



Scientific Controversies and Epistemological Sensitization - Effects of an Intervention on Psychology Students' Epistemological Beliefs and Argumentation Skills

OPEN ACCESS

Eric Klopp* and Robin Stark

Department of Education, Saarland University, Saarbrücken, Germany

Edited by:

Michael Weinstock,
Ben-Gurion University of the Negev,
Israel

Reviewed by:

Dimitrios Stamovlasis,
Aristotle University of Thessaloniki,
Greece
Hui Luan,
National Taiwan Normal University,
Taiwan

*Correspondence:

Eric Klopp
e.klopp@mx.uni-saarland.de

Specialty section:

This article was submitted to
Higher Education,
a section of the journal
Frontiers in Education

Received: 28 September 2021

Accepted: 30 December 2021

Published: 01 February 2022

Citation:

Klopp E and Stark R (2022) Scientific Controversies and Epistemological Sensitization - Effects of an Intervention on Psychology Students' Epistemological Beliefs and Argumentation Skills. *Front. Educ.* 6:785241. doi: 10.3389/feduc.2021.785241

The present study investigates the effects of an intervention presenting resolvable, scientific controversies and an epistemological sensitization measure on the changes in psychology students' epistemological beliefs. Drawing on the notion that the presentation of resolvable scientific controversies induces epistemological doubt and the notion that inducing epistemological doubt is eased in the presence of an epistemological sensitization, we used an epistemological beliefs intervention consisting of five resolvable controversies that were applied in a sample consisting of psychology students. We hypothesized that the intervention would reduce absolutist and multiplist epistemological beliefs while, at the same time, increasing evaluativist beliefs. We also assumed that the epistemological sensitization would enhance the effect of the intervention. For a domain-specific questionnaire, the results indicated a reduction of absolutist epistemological beliefs regardless of the presence of the epistemological sensitization. Unexpectedly, there was a backfire effect indicated by a rise of multiplist beliefs. For a domain- and topic-specific questionnaire, there was no significant reduction of absolutist and multiplist beliefs but a significant increase in evaluativist beliefs when the epistemological sensitization was present. A measure assessing argumentation skills revealed an increase in argumentation skills only when the epistemological sensitization is present. Finally, we discuss limitations, educational implications, and directions for future research.

Keywords: epistemological beliefs, sensitization, resolvable controversies, epistemological change, argumentation

1 INTRODUCTION

The ability to develop a scientific style of argumentation is an essential part of the field of social sciences and is also deeply related to scientific thinking itself (Fischer et al., 2014). Various authors have demonstrated the importance of epistemological beliefs¹, i.e., beliefs about the nature of knowledge and the process of knowledge acquisition, for proper scientific argumentation (cf., Iordanou et al., 2016). Especially the work of Kuhn (1991), Kuhn (2001) demonstrated that evaluativist epistemological beliefs, i.e., the belief that knowledge is based on weighted evidence, are a prerequisite for an advanced level of scientific argumentation. Therefore, it is essential for students to develop evaluativist epistemological beliefs. Because scientific argumentation skills are a centerpiece of scientific competencies in domains like psychology (Fischer et al., 2014; Dietrich et al., 2015), it seems therefore necessary to foster psychology students' evaluativist epistemological beliefs systematically.

However, psychology has an "ill-defined" knowledge structure (cf., Rosman et al., 2017). This ill-defined knowledge structure consists in seemingly inconsistent theories, definitions, paradigms, and empirical results. Usually, epistemological beliefs develop during the enculturation within a certain domain and its respective scientific community (Muis et al., 2006; Palmer and Marra, 2008; Klopp and Stark, 2016). However, such an enculturation process may also be problematic as it is the case of psychology and its ill-defined knowledge structure. Rosman et al. (2017) investigated the development of psychology students' epistemological beliefs during the first four semesters. Their results indicate an overall high level of multiplist epistemological beliefs, i.e., the belief that knowledge is arbitrary and each account is true in its own right, in psychology students. Especially, these authors observed an increase in multiplist epistemological beliefs during the first semester. The enculturation in the domain of psychology with its inconsistent theories, definitions, paradigms, and empirical results leads to the development of multiplicitist epistemological

beliefs that are not favorable for argumentation skills (Kuhn, 2001). Therefore, there is a need for interventions that counteract the development of multiplicitist beliefs and that foster the development of evaluativist beliefs.

This article presents the evaluation of a computer-based, adapted version of the epistemological belief intervention developed by Rosman et al. (2016) to foster evaluativist epistemological beliefs in psychology students. Along with this, we examined the effects of an epistemological sensitization measure in the context of an epistemological beliefs intervention aiming at psychology students' epistemological beliefs.

1.1 Approaches to Epistemological Beliefs

There are two different approaches in the research on epistemological beliefs (cf., Hofer and Bendixen, 2012): the developmental approach as well as the dimensional approach. In the following, we first provide a short description of each approach and, afterward, describe how the developmental and dimensional approach can be integrated in a common framework.

