# The scientific impotence excuse in education – Disentangling potency and pertinence assessments of educational research

Holger Futterleib*, Eva Thomm and Johannes Bauer

Faculty of Education, University of Erfurt, Erfurt, Germany

When facing belief-contradictory scientific evidence, preservice teachers tend to doubt the potency of science and consult scientific sources less frequently. Thus, individuals run the risk not only to maintain questionable assumptions but also to develop dysfunctional stances toward research as a reliable source of knowledge. In two studies, we (a) replicated findings on the so-called *scientific impotence excuse* (SIE) in education and (b) differentiated the effects on the potency and pertinence of science to investigate educational topics to better understand the nature of SIE-related science devaluation. Both studies followed a 2 × 2 mixed experimental design: Preservice teachers assessed their prior belief about an educational topic (i.e., effectiveness of grade retention) before and after reading either confirming or disconfirming scientific evidence concerning the topic. Study 1 ($N$ = 147 preservice teachers; direct replication) confirmed the central prior findings of science devaluation when belief-evidence conflicts occur. In contrast, the results of Study 2 ($N$ = 152; follow-up study) revealed no systematic devaluations of science when disentangling the facets of potency and pertinence. Despite partial devaluation tendencies, both studies revealed that preservice teachers adapted their prior beliefs to the evidence presented. These findings extend previous research by providing insights into the conditions of science devaluation.

KEYWORDS

preservice teacher education, prior beliefs, motivated reasoning, science devaluation, evidence-based practice

## Introduction

When people are confronted with scientific evidence that contradicts their entrenched prior beliefs, they often engage in a variety of defensive responses in order to brush aside the evidence rather than to revise their personal assumptions (Kunda, 1990; Chinn and Brewer, 1993, 1998; Nauroth et al., 2014). One defensive response of such motivated reasoning can be to devalue the potency of science to study a given issue, at all. Munro (2010) dubbed this response the *scientific impotence excuse* (SIE): Facing belief-inconsistent scientific evidence, individuals dismiss the information on the grounds "that the topic of study is not amenable to scientific investigation" (Munro, 2010, p. 579). This tendency is highly problematic, because people do not only discount the particular piece of evidence; worse, they devalue the potency of science as an epistemic enterprise to attain valid knowledge about the topic. Employing the SIE may, thus, pave the way to generalized science denial (e.g., Nisbet et al., 2015; Lewandowsky and Oberauer, 2016; Hornsey, 2020).

Recently, Thomm et al. (2021a) showed that the SIE could also account for why preservice teachers develop critical stances toward educational research findings—a crucial barrier to implement evidence-informed practices as early as in initial teacher education (e.g., van Schaik et al., 2018). Despite reporting generally positive attitudes toward educational research, preservice teachers began to question its potency to examine the topic at stake when facing belief-inconsistent educational research findings. However, it remains unclear whether this devaluation concerns only participants' doubt on the *epistemic value* of research (i.e., its *potency* to provide valid knowledge) or extends to their doubt on the *pertinence* of research to investigate the topic at hand. The latter would be worse, because it strips educational research of its role as a relevant social institution to deliver reliable and valid knowledge on the topic.

In two experiments, we aimed to inspect the stability of prior findings and to disentangle the nature of devaluation in the context of preservice teachers' evaluation of educational research and its findings. Study 1 sought to replicate directly the main findings of Thomm et al. (2021a). The direct replication was chosen to test the previously found and surprising pattern showing that preservice teachers devalued the potency of science while adjusting their beliefs in the direction of the evidence presented. We also intended to corroborate the findings of the prior study, especially as it is one of the few to address preservice teachers' science devaluation. Study 2 complemented previous studies by investigating the potential effects of devaluation on both the potency and pertinence of educational research. This contribution extends prior research, as it evaluated the stability of the SIE as a mechanism of science devaluation and offers a differentiation of its effects on educational research's potency and pertinence. Thereby, the studies provide important insights into the pitfalls that must be considered when preservice teachers engage with scientific evidence in research-based teacher education.

## Understanding devaluation of educational research

Research on motivated reasoning indicates that individuals tend to argue away information that threatens their prior beliefs (Kunda, 1990; Chinn and Brewer, 1993, 1998; Nauroth et al., 2014; Britt et al., 2019). Chinn and Brewer (1993, 1998) identified several ways in which people react to such belief-discrepant evidence. Instead of revising prior assumptions, individuals may turn to ignore, reject, or reinterpret anomalous information to protect their beliefs. The SIE, as posed by Munro (2010) on the basis of cognitive dissonance theory (Festinger, 1957), complements this array of protective mechanisms, but goes beyond the mere rejection of a piece of scientific evidence. Using the SIE, individuals resolve the belief-evidence discrepancy by devaluating the ability of science to study the topic, that is, science's *potency*. Thus, individuals justify devaluation by claiming that the issue cannot be investigated by the means of science, and run the risk to develop unfavorable, generalized attitudes toward it. Munro (2010) also suggests that people are particularly prone to employ the SIE if the scientific information is strong and, thus, cannot be argued away easily (e.g., by referring to flawed methods). Indeed, SIE arguments are apparent in public debates, for example, when proponents of homeopathy claim that its effects cannot be studied by standard scientific methods such as randomized controlled trials.

Across two experiments, Munro (2010) provided evidence for the SIE. After indicating their prior beliefs about a specific medical claim, participants read scientific evidence (i.e., short abstracts) that either confirmed or disconfirmed this claim. In line with the SIE, the results showed that participants systematically discounted the potency of science to study the topic if the read evidence contradicted their prior beliefs. Moreover, they generalized their doubt to the investigation of other unrelated scientific topics and were even less inclined to choose scientific sources to inform themselves. Complementing the overall picture, participants also resisted changing their prior beliefs. Drawing on these studies, Thomm et al. (2021a) examined whether the SIE could also be observed when preservice teachers faced belief-discrepant evidence from educational research. Education is a particularly interesting field to study the SIE, first, because there is a sharp contrast between the developments to make teaching a more research-based profession (Bauer and Prenzel, 2012; Rousseau and Gunia, 2016), and the empirical observations that teachers rarely draw on educational research and, rather, rely on personal observations and common sense (Dagenais et al., 2012; Lysenko et al., 2014; Pieschl et al., 2021).

Second, the social sciences may be particularly vulnerable to devaluation, as they are often perceived as "soft" and unreliable (Berliner, 2002). Hence, people may overestimate their own abilities in judging and explaining educational issues—a tendency exacerbated by the seeming verification of beliefs through everyday observations (Thomm et al., 2021b).

In line with Munro's (2010) results, Thomm et al. (2021a) found that preservice teachers facing belief-discrepant evidence on an educational issue (i.e., the effectiveness of grade retention on low-achieving students' academic progress) devalued the potency of science, and showed a lower preference for scientific sources to further inform themselves about this topic. This devaluation occurred even though participants reported overall positive attitudes toward educational research. However, unlike the evidence reported in Munro (2010), participants did not expand their doubt about scientific potency to the investigation of other educational or unrelated topics. Moreover, participants tended to change their beliefs in the direction of the presented evidence. Apparently, this evidence worked as a refutation of participants' prior beliefs (cf. Tippett, 2010; Kendeou et al., 2014), even though the scientific abstracts used in the study were not designed according to refutation text principles.

Though these studies corroborated the main hypotheses regarding the SIE, the stability and nature of the effect require further inquiry. First, given its significance in the educational context, replication of the effect is important to evaluate its consistency and strength. Second, different aspects of the devaluation need to be disentangled more thoroughly, as elaborated below.

## Disentangling devaluation of potency and pertinence of educational research

Since many educational topics are accessible to one's own experiences and observations (Calderhead and Robson, 1991; Pajares, 1992; Richardson, 1996; Menz et al., 2021a), it may not be immediately obvious why these topics are subject to research, at all, and that research knowledge might be useful for teachers. Consequently, devaluation may not be confined to the potency of research to study educational topics; people may also contest that educational research is *pertinent* to do so (cf. Bromme and Thomm, 2016). While potency refers to the assigned epistemic value of research, pertinence represents a normative ascription of the relevant expertise to it and a mandate to contribute valuable knowledge about the domain at stake (Kitcher, 2011; Bromme and Gierth, 2021). Thus, questioning the pertinence of educational research is a more fundamental form of science rejection than doubting its potency. Discounting pertinence would allow dismissing

educational research simply as "fishing in foreign waters," even if one had to admit that scientific methods principally can contribute valid knowledge.

