Mechanisms of Epistemic Change: The Roles of Reflection and Social Interaction

Tom Rosman & Martin Kerwer

Leibniz Institute for Psychology (ZPID), Trier, Germany

Abstract

Fostering students' epistemic beliefs is key for achieving a more nuanced approach to psychological knowledge. The Bendixen-Rule model on epistemic change posits epistemic doubt (questioning one's prior epistemic beliefs), epistemic volition (the will to change one's beliefs) and resolution strategies (strategies to overcome epistemic doubt by epistemic change) as three interrelated process components that lead to the development of more advanced epistemic beliefs. However, while the model has risen to relative prominence over the last years, the postulated mechanisms of change still lack empirical backing. In this article, we report on an experimental study with $N = 153$ psychology students that aimed at testing the effects of two specific resolution strategies—reflection and social interaction. This was realized by developing intervention components that target the two strategies, and by analyzing these components' incremental effects on epistemic change. Results showed that reflection and social interaction might be promising strategies to address epistemic doubt.

Psychology lecturers should thus give students room for reflecting on and discussing their beliefs once doubt has arisen.

Keywords: Epistemic beliefs, epistemic change, reflection, social interaction, experimental study

# 1	Background

Epistemic beliefs are individual beliefs about the nature of knowledge and knowing: A psychology freshman, let us call her Lena, may conceive psychological knowledge as a body of certain and undisputable 'facts'—an absolute belief according to the well-known framework by Kuhn and Weinstock (2002). In contrast, her roommate Julia may view psychological knowledge as tentative and subjective, up to the point where she reduces all psychological findings to mere 'opinions' (a multiplistic belief; Kuhn & Weinstock, 2002). Finally, a more advanced student—Sarah—may recognize that while some psychological theories are backed by considerable evidence, others are based on shaky foundations. As Sarah's perspective—a so-called evaluativistic belief (Kuhn & Weinstock, 2002)—implies an evidence-based approach to knowledge and knowing, it comes as no surprise that advanced epistemic beliefs have been shown to positively affect learning and information processing in psychology and beyond (e.g., Barzilai & Zohar, 2012; Greene et al., 2018; Kardash & Howell, 2000; Rosman et al., 2018). For this reason, a multitude of epistemic change interventions have been developed, striving towards the goal of fostering advanced epistemic beliefs (see Muis et al., 2016).

As a theoretical underpinning, most interventions on epistemic change refer to the Process Model for Personal Epistemology Development (Bendixen & Rule, 2004; Rule &

Bendixen, 2010). This framework suggests three interrelated process components that interactively influence epistemic change: For epistemic change to occur, individuals must recognize a discrepancy between their actual beliefs and new information (*epistemic doubt*), they need motivation or willingness to act on their beliefs (*epistemic volition*), and they have to employ adequate strategies to resolve epistemic doubt by changing their beliefs, such as reflection or social interaction (*resolution strategies*; Bendixen & Rule, 2004; Rule & Bendixen, 2010). However, given the relative prominence of the Bendixen-Rule framework in studies on epistemic change, it is striking that—as Bråten (2016) concluded reviewing the literature on epistemic change interventions—its "mechanism [of change] essentially lacks empirical backing" (p. 361). In fact, while the role of epistemic doubt is comparatively well-studied (e.g., Ferguson et al., 2012; Kerwer & Rosman, 2018), to our knowledge, only one experimental intervention specifically addresses epistemic volition (Kerwer et al., 2021), and the resolution strategies component has not yet been put to experimental scrutiny at all (but see Bendixen, 2002 and Ferguson et al., 2012). In terms of the Bendixen-Rule framework, resolution strategies are however of crucial importance as they establish the link between epistemic doubt and epistemic change: Without adequate resolution strategies, individuals might, for example, just ignore their doubt, resulting in little to no epistemic change.

Considering the importance of resolving epistemic doubt, the present article thus empirically investigates the resolution strategies component of the Bendixen-Rule framework. We expect that more knowledge on this component will benefit theory building, but also impact the design of better epistemic change interventions. However, it should be pointed out that the resolution strategies component still is the least developed one in the framework in terms of its definition, its suggested properties, and its underlying psychological mechanisms. Instead of adding more complexity to the framework by specifying additional resolution strategies, we therefore focus on *reflection* and *social*

interaction as the two probably most prototypical—and certainly most frequently

mentioned—strategies.

According to Bendixen and Rule (2004), reflection "involves reviewing the past,

analyzing belief implications, and making educated choices" (p. 73). This requires a deep and

elaborative processing of the information that has led to epistemic doubt, such as contrasting

this information and its epistemological "nature" to one's prior beliefs, and elaborating on

why this might be so. Hence, referring to Craik and Lockhart's (1972) distinction between

deep and shallow processing, epistemic change should be more pronounced when the

information that is evoking epistemic doubt is processed at a deeper level. Social interaction,

in contrast, is strongly related to the PACES framework on epistemic change (Muis et al.,

2016). In this framework, Muis et al. (2016) suggest five aspects of classroom environments

that may foster epistemic change. *Pedagogy* describes a constructivist approach to learning,

where "teachers can repeatedly expose students to conflicting or paradoxical points of view,

and ask students whether and how this conflict could be resolved" (p. 350), or engage

students in group discussions where multiple perspectives on an issue are presented.

Authority means that teachers should reduce their authoritative role and consider the

individual students as sources and justifiers of knowledge. *Curriculum* is about including

materials suited to elicit epistemic change into the courses, such as texts that containing

conflicting and contradictory information. *Evaluation* describes constructivist assessment

practices (e.g., open-ended questions versus multiple-choice) and formative feedback.

Finally, *Support* is about the "scaffolding that teachers provide during the change process,

such as explicit modeling of critical thinking or resolution strategies" (p. 353),

acknowledgment of negative affect elicited by epistemic doubt, or sharing personal

experiences with students (Muis et al., 2016). We argue that social interaction activities led

by the principles of Pedagogy, Authority and Support may be best in eliciting epistemic

change since they are most strongly related to argumentation activities, which, in turn, is a key component of social interaction as specified by Bendixen and Rule (2004).

Since engaging in argumentation and discussion to resolve one's epistemic doubt is not possible without a certain amount of reflection, we thereby see reflection as a prerequisite for social interaction (and probably as the most central resolution strategy). In addition, we argue that social interaction may serve as an additional mechanism that will possibly lead to stronger epistemic change since individuals get to know additional perspectives or methods to resolve their doubt. Furthermore, social interaction may also further foster deep and elaborative processing of the intervention materials (through techniques such as asking students how a specific conflict may be resolved, for example), which is why it may also increase the effects of interventions targeting reflection alone. An example: Imagine Lena (who we described earlier as an absolutist thinker) getting confronted with conflicting findings on whether grade retention leads to better learning. This will likely trigger epistemic doubt regarding her absolute epistemic beliefs, which she will try to resolve by employing certain resolution strategies. For example, she may engage in reflection by herself and come to the conclusion that psychological knowledge is inherently subjective (multiplistic beliefs), or she might understand that conflicting knowledge claims may be integrated through evaluation (evaluativistic beliefs). Such evaluativistic beliefs might be fostered even more when she engages in argumentative discourse with Sarah, our evaluativist thinker who recognizes the value of evidence in the generation and evaluation of knowledge.