In the developmental approach, the trajectory of the development of epistemological beliefs is characterized as a sequence of qualitatively different levels. One of the prominent models in this tradition is (Kuhn, 1991) model describing three levels in the development of epistemological beliefs: absolutism, multiplism, and evaluativism. The stage theory of epistemological development, which dates back to Perry (1970), draws on the notion of developmental stages. The notion of developmental stages is characterized by the following characteristics (cf., Hayslip et al., 2006): Firstly, there is the assumption that these stages are qualitatively different levels of epistemological development. These stages are discontinuous, i.e., that achieving a certain stage means a move from the former stage to the following stage. Secondly, achieving a lower stage is a prerequisite for reaching the following stage and it is normatively assumed that the following stages are superior to the former stages. Thirdly, the sequence of the stages is universal. Weinstock (2006) characterizes the epistemological beliefs on each level as follows: On the first level, the absolutist level, there is only one correct account of knowledge. Other accounts of knowledge fail because they are the result of erroneous or biased thinking. On the second level, the multiplist level, there are many possible accounts of knowledge, either of which is seen as correct. The different accounts of knowledge are, therefore, purely subjective and the justification of knowledge claims refers to assertions of opinions. On the third level, the evaluativist level, accounts of knowledge are constructed on evidence. The evidence for one account is weighted against evidence for the other accounts and a reasoned decision for one account is made. The justification of knowledge claims on the evaluativist level refers to the evaluation of various claims. Evaluativist epistemological beliefs are normatively the most sophisticated level of epistemological beliefs.

In the dimensional approach, epistemological beliefs are conceptualized as dimensions of interindividual differences (Hammer and Elby, 2002) referring to subjective assumptions

¹In this study, we use the term epistemological beliefs despite the fact that the term epistemic beliefs is widely used (e.g., Rosman et al., 2016). The term epistemological relates to the theory of knowledge, especially with regard to its methods, validity, and scope (Lexico, 2021a), whereas the term epistemic relates knowledge or to the degree of its validation (Lexico, 2021b). In consequence, epistemological beliefs characterize a personal theory of knowledge and the process of knowledge acquisition (cf., Hofer and Pintrich, 1997). In contrast, the term epistemic should be used in relation to (cognitive) processes concerning the validation of knowledge, e.g., epistemic strategies (cf., Richter and Schmid, 2010, for an example in this respect). We also use the term beliefs to refer to an individual's association of the characteristics or attributes of knowledge and the process of knowledge acquisition as an umbrella term for the various concepts mentioned hereafter, e.g., epistemological understanding (Kuhn, 2001), epistemological position (Weinstock, 2006), or epistemic thinking (Barzilai and Weinstock, 2015). With Hammer and Elby (2002, pp-171–173), we conceive all these concepts as representations of interindividual differences in epistemological beliefs that are measured using questionnaires.

TABLE 1 | Profiles of the nine epistemological belief dimensions for each developmental stage [adapted from Barzilai and Weinstock (2015)].

Dimension	Absolutism	Multiplicism	Evaluatism
Source of knowledge	Perceived from outside reality	Constructed by human mind	Constructed by human mind but according to standards of knowledge construction
Nature of knowledge	Facts	Personal opinions	Theories and arguments
Structure of knowledge	Simple facts that are true	Equally right personal opinions	Multiple accounts that can be judged about their degree of truth
Role of multiple perspectives	Multiple perspectives are an obstacle for knowledge	Multiple perspective are an obstacle for knowledge that cannot be mastered	Considering multiple perspectives can improve knowledge generation
Justification of knowing	Draws on reality	Draws on personal preferences and judgments	Coordination of theory and evidence, drawing on shared norms and standards
Reliable explanation	Based on facts	Based on personal knowledge and experience	Based on theory and the available information
Certainty of knowledge	Knowledge is certain and certainty is an achievable goal	Knowledge is fundamentally uncertain	Knowledge is fundamentally uncertain but it is possible to improve the degree of certainty
Attainability of truth	Truth is attainable	Truth is unattainable	Approximate truth is attainable by argumentation
Expertise	Expert knowledge is certain	Experts can differ in their opinion	Consideration and evaluation of multiple expert opinions

about knowledge and the process of knowledge acquisition (Hofer and Pintrich, 1997). This approach dates back to the work of Schommer (1990) and a prominent model therein is the one of Hofer and Pintrich (1997). These authors describe the four dimensions Certainty of knowledge, Simplicity of knowledge, Source of knowledge, and Justification of knowledge.

Recently, the developmental and the dimensional approach were integrated (Weinstock, 2006; Greene et al., 2008; Barzilai and Weinstock, 2015) into one framework. This integration means that these two approaches are not different but rather two sides of the same coin. According to Weinstock (2006), the developmental levels of epistemological beliefs may be characterized as a certain profile of the various dimensions of epistemological beliefs. A further attempt to merge the developmental and the dimensional approach was presented by Barzilai and Weinstock (2015). This attempt consists in describing a certain level of epistemological development using a profile of a given set of beliefs. After reviewing the current theoretical and empirical research, they describe the three developmental levels of epistemological beliefs as profiles of the following nine dimensions: Certainty of knowledge, Source of knowledge, Nature of knowledge, Structure of knowledge, The role of multiple perspectives, Justification for knowing, Reliable explanation, Attainability of truth, and Expertise. **Table 1** (adapted from Barzilai and Weinstock, 2015) provides an overview of how a specific profile of these dimensions characterizes the levels of epistemological development. Schommer-Aikins (2002) describes epistemological development as an asynchronous change in individual profiles of epistemological belief dimensions.

In this way, the developmental and the dimensional approach are not two distinct theories of epistemological beliefs, instead, they are different perspectives on the same phenomenon. The integrated approach offers a new view of the developmental approach. Whereas the classical stage theory assumes discontinuous, qualitatively different stages, the conceptualization of the stages as a combination of dimensional profiles understands development as a continuous process. Instead of being qualitatively different stages that have abrupt transitions, developmental stages are different quantitative

profiles with steady transitions. Especially, the assessment of stages of epistemological development differs. Whereas in the classical developmental approach, an individual is classified as either being an absolutist, multiplist, or evaluativist (e.g., Kuhn et al., 2000), the integrated approach enables to assess absolutism, multiplism, and evaluativism simultaneously depending on the individual degree of each dimension (e.g., Barzilai and Weinstock, 2015; Peter et al., 2016).