Existing studies have not yet differentiated (preservice) teachers' appraisal of the potency and pertinence of educational research. However, some studies implicitly have addressed aspects related to pertinence. As mentioned above, there is a multitude of studies indicating that (preservice) teachers frequently judge educational research as irrelevant and detached from their practice (McIntyre, 2005; Hammersley, 2013; Winch et al., 2015; Farley-Ripple et al., 2018; van Schaik et al., 2018; Thomm et al., 2021b) and favor experience-based knowledge, instead (e.g., Bråten and Ferguson, 2015; van Schaik et al., 2018; Kiemer and Kollar, 2021). Cain (2017) found that teachers did not only question the validity of findings but assigned science "no greater authority than their own experiences or other forms of information" (p. 13). Though such findings shed some light on teachers' perceptions of the pertinence of educational research, it is still an open issue how such perceptions are influenced by belief-evidence conflicts and how this relates to potency appraisals.

## Overview of the studies

The present contribution aimed to provide to a better understanding of the nature of the devaluation of educational research through preservice teachers by examining the SIE in two studies. Study 1 was a direct replication of Thomm et al. (2021a) that aimed to inspect the stability and strength of the effect of belief-evidence conflicts on the SIE. Study 2 was a follow-up (Schmidt, 2009) that aimed to differentiate the assessments of the perceived potency and pertinence of educational research as facets of devaluation, and to increase the external validity of prior findings. Both studies were preregistered[1]. To prevent possible cross-participation, they were conducted simultaneously with participants being assigned randomly to one of the respective studies.

## Study 1

For the direct replication, we stated the same hypotheses as Thomm et al. (2021a). Though, as elaborated above, some of their results were inconsistent with the theoretical predictions and Munro's (2010) results (i.e., hypotheses H2a and H3, below), we decided to retain the original hypotheses because they reflect

---

1    Study 1: https://osf.io/prj87
Study 2: https://osf.io/m4eaj

the theoretical reasoning behind the SIE (Munro, 2010). Hence, we examined the following hypotheses:

> H1a: Preservice teachers are more critical about the potency of educational research to study a specific educational topic when scientific evidence contradicts rather than confirms preservice teachers' prior beliefs.

> H1b: Preservice teachers reading scientific evidence that contradicts rather than confirms their prior beliefs will show a decreased preference for scientific sources and, conversely, an increased preference for non-scientific sources.

> H1c: Preservice teachers choose scientific sources less often than non-scientific sources to seek additional information about the specific educational topic.

> H2a: Preservice teachers generalize their devaluation of educational research by doubting educational research's potency to study further educational topics.

> H2b: There are no carry-over effects of science devaluation to topics from other unrelated domains (medicine and pseudo-scientific).

> H3: Preservice teachers retain their prior beliefs although scientific evidence might contradict them.

## Methods

For the direct replications, all methods followed the design, materials, and procedures of Thomm et al. (2021a).

### Design

Study 1 was a $2 \times 2$ mixed experiment with the within-participants factor prior belief (before vs. after reading the evidence) and the between-participants factor evidence (confirming vs. disconfirming the effectiveness of grade retention). Accordingly, preservice teachers were randomly assigned to one of two conditions, reading either confirming or disconfirming scientific evidence on the effectiveness of grade retention (GR) to reduce potential deficits in students' school achievement. Before and after reading the assigned evidence, participants reported their respective beliefs about GR effectiveness.

### Participants

Based on an *a priori* power analysis with GPower 3.1 (Faul et al., 2009), we aimed at an effective sample size of $N = 202$

preservice teachers to detect the effects of at least $f^2 = 0.087$ (cf. Thomm et al., 2021a) with 95% power ($\alpha = .05$) in a multiple linear regression[2].

Participants were recruited online through invitation via participant databases, advertisements at university lectures, and mailing lists in Germany. Participation was voluntary and participants could withdraw at any time without giving reasons or experiencing any consequences. As an incentive, participants could enroll in a lottery with winnings in the amount of 10–20€.

A total of $N = 237$ participants completed the study. During data cleaning, we deleted cases according to preregistered exclusion criteria as follows: $n = 15$ participants who had not provided informed consent or had withdrawn it; one case who was not enrolled in a teacher education program; $n = 30$ participants with unreasonable response times for completing the experiment (i.e., $<5$ or $>120$ mins); and $n = 44$ participants who had spent less than 1 min on the evidence reading task. This reflects an exclusion rate of 38%. The final sample of $N = 147$ still provided 85.4% power and allowed detecting effects as low as $f^2 = 0.076$ with 80% probability. Preservice teachers in the final sample were mostly female (83.7%) and $M = 21.9$ years old ($SD = 2.77$ years). They were mostly studying in a bachelor's degree program (49.7%) and had completed, on average, 3.82 semesters ($SD = 1.92$). Further, 31.3% were enrolled in a master's degree program with a duration of study of $M = 8.22$ semesters ($SD = 1.71$). The remaining 19% studied in traditional state examination programs (with no separate Bachelor's and Master's phase) with a duration of study of $M = 4.9$ semesters ($SD = 2.7$).

### Procedure

The study was realized as an online experiment. After the general introduction and giving informed consent, participants reported demographic information (i.e., age, gender, study program). Subsequently, they rated their prior *belief about the effectiveness of GR*. Then, they read an introductory text on the topic of GR and were randomly assigned to one of two conditions presenting either confirming or disconfirming scientific evidence (i.e., five short abstracts) on the effectiveness of GR. After reading the evidence, participants rated the potency of science to study the topic at stake (*topic-specific potency*), as well as the potency of science to study further domain-related and domain-unrelated topics (*domain-related* and *domain-unrelated potency*). They further judged their preferences for scientific and non-scientific sources to learn more about GR effects (*source preference*) and were asked to select their most preferred source (*source choice*). Finally, participants reassessed their beliefs about the effectiveness of GR. Having finished the experiment, participants had the option to withdraw their consent for data usage and received a thorough debriefing.

---

2   Due to a technical error, the preregistered power analysis indicated a recommended sample size of $N = 181$, but $N = 202$ participants would have been needed to achieve the power of 95%.

## Materials

Preservice teachers first read a short introductory text describing a recently published review summarizing scientific studies about the effects of GR on remedying low-performing students' deficits in school achievement. Next, each experimental group received five abstracts, each summarizing an empirical study on GR. The abstracts had been designed and tested by Thomm et al. (2021a) for the target group and to provide equivalent levels of length and complexity across the experimental conditions. The abstracts had a standardized form equivalent to typical study abstracts and ended with a clear final conclusion on the effectiveness of GR. Overall, they were representative of scientific research in this field regarding the applied methods, results, and conclusions. Across the evidence conditions, results and conclusions varied from confirming to disconfirming the effectiveness of grade retention. For more information on the development of the materials, see Thomm et al. (2021a); materials are available in the corresponding Appendix S1.

## Measures

Unless indicated otherwise, the answer format for all measures described below was a 9-point rating scale (1 = do not agree at all, 9 = very much agree).

*Prior belief* was measured by rating the statement "Repeating a grade helps struggling students to compensate for their achievement deficits."

*Topic-specific doubt on the potency of science* to study the effectiveness of GR was assessed by rating the statement "The question whether GR helps struggling students to compensate their deficits in achievement is one that <u>cannot</u> be answered using scientific methods" (cf. Munro, 2010).

To measure the *generalization of science devaluation*, participants assessed the potency of science to study six additional educational topics (e.g., impact of class size on learning outcomes) and eight unrelated topics. The unrelated topics covered health issues (e.g., cell phone radiation causing cancer) and pseudo-scientific topics (e.g., astrology as a possible predictor of personality; Munro, 2010). Per topic, the participants rated the statement "How far can scientific methodologies be used to determine whether (e.g., computer-based learning supports students' knowledge acquisition)?" (1 = not at all, 9 = very well). Items were averaged per domain and resulted in sufficiently reliable scores (education: $\alpha = 0.68$, medicine: $\alpha = 0.81$, and pseudo-science: $\alpha = 0.73$).

As a measure of *source preferences* participants received a list of seven scientific (i.e., findings from scientific studies; educational scientist) and non-scientific sources (i.e., opinions and experiences of teacher, school student who repeated a class, teacher association, proponent and opponent of GR). They judged how likely they would be to seek information from each source. Both scales yielded acceptable reliability (preference for scientific sources, $\alpha = 0.62$; preference for non-scientific sources,

$\alpha = 0.63$)[3]. In addition to their source preferences, participants had to choose one source from the list that they would finally consult (*source choice*).