In the present work, we report on the development and empirical evaluation of two intervention components that explicitly target the reflection and social interaction components of the Bendixen-Rule framework. The intervention components thereby extend an existing intervention concept targeting epistemic doubt and volition, which is realized as

an experimental task in which subjects deal with conflicting evidence (Kerwer et al., 2021; Kerwer & Rosman, 2018). We expected that an intervention including the reflection (R) and social interaction (RSI) components would lead to stronger epistemic change compared to a control intervention (C), and compared to an intervention only targeting epistemic doubt and epistemic volition (DV). Contrasting these components thus allows to generate evidence for the beneficial effects of reflection and social interaction in epistemic change, supporting a key assumption of the Bendixen-Rule framework. Furthermore, we expected that an intervention including both a reflection and a social interaction component would be superior compared to an intervention only including a reflection component. Finally, numerous studies have found positive relationships between epistemic beliefs and multiple-documents comprehension (Barzilai & Strømsø, 2018; Bråten et al., 2011; Bråten et al., 2013, 2014). This may be because "awareness that problems have multiple aspects, that there can be more than one possible explanation and solution, and that knowledge is uncertain and evolving, may lead readers to pay greater attention to multiple viewpoints on the issue and to define the task as one that requires integration of information from diverse sources" (Barzilai & Strømsø, 2018, p. 104). Hence, epistemic beliefs seem to affect sourcing behavior, which is why we also expected positive effects of our intervention on source evaluation task performance. Correspondingly, we preregistered the following set of hypotheses (Rosman & Kerwer, 2019):

Hypothesis 1: For epistemic change on self-report measures (FREE-GST[1]; see below), we expect the following differences between intervention conditions:

---

[1] It should be noted that the preregistration also mentions "justification beliefs" at this point. However, since all planned analyses on justification beliefs are exploratory (see preregistration),

- *H1a:* All epistemic change interventions (RSI, R, DV) will promote stronger change towards advanced epistemic beliefs compared to the control condition (C).

- *H1b:* Epistemic change interventions targeting epistemic doubt, epistemic volition and resolution strategies (RSI and R) will promote stronger epistemic change compared to an intervention targeting epistemic doubt and volition only (DV).

- *H1c:* An epistemic change intervention targeting epistemic doubt, epistemic volition, as well as both reflection and social interaction (RSI), will have higher effects on epistemic change compared to an intervention targeting epistemic doubt, epistemic volition, and reflection only (R).

Hypothesis 2 (H2a to H2c): The pattern of effects described in H1a to H1c does not only apply to change in epistemic beliefs, but also to source evaluation task performance.

2 Method

2.1 Intervention Components

Intervention components for doubt and volition were identical to those employed by Kerwer et al. (2021), and were administered in their computer-based versions (see Kerwer & Rosman, 2020b). All intervention materials were specifically tailored to psychology students and revolved around a psychological topic (gender stereotyping in schools).

Epistemic doubt. To evoke epistemic doubt, participants read 18 short texts describing fictitious studies on whether boys or girls are disadvantaged by secondary school teachers. One third of the texts indicate that girls are disadvantaged, one third indicate that boys are

---

justification beliefs analyses are only reported in Supplementary Material 1 for reasons of manuscript length, and the term "justification beliefs" is omitted at this point.

disadvantaged, and one third of the fictitious studies indicate no significant gender bias (e.g., in awarded grades; for details see Kerwer et al., 2021 as well as Rosman et al., 2019). Since these apparent contradictions can be resolved by identifying moderator[2] variables (hence the name *resolvable controversies*), they are suited to evoke epistemic doubt regarding absolute beliefs (an absolutist would neglect the existence of controversies) and multiplistic beliefs (a multiplist would neglect the possibility of resolving controversies; Kerwer et al., 2021).

Epistemic Volition. The epistemic volition component includes contents such as a written introduction to the concept of epistemic beliefs (based on the framework by Kuhn and Weinstock, 2022), empirical evidence on the benefits of advanced epistemic beliefs, written case studies exemplifying different types of beliefs, fictitious feedback on students' epistemic beliefs as well as a fictitious individual change prognosis (for details see Kerwer et al., 2021). It should be noted that the volition component is administered *before* the epistemic doubt component, which might seem counterintuitive at first, but is in line with Rule and Bendixen's (2010) argument that epistemic volition must already be present when individuals experience epistemic doubt (or at least appear simultaneously with the emergence of doubt). In our view of the Bendixen-Rule framework, this is because individuals must be aware of their epistemic beliefs to experience doubt—if this is not the case, they will not experience a dissonance between their actual beliefs and new experiences, and thus no negative affective state to be reduced via epistemic change (see also Kerwer et al., 2021).

Reflection. The reflection component uses the exact same text materials as the epistemic doubt component, but participants additionally verbalize, instructed by a student assistant,

---

[2] It should be noted that other aspects such as variations in study quality might also allow a resolution of controversial findings. However, to keep the study design as simple as possible, only moderators were included in the present study.

their task-related thoughts and beliefs using think-aloud methodology. As the main goal of this component was to trigger deep and elaborative processing, the intervention draws on level 3 verbalizations, in which participants are instructed to explain and justify their thoughts instead of just verbalizing them (Ericsson & Simon, 1993). Level 3 verbalizations are triggered by several means. First, a written instruction is given, which asks participants to actively reflect on their epistemic beliefs in the task to come ("Instruction: I try to become aware of my own epistemic beliefs regarding psychological research while working on the task. When doubts about my epistemic beliefs arise, I consciously address them. If my motivation fades, I remember that advanced epistemic beliefs will prove beneficial to my studies). Second, the reflection of conflicting knowledge claims is supported by specific questions that are shown after 6, 12 and 18 texts ("First, think out loud about the extent to which, according to the studies you've read, schools are experiencing gender discrimination. How do you feel about these findings? Think aloud about whether and to what extent you find the studies contradictory! In doing so, try to stay actively aware of your epistemic beliefs. Does this evidence fit with your beliefs about psychological knowledge?"). Finally, there is a concluding reflection task after the last text on the intervention's implications for participants' personal epistemic beliefs ("Think aloud about whether a differentiated set of findings, as reflected in the texts, fits your beliefs on psychological science. Perhaps you thought quite differently yesterday?"; cf. Bendixen and Rule's [2004] definition of reflection as "reviewing the past, analyzing belief implications, and making educated choices", p. 73). The purpose of using think-aloud methodology is twofold. First, the process of thinking aloud itself can be seen as a reflective process that boosts students' reflections over their beliefs since it helps them to "use previously acquired mental representations to plan, evaluate, and reason between alternative solutions" (Epler et al., 2013, p. 49; Ericsson & Simon, 1998). Hence, it is well-suited to evoke reflection processes and deep processing, especially when

the think-aloud methodology uses level 3 verbalizations. Second, it allows to monitor whether participants adhere to the instruction of reflecting over their beliefs—the instructor will thus provide additional reflection cues if participants' spontaneous reflection processes come to a halt. Aside from the think-aloud instructions and these additional cues, instructors, however, do not interact with the participants, thereby adopting a rather 'professional', impersonal stance.