In this paper, we focus on the developmental stages from the perspective of the integrated approach because the main topic of this study is the development of epistemological beliefs that are beneficial for scientific argumentation. As shown below, there is a strong relation between the stages of epistemological development and the research on argumentation skills. In the following section we discuss the importance of epistemological beliefs for psychology, report empirical results regarding the development of psychology students' epistemological beliefs and demonstrate why especially evaluativist epistemological beliefs are beneficial for psychology students.

1.2 Epistemological Beliefs in Psychology

In a recent review, Green and Hood (2013) highlighted the importance of epistemological beliefs for psychology. The authors summarize findings indicating strong relations of epistemological beliefs with learning approaches and academic achievement. Adequate epistemological beliefs are an important precursor of argumentation competencies (Hefter et al., 2015) and (Kuhn, 2001) states that evaluativist epistemological beliefs are one of the main prerequisites for properly applying argumentation strategies. Mason and Scirica (2006) report that evaluativist epistemological beliefs are a significant predictor for student's argumentation skills. When arguing about a controversial topic, participants with evaluativist epistemological beliefs generate arguments, counterarguments, and rebuttals of a higher quality than participants with lower levels of epistemological beliefs. Moreover, the authors report that topic knowledge was a significant moderator in the production of rebuttals.

As stated in their Theory of Integrated Domains in Epistemology (TIDE), Muis et al. (2006) argue that

epistemological beliefs have both domain-specific and domain-general components. In a recent extension to the TIDE, Merk et al. (2018) argue that epistemological beliefs cover a range from domain-general beliefs to domain-specific beliefs to topic-specific beliefs. Thus, there may be psychology-specific epistemological beliefs and topic-specific epistemological beliefs (e.g., regarding a specific learning theory) at the same time. Besides the relevance of domain-specific epistemological beliefs, Merk et al. (2018) provide evidence that topic-specificity plays a significant role in the evaluation of psychological topics, too. Thus, the development of domain-specific epistemological beliefs of psychology students during their course of study is of utmost interest. At the same time, topic-specific epistemological beliefs must not be neglected.

Rosman et al. (2017) investigated the development of domain-specific epistemological beliefs of psychology students during the first four semesters of their course of studies. In a longitudinal design, epistemological beliefs were measured by means of the Epistemological Beliefs Inventory—Absolutism, Multiplicism (EBI-AM; Peter et al., 2016), a questionnaire that assesses psychology-specific absolutist and multiplist epistemological beliefs. Epistemological beliefs were measured at the beginning of the first, second, third, and fourth semester. The authors used growth curve models to investigate epistemological development over time. Their results indicate a relatively low level of absolutism that, moreover, does not change over time. Regarding multiplism, their results indicate that a quadratic growth model in combination with a cubic trend is the best fitting model. There is a raise of multiplism in the first semester that afterward decreases slightly below the initial level in the second semester and slightly raises again in the third semester. Moreover, the level of multiplism is generally higher than the level of absolutism at all measurement points. Regarding multiplism, Rosman et al. (2017) attribute their results to the structure of psychological knowledge. Psychological knowledge has an “ill-defined knowledge structure” that finds its expression in numerous inconsistent theories, definitions, paradigms, and empirical results (cf., Muis et al., 2006). This ill-defined knowledge structure is a challenge for psychology students: Beginners are not able to cope with these inconsistencies. This inability, in turn, yields the beginners to acquire multiplist epistemological beliefs. Ideally, psychology students would develop skills to cope with these inconsistencies later on; e.g., they acquire methodological skills and information literacy and learn to evaluate and weight evidence. Consequently, with increasing skills to cope with these inconsistencies, multiplism declines and evaluativism increases. A shortcoming of Rosman et al. (2017) study is that the instrument they used only allows assessing absolutism and multiplism and, consequently, there is no direct evidence for the development of evaluativism. Despite these shortcomings, the results indicate that psychology students show an overall high level of multiplism in the first three semesters.

Rosman et al.’s results and their interpretation in terms of the structure of psychological knowledge provide insights and valuable reasons why evaluativist beliefs are adequate, i.e., are satisfactory in their quality, to cope with the knowledge structure

found in psychology². According to the description of evaluativism by Barzilai and Weinstock (2015; **Table 1**), especially the characteristic profile of the dimensions Nature of knowledge, The role of multiple perspectives, Structure of knowledge, and Justification for knowing explains this fact. Evaluativists acknowledge that knowledge is constructed using theories, arguments and interpretations. They believe that taking multiple perspectives into account can improve knowledge construction and that there are multiple right accounts, some being more right than others. They also believe that the coordination of theory and evidence and the use of shared norms and standards are necessary for appropriate knowledge construction. Thus, only evaluativist can handle the ill-defined knowledge structure in psychology. Solely, evaluativist epistemological beliefs allow for coping with inconsistent theories and empirical results, as only evaluativists consider multiple accounts to be dependent on the respective context.