## Analyses

We performed all analyses using R 4.0.2 (R Core Team, 2022). To test hypotheses H1a, H1b, H2a, and H2b (i.e., moderating effects of prior belief on the relation of scientific evidence and doubt of scientific potency, respectively their source preferences), we conducted multiple regression analyses using Hayes' PROCESS macro 4.1 for R (Hayes, 2022). The evidence condition was dummy-coded (0 = confirming vs. 1 = disconfirming the effectiveness of GR), and prior belief was centered at the grand mean. For testing H2b, we set $\alpha = 0.20$ because the null hypothesis was the target (i.e., no generalization of doubting science to topics from unrelated domains).

To test H1c, we ran a binary logistic regression regarding whether belief-evidence conflicts decrease the choice of a scientific source (coded as 1) over a non-scientific one (coded as 0). H3 (i.e., belief change) was tested by a repeated-measures ANOVA.

As preregistered, we applied transformations to variables exhibiting highly asymmetric distributions prior to analysis in order to avoid biased standard errors and significance tests (Tabachnick and Fidell, 2014; Fox, 2016). Highly asymmetric distributions were characterized by both P-P plots and significant tests of skewness ($z \geq |2.58|$, indicating $p \leq 0.01$; Field et al., 2013). We applied log transformation to variables with positive moderate skew (i.e., topic-specific potency), inverse transformation to variables with extreme positive skew (i.e., pseudo-scientific topics potency), and a reflect-and-log transformation for variables exhibiting moderate negative skew (i.e., domain-specific potency, preference for scientific sources; Tabachnick and Fidell, 2014).

## Results

Table 1 provides an overview of the descriptive statistics. Scale means indicated that participants altogether had favorable beliefs about the potency of science and a noteworthy preference for scientific sources to inform themselves about educational topics.

## Devaluation of the potency of science and its sources (hypotheses 1a−1c)

To test H1a, we regressed doubt over the potency of science on evidence condition, prior beliefs, and their interaction. Analyses yielded a statistically significant overall effect, $F$(3,

---

3  Like Thomm et al. (2021a), we averaged the preference ratings for proponent and opponent of GR because of their high correlation ($r = 0.89$, $p < 0.001$).

TABLE 1   Study 1: Descriptive statistics by evidence condition.

|  | M (SD) | Skew (SE) | Kurtosis (SE) |
|---|---|---|---|
| **Evidence confirming effectiveness of grade retention** |  |  |  |
| Belief about GR effectiveness (T1) | 5.63 (2.08) | −0.53 (0.28) | −0.52 (0.55) |
| Belief about GR effectiveness (T2) | 7.40 (1.36) | −1.27 (0.28)* | 2.47 (0.55) |
| Doubt on science's potency to study GR | 2.16 (1.55) | 2.09 (0.28)* | 4.38 (0.55) |
| Potency to study related educational topics | 6.40 (1.38) | −1.21 (0.28)* | 2.50 (0.55) |
| *Potency to study unrelated topics* |  |  |  |
| Medicine | 5.81 (1.84) | −0.67 (0.28) | 0.06 (0.55) |
| Pseudo-science | 2.11 (1.36) | 1.79 (0.28)* | 3.69 (0.55) |
| *Source Preference* |  |  |  |
| Scientific sources GR | 7.73 (1.00) | −0.76 (0.28)* | 0.05 (0.55) |
| Non-scientific sources GR | 6.06 (1.34) | −0.28 (0.28) | 0.37 (0.55) |
| **Evidence disconfirming effectiveness of grade retention** |  |  |  |
| Belief about GR effectiveness (T1) | 5.90 (2.13) | −0.40 (0.28) | −1.03 (0.56) |
| Belief about GR effectiveness (T2) | 3.35 (1.85) | 0.90 (0.28)* | −0.10 (0.56) |
| Doubt on science's potency to study GR | 3.11 (2.10) | 1.04 (0.28)* | 0.19 (0.56) |
| Potency to study related educational topics | 6.18 (1.33) | −1.43 (0.28)* | 4.37 (0.56) |
| *Potency to study unrelated topics* |  |  |  |
| Medicine | 5.45 (1.54) | −0.11 (0.28) | −0.14 (0.56) |
| Pseudo-science | 2.65 (2.09) | 1.56 (0.28)* | 1.64 (0.56) |
| *Source Preference* |  |  |  |
| Scientific sources GR | 7.36 (1.52) | −0.96 (0.28)* | 0.29 (0.56) |
| Non-scientific sources GR | 5.71 (1.42) | −0.28 (0.28) | 0.00 (0.56) |

Values represent untransformed data; GR = grade retention; *P-P-plots and significant tests of skew ($z \geq |2.58|$, $p \leq 0.01$) indicated serious skewness.
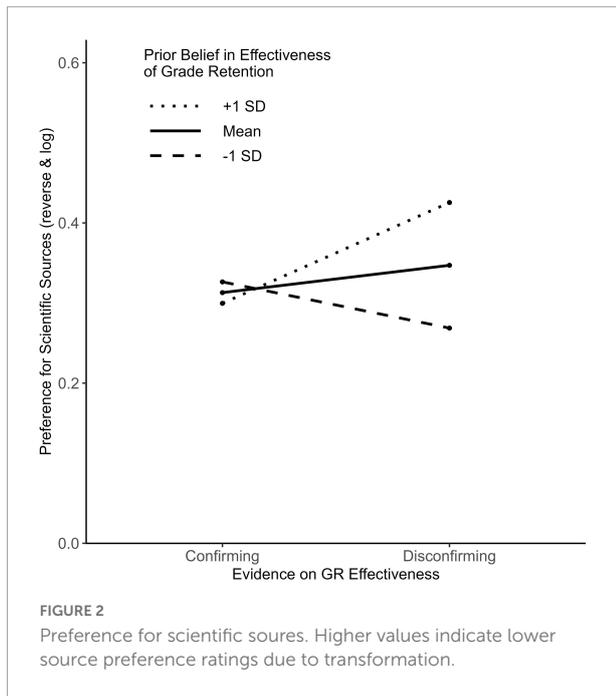
143) = 4.90, $p = 0.003$, $R^2 = 0.09$. In line with our assumption, the interaction term of evidence condition and prior beliefs confirmed a significant moderation effect of prior belief, $b = 0.04$, $SE(b) = 0.02$, $t = 2.02$, $p = 0.046$. **Figure 1** depicts the crossover interaction entailed by H1a. Additionally, the regression model yielded a statistically significant main effect of evidence condition, $b = 0.14$, $SE(b) = 0.05$, $t = 3.15$, $p = 0.002$.

Probing the interaction, we conducted a simple slopes analysis. Findings revealed a statistically significant simple effect of evidence condition for participants with both high [i.e., 1 SD above the sample mean; $b = 0.23$, $SE(b) = 0.06$, $t = 3.66$, $p < 0.001$] and average prior belief [i.e., at the sample mean; $b = 0.14$, $SE(b) = 0.05$, $t = 3.15$, $p = 0.002$], but not for participants with low prior belief [i.e., 1 SD below the sample mean; $b = 0.05$, $SE(b) = 0.06$, $t = 0.80$, $p = 0.452$]. That is, particularly participants with strong or average prior beliefs in GR effectiveness tended to doubt the potency of science when evidence contradicted their prior beliefs.

To test H1b, we regressed the preferences for scientific sources on evidence condition, prior beliefs, and their interaction. The overall model was statistically significant, $F(3, 142) = 3.64$, $p = 0.014$, $R^2 = 0.07$, and yielded a statistically significant interaction, $b = 0.04$, $SE(b) = 0.02$, $t = 2.54$, $p = 0.012$ (**Figure 2**).