Reflection and Social Interaction. This component builds upon the reflection component presented above, employing the same think-aloud methodology, reflection questions, and a concluding reflection task (see above). However, it differs in that the instructor engages in a more natural conversation with the participants and tries to actively shape their epistemic beliefs through social interaction, thereby employing many of the techniques suggested in the PACES framework (Muis et al., 2016). For example, she (or he) introduces herself with her first name, mentions that she studies psychology, too, and acts as role model by claiming that she has already dealt successfully with the challenge of doubting one's beliefs. Rather than acting as a 'nameless' examiner (cf. epistemic climate, Muis & Duffy, 2013), the instructor offers support if participants question existing beliefs (cf. the notion of compassion/modeling, Rule & Bendixen, 2010), but also challenges participants' existing naive beliefs to some extent. To ensure a standardized procedure, the instructor is handed a document which outlines possible responses to a number of scenarios that may arise during the session. For example, if a participant mentions that the findings described in the texts seem very inconsistent, the instructor prompts her to look at differences between the studies, and if a participant verbalizes his doubt, the instructor offers social support by mentioning that she also had experienced such doubt beforehand, which was frustrating at first, but beneficial in the long term.

2.2   Design and Experimental Conditions

The study used a 4x2 pre-post design with one within-subjects factor (levels: pre- and post-intervention) and one between-subjects factor (intervention type with 4 levels). The intervention components presented above were thereby administered as follows: One group only received the doubt and volition components (DV group), one group additionally received the reflection component (R group), one group additionally received the reflection and social interaction components (RSI group), and one group was subjected to a control task on learning strategies (C group), hence not receiving any of the above-mentioned components. This latter task was identical to the control task by Kerwer and Rosman (2020a), and consisted of students reading non-conflicting texts on students employing different learning strategies. In addition, to keep the study duration constant across conditions, participants in this condition received a number of additional questionnaires on learning strategies (see preregistration for details; Rosman & Kerwer, 2019). It is of note that for practical reasons, the DV and C interventions took place in a group setting (several participants working individually on the respective materials in a computer lab, supervised by an instructor), whereas participants from the R and RSI conditions received the interventions in one-to-one sessions with an instructor. Participants were randomly assigned to these treatment conditions using a two-step process (see preregistration for details).

2.3   Participants and Procedure

Sample size estimation was based on effect sizes taken from Kerwer and Rosman (2018), whose smallest intervention effect was $f = .094$ ($r = .66$) in a study with a design similar to ours. We argue that in the present study, group differences should be more pronounced since the resolution strategy groups receive a considerably more elaborate (and thus more 'intense') intervention. Hence, a power analysis in GPower (Faul et al., 2009), in

11

which we set α to .05 in a four-group repeated measures design, indicated that with a total sample size of $N = 180$ participants ($n = 45$ per group), statistical power ($1 - \beta$) would be .78 for $f = .1$ and .92 for $f = .12$. We therefore decided to recruit $N = 180$ participants.

Since our intervention focused on psychology and since we were primarily interested in participants with less advanced (and perhaps more malleable) epistemic beliefs, we decided to only recruit psychology students enrolled at Trier University for less than four semesters. Participants were paid for their participation and were recruited by means of flyers, mailing lists and Facebook groups. The study took place from October to December 2019. Pretest measurements were taken online, with participants completing the data collection independently (without instructor) using their own device. One week after completing the pretest, participants were invited for intervention and posttest, which took place in two labs at Trier University. The study was supervised by two female student assistants. Both held a BSc degree in psychology and had some background knowledge on epistemic beliefs. Two training sessions on the reflection and social interaction components took place before the study was conducted. In these sessions, several intervention scenarios were played through, with the authors of the present paper engaging in different roles (e.g., study participants or role models). Additionally, the student assistants were handed a document with detailed instructions, which was refined after the training sessions.

Sample size calculations yielded a target sample size of $N = 180$ (see preregistration). However, despite a lot of advertising, participation rates were lower than expected, which is why we decided to additionally recruit Master's students and students with psychology as a minor subject in their first four semesters. Even with this deviation from the preregistration, we, however, were only able to recruit $N = 153$ participants, which is 15 % less than

expected. Participants were 88.7 % female, and were $M = 22.88$ ($SD = 4.28$) years old. Study subjects were 76.8 % BSc psychology, 22.5 % MSc psychology, and 0.7 % minor.

2.4   Measures

Primary outcome. Topic-specific epistemic beliefs, our primary outcome, were measured using the FREE-GST questionnaire (Rosman et al., 2019). The FREE-GST is based on the domain-general FREE questionnaire by Krettenauer (2005), but focuses on the topic of gender stereotyping in secondary schools, thereby fitting the scope of our epistemic doubt intervention. Participants are first presented with a scenario including three controversial positions on gender discrimination in secondary schools. Subsequently, 15 either absolute, multiplistic, or evaluativistic statements on the controversy are to be rated on 6-point Likert scales. A sample item for high absolutism is 'The future will show which position is definitely correct' (Rosman et al., 2019). On this basis, we calculated our primary outcome, the D-Index, which reflects a multidimensional change towards advanced beliefs (Krettenauer, 2005) and is computed as 2*Evaluativism – (Absolutism + Multiplism). Scale reliability of the D-Index was calculated by assigning the respective absolutism, multiplism and evaluativism items to triplets and by subsequently averaging these triplets' reliabilities. In the original FREE questionnaire, each such triplet would consist of one absolute, one multiplistic, and one evaluativistic item for each of the 15 sub-topics of the questionnaire, hence resulting in 15 triplet combinations in total. However, in our version of the questionnaire, it is not possible to structure the triplets by sub-topics since the questionnaire focuses on one topic only. Therefore, we computed the D-Index' mean reliability as an average of 3000 randomly selected triplet combinations, of whom each included one absolute, one multiplistic, and one evaluativistic item (see Kerwer & Rosman, 2020b). Averaging Omega over this high number of triplet combinations ensures that the amount of

"arbitrariness" (or bias) that may result from the random assignment to triplet combinations is reduced to a minimum. These calculations resulted in a reliability of $\omega = .55$ for the first and $\omega = .74$ for the second measurement point. In addition, we calculated arithmetic means for the absolutism, multiplism, and evaluativism scales, which we used for exploratory analyses.