Especially, domain-specific epistemological beliefs seem to be important for the application of advanced argumentation strategies (Iordanou et al., 2016). A recent study by Klopp and Stark (2020) revealed that undergraduate psychology students have rather low scientific argumentation skills (cf., Astleitner et al., 2003). As summarized by Klopp and Stark (2020), students make typical argumentation errors, some of which also relate to epistemological issues, e.g., not giving any justification or evidence for a claim or deficits in the acknowledgment of different perspectives on a certain topic (cf., Fischer et al., 2014). Sadler (2004) found that students cannot recognize contrasting arguments. In general, the work of Kuhn (2001) has shown that evaluativist epistemological beliefs are beneficial for advanced argumentation skills. In light of Rosman et al. (2017) analysis of the structure of psychological knowledge, it is obvious that evaluativist beliefs are also beneficial for scientific argumentation in psychology. Only on the evaluativist level, persons acknowledge the argumentative knowledge-building process that, in turn, enables the construction of valid arguments. Additionally, this perspective also sheds light on the typical erroneous arguments reported above. For instance, on the absolutist level, there is no need to justify a claim because only one account is true and thus, there is no need to provide a justification. Additionally, absolutism may hinder to acknowledge different perspectives. Analogous considerations apply on the multiplist level: there are many possible accounts of knowledge, either of which is correct and consequently, there is no need to provide evidence. Thus,

²In this paper, we consciously opted not to use the common term sophisticated epistemological beliefs. Elby and Hammer (2001) criticized the normative concept of sophisticated epistemological beliefs. Epistemological beliefs that are commonly termed sophisticated may either not be productive concerning a specific learning goal or may not be correct with regard to the body of knowledge considered by a scientific community. Given this critique, we therefore use the term adequate, i.e., something is satisfactory in quantity and quality (cf., Lexico, 2021c). With regard to evaluativist epistemological beliefs, as mentioned in the main text, they are satisfactory to cope with the knowledge structure in psychology. However, evaluativist beliefs may not always be adequate, please see our consideration in the last paragraph of the Limitations section.

epistemological beliefs and argumentation skills are interrelated. On the one side, epistemological beliefs are predictors of argumentation skills. For instance, Mason and Scirica (2006) found that epistemological beliefs predicted the generation of arguments, counterarguments, and rebuttals when arguing about controversial issues. Moreover, these authors found that persons on the evaluativist level generated arguments of higher quality than those on the multiplist level. On the other side, argumentation may also help to foster epistemological development (Iordanou et al., 2016, for an overview). For instance, Iordanou (2016) found that engaging in argumentation on socio-scientific issues fosters evaluativist epistemological beliefs. Thus, epistemological beliefs and argumentation are twisted and it seems that epistemological beliefs are important prerequisites for the acquisition of scientific argumentation skills.

In summary, Rosman et al. (2017) results are not merely describing psychology students' development of epistemological beliefs. In combination with the intrinsic relation of epistemological beliefs and argumentation skills, they also have important instructional consequences and consequences for the development of scientific competences like argumentation skills. Concerning these findings, it seems indicated to foster evaluativist beliefs of psychology students, while at the same time, measures should be implemented to reduce absolutism and multiplism. From an instructional perspective, a steady decline of multiplism is desirable, regardless of the actual form of the developmental trajectory. In an ideal case, evaluativist beliefs are fostered before multiplist beliefs are developed during the initial enculturation in psychology to prevent the observed raise of multiplism in the first semester or, respectively, lower the overall high level of multiplism that corresponds to the decline multiplism mentioned earlier. In the next section, we describe a process model of epistemological belief change that is used as a theoretical framework to develop strategies to foster advanced psychology-specific epistemological beliefs.

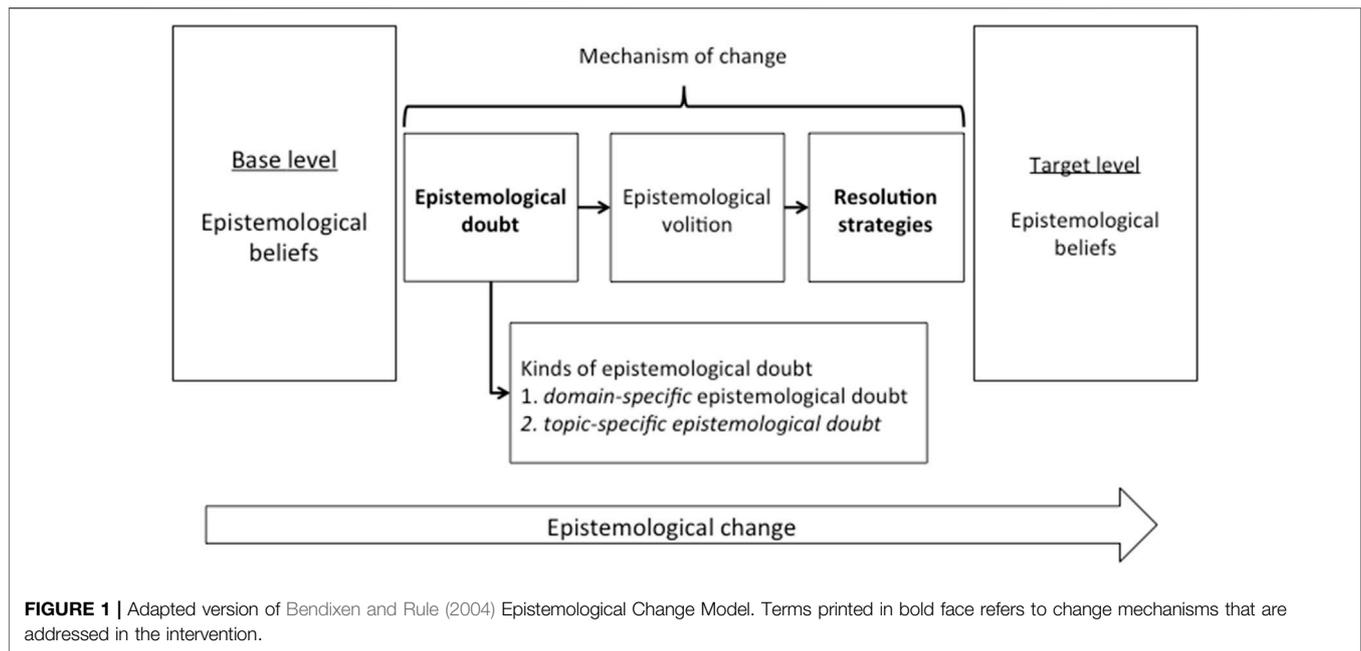
1.3 A Process Model of Epistemological Change and an Intervention