FIGURE 1

Doubt on the potency of science to study the effectiveness of grade retention.
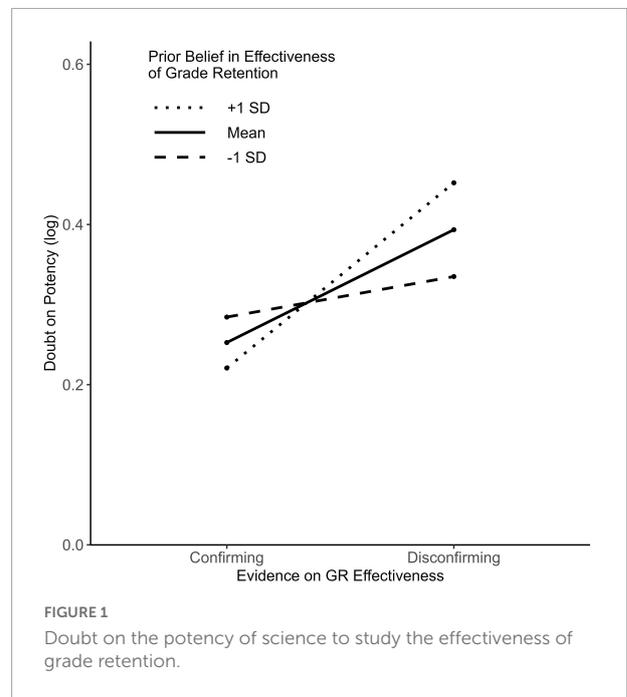
Subsequent simple slopes analysis revealed a statistically significant effect of evidence on participants with high prior belief [i.e., 1 SD above the sample mean; $b = 0.13$, $SE(b) = 0.05$,

**FIGURE 2**
Preference for scientific soures. Higher values indicate lower source preference ratings due to transformation.



**FIGURE 3**
Belief change regarding the effectiveness of grade retention (Study 1).

$t = 2.47$, $p = 0.015$] but not for participants with average [i.e., at the sample mean; $b = 0.03$, $SE(b) = 0.04$, $t = 0.95$, $p = 0.346$] and low prior belief [i.e., 1 $SD$ below sample mean; $b = -0.06$, $SE(b) = 0.05$, $t = -1.13$, $p = 0.260$]. Thus, facing belief-discrepant evidence, participants with high prior belief in the effectiveness of GR tended to have a decreased preference for scientific sources.

An analogously performed regression with preference for non-scientific sources did not attain significance, $F(3, 142) = 0.87$, $p = 0.456$, $R^2 = 0.02$.

The logistic regression for source choice (H1c) failed to attain statistical significance for the overall model [$\chi^2(3) = 7.60$, $p = 0.055$], even though the interaction effect entailed by the hypothesis yielded statistical significance [$b = -0.40$, $\chi^2(1) = 4.81$, $p = 0.028$].

## Generalization of science devaluation to further domain-related and unrelated topics (hypotheses 2a and 2b)

Contrary to H2a, the regression model for the potency of science to investigate further educational topics (H2a) did not yield statistical significance, $F(3, 143) = 0.79$, $p = 0.501$, $R^2 = 0.02$. The same was true for the regression models for unrelated medical [$F(3, 143) = 0.63$, $p = 0.600$, $R^2 = 0.01$] and pseudo-scientific topics [$F(3, 143) = 0.52$, $p = 0.671$, $R^2 = 0.01$]. The latter two findings, however, are in line with the expected null effects for these dependent variables (H2b). Hence, participants did not generalize their doubt over the potency of science to other related or unrelated topics.
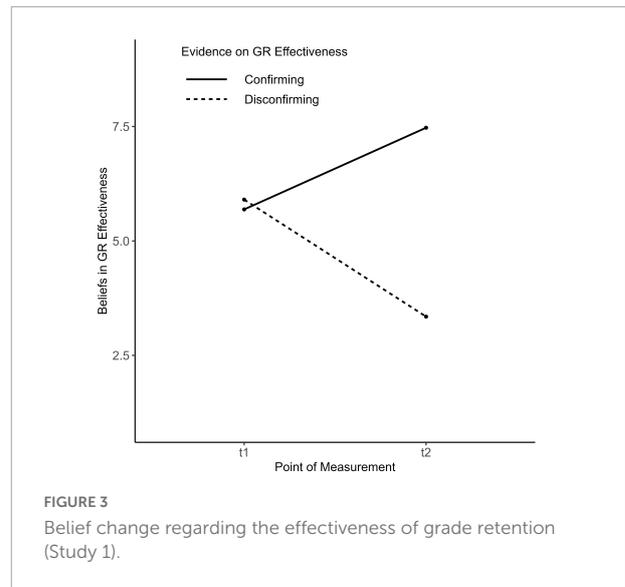
## Belief change in the face of belief-contradictory evidence (hypothesis 3)

A mixed ANOVA with the within-participants factor belief (before vs. after reading the evidence) and the between-participants factor evidence condition (confirming vs. disconfirming evidence) yielded statistically significant main effects of evidence condition [$F(1, 144) = 54.14$, $p < 0.001$, $\eta^2 = 0.22$] and prior belief [$F(1, 144) = 6.92$, $p = 0.009$, $\eta^2 = 0.01$], as well as a statistically significant interaction [$F(1, 144) = 218.76$, $p < 0.001$, $\eta^2 = 0.26$] (**Figure 3**). Follow-up dependent samples $t$-tests indicated that participants changed their beliefs in the direction of the presented evidence. Participants who had read confirming evidence increased their belief in the effectiveness of GR, $t(73) = 9.40$, $p < 0.001$, $d = 1.09$, whereas participants reading disconfirming evidence decreased it, $t(71) = 11.38$, $p < 0.001$, $d = 1.34$.

## Conclusion

Study 1 aimed to directly replicate the findings of Thomm et al. (2021a). We found the same pattern of findings as the original study with comparable (small) effect sizes, the only exception being the non-significant overall model of source choice. Both studies support the SIE (Munro, 2010): In the face of strong, belief-threatening evidence that cannot be brushed aside easily, people may turn to devalue the potency of science to study the issue to protect their beliefs (H1a). This discounting also affected preferences for scientific sources (H1b). However, concerning source choice (H1c), we only found a tendency in the expected direction: The significant interaction effect was invalidated by the non-significant overall test such that this result should not be interpreted. These discrepant findings suggest that a lower preference for scientific sources does not

translate into increased preference for non-scientific sources, *per se*, hence, favoring the choice of the latter. Preservice teachers are, instead, more likely to consider scientific sources to answer a (scientific) question.

In contrast to the findings by Munro (2010), the results of this study and Thomm et al. (2021a) unanimously suggest that individuals do not necessarily transfer devaluation to further related and unrelated topics. Our results indicate that preservice teachers did not generalize their doubt about the potency of science to other educational research topics (H2a). There was also no evidence of carry-over effects to unrelated medical or pseudo-scientific topics. Thomm et al. (2021a) already discussed that individuals perceive knowledge domains differently and, therefore, may not automatically generalize devaluation across unrelated domains. Overall, this may indicate that the devaluation implied by the SIE seems to be a topic-related phenomenon, at least at first encounter.

Moreover, despite the reported difficulty to initiate belief revision (e.g., Richardson, 1996; Lewandowsky et al., 2012), both Thomm et al. (2021a) and this replication found that participants' prior beliefs shifted toward the conclusion supported by the read evidence. Though unexpected, these findings can possibly be explained by literature on knowledge revision and refutational texts (Tippett, 2010; Kendeou et al., 2014; Butterfuss and Kendeou, 2021). We will elaborate more deeply on both issues in the general discussion.

In summary, despite the discrepant result regarding H1c, overall, we conclude that the present study successfully replicated the main findings reported in Thomm et al. (2021a).

# Study 2

It remained an open question whether the devaluation found from Study 1 mainly related to doubt regarding the epistemic value of research, or whether it extends to the perceived pertinence of science to investigate a topic. These aspects of devaluation can occur independently or in combination. For example, individuals can question the potency of science to investigate a specific issue, while still considering science as generally pertinent to providing reliable and valid knowledge on it. As discussed above, devaluating the pertinence of science would constitute an even stronger case of science devaluation compared to discounting potency, solely. Thus, to better understand the nature of SIE-related devaluation, Study 2 examined whether belief-evidence conflicts affect both the perceived potency and pertinence of educational research to study educational topics. Regarding potency, we tested the same hypotheses as in Study 1. Concerning pertinence, we stated two research questions focused on the specific educational topic (i.e., GR effects) and on generalization to other relevant topics. We did not expect or test for generalization to other unrelated (i.e., medical and pseudo-scientific) topics, because we considered

pertinence assessments as a strongly topic-related phenomenon (i.e., educational research is pertinent to answer educational research questions, but not for questions from other domains).

In addition to these substantive issues, in Study 2, we also sought to enhance the external validity of the experimental materials. To assure experimental control, Thomm et al. (2021a) and Study 1 had presented participants with the same texts across both evidence conditions, manipulating only the direction of the results and conclusion. In Study 2, we presented summaries of original published studies on grade retention effects as evidence. Consequently, participants read different studies in the respective condition.