Secondary outcome. To measure source evaluation performance, students worked on a scenario-based task requiring them to investigate whether achievement motivation is associated with heart diseases or not (details see Kerwer et al., 2021). For their inquiry, students were presented with four fictitious search engine results pages. According to Kerwer et al. (2021), "each of these four outputs contained four fictitious results—two high-quality and two low-quality sites, while the valence (i.e., the supposed direction of the relationship—achievement motivation increases or decreases the risk of heart diseases) differs within each pair (e.g., within high quality sites). For each output, subjects chose which site they would like to inspect in more detail during a later stage of the study" (p. 12). Two indices were calculated based on this task—one on the quality of the chosen links and one on the diversity of the search. The quality score thereby simply reflected the number of high-quality sources selected (range: 0-4). The diversity score, in contrast, reflected whether participants only selected links suggesting a positive relationship between motivation and heart diseases, or whether they opted for a more balanced approach (balanced = two selected links suggest a positive and two links suggest a negative relationship between motivation and heart diseases; unbalanced = three of four selected links either suggest a positive or a negative relationship; highly unbalanced = either all four links suggest a positive relationship, or all four links suggest a negative relationship).

14

2.5   *Statistical Analyses*

Prior to the data analyses, pre- and posttest scores on the D-Index were, as outlined in the preregistration, standardized using their pre-intervention means and standard deviations. Subsequently, Hypothesis 1 was tested using latent difference score modeling (McArdle, 2009). Latent difference score models express within-subject changes of variables using constrained structural equation models (i.e., by fixing autoregressive paths and residual variances, see McArdle, 2009). In our specific case, we analyzed change on one observed variable (i.e., the D-Index) that was measured on two occasions (pre-intervention, post-intervention). Thus, one latent change score variable was built which represents pre-post changes in this observed variable. A major advantage of latent change score models is that they incorporate the full flexibility of structural equation modeling. For instance, effects of predictor variables on these latent change scores can be estimated. As independent variables for our model, we created indicator variables that specify whether an intervention component was applied or not (i.e., doubt and volition: $C = 0$; DV, R, RSI $= 1$; reflection: C, DV $= 0$; R, RSI $= 1$; social interaction: C, DV, R $= 0$; RSI $= 1$). For example, subjects in the RSI condition received all intervention components and, thus, all indicator variables are assigned the value 1 in this group. We analyzed the corresponding structural equation model using the R package lavaan (Rosseel, 2012). Hypothesized intervention effects were tested based on new parameters that were specified as (functions of) the effects of the abovementioned indicator variables (see Supplemental Material 1)[3]. The following parameters (that represent linear contrasts) were specified: H1a was tested by comparing the latent change in the C condition (M_C) to mean changes across the DV, R, and RSI conditions ($[M_{DV} + M_R + M_{RSI}]/3$

---

[3] We thank an anonymous reviewer for her or his valuable suggestions on improving our analysis during the revision stage.

– M_C). H1b was tested by comparing the latent change in the DV condition to mean changes across the R and RSI condition ($[M_R + M_{RSI}]/2 – M_{DV}$), and H1c was tested by comparing mean changes in the R and RSI condition ($M_R - M_{RSI}$). To compute these contrasts, we used the ":=" operator (Rosseel, 2021). This operator includes additional parameters in a model which are based on functions of already existing model parameters (i.e., effect estimates for our indicator variables in our case). Standard errors for these new parameters were estimated by means of the delta method (Rosseel, 2021). Since the resulting latent change score model is saturated, no model fit coefficients are reported.

To test our confirmatory hypotheses on source evaluation task performance, we conducted ordinal logistic regression analyses in R using the polr function of package mass. As a predictor, intervention group was included in the model. Positive regression coefficient estimates in this ordinal model indicate that the likelihood to score higher on an ordinal outcome variable (in our case, to select more high quality sources or more diverse sources) increases if the value of a predictor increases compared to the reference group (in our case, if a subject was assigned to a specific intervention compared to the control condition). Subsequently, analogous to our procedures for testing H1, we specified contrasts based on our preregistered hypotheses using the glht function of package multcomp (Hothorn et al., 2008).

2.6   *Manipulation Checks*

Detailed information on all analyses be found in the analysis markdown (see Supplemental Material 1). As a first manipulation check, we tested whether participants realized that the texts from the epistemic doubt intervention contained conflicting knowledge claims. This was done by administering the perceived contradictoriness measure by Kerwer and Rosman (2018; sample item: "Upon reading the texts... findings seemed to be very

contradictory"). A one-factorial ANOVA revealed significant differences between experimental conditions on this variable ($F_{3,149} = 6.648$; $p < .001$), and Tukey HSD post hoc tests revealed significantly higher scores in the DV ($M = 3.17$; $SD = 1.04$) and R ($M = 3.15$; $SD = 1.12$) conditions compared to the C ($M = 2.26$; $SD = 0.89$) condition (all $p < .01$), but no significant difference between the RSI and C conditions ($p = .263$). Except for the latter result, this is in line with our expectations formulated in the preregistration.

Moreover, to test if the epistemic volition intervention worked out as intended, we administered the state measure on epistemic volition by Kerwer et al. (2021; sample item: "I am feeling highly motivated to reconsider my current understanding of psychological knowledge"). In a one-factorial ANOVA, no differences between experimental conditions were found ($F_{3,149} = 1.545$; $p = .205$). However, an additional Welch's unequal variances t-test revealed higher scores in the three intervention groups (which were all subjected to the volition intervention) compared to the control group ($t_{df=62.616} = 2.029$; $p < .05$). Given that the corresponding ANOVA came out significant in Kerwer et al. (2021)—who used the exact same measure and intervention but a larger sample—the inconsistencies in our data might be explained by issues of statistical power.

Additionally, to test if the reflection intervention fostered a deeper processing of the conflicting evidence presented in the doubt intervention, we administered a self-report measure on deliberate integration efforts (sample item: "Upon reading the texts… I deliberately tried to figure out under which conditions contradictions between the short texts occurred.") in the R, RSI, and DV conditions. A one-factorial ANOVA revealed significant differences between these three groups ($F_{2,111} = 5.947$; $p < .05$), and Tukey HSD post hoc tests revealed significantly higher scores in the R condition ($M = 4.41$; $SD = 0.89$) compared to the DV condition ($M = 3.65$; $SD = 1.43$; $p < .05$), but no significant difference between the

RSI and DV conditions (p = .135). Again, this is in line with our expectations, except for the latter result.

Finally, we expected that participants in the RSI condition would perceive their interaction with the instructor during the reflection task to be more similar to a normal conversation compared to participants in the R condition. We therefore administered a corresponding measure (sample item: "While thinking aloud… interacting with the instructor felt like an ordinary conversation."). As expected, we found higher scores on this variable in RSI condition (M = 3.70; SD = 1.02) compared to the R condition (M = 3.17; SD = 1.00; $t_{df=71.959}$ = 2.239; p < .05) in a Welch's unequal variances t-test.

3   Results

An overview of the study results can be found in Figure 1 and Table 1. Study data is available in Supplemental Material 2, and the corresponding analysis code can be found in Supplemental Material 1. One-tailed tests were used where appropriate (see preregistration; Rosman & Kerwer, 2019). A one-factorial ANOVA revealed no significant group differences in topic-specific epistemic beliefs at pretest ($F_{3,147}$ = 1.976; p = .120)[4]. Moreover, there were no univariate outliers for topic-specific beliefs according to the preregistered criterion (z-scores with $p(z)$ < .001).