A model describing changes in epistemological beliefs is the Process Model of Personal Epistemology Development (Bendixen, 2002; Bendixen and Rule, 2004). It describes the necessary mechanisms to induce changes from current to more adequate epistemological beliefs. These mechanisms are epistemological doubt, epistemological volition, and resolution strategies (Figure 1). Epistemological doubt refers to questioning one's own epistemological beliefs. For epistemological doubt to arise, an individual has to recognize a dissonance between his or her current beliefs and new experiences (Rosman et al., 2016) and also has to recognize that the current epistemological beliefs do not work well with the new experience (Bendixen and Rule, 2004). Furthermore, the new experience must be of personal relevance for the individual. Personal relevance means that the individual has a stake in the outcome or an interest in the topic. Epistemological volition is a further important factor for epistemological change. Epistemological volition refers to a

concentrated effort to change one's current beliefs to the affordances and constraints of the new experience. After experiencing epistemological doubt and volition, the last factor for epistemological change is resolution strategies. According to Bendixen (2002), important resolution strategies are reflection and social interaction. Reflection involves reviewing past experiences and one's current epistemological beliefs and analyzing implications. Afterward, an educated choice to change one's epistemological beliefs is made. In social interaction, individuals argue with other individuals, and this can lead to a revision of current beliefs.

To foster adequate epistemological beliefs, i.e., evaluativist beliefs, Rosman et al. (2016) developed an intervention drawing on this model. The intervention aims at eliciting epistemological doubt by the presentation of resolvable psychological controversies. Resolvable controversies are contradicting psychological theories or empirical findings (see also Kienhues et al., 2016) that can be resolved if contextual variables are taken into account. Thus, this invention builds on the structure of psychological knowledge. The intervention used a multiple-texts approach to present the controversies. In total, Rosman et al. (2016) developed six controversies. All controversies were fictitious examples to ensure that prior knowledge and beliefs do not interfere with the epistemological change mechanisms. Even though the examples were fictitious, their content was related to psychology, e.g., one example described the evaluation results of a psychotherapeutic measure and another example referred to the domain of learning and instruction. Each controversy contained cues indicating how the seemingly contradictory claims can be resolved. After the presentation of each controversy, the participants received the resolution of the controversies. As many of the notable controversies in psychology originate from the use of different methods, Rosman et al. (2016) focused on methodological reasons (e.g., moderator effects) for the controversies to arise.

The controversies aimed to induce epistemological doubt. The appearance of epistemological doubt depends on the present epistemological beliefs. If the participants have absolutist or multiplist epistemological beliefs, the presentation should create epistemological doubt by making them aware of the dissonance between their current epistemological beliefs and the beliefs required to cope with the presented controversies. It is implicitly assumed that the presented controversies are of personal relevance for the participants because the examples referred to psychological content, even though this content is fictitious. The goal of the presentation of the resolution strategies is to finally induce epistemological change. The driving mechanism behind epistemological change is that the resolution of the controversies is incompatible with absolutism and multiplism (Rosman et al., 2017). Absolutists would deny the mere possibility of the resolution of scientific controversies because only one account can be correct. Multiplists would deny that controversies even exist because scientific claims are scientists' personal opinions. Thus, the resolution of the controversies is only compatible with evaluativism and the intervention should, therefore, reduce absolutist and multiplist



epistemological beliefs. The intervention took place in a group setting. An instructor prompted the participants to reflect on the controversies, to reason why they emerged and to discuss the controversies in the group. The instructor also presented the resolution strategies. The instructor functioned as a role model and had the crucial task to moderate the change process to avoid backfire effects (e.g., unintentionally increasing absolutism by reducing multiplism, Trevors et al., 2016).

Rosman et al. (2016) provided evidence for the effectiveness of this intervention. In a pre-posttest-design, they showed that the intervention indeed reduces absolutist and multiplist epistemological beliefs significantly. As the authors used the above-mentioned EBI-AM questionnaire, they cannot provide results regarding the change of evaluativist beliefs. Thus, the findings, so far, do not directly indicate a change towards evaluativism as a decrease in absolutism und multiplism is not sufficient to imply an increase in evaluativism. It seems necessary to investigate the intervention of Rosman et al. (2016) with a dedicated focus on a possible change in evaluativist epistemological beliefs.

Moreover, as an instructor presented the resolution strategies, the effectiveness of the resolution strategies to change epistemological beliefs might depend on the instructor, i.e., the effectiveness of the intervention depends on the instructor. As in Rosman et al. (2016) intervention, there was only one instructor, no conclusion about the role of the instructor in the change of epistemological beliefs is possible because the instructor was not an experimental factor. Also, the group discussion may have had effects on the change of epistemological beliefs that could not be disentangled in the design of Rosman et al. (2016).