# Methods

All methods and procedures were identical to Study 1 unless indicated otherwise, below.

## Design and procedure

Perception of the pertinence of educational research was added as a dependent variable to the design. This resulted in minor modifications of the procedure: After reading the introductory text and the scientific evidence, participants assessed the topic-specific potency and topic-specific pertinence of science, both displayed on one page of the online questionnaire. Next, participants rated the potency of science to study educational topics (domain-related) and other topics (domain-unrelated). Afterward, participants answered the pertinence questions on further educational topics. We deliberately placed potency before pertinence assessments for each respective topic, instead of balancing their order. This was done to ensure that potency measures remained unaffected by the respective pertinence measures and to maintain comparability to the effects on potency across Study 1 and Study 2. We judged this as being more important than ruling out potential position effects (see general discussion).

## Participants

A total of $N = 221$ participants completed Study 2. We removed cases according to the preregistered exclusion criteria as follows: $n = 10$ participants with missing consent declarations; $n = 2$ participants not enrolled in a teacher education program; $n = 9$ for unreasonable response times (i.e., $<5$ or $>120$ mins); and $n = 48$ participants who spent less than 1 min on the page with the evidence (experimental manipulation); overall exclusion rate 31.2%. The final sample ($N = 152$) still yielded sufficient statistical power (86.7%) and allowed detecting effects as low as $f^2 = 0.074$ with 80% probability. Participants in the final sample were mostly female (84.1%), $M = 22.23$ years old ($SD = 3.24$ years) and mostly studying in a bachelor's degree program (49.3%) with a duration of study of $M = 3.88$ semesters ($SD = 1.91$). Of the participants, 33.6% were enrolled in a

master's degree program ($M$ = 8.61 semesters, $SD$ = 2.74), and 17.1% in a state examination program ($M$ = 5.23 semesters, $SD$ = 2.69).

## Materials

To create a new set of scientific abstracts, either confirming or disconfirming the effectiveness of GR, that were analogous in design to Study 1, we carried out an extensive literature research to identify appropriate original research articles addressing the effectiveness of GR. We then formed pairs of studies, one confirming and one disconfirming the effectiveness of GR, that were comparable in publication year, outcome variables, and/or the methods used. Due to the use of original studies as a basis, the abstracts were not equivalent across conditions, as they had been in Study 1. Specifically, the evidence presented to disconfirm the effectiveness of GR was more diverse than it was in the previous studies, reporting both null findings and negative effects of GR. We harmonized the abstracts in structure, length ($M$ = 168.2, $SD$ = 18.8 words), and complexity (e.g., statistical methods) to enhance comparability across conditions. Moreover, we added the original author names and the year of publication. The material was pretested and revised via cognitive interviews with $N$ = 10 students from different subjects (i.e., psychology, education, and teaching) regarding comprehensibility, consistency, and methodological soundness. The abstracts are available in the respective OSF directory.

## Measures

The measures of *prior belief* (both measurement points), *topic-specific potency*, and *potency for domain-related* and *domain-unrelated* topics were identical to Study 1. However, the reliabilities of the potency scales for related and unrelated topics were weaker than in Study 1 (education: α = 0.55, medicine: α = 0.76, and pseudo-science: α = 0.63). To maintain the consistency of the measures across studies, we decided to retain these scales as they were. The results for potency for further educational topics should be interpreted cautiously due to the low reliability of this scale.

We made minor adjustments in *source preference and choice* to cover a broader range of scientific sources (i.e., research results in scientific journals, scientific textbooks, opinion of an educational scientist, applied educational, or popular science journals) and non-scientific sources (i.e., the education section of the daily press, educational guidebooks, the experiences of a seasoned teacher, the experiences of other preservice teachers, the experiences of family or friends). Because of these changes, we inspected the factorial structure with exploratory factor analysis and built scales on this basis. Two items had to be deleted due to high cross-loadings. The resulting scales had acceptable reliabilities (preference for scientific source α = 0.70; preference for non-scientific sources α = 0.80).

Topic-specific doubt on the *pertinence of science* was measured with the statement "Science is not pertinent

to answer the question of whether GR helps struggling students to compensate for their deficits in achievement." Pertinence devaluation items regarding further educational topics referred to the same topics as the potency assessments. Participants rated, for each topic, the statement "To what extent is science pertinent to investigate whether (e.g., computer-based learning supports students' knowledge acquisition)?" The average scores displayed acceptable reliability: α = 0.83.

## Results

Table 2 provides an overview of the descriptive statistics. Scale means indicated that participants had favorable beliefs about the potency and pertinence of science and a high preference for scientific sources to inform themselves about the educational topic at stake. The ratings of potency and pertinence were positively correlated for both grade retention ($r$ = 0.46, $p$ < 0.001) and the further educational topics ($r$ = 0.45, $p$ < 0.001).

## Devaluation of potency and pertinence of science, scientific source preferences, and source choice (hypotheses 1a, 1b, and 1c; Research question 1)

To inspect potential devaluation, we tested the impact of belief-evidence conflicts on the assessments of both the potency and pertinence of science. The overall regression model for topic-specific potency (log-transformed) failed statistical significance, $F(3, 148)$ = 2.19, $p$ = 0.092, $R^2$ = 0.04, even though the interaction effect entailed by H1a was significant, $b$ = 0.06, $SE(b)$ = 0.02, $t$ = 2.34, $p$ = 0.02. The regression model for topic-specific pertinence (log-transformed) on evidence condition, prior beliefs, and their interaction also failed statistical significance, $F(3, 148)$ = 1.84, $p$ = 0.142, $R^2$ = 0.04.

Also, the regression model for scientific source preference (H1b) was not statistically significant, $F(3, 148)$ = 0.07, $p$ = 0.974, $R^2$ = 0.001. The analogously performed regression for non-scientific source preferences (reflected and square-root-transformed) yielded a statistically significant overall model [$F(3, 148)$ = 2.94, $p$ = 0.035, $R^2$ = 0.04], but no significant individual effects. Finally, the logistic regression for source choice (H1c) also remained non-significant, $\chi^2(3)$ = 3.23, $p$ = 0.308.

## Generalizing science devaluation to further educational and unrelated topics (hypotheses 2a and 2b; Research question 2)

Contrary to H2a, participants showed no systematic devaluation of the potency of science to investigate other educational topics, $F(3, 148)$ = 0.18, $p$ = 0.905,

TABLE 2  Study 2: Descriptive statistics by evidence condition.

| | M (SD) | Skew (SE) | Kurtosis (SE) |
|---|---|---|---|
| **Evidence confirming effectiveness of grade retention** | | | |
| Belief about GR effectiveness (T1) | 5.77 (1.74) | −0.42 (0.30) | −0.13 (0.60) |
| Belief about GR effectiveness (T2) | 7.47 (1.18) | −1.09 (0.30) | 1.86 (0.60) |
| Doubt on science's potency to study GR | 2.37 (1.51) | 1.35 (0.30)* | 1.46 (0.60) |
| Doubt on science's pertinence to study GR | 2.31 (1.43) | 0.73 (0.30) | −0.64 (0.60) |
| Potency to study related educational topics | 6.60 (1.12) | 0.09 (0.30) | −0.49 (0.60) |
| Pertinence to study related educational topics | 6.89 (1.44) | −0.58 (0.30) | 0.17 (0.60) |
| *Potency to study unrelated topics* | | | |
| Medicine | 6.06 (1.31) | −0.04 (0.30) | −0.82 (0.60) |
| Pseudo-science | 2.61 (1.52) | 1.17 (0.30)* | 1.31 (0.60) |
| *Source Preference* | | | |
| Scientific sources GR | 7.54 (1.25) | −0.97 (0.30)* | 0.42 (0.60) |
| Non-scientific sources GR | 6.01 (1.60) | −0.11 (0.30) | −0.39 (0.60) |
| **Evidence disconfirming effectiveness of grade retention** | | | |
| Belief about GR effectiveness (T1) | 5.79 (1.88) | −0.45 (0.25) | −0.52 (0.50) |
| Belief about GR effectiveness (T2) | 2.94 (1.54) | 1.01 (0.25) | 0.51 (0.50) |
| Doubt on science's potency to study GR | 2.57 (1.62) | 1.10 (0.25)* | 0.64 (0.50) |
| Doubt on science's pertinence to study GR | 2.09 (1.47) | 1.58 (0.25)* | 2.61 (0.50) |
| Potency to study related educational topics | 6.67 (1.08) | −0.33 (0.25) | 0.75 (0.50) |
| Pertinence to study related educational topics | 7.08 (1.38) | −0.65 (0.25) | 0.20 (0.50) |
| *Potency to study unrelated topics* | | | |
| Medicine | 6.28 (1.42) | −0.24 (0.25) | 0.13 (0.50) |
| Pseudo-science | 3.12 (2.09) | 1.10 (0.25)* | 0.32 (0.50) |
| *Source Preference* | | | |
| Scientific sources GR | 7.57 (1.08) | −0.72 (0.25)* | 0.31 (0.50) |
| Non-scientific sources GR | 5.92 (1.86) | −0.33 (0.25) | −0.74 (0.50) |

Values represent untransformed data; GR = grade retention; *both P-P-plots and significant tests of skew ($z \geq |2.58|$, $p \leq 0.01$) indicated serious skewness.