3.1 Hypothesis 1: Epistemic Change

Descriptively, the observed pattern of epistemic change was in line with our expectations (see Figure 1 and Table 1): Epistemic change was stronger in the RSI condition

---

[4] Since we were not able to match pre- and posttest data for two participants in our codeword-based matching procedure, the sample size for these calculations is N = 151.

compared to the R condition, the DV condition, and the C condition (see Table 2 for means of latent change scores and differences in mean change scores between conditions). However, postulated differences between the experimental conditions did not reach significance in all cases. More specifically, H1a was confirmed—epistemic change was significantly higher across the DV, RSI and R conditions compared to the C condition ($(M_{DV} + M_R + M_{RSI})/3 - M_C = .372$, $SE = .183$, $p = .021$; one-sided test). However, the contrast comparing the C and DV groups was not significant, while the contrasts comparing the R and RSI groups to the C group were significant (see Table 2). In addition, our tests on H1b ($(M_R + M_{RSI})/2 - M_{DV} = .257$, $SE = .192$, $p = .090$; one-sided test) and H1c were non-significant ($M_{RSI} - M_R = .081$, $SE = .227$, $p = .360$; one-sided test). Thus, H1b and H1c are not supported.

3.2 Hypothesis 2: Sourcing

Descriptive statistics on source evaluation task performance (i.e., on the diversity and quality indices) are reported in Tables 3 and 4. Descriptively, differences between groups in sourcing behaviors measured by the diversity index (see Table 3) were very small. Ordinal logistic regression analyses revealed no significant differences between groups and a very small Nagelkerke R^2 statistic of .007 (see Table 5). Moreover, all contrasts on H2a to H2c were non-significant (see Table 5). Thus, Hypothesis 2 is fully rejected with regard to the diversity index.

Regarding source evaluation task performance on the quality index, participants in the R condition chose the largest amount of high quality sources on a descriptive level (see Table 4). Despite that group differences were considerable larger for the quality index compared to the diversity index (Nagelkerke $R^2 = .068$), all contrasts for testing H2a to H2c were non-significant (see Table 5). Thus, Hypothesis 2 is also fully rejected with regard to the quality index. However, additional exploratory ordinal logistic regression analyses point towards

significantly higher quality index scores in the R condition compared to all other conditions
(see Supplemental Material 1).

4   Discussion

In the process of developing more advanced epistemic beliefs, there is a point where
individuals begin to doubt their current beliefs (Bendixen, 2002; Bendixen & Rule, 2004;
Rule & Bendixen, 2010). However, for this doubt to lead to epistemic change, certain
resolution strategies, which allow to resolve epistemic doubt by adopting more advanced
beliefs, are necessary. According to the Bendixen-Rule framework of epistemic change, two
such strategies are reflection and social interaction—individuals could for example reflect on
the implications of certain epistemic beliefs, or they could discuss their doubts with peers and
lecturers (Bendixen, 2002; Bendixen & Rule, 2004; Rule & Bendixen, 2010). Since both
strategies are still under-developed in the conceptual model by Bendixen and Rule, the
present study aimed at analyzing them in more detail. This was realized by developing
intervention components that specifically target the two strategies, and by analyzing these
components' effects on epistemic change that go beyond the effects of intervention
components targeting other process components of the model (i.e., epistemic doubt and
epistemic volition). Since epistemic beliefs and sourcing behavior are strongly associated, we
additionally tested the intervention components' effects on source evaluation performance.

In H1a, we expected that all intervention components would lead to stronger
epistemic change compared to a control intervention. This hypothesis was generally
supported. However, additional analyses yielded no significant differences between the
control group and the epistemic doubt and volition group. Given the lower-than-expected
sample size, this may have been caused by power issues, but nevertheless, the findings on the
two other components suggest that resolution strategies are important for epistemic change. It

should thereby be noted that participants in the doubt and volition condition might also have employed certain resolution strategies (i.e., reflecting on their beliefs without explicitly being asked to do so). However, if this was the case, the amount of evoked epistemic change was too small to show in our data—hence explicitly targeting resolution strategies during an epistemic belief intervention seems worthwhile. Another potential explanation for the non-significant results regarding doubt and volition is our manipulation check on epistemic volition. In fact, the wording of the corresponding items may reveal to participants that our study was about changing their understanding of psychological knowledge (cf. sample item in section 2.5). Since the manipulation check was administered in all four conditions, social desirability might thus have introduced bias on our dependent variables, leading to an underestimation of the differences between the four experimental conditions.

H1b posits that epistemic change is stronger in the two resolution strategies groups compared to the doubt and volition group, and H1c suggests that epistemic change is stronger in the group including both resolution strategies (i.e., reflection *and* social interaction) compared to reflection only. Unexpectedly, neither of these hypotheses were supported, which might again be caused by the above-mentioned power issues. Furthermore, as outlined above, the social interaction intervention consisted of a rather small change to the protocol of the reflection intervention: Students were not only prompted to think aloud, but they were actively engaged in a discussion about their beliefs by the instructor. Possibly, this may have been caused by the discussion increasing cognitive load, thus making it harder for participants to reflect on their beliefs. Another possible explanation for this unexpected finding is that the experimenter might have been considered as an authoritative figure (even though our experimenters were student assistants), thus hindering (instead of fostering) reflective processes. Our manipulation check on reflection lends support to this argument, since we did not find a significant difference in reflective processes between the social

interaction and the doubt and volition condition, but still found a corresponding difference between the reflection and the doubt and volition condition. To address this issue, future research might try to implement discussions between different participants instead of a discussion between the experimenter and the participant, but it should also be noted that such a manipulation would exhibit much less controllability. For example, participants might reinforce each other in their multiplistic views, which obviously is the opposite of what we want to achieve.

Hypothesis 2, which posited positive effects of the intervention on source evaluation, was not supported. However, additional analyses using contrasts revealed that the reflection condition is superior over social interaction, as well as over doubt and volition—at least with regard to the quality of selected sources. This further supports our argument that the social interaction component might have hampered reflective processes. Future research might try to design intervention components targeting reflection and social interaction separately, and investigate potential interactions between them. Furthermore, it should be pointed out that we did not find any effects on the diversity of the selected sources. Taken together, these findings suggest that reflecting over one's beliefs might affect the will or skill to assess the quality of specific sources, but does not seem to significantly impact the 'balancedness' of an information search.