Thus, it is worthwhile to replicate Rosman et al.'s results without the presentation of the resolution strategies by an instructor and with participants that work individually on the

controversies. A possible intervention format, in which the presentation of the resolution strategies is the same for all participants, consists in a computer-based presentation of the intervention.

1.4 Epistemological Sensitization

One of the main components of the process model of epistemological change is epistemological doubt. The main feature of the controversies in Rosman et al. (2016) study is that they induce epistemological doubt concerning the topic. Thus, this kind of epistemological doubt is topic-specific. As epistemological beliefs span a range from domain-general to topic-specific, it is evident that epistemological doubt can be represented on a continuum from domain-general to topic-specific. Thus, the question emerges if the induction of domain-specific epistemological doubt may increase the effectiveness of the intervention. Moreover, the question how domain-general epistemological doubt arises.

One possibility is to use a so-called epistemological sensitization measure. Porsch and Bromme (2011) introduced the idea of epistemological sensitization in the context of the research on source choices as a heuristic concept that aims to elicit certain epistemological features in a given context. In their study, the authors investigated the effects of epistemological sensitization on the number of sources used in the evaluation of texts on the subject of tides. Porsch and Bromme (2011) assumed that "sophisticated" epistemological beliefs go along with using more sources than naïve epistemological beliefs. In their experiment, the authors used two versions of an epistemological sensitization measure: a naïve sensitization, in which the knowledge about tides was characterized as structured and static, and a sophisticated sensitization, in which the epistemological nature of selected facts of the science of tides was highlighted. The goal of this sensitization was to elicit either

naïve or sophisticated epistemological beliefs. The authors found that participants in the sophisticated sensitization condition used more sources than the participants in the naïve condition. Thus, the epistemological sensitization measure indeed elicited more sophisticated epistemological beliefs.

Although the epistemological sensitization in Porsch and Bromme (2011) is topic-specific, the principle of sensitization also applies in a domain-specific setting. Following Muis et al. (2006) TIDE model, we have a hierarchical relation in the sense of a topic that is embedded in a given domain, and therefore both are governed by the same epistemological features. An epistemological sensitization yields deeper elaboration and more elaborate learning processes, e.g., metacognitive planning and spending more time on a complex task (Pieschl et al., 2008).

In this way, an epistemological sensitization could be implemented to induce domain-specific epistemological doubt before the controversies are presented. Firstly, presenting the epistemological features of a domain activates prior knowledge about these issues or, in case a person does not have prior knowledge, provides some basic facts about the domain's epistemological features. Thus, depending on a person's epistemological beliefs, the sensitization may cause epistemological doubt if the presented epistemological features are not consistent with the beliefs, e.g., when a person holds absolutist beliefs about psychology, then providing these reasons is in dissonance with absolutism, and may yield domain-specific epistemic doubt. Secondly, the epistemological sensitization potentially leads to a deeper and more planned elaboration of the controversies, making the development of topic-specific doubt more likely, which is beneficial for epistemological change later on.

A proper sensitization measure to induce domain-specific epistemological doubt consists of presenting some basic epistemological features of psychology, especially the main reason why controversies exist in a specific field of psychology. Such a discussion should focus on general epistemological aspects of psychological knowledge in contrast to the controversies' presentation of topic-specific epistemological features. However, the controversies should always reflect instances of the epistemological features presented in the sensitization measure.

To sum up, we assume that an epistemological sensitization should induce domain-specific epistemological doubt if the participant's epistemological beliefs are inconsistent with the epistemological features present in the sensitization measure. Suppose the participants have absolutist or multiplicit domain-specific epistemological beliefs. In that case, the sensitization should raise epistemological doubt by making them aware of the dissonance between their current epistemological beliefs and the epistemological features of psychological science. In turn, raising domain-specific epistemological doubt before presenting the controversies should draw the participants' attention to their current topic-specific epistemological beliefs. This awareness makes it more likely that the dissonance between their current epistemological beliefs and the beliefs required to cope with the presented

controversies are recognized and creates a deeper elaboration of the controversies more likely. In short, we assume that epistemological sensitization makes the intervention more effective.

1.5 Hypotheses

The goal of the present study is to conceptually replicate the effects of Rosman et al. (2016) intervention on the change of epistemological beliefs. The present study aims at investigating the effects of a computerized version of the intervention on which the participants work individually in contrast to the social format of the intervention reported in Rosman et al. (2016). In addition, we want to extend Rosman et al. (2016) results in using a measure of epistemological beliefs that explicitly assesses evaluativist epistemological beliefs. Furthermore, we intend to investigate if the effects of the intervention can be strengthened by an epistemological sensitization. Another important extension of our study is to include a measure of students' argumentation skills. To sum up, regarding epistemological beliefs, we hypothesized that

- (H1) the intervention reduces absolutism and multiplism,
- (H2) the intervention fosters evaluativism, and that
- (H3) an epistemological sensitization measure taking place before the intervention will strengthen the effects of the intervention.

Finally, as epistemological beliefs are a prerequisite for scientific argumentation, we hypothesize that.

- (H4) the intervention fosters the participants' argumentation skills and
- (H5) an epistemological sensitization measure taking place before the intervention will strengthen the effect of the intervention on argumentation skills.

2 METHODS

2.1 Sample, Design, and Procedure

The sample consists of 68 (42 female, 26 male) psychology students from a university in the southwest of Germany. The mean was 23.18 age ($sd = 2.50$) and the median semester was 4 (range 10). The subjects were recruited through social networks and bill-board postings. They received a two-hour time credit for the required participation in psychological experiments.