$R^2 = 0.004$. Analogously, there were no effects on pertinence assessments (reflected and square-root-transformed) for these additional educational topics, $F(3, 148) = 1.22$, $p = 0.303$, $R^2 = 0.02$.

In line with H2b, regression models for potency regarding medical [$F(3, 148) = 0.58$, $p = 0.626$, $R^2 = 0.001$] and pseudo-scientific topics [log-transformed; $F(3, 148) = 1.05$, $p = 0.374$, $R^2 = 0.02$] were non-significant.

### Belief change in the face of belief-contradictory evidence (hypothesis 3)

The mixed ANOVA for belief change yielded statistically significant main effects of evidence condition [$F(1, 149) = 115.77$, $p < 0.001$, $\eta^2 = 0.35$] and prior belief [$F(1, 149) = 47.63$, $p < 0.001$, $\eta^2 = 0.09$], as well as a statistically significant interaction [$F(1, 149) = 229.56$, $p < 0.001$, $\eta^2 = 0.33$] (Figure 4). Follow-up $t$-tests for dependent samples indicated that participants changed their beliefs in the direction of the evidence presented. Participants who had read confirming evidence increased their belief in the effectiveness of GR, $t(60) = 8.31$,

$p < 0.001$, $d = 1.06$, and vice versa, $t(90) = 13.9$, $p < 0.001$, $d = 1.47$.

## Conclusion

Study 2 provided mixed findings. In contrast to Study 1, it did not confirm the effect of topic-specific devaluation of science. Further, the findings from Study 2 did not indicate that participants tended to decrease their preferences for scientific sources and we did not observe any generalization to the study of other topics. The latter finding is in line with Study 1 and Thomm et al. (2021a). Regarding the perceived pertinence of science, a comparable pattern emerged: Despite being confronted with belief-challenging evidence, preservice teachers did not devalue the pertinence of science systematically. Finally, as in these prior studies, there was evidence of belief change in the direction of the presented evidence.

Overall, the results of Study 2 suggest that there was no systematic devaluation of science, thus raising the question of how stable and far-reaching the effects of belief-evidence

**FIGURE 4**
Belief change regarding the effectiveness of grade retention (Study 2).
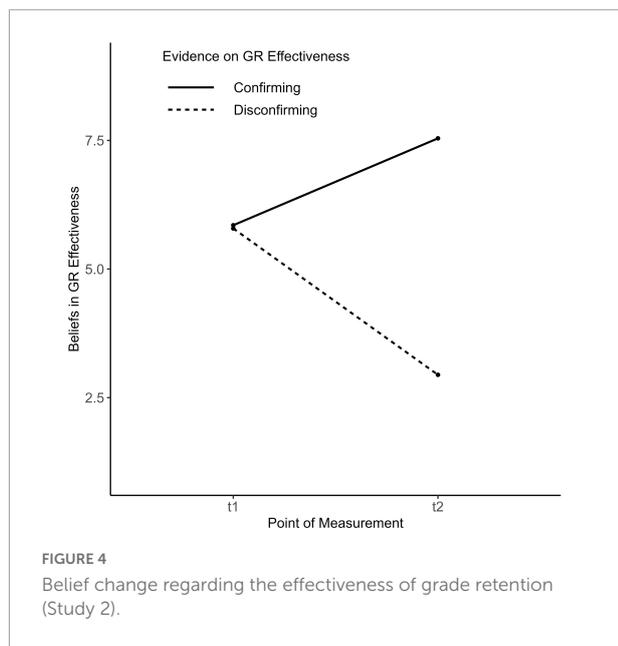
conflicts on science devaluation are. However, as elaborated in more detail below, the discrepant pattern of results in Study 2 may be due to differences in the newly developed abstracts. The more heterogeneous evidence provided in the abstracts might have seemed less conclusive to the participants and left them with more opportunity to argue away the evidence.

# General discussion

## Discussion

Understanding the mechanisms of science devaluation is of great societal importance (Chinn and Brewer, 1998; Lewandowsky and Oberauer, 2016; Britt et al., 2019; Hornsey, 2020; Kienhues et al., 2020). The present study addressed such a mechanism, the SIE, that occurs when individuals encounter belief-threatening scientific evidence (Munro, 2010). As indicated above, this may frequently occur with educational topics (Asberger et al., 2021; Thomm et al., 2021a). Moreover, tendencies to devalue knowledge from educational research seem to be already prevalent in preservice teachers (van Schaik et al., 2018). To advance our understanding of the SIE and the related science devaluation, Study 1 replicated the preliminary findings from Thomm et al. (2021a). In Study 2, we aimed to draw a more detailed picture by distinguishing the effects on appraisals of the potency and the pertinence of science. To integrate the results, **Table 3** provides an overview of how the findings from the two present studies relate to prior research by Munro (2010) and Thomm et al. (2021a). The below discussion follows the order of the main hypotheses around the SIE, that is, the effects on the devaluation of science's potency and sources

(H1a–H1c), the generalization of devaluation (H2a–H2b), and the resistance to belief change (H3). Subsequently, we address the findings on pertinence from Study 2.

*Devaluation of science's potency and sources.* The core hypothesis associated with the SIE is that people tend to devalue the potency of science to address a scientific issue when facing strong belief-threatening evidence. In addition to the direct assessment of the topic-related potency (H1a), source preferences (H1b), and choice (H1c) are indicators of such devaluation. The three studies that implemented the original paradigm (i.e., Munro, 2010; Thomm et al., 2021a; Study 1) delivered corroborating evidence on these hypotheses. One apparent discrepancy is that Study 1 identified only a descriptive tendency regarding the effects on source choice (i.e., non-significant overall model despite a significant interaction). This result needs to be contextualized in the mostly small effect sizes identified in the existing studies. Generally, the effects of encountering a single belief-evidence conflict on science devaluation seem to be small when judged by conventional rules of thumb. That being said, Thomm et al. (2021a) argue that these small effects should not be underestimated because, so far, we know little about the accumulative effects of repeated conflict experiences. In summary, despite the mentioned constraints, Study 1 indicates, in concert with prior research, that the SIE is a valid mechanism of science devaluation.

In contrast to these findings, Study 2 did not confirm the hypothesized devaluation tendencies. As hinted above, we are of the opinion that this discrepancy is most likely due to the changes in the materials applied in Study 2 (though other explanations are possible, such as chance, small/instable effects, or the interference of an additional pertinence assessment). Whereas, in the former studies, participants received methodologically strong and consistent evidence, the abstracts in Study 2 offered participants more opportunities to challenge the evidence itself. As Munro (2010) suggested, employment of the SIE may be more likely when there is very little opportunity to dismiss the validity of the evidence. Thus, the participants in Study 2 may have been able to reduce the cognitive conflict in other ways, rather than turning to the SIE. In this light, one might argue that changing the experimental materials in Study 2 was unfortunate because this may have diminished the likelihood to observe the effect of interest. We would respond, however, by saying that these materials more closely resembled the results teachers might encounter from a real literature search. The results from (educational) research rarely point unanimously toward a consistent answer. In contrast, though from this point of view the materials used in Study 1 and Thomm et al. (2021a) may seem somewhat artificial, they also represent the evidence teachers might encounter in real life. For example, journalistic media reports frequently present research findings as conclusive and consistent, without

TABLE 3 Comparison of findings on hypotheses about effects of the scientific impotence excuse.

