652. doi: 10.1007/s11145-016-9638-8

Thomm, E., Gold, B., Betsch, T., and Bauer, J. (2021a). When preservice teachers' prior beliefs contradict evidence from educational research. *Br. J. Educ. Psychol.* 91, 1055–1072. doi: 10.1111/bjep.12407

Thomm, E., Sälzer, C., Prenzel, M., and Bauer, J. (2021b). Predictors of teachers' appreciation of evidence-informed practice and educational research findings. *German J. Educ. Psychol.* 35, 173–184. doi: 10.1024/1010-0652/a000301

Thomm, E., Seifried, E., and Bauer, J. (2021c). Informing professional practice: (future) Teachers' choice, use, and evaluation of (non-)scientific sources of educational topics. *German J. Educ. Psychol.* 35, 121–126. doi: 10.1024/1010-0652/a000309

Toulmin, S. E. (2003). The uses of argument. Cambridge: Cambridge University Press.

Uçar, B., and Cevik, Y. D. (2020). The effect of argument mapping supported with peer feedback on pre-service teachers' argumentation skills. *J. Digit. Learn. Teach. Educ.* 37, 6–29. doi: 10.1080/21532974.2020.1815107

Van Schaik, P., Volman, M., Admiraal, W., and Schenke, W. (2018). Barriers and conditions for teachers' utilisation of academic knowledge. *Int. J. Educ. Res.* 90, 50–63. doi: 10.1016/j.ijer.2018.05.003

Voss, T. (2022). Not useful to inform teaching practice? Student teachers hold skeptical beliefs about evidence from education science. *Front. Educ.* 7:976791. doi: 10.3389/feduc.2022.976791

Wenglein, S. (2018). Studien zur Entwicklung und Evaluation eines Trainings für angehende Lehrkräfte zum Nutzen empirischer Studien. Doctoral dissertation, Technische Universität München.

Wenglein, S., Bauer, J., Heininger, S., and Prenzel, M. (2015). Kompetenz angehender Lehrkräfte zum Argumentieren mit Evidenz: Erhöht ein Training von Heuristiken die Argumentationsqualität. *Unterrichtswissenschaft* 43, 209–224. doi: 10.3262/UW1503209

Westbroek, H., Janssen, F., Mathijsen, I., and Doyle, W. (2022). Teachers as researchers and the issue of practicality. *Eur. J. Teach. Educ.* 45, 60–76. doi: 10.1080/02619768.2020.1803268

Williams, D., and Coles, L. (2007). Teachers' approaches to finding and using research evidence: an information literacy perspective. *Educ. Res.* 49, 185–206. doi: 10.1080/0013188070136 9719

Wolfe, C. R., Britt, M. A., and Butler, J. A. (2009). Argumentation Schema and the myside Bias in written argumentation. *Writ. Commun.* 26, 183–209. doi: 10.1177/0741088309333019

Yang, W. T., Lin, Y. R., She, H. C., and Huang, K. Y. (2015). The effects of prior-knowledge and online learning approaches on students' inquiry and argumentation abilities. *Int. J. Sci. Educ.* 37, 1564–1589. doi: 10.1080/09500693.2015.1045957

Zimmermann, M., and Mayweg-Paus, E. (2021). The role of collaborative argumentation in future teachers' selection of online information. *Z. Pädagog. Psychol.* 35, 185–198. doi: 10.1024/1010-0652/a000307