The study had a pre-posttest-design and the subjects were randomly assigned to the experimental conditions ($N_{EC} = 47$) and a control condition ($N_{CC} = 21$). The experimental condition received the training intervention. There were two groups in the experimental condition: one group received the epistemological sensitization [experimental group 1 (EC1); $N_{EC1} = 23$] whereas the other group [experimental group 2 (EC2); $N_{EC2} = 24$] instead had to read a neutral text. The participants in the control group (CG) received a bogus training intervention not related to the content of the real training intervention.

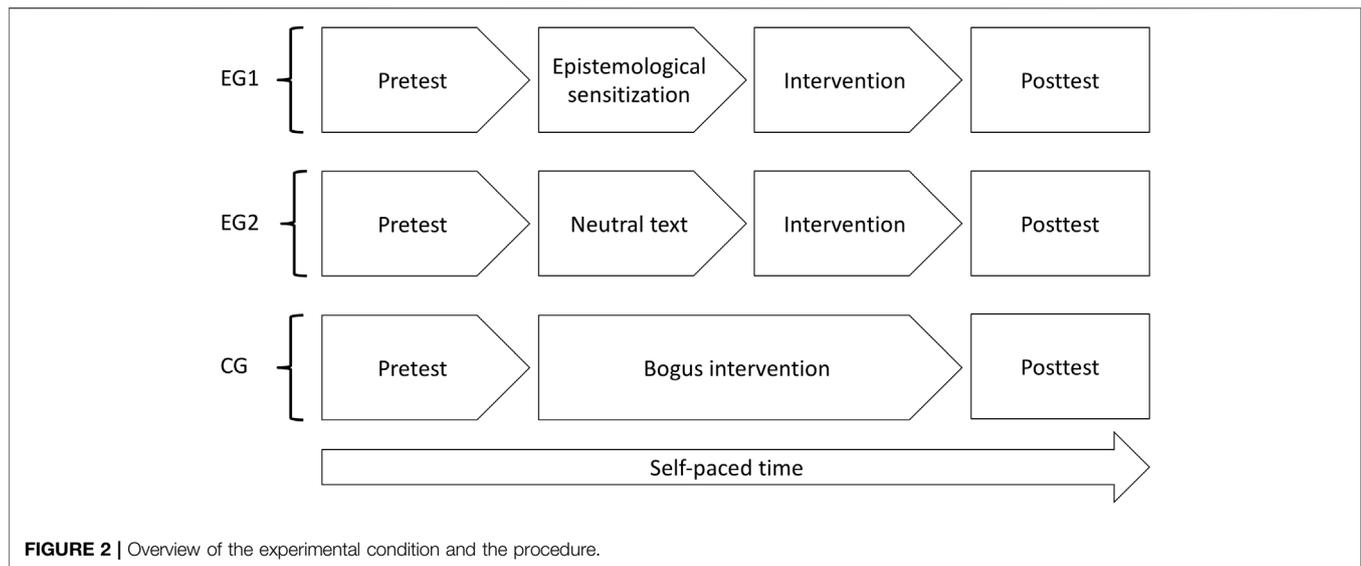


FIGURE 2 | Overview of the experimental condition and the procedure.

The procedure was as follows: First, all participants answered a short demographic questionnaire and then took pretests consisting of measures of epistemological beliefs and an argumentation task. After that, the participants in the experimental conditions either received the epistemological sensitization or read the neutral text. Then, the actual training intervention started. The participants in the control condition started with the bogus training intervention after the pretest. When the training or bogus training intervention was finished, all participants received the posttest. The pre- and posttest as well as the training and bogus training intervention, were computer-based and one session was scheduled for 120 min. However, the time was not limited to ensure ecological validity. On average, a session lasted 81.98 min ($sd = 21.01$). The shortest session lasted 31.40 min and the longest 137.37 min. The participants worked in a self-paced manner through the intervention. **Figure 2** presents an overview of the experimental designs and the procedure.

2.2 Intervention

The intervention consisted of two parts: the first part was the epistemological sensitization measure and the second part was the training intervention. The epistemological sensitization measure, which was applied in EC1, consisted of the adaption of a textbook chapter covering the topic of conflicting scientific claims in psychology (Bromme and Kienhues, 2014). The text addresses the issue either in the form of conflicting empirical results or conflicting theories. Firstly, the text discussed the revision of scientific knowledge. Knowledge depends on a consensus in the scientific community, so that there is a body of knowledge for which consensus is given and a body of knowledge that is under the scrutiny of ongoing research and, therefore, yields conflicting claims. Secondly, the text discussed scientific methods as a potential source of conflicting claims. In particular, the theoretical assumptions underlying various methods were mentioned as a source of potential conflicting claims. Thirdly, the text introduced different research paradigms

as a reason for conflicting claims. Lastly, the difference between knowledge to explain psychological phenomena and knowledge to conceptualize interventions was discussed. The epistemological sensitization text had 557 words and a Flesch reading score of 26, which can be found in the (**Supplementary Material SA**).

The neutral text, which was applied in EC2, featured a definition of psychology as the science that is concerned with the description, explanation, and prediction of human behavior. Each facet of the definition was discussed and afterward, the various subdisciplines of psychology were described shortly. The neutral text had approximately the same length as the epistemological sensitization text; it had 574 words and a Flesch reading score of 22.