| Study | Increased doubt on potency of science for specific topic (H1a) | Decreased preference for scientific sources (H1b) | Decreased probability for choice of scientific sources (H1c) | Increased doubt on potency of science for other topics | | Belief maintenance (H3) |
|---|---|---|---|---|---|---|
| | | | | Related (H2a) | Unrelated (H2b) | |
| Munro (2010) | Yes | Yes | Yes | — | Yes | Yes |
| Thomm et al. (2021a) | Yes | Yes | Yes | No | No | No |
| Study 1 (replication) | Yes | Yes | No | No | No | No |
| Study 2 (follow-up)† | No* | No | No | No* | No | No |

†Only study that provided participants with evidence presenting heterogeneous effects regarding the topic; *findings are identical for pertinence assessments.

communicating uncertainty (e.g., van der Bles et al., 2020). Hence, the different abstracts used in Studies 1 and 2 may pertain to different situations in which individuals encounter research. The findings from Study 2 inspire further research to evaluate more closely the *conditions* under which individuals are inclined to employ the SIE over other potential responses (Chinn and Brewer, 1998). Specifically, future studies might examine the effects of the presence and types of (a) methodological information contained in the abstracts and (b) information about the uncertainty associated with the evidence (e.g., different numerical or verbal formats for communicating uncertainty; van der Bles et al., 2020). Using such features to vary the strength and conclusiveness of the evidence would permit the creation of settings that should be differentially prone to elicit the SIE. Next to contributing to a better understanding of science devaluation, such studies could also make a valuable contribution to enhancing science communication (Kienhues et al., 2020).

*Generalization to further related and unrelated topics.* In contrast to Munro (2010), neither Thomm et al. (2021a) nor the present experiments delivered any evidence of generalization of devaluation to other related (i.e., educational) and unrelated (i.e., medical and pseudo-scientific) topics. The reasoning behind the respective hypotheses was that, if people discount the potency of science to deliver knowledge on a topic, this doubt would likely extend to similar topics, if not to all of (empirical) science. However, it appears that the generalization effects identified by Munro (2010) do not replicate, at least in applications with educational topics. Hence, we conclude that devaluation may not transfer easily, at least from single belief-evidence conflicts. This might be good news, at first glance. However, the experiences of belief-evidence conflicts may be aggravated by additional factors, such as tensions between the information and the individual's social identity. For example, Nauroth et al. (2014) found that computer gamers devaluated the scientific evidence on the negative effects of gaming that threatened their social identity. Similar effects may

occur for teachers when the evidence contradicts not only their topic-related beliefs but also their professional identities or the values and practices of their community of practice. These issues should be addressed in further studies.

*Resistance to belief change.* Unlike Munro (2010), Thomm et al. (2021a) and both present studies found belief change in the direction of the evidence read with sizeable effects. This occurred even though beliefs are notoriously difficult to change (Richardson, 1996; Lewandowsky et al., 2012; Swire et al., 2017; Menz et al., 2021b). This belief change may seem surprising in light of the observed tendencies to devalue science. Two issues arise in this regard. First, the question regarding how the belief change might have been initiated can be answered by drawing upon the Knowledge Revision Components Framework–Multiple Documents (KReC-MD; Butterfuss and Kendeou, 2021). Though not explicitly designed that way, the studies' materials may have served the central principles of knowledge revision. For example, the introductory text on GR and the subsequent study abstracts may have simultaneously activated and contrasted prior and new knowledge and, thus, served the KReC-MD's principles of co-activation and competing activation. Second, one might wonder about the depth and stability of the belief change. The current data provide no evidence on this, unfortunately. While the result might reflect a true belief revision, it might as well be an instance of what Chinn and Brewer (1993, 1998) call *peripheral theory change*. That is, individuals might have provisionally changed their espoused beliefs while still preserving the original one. According to Thomm et al. (2021a), peripheral theory change is one explanation for the seeming contradiction between simultaneous science devaluation, as found in Study 1, and belief change. Being confronted with five pieces of fully (Study 1; Thomm et al., 2021a) or quite (Study 2) consistent evidence that covered various educational contexts and methodological approaches might have been strongly persuasive to participants. Hence, in the post assessment, they may have reported what they *should* believe according to science without

actually believing it, as expressed in their devaluation of science's potency. Another explanation is that a true belief change may have occurred, but under epistemic vigilance (cf. Sperber et al., 2010). That is, participants might not have changed their beliefs blindly but added a doubt over science as a cognitive marker that the experienced epistemic conflict had not been solved satisfactorily (Thomm et al., 2021a). Despite these open questions, the finding that reading multiple science texts can initiate belief change may be promising. As a potential implication for teacher education, it may be helpful to present multiple sources of evidence to back up positions that may conflict with students' prior beliefs. Moreover, teacher educators might address rejection strategies like the SIE explicitly to make students aware of problematic reactions to belief-inconsistent information.

*Devaluation of science's pertinence.* Study 2 did not deliver any evidence of SIE-related effects on the participants' appraisal of the pertinence of educational science to investigate GR effectiveness or further educational topics. That is, the experiences of belief-evidence conflicts do not seem to raise preservice teachers' doubts regarding educational research as a relevant societal institution for contributing knowledge about educational issues. This result may be seen as reassuring. However, because pertinence was investigated only in Study 2, we do not know whether the effects on pertinence would have occurred had participants faced the more consistent evidence materials from Study 1. The limitations regarding the Study 2 materials discussed above may apply to pertinence, too. Moreover, as aforementioned, we cannot rule out that the potency and pertinence assessments interfered with each other. Hence, despite the demonstrated null effects, we suggest that the conditions of pertinence devaluation (as well as its relation to potency devaluation) are a worthwhile subject for further investigation.

## Limitations

Beyond the issues discussed already, we acknowledge the following limitations. First, both studies evaluated science devaluation by examining the exemplary educational topic of GR effectiveness. Investigating the generalizability of the results to other educational topics would be warranted. Second, applying the preregistered exclusion criteria led to a substantial reduction in both studies' sample sizes. Though unfortunate, this is a common problem in online research. The resulting statistical power was still sufficiently high, but since the effects of the SIE seem to be small, future studies should use larger samples. Third, the minimum reading time criterion may have disadvantaged fast readers. In the

same vein, though excluding participants without sufficient exposure to the treatment is reasonable, time criteria are always somewhat arbitrary. Latent class analysis of time on task might provide a more principled approach for this purpose (Bauer, 2022). Finally, the reliabilities of the source preference scales proved quite low in Study 2, as compared with Study 1 and Thomm et al. (2021a). Low reliabilities may have led to less precise and attenuated estimates of the regression coefficients.

## Conclusion

Despite the discussed limitations, our contribution advances prior research to gain a deeper understanding of the mechanisms of science devaluation and, more generally, preservice teachers' attitudes toward and interactions with scientific evidence (van Schaik et al., 2018; Thomm et al., 2021b; Ferguson et al., 2022). Study 1 provided an overall successful direct replication of earlier findings regarding SIE-related science devaluation. Though Study 2 delivered no indication of science devaluation, it supported Munro's (2010) assumptions about the conditions under which the SIE occurs. Interestingly, in line with Thomm et al. (2021a), both studies hint that preservice teachers may have overall favorable attitudes toward science that, however, may be damaged under certain conditions. Moreover, we added evidence that belief change may be initiated (at least provisionally) by confronting preservice teachers with multiple pieces of scientific evidence, even when the texts do not fully satisfy the principles proposed in the refutation and knowledge revision literature (Tippett, 2010; Kendeou et al., 2014; Butterfuss and Kendeou, 2021).

Regarding practical implications, our findings raise the question of how teacher education can help to mitigate potential devaluation mechanisms. To this end, teacher educators need to explicitly address (preservice) teachers' prior beliefs about course-related topics that can shape how they interact with the research-based knowledge they are required to learn (Fives and Buehl, 2012; Asberger et al., 2021; Menz et al., 2021b). Making participants aware of the likelihood of conflicts between their prior beliefs and course contents, as well as of typical devaluative responses to such conflicts, may be effective preemptive measures in advance of learning. By informing participants about cognitive biases and potential devaluation mechanisms in advance, it could help them to expect and resolve the conflicts that sometimes arise, as well as to foster an open attitude toward conflicting information. Such techniques of inoculation, related refutation, and debunking have been found effective in other contexts and can be adopted by teacher educators (Kendeou et al., 2014; Cook et al., 2017; Lewandowsky and van der Linden, 2021; Pieschl et al., 2021). However, this requires equipping them with the knowledge of how to implement such methods in their lectures.