As for the limitations of the present study, it should first be mentioned that we missed our target sample size criterion by around fifteen percent, and that the interpretation of our data is therefore hampered by issues of statistical power. Future research might try to invite psychology students from different universities to circumvent recruiting problems. However, we also would like to point out that standardized one-to-one sessions with specifically trained experimenters are comparatively costly, which certainly adds to the value of our research

considering that not many comparable studies exist. A second limitation is that the reliability of our primary outcome measure, the D-Index, was on the lower bound of what is considered acceptable. The reasons for this are somewhat unclear since the instrument worked well in prior studies, which used the exact same test administration (e.g., Kerwer & Rosman, 2020b). In addition, using the D-Index might have been problematic in itself since its formula gives more weight to evaluativism compared to absolutism and multiplism, and since different patterns of belief change may result in the same D-Index change. However, since our hypotheses posit changes in absolutism, multiplism, *and* evaluativism, assessing multidimensional changes towards advanced beliefs seems purposeful, especially since conducting three separate analyses (as we did in exploratory analyses; see Supplemental Material 1) considerably reduces statistical power when correcting for multiple testing. Third, our assumption of reflection as a prerequisite for social interaction led us to not including a group which only received the social interaction component. This was also for practical reasons since a moderated discussion on one's epistemic beliefs requires a certain amount of reflection on these beliefs beforehand. In a real-world scenario (as compared to our study context), reflection and social interaction might be more strongly intertwined, and, in addition, it is also conceivable that reflection is further fostered through interacting with one's peers. Finally, our manipulation checks might have led to participants figuring out what the study was about, thus lowering the differences between experimental conditions due to social desirability. This might be especially true for the manipulation check on epistemic volition (see above), but also for the check on epistemic doubt, which might have acted as a trigger and made intervention group participants become aware of conflicting information. Removing the manipulation checks in future studies might be an option to circumvent this issue.

In conclusion, our results show that especially reflection might be a promising strategy to address epistemic doubt. Psychology lecturers should thus give students room for reflecting on their beliefs once doubt on their current epistemic beliefs has arisen. For example, they might present their students with conflicting findings on a scientific issue and ask them to verbalize their thoughts on how the contradictions may have come about, or whether such a pattern of findings is in line with their beliefs on psychological knowledge. As for the implications for future research, the social interaction component warrants further investigation, as do other potential resolution strategies not considered in the present study. In addition, future research might strive to develop scales for epistemic doubt, epistemic volition, and for specific resolution strategies, possibly with our manipulation check measures as a basis. These variables could then be introduced into a mediation model in order to test whether reflection and social interaction (and other resolution strategies) mediate the relationship between epistemic doubt and epistemic change. This would allow a more nuanced view on the effects of and interactions between different resolution strategies, which is especially important for future theory building. Another aspect worthy of further investigation is a closer analysis of the social interaction component. For example, one might investigate individual variables (e.g., extraversion or agreeableness) that interact with the intervention, test the effects of different social interaction formats (e.g., moderated versus open discussion), or even test interactions between participant characteristics and intervention format (e.g., testing whether more extravert participants benefit more from an open discussion). Hence, as Kerwer et al. (2021) already point out, a lot of work on the Bendixen-Rule framework remains to be done—which we expect to be worth the effort due to the model's suitability to provide guidance on enhancing psychology learning and teaching.

References

Barzilai, S., & Strømsø, H. I. (2018). Individual differences in multiple document comprehension. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 99–116). Routledge.

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. *Cognition and Instruction*, *30*(1), 39–85. https://doi.org/10.1080/07370008.2011.636495

Bendixen, L. D. (2002). A process model of epistemic belief change. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 191–208). Lawrence Erlbaum Associates.

Bendixen, L. D., & Rule, D. C. (2004). An integrative approach to personal epistemology: A guiding model. *Educational Psychologist*, *39*(1), 69–80. https://doi.org/10.1207/s15326985ep3901_7

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, *46*(1), 48–70. https://doi.org/10.1080/00461520.2011.538647

Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2013). Justification beliefs and multiple-documents comprehension. *European Journal of Psychology of Education*, *28*(3), 879–902. https://doi.org/10.1007/s10212-012-0145-2

Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2014). Students working with multiple conflicting documents on a scientific issue: Relations between epistemic cognition while reading and sourcing and argumentation in essays. *British Journal of Educational Psychology*, *84*, 58–85. https://doi.org/10.1111/bjep.12005

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Epler, C., Drape, T., Broyles, T., & Rudd, R. (2013). The influence of collaborative reflection and think-aloud protocols on pre-service teachers' reflection: A mixed methods approach. *Journal of Agricultural Education*, *54*(1), 47–59. https://doi.org/10.5032/jae.2013.01047

Ericsson, K. A., & Simon, H. A. (1998). How to Study Thinking in Everyday Life: Contrasting Think-Aloud Protocols With Descriptions and Explanations of Thinking. *Mind, Culture, and Activity*, *5*(3), 178–186. https://doi.org/10.1207/s15327884mca0503_3

Ferguson, L. E., Bråten, I., & Strømsø, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learning and Instruction*, *22*(2), 103–120. https://doi.org/10.1016/j.learninstruc.2011.08.002

Greene, J. A., Cartiff, B. M., & Duke, R. F. (2018). A meta-analytic review of the relationship between epistemic cognition and academic achievement. *Journal of Educational Psychology*, *110*(8), 1084–1111. https://doi.org/10.1037/edu0000263

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Kardash, C. M., & Howell, K. L. (2000). Effects of epistemological beliefs and topic-specific beliefs on undergraduates' cognitive and strategic processing of dual-positional text. *Journal of Educational Psychology*, *92*(3), 524–535. https://doi.org/10.1037//0022-0663.92.3.524

Kerwer, M., & Rosman, T. (2018). Mechanisms of epistemic change: Under which circumstances does diverging information support epistemic development? *Frontiers in Psychology*, *9*(2278). https://doi.org/10.3389/fpsyg.2018.02278

Kerwer, M., & Rosman, T. (2020a). Epistemic change and diverging information: How do prior epistemic beliefs affect the efficacy of short-term interventions? *Learning and Individual Differences*, *80*, 101886. https://doi.org/10.1016/j.lindif.2020.101886

Kerwer, M., & Rosman, T. (2020b). Epistemic change and diverging information: How do prior epistemic beliefs affect the efficacy of short-term interventions? *Learning and Individual Differences*, *80*, 101886. https://doi.org/10.1016/j.lindif.2020.101886

Kerwer, M., Rosman, T., Wedderhoff, O., & Chasiotis, A. (2021). Disentangling the process of epistemic change: The role of epistemic volition. *British Journal of Educational Psychology*, *91*(1), 1–26. https://doi.org/10.1111/bjep.12372

Krettenauer, T. (2005). Die Erfassung des Entwicklungsniveaus epistemologischer Überzeugungen und das Problem der Übertragbarkeit von Interviewverfahren in standardisierte Fragebogenmethoden [Measuring the developmental level of epistemological beliefs and the problem of transfering interview procedures to standardized questionnaire methods]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *37*(2), 69–79. https://doi.org/10.1026/0049-8637.37.2.69

Kuhn, D., & Weinstock, M. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 121–144). Lawrence Erlbaum Associates.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605. https://doi.org/10.1146/annurev.psych.60.110707.163612

Muis, K. R., & Duffy, M. C. (2013). Epistemic climate and epistemic change: Instruction designed to change students' beliefs and learning strategies and improve achievement. *Journal of Educational Psychology*, *105*(1), 213–225. https://doi.org/10.1037/a0029690

Muis, K. R., Trevors, G. J., & Chevrier, M. (2016). Epistemic climate for epistemic change. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 331–359). Routledge.