## Data availability statement

## Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the University of Erfurt (No. 20211123). The patients/participants provided their written informed consent to participate in this study.

## Author contributions

HF contributed to the study conceptualization, methodology, project administration, data collection and curation, analysis and visualization, interpretation of results, and original draft of the manuscript. ET and JB contributed to the study conceptualization, methodology, supervision, interpretation of results, and reviewing and editing the manuscript. All authors contributed to the article and approved the submitted version.

## 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.

## References

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

Bauer, J. (2022). "A primer to latent profile and latent class analysis," in *Methods for Researching Professional Learning and Development: Challenges, Applications and Empirical illustrations*, eds M. Goller, E. Kyndt, S. Paloniemi, and C. Damsa (Cham: Springer International Publishing), 243–268. doi: 10.1007/978-3-031-08518-5_11

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

Berliner, D. C. (2002). Comment: Educational research: The hardest science of all. *Educ. Res.* 31, 18–20. doi: 10.3102/0013189x031008018

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

Britt, M. A., Rouet, J. -F., Blaum, D., and Millis, K. (2019). A reasoned approach to dealing with fake news. *Policy Insights Behav. Brain Sci.* 6, 94–101. doi: 10.1177/2372732218814855

Bromme, R., and Gierth, L. (2021). "Rationality and the public understanding of science," in *The Handbook of Rationality*, eds M. Knauff and W. Spohn (Cambridge, MA: The MIT Press), 767–776.

Bromme, R., and Thomm, E. (2016). Knowing who knows: Laypersons' capabilities to judge experts' pertinence for science topics. *Cogn. Sci.* 40, 241–252. doi: 10.1111/cogs.12252

Butterfuss, R., and Kendeou, P. (2021). Krec-MD: Knowledge revision with multiple documents. *Educ. Psychol. Rev.* 33, 1475–1497. doi: 10.1007/s10648-021-09603-y

Cain, T. (2017). Denial, opposition, rejection or dissent: Why do teachers contest research evidence? *Res. Pap. Educ.* 32, 611–625. doi: 10.1080/02671522.2016.1225807

Calderhead, J., and Robson, M. (1991). Images of teaching: Student teachers' early conceptions of classroom practice. *Teach. Teach. Educ.* 7, 1–8. doi: 10.1016/0742-051X(91)90053-R

Chinn, C. A., and Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Rev. Educ. Res.* 63, 1–49. doi: 10.3102/00346543063001001

Chinn, C. A., and Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *J. Res. Sci. Teach.* 35, 623–654. doi: 10.1002/(SICI)1098-2736(199808)35:6<623:AID-TEA3>3.0.CO;2-O

Cook, J., Lewandowsky, S., and Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: exposing misleading argumentation techniques reduces their influence. *PLoS One* 12:e0175799. doi: 10.1371/journal.pone.0175799

Dagenais, C., Lysenko, L., Abrami, P. C., Bernard, R. M., Ramde, J., and Janosz, M. (2012). Use of research-based information by school practitioners and determinants of use: A review of empirical research. *Evid. Policy* 8, 285–309. doi: 10.1332/174426412X654031

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., Buchner, A., and Lang, A. -G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Ferguson, L. E., Bråten, I., Skibsted Jensen, M., and Andreassen, U. R. (2022). 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.* doi: 10.1177/00224871221105813 [Epub ahead of print].

Festinger, L. (1957). *A Theory of Cognitive Dissonance (vol. 2).* Redwood City, CA: Stanford University Press

Field, A., Miles, J., and Field, Z. (2013). *Discovering Statistics Using R.* Thousand Oaks, CA: SAGE.

Fives, H., and Buehl, M. M. (2012). "Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us?," in *APA Educational Psychology Handbook, Vol 2: Individual Differences and Cultural and Contextual Factors*, eds K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, and M. Zeidner (Washington, DC: American Psychological Association), 471–499.

Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models*, 3rd Edn. Los Angeles, CA: SAGE.

Hammersley, M. (2013). *The Myth of Research-Based Policy and Practice.* London: SAGE.

Hayes, A. F. (2022). "Introduction to mediation, moderation, and conditional process analysis: A regression-based approach," in *Methodology in the Social Sciences*, 3rd Edn, (New York, NY: The Guilford Press).

Hornsey, M. J. (2020). Why facts are not enough: Understanding and managing the motivated rejection of science. *Curr. Dir. Psychol. Sci.* 29, 583–591. doi: 10.1177/0963721420969364

Kendeou, P., Walsh, E. K., Smith, E. R., and O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Process.* 51, 374–397. doi: 10.1080/0163853X.2014.913961

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

Kienhues, D., Jucks, R., and Bromme, R. (2020). Sealing the gateways for post-truthism: Reestablishing the epistemic authority of science. *Educ. Psychol.* 55, 144–154. doi: 10.1080/00461520.2020.1784012

Kitcher, P. (2011). Public knowledge and its discontents. *Theory Res. Educ.* 9, 103–124. doi: 10.1177/1477878511409618

Kunda, Z. (1990). The case for motivated reasoning. *Psychol. Bull.* 108, 480–498. doi: 10.1037/0033-2909.108.3.480

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., and Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychol. Sci. Public Interest* 13, 106–131. doi: 10.1177/1529100612451018

Lewandowsky, S., and Oberauer, K. (2016). Motivated rejection of science. *Curr. Dir. Psychol. Sci.* 25, 217–222. doi: 10.1177/0963721416654436

Lewandowsky, S., and van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *Eur. Rev. Soc. Psychol.* 32, 348–384. doi: 10.1080/10463283.2021.1876983

Lysenko, L., Abrami, P. C., Bernard, R. M., Dagenais, C., and Janosz, M. (2014). Educational research in educational practice: Predictors of use. *Can. J. Educ.* 37, 1–26.

McIntyre, D. (2005). Bridging the gap between research and practice. *Camb. J. Educ.* 35, 357–382. doi: 10.1080/03057640500319065

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

Menz, C., Spinath, B., and Seifried, E. (2021b). 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

Munro, G. D. (2010). The scientific impotence excuse: Discounting belief-threatening scientific abstracts. *J. Appl. Soc. Sci.* 40, 579–600. doi: 10.1111/j.1559-1816.2010.00588.x

Nauroth, P., Gollwitzer, M., Bender, J., and Rothmund, T. (2014). Gamers against science: The case of the violent video games debate. *Eur. J. Soc. Psychol.* 44, 104–116. doi: 10.1002/ejsp.1998

Nisbet, E. C., Cooper, K. E., and Garrett, R. K. (2015). The partisan brain. *Ann. Am. Acad. Pol. Soc. Sci.* 658, 36–66. doi: 10.1177/0002716214555474

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Rev. Educ. Res.* 62, 307–332. doi: 10.3102/00346543062003307

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

R Core Team (2022). *R: A Language and Environment for Statistical Computing.* Vienna: R Core Team.

Richardson, V. (1996). "The role of attitudes and beliefs in learning to teach," in *Handbook of Research on Teacher Education*, 2nd Edn, eds J. Sikula, T. J. Buttery, and E. Guyton (New York, NY: Macmillan), 102–119.

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

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Rev. Gen. Psychol.* 13, 90–100. doi: 10.1037/a0015108

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., et al. (2010). Epistemic vigilance. *Mind Lang.* 25, 359–393. doi: 10.1111/j.1468-0017.2010.01394.x

Swire, B., Ecker, U. K. H., and Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 1948–1961. doi: 10.1037/xlm0000422

Tabachnick, B. G., and Fidell, L. S. (2014). *Using Multivariate Statistics (6th ed., Pearson New International Edition).* Harlow: Pearson.

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-based practice and educational research findings. *Zeitschrift Für Pädagogische Psychologie* 35, 173–184. doi: 10.1024/1010-0652/a000301

Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *Int. J. Sci. Math. Educ.* 8, 951–970. doi: 10.1007/s10763-010-9203-x

van der Bles, A. M., van der Linden, S., Freeman, A. L. J., and Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proc. Natl. Acad. Sci. U.S.A.* 117, 7672–7683. doi: 10.1073/pnas.1913678117

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

Winch, C., Oancea, A., and Orchard, J. (2015). The contribution of educational research to teachers' professional learning: Philosophical understandings. *Oxf. Rev. Educ.* 41, 202–216. doi: 10.1080/03054985.2015.1017406