Rosman, T., Mayer, A.-K., Merk, S., & Kerwer, M. (2019). On the benefits of 'doing science': Does integrative writing about scientific controversies foster epistemic beliefs? *Contemporary Educational Psychology*, *58*, 85–101. https://doi.org/10.1016/j.cedpsych.2019.02.007

Rosman, T., Peter, J., Mayer, A.-K., & Krampen, G. (2018). Conceptions of scientific knowledge influence learning of academic skills: Epistemic beliefs and the efficacy of information literacy instruction. *Studies in Higher Education*, *43*(1), 96–113. https://doi.org/10.1080/03075079.2016.1156666

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y. (2021). The lavaan tutorial. http://lavaan.ugent.be/tutorial/tutorial.pdf

Rule, D. C., & Bendixen, L. D. (2010). The integrative model of personal epistemology development: Theoretical underpinnings and implications for education. In L. D. Bendixen & F. C. Feucht (Eds.), *Personal epistemology in the classroom: Theory, research, and implications for practice* (pp. 94–123). Cambridge University Press.

Tables

Table 1

Descriptive statistics on epistemic beliefs by intervention group

	Control		Volition and Doubt		Reflection		Social Interaction	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Absolutism (pre)	2.91	0.64	2.67	0.81	2.54	0.68	2.70	0.73
Multiplism (pre)	3.25	0.63	3.18	0.73	3.07	0.77	3.09	0.73
Evaluativism (pre)	4.77	0.50	4.83	0.58	4.77	0.56	4.99	0.43
D-Index (pre)	1.69	0.70	1.90	0.79	1.96	0.80	2.09	0.66
Absolutism (post)	2.54	0.66	2.21	0.82	1.86	0.69	1.99	0.77
Multiplism (post)	3.24	0.92	3.28	0.75	3.25	0.82	3.01	0.82
Evaluativism (post)	4.92	0.53	5.03	0.54	5.04	0.55	5.12	0.51
D-Index (post)	2.04	0.79	2.29	0.88	2.48	0.90	2.62	0.76

Note. M = arithmetic mean; *SD* = standard deviation; pre = pretest; post = posttest.

Table 2

Means of latent change scores by intervention group, mean differences of latent change

scores and hypothesis tests on H1

Label	*EST*	*SE*	*p*	CI_{lower}	CI_{upper}
M_C	0.325	0.157	.039	0.017	0.633
M_{DV}	0.525	0.154	.001	0.224	0.827
M_R	0.742	0.160	> .001	0.428	1.056
M_{RSI}	0.823	0.161	> .001	0.508	1.138
test H1a	0.372	0.183	.042	0.013	0.730
test H1b	0.257	0.192	.180	-0.119	0.633
test H1c	0.081	0.227	.720	-0.363	0.526
C vs. DV	0.200	0.220	.362	-0.230	0.631
C vs. R	0.417	0.225	.064	-0.024	0.857
C vs. SI	0.498	0.227	.028	0.053	0.943
DV vs. R	0.216	0.222	.330	-0.219	0.652
DV vs. RSI	0.298	0.223	.182	-0.140	0.735
R vs. RSI	0.081	0.227	.720	-0.363	0.526

Note. EST = parameter value for latent change score means/group differences in

means/contrasts; *CI* = 95% confidence interval; *SE* = standard error; C = control; DV = doubt

and volition; R = reflection; RSI = reflection and social interaction; contrasts: H1a = (DV + R

+ RSI) / 3 – C, H1b = (R + RSI) / 2 - DV, H1c = RSI - R; all *p*-values are based on two-sided

tests.

Table 3

Descriptive statistics on the diversity score of the source evaluation task by intervention group.

Diversity Score	Control		Volition and Doubt		Reflection		Social Interaction	
	n	%	*n*	%	*n*	%	*n*	%
Highly Unbalanced	3	7.69	5	12.50	2	5.56	1	2.70
Slightly Unbalanced	18	46.15	19	47.50	20	55.56	24	64.86
Balanced	18	46.15	16	40.00	14	38.89	12	32.43

Table 4

Descriptive statistics on the quality score of the source evaluation task by intervention group.

Number of High Quality Sources Chosen	Control		Volition and Doubt		Reflection		Social Interaction	
	n	%	*n*	%	*n*	%	*n*	%
0	0	0.00	0	0.00	0	0.00	0	0.00
1	0	0.00	2	5.00	0	0.00	0	0.00
2	3	7.69	2	5.00	0	0.00	1	2.70
3	4	10.26	3	7.50	1	2.78	6	16.22
4	32	82.05	33	82.50	35	97.22	30	81.08

Table 5

Ordinal logistic regression of diversity and quality scores on intervention group.

Diversity Score				Quality Score			
Predictor	*EST*	*CI*	*OR*	Predictor	*EST*	*CI*	*OR*
DV	-0.339	[-1.224, 0.538]	0.71	DV	-0.020	[-1.191, 1.151]	0.98
R	-0.212	[-1.103, 0.673]	0.81	R	2.050	[0.251, 5.006]	7.77
RSI	-0.367	[-1.247, 0.503]	0.69	RSI	-0.006	[-1.176, 1.164]	0.99
Nagelkerke R^2	.007			Nagelkerke R^2	0.068		

Contrast	*EST*	*SE*	*p*	Contrast	*EST*	*SE*	*p*
test H2a	-0.306	0.365	.402	test H2a	0.675	0.570	.237
test H2b	0.049	0.389	.900	test H2b	1.042	0.688	.130
test H2c	-0.155	0.450	.731	test H2c	-2.056	1.096	.061

Note: EST = regression weight; *CI* = 95% confidence interval; *OR* = odds ratio, *SE* = standard error; C = control; DV = doubt and volition; R = reflection; RSI = reflection and social interaction; reference category is the control condition; contrasts: H2a = (DV + R + RSI) / 3 – C, H2b = (R + RSI) / 2 - DV, H2c = RSI - R; all p-values are based on two-sided tests.
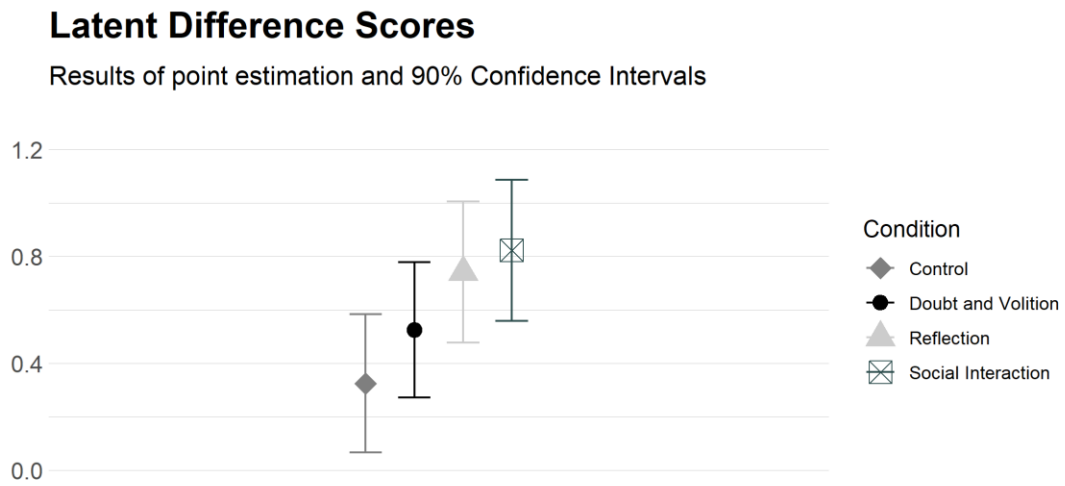
Figures

Latent Difference Scores
Results of point estimation and 90% Confidence Intervals

Figure 1. Changes in epistemic beliefs from pre- to posttest. Positive values indicate an

increase on the D-Index.

References

Barzilai, S., & Strømsø, H. I. (2018). Individual differences in multiple document comprehension. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 99–116). Routledge.

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. *Cognition and Instruction*, *30*(1), 39–85. https://doi.org/10.1080/07370008.2011.636495

Bendixen, L. D. (2002). A process model of epistemic belief change. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 191–208). Lawrence Erlbaum Associates.

Bendixen, L. D., & Rule, D. C. (2004). An integrative approach to personal epistemology: A guiding model. *Educational Psychologist*, *39*(1), 69–80. https://doi.org/10.1207/s15326985ep3901_7

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, *46*(1), 48–70. https://doi.org/10.1080/00461520.2011.538647

Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2013). Justification beliefs and multiple-documents comprehension. *European Journal of Psychology of Education*, *28*(3), 879–902. https://doi.org/10.1007/s10212-012-0145-2

Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2014). Students working with multiple conflicting documents on a scientific issue: Relations between epistemic cognition while reading and sourcing and argumentation in essays. *British Journal of Educational Psychology*, *84*, 58–85. https://doi.org/10.1111/bjep.12005

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory

    research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

    https://doi.org/10.1016/S0022-5371(72)80001-X

Epler, C., Drape, T., Broyles, T., & Rudd, R. (2013). The influence of collaborative reflection

    and think-aloud protocols on pre-service teachers' reflection: A mixed methods

    approach. *Journal of Agricultural Education*, *54*(1), 47–59.

    https://doi.org/10.5032/jae.2013.01047

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

    MIT Press.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting

    think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture,*

    and Activity, *5*(3), 178–186. https://doi.org/10.1207/s15327884mca0503_3

Ferguson, L. E., Bråten, I., & Strømsø, H. I. (2012). Epistemic cognition when students read

    multiple documents containing conflicting scientific evidence: A think-aloud study.

    Learning and Instruction, *22*(2), 103–120.

    https://doi.org/10.1016/j.learninstruc.2011.08.002

Greene, J. A., Cartiff, B. M., & Duke, R. F. (2018). A meta-analytic review of the

    relationship between epistemic cognition and academic achievement. *Journal of*

    Educational Psychology, *110*(8), 1084–1111. https://doi.org/10.1037/edu0000263

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric

    models. *Biometrical Journal*, *50*(3), 346–363.

    https://doi.org/10.1002/bimj.200810425

Kardash, C. M., & Howell, K. L. (2000). Effects of epistemological beliefs and topic-specific

    beliefs on undergraduates' cognitive and strategic processing of dual-positional text.

Journal of Educational Psychology, *92*(3), 524–535. https://doi.org/10.1037//0022-0663.92.3.524

Kerwer, M., & Rosman, T. (2018). Mechanisms of epistemic change: Under which circumstances does diverging information support epistemic development? *Frontiers in Psychology*, *9*(2278). https://doi.org/10.3389/fpsyg.2018.02278

Kerwer, M., & Rosman, T. (2020a). Epistemic change and diverging information: How do prior epistemic beliefs affect the efficacy of short-term interventions? *Learning and Individual Differences*, *80*, 101886. https://doi.org/10.1016/j.lindif.2020.101886

Kerwer, M., & Rosman, T. (2020b). Epistemic change and diverging information: How do prior epistemic beliefs affect the efficacy of short-term interventions? *Learning and Individual Differences*, *80*, 101886. https://doi.org/10.1016/j.lindif.2020.101886

Kerwer, M., Rosman, T., Wedderhoff, O., & Chasiotis, A. (2021). Disentangling the process of epistemic change: The role of epistemic volition. *British Journal of Educational Psychology*, *91*(1), 1–26. https://doi.org/10.1111/bjep.12372

Krettenauer, T. (2005). Die Erfassung des Entwicklungsniveaus epistemologischer Überzeugungen und das Problem der Übertragbarkeit von Interviewverfahren in standardisierte Fragebogenmethoden [Measuring the developmental level of epistemological beliefs and the problem of transfering interview procedures to standardized questionnaire methods]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *37*(2), 69–79. https://doi.org/10.1026/0049-8637.37.2.69

Kuhn, D., & Weinstock, M. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 121–144). Lawrence Erlbaum Associates.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605. https://doi.org/10.1146/annurev.psych.60.110707.163612

Muis, K. R., & Duffy, M. C. (2013). Epistemic climate and epistemic change: Instruction designed to change students' beliefs and learning strategies and improve achievement. *Journal of Educational Psychology*, *105*(1), 213–225. https://doi.org/10.1037/a0029690

Muis, K. R., Trevors, G. J., & Chevrier, M. (2016). Epistemic climate for epistemic change. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 331–359). Routledge.

Rosman, T., & Kerwer, M. (2019). Preregistration: Mechanisms of Epistemic Change – The Role of Reflection and Social Interaction. PsychArchives. https://doi.org/10.23668/psycharchives.2621.

Rosman, T., Mayer, A.-K., Merk, S., & Kerwer, M. (2019). On the benefits of 'doing science': Does integrative writing about scientific controversies foster epistemic beliefs? *Contemporary Educational Psychology*, *58*, 85–101. https://doi.org/10.1016/j.cedpsych.2019.02.007

Rosman, T., Peter, J., Mayer, A.-K., & Krampen, G. (2018). Conceptions of scientific knowledge influence learning of academic skills: Epistemic beliefs and the efficacy of information literacy instruction. *Studies in Higher Education*, *43*(1), 96–113. https://doi.org/10.1080/03075079.2016.1156666

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y. (2021). *The lavaan tutorial*. http://lavaan.ugent.be/tutorial/tutorial.pdf

Rule, D. C., & Bendixen, L. D. (2010). The integrative model of personal epistemology development: Theoretical underpinnings and implications for education. In L. D. Bendixen & F. C. Feucht (Eds.), *Personal epistemology in the classroom: Theory, research, and implications for practice* (pp. 94–123). Cambridge University Press.