The training intervention was structured analogously to the intervention of Rosman et al. (2016). We used the same six artificial examples as in Rosman et al. (2016). The examples were slightly altered and adapted to the computerized administration. Each example started with the learning objectives. Then, two conflicting claims about a topic were presented, followed by a table summarizing the main points of the two texts. The table was designed to facilitate the discovery of the reason for the conflicting claims. Afterward, the participants had to answer two reflection questions focusing on reasons for the conflicting claims. Responding to these questions was mandatory, otherwise, the subjects could not continue. For each reflection question, feedback was provided and the participants were instructed to compare this feedback with their solution. Each example concluded with a synopsis summarizing the reasons for the conflicting claims. **Figure 3** depicts the first example and **Table 2** shows the content of the controversy, the resolution, and the reflection questions of each example (**Supplementary Material SB**). After the posttest, the participants were informed that the examples were fictitious.

The bogus intervention in the control group consisted of six short stories featuring the discussion between two fictional professors. The discussion referred to topics not directly related

Example 1: Learning from pictures

Learning objectives:

- Recognize that theories have a limited range of validity
- Learn that an important goal in research is to delineate the range of validity
- Recognize that contradictions between research findings may be caused by certain boundary condition or moderator variables

Mister Holzmer works at the Department of Education Psychology at the University of Franzenheim. He designs a new teaching method (“Learning from pictures”) and evaluates its effectiveness with a sample of 120 primary school students in the second year taught by the same teacher. In comparison to a conventional teaching method, the results reveal positive effects of the new teaching method. Mister Holzmer therefore recommends teachers to use the new teaching method.

His colleague, **Mister Mertes**, evaluates the new teaching methods again using two samples of secondary school students between twelve and fourteen years. In a sample of 80 twelve year-old secondary students as well as in a sample of 70 fourteen year old students (all taught by the same teacher) there is no evidence of the effectiveness of the new teaching method. In both samples, the learning outcomes were not higher in comparison to a conventional teaching method. On the contrary, there are cases in which the learning outcome decreased. Mister Mertes there highly recommends not using the new teaching method.

Comparison of the two studies

Mr. Holzmer’s study (primary school students):	Mr. Mertes’s study (secondary school students)
<i>Methods:</i> large sample size but small effect size; experimental design with control group	<i>Methods:</i> large sample size but small effect size; experimental design with control group
<i>Sample:</i> only primary school students in the second year	<i>Sample:</i> secondary school students between twelve and fourteen years
<i>Conclusion:</i> recommended to use the method but without regarding the target group	<i>Conclusion:</i> recommended not to use the method but without regarding the different target group in this study

Please reflect on the following questions:

- Which researcher is right and why?
- Is the teaching method recommended?

FIGURE 3 | First fictitious example from the intervention.

to psychological controversies, e.g., the first discussion was about the construction of the pyramids in ancient Egypt. After each discussion, the participants were instructed to rate the strength of each professor’s argumentation, i.e., to rate to which degree he or she agrees with each professor’s argumentation, and finally to decide on which professor’s claim they consent to.

2.3 Dependent Variables

2.3.1 Epistemological Beliefs

Epistemological beliefs were measured using two questionnaires. The first questionnaire was the EBI-AM (Peter et al., 2016). This questionnaire measures domain-specific absolutism and multiplism drawing on the developmental approach. One of the shortcomings of the EBI-AM is that it does not measure

evaluativist epistemological beliefs. Therefore, we also applied the scenario-based Epistemic Thinking Assessment (ETA; Barzilai and Weinstock, 2015) which assess epistemological beliefs in the sense of the developmental approach, too. In contrast to the EBI-AM, the ETA measures domain-specific absolutism, multiplism, and evaluativism in conjunction with topic-specific aspects (Barzilai and Weinstock, 2015).

The EBI-AM measures absolutism with 12 and multiplism with 11 items. The items refer to the domain of psychology and ask the participants to give their ratings of epistemological features of the domain psychology. The items were administered in conjunction with a six-point rating scale on which the subjects had to indicate their agreement with the item statement. We calculated the mean scores for each scale

TABLE 2 | Description of the controversies in the intervention.

Description	Controversy	Resolution	Reflection questions
Learning from pictures	Diverging results concerning the effectiveness of a learning intervention	Age of the subjects as a moderator, the intervention works for younger but not for older students	Which researcher is right and why? Is the teaching method recommended?
Phantasy therapy	Divergent results concerning the effectiveness of a new therapy	Different experimental design and different measurements provide different results	Which study would you believe more and why? What role does a researcher's reputation play in the quality assessment of his or her scientific work
Joy and anger	Divergent correlation patterns (positive vs. negative) between joy and anger in two studies	Cross-sectional vs. longitudinal designs	Which study would you believe more and why? In your opinion, what are the decisive criteria for evaluating the present findings?
Job satisfaction and self-determination	Divergent findings about the relation between job satisfaction and self-determination	Quantitative vs. qualitative study	Which study is better and why? In your opinion, what are the decisive criteria for evaluating the present findings?
Family cohesion and job performance	Meta-analysis with inconsistent results about the relation between family cohesion and job performance vs. a study demonstrating a consistent but moderated relation	Meta-analysis as summary of inconsistent studies vs. a study that aims to clarify the reasons for the inconsistent studies, the reason is a moderated relation	Are contradictions in research findings an indicator of poor research? How should research deal with such contradictions?
Binge eating	Binge eating as a result of suppressed conflicts vs. binge eating as learned behavior	Different paradigms providing different theoretical explanations	Which explanation is preferable and why? Which of the two explanations would you say is correct?

Note. The entries in the column description were the headings for the respective controversy (see Figure 3).

as a measure for absolutism and multiplism. The absolutism and multiplism scales yielded good internal consistencies in terms of Guttman's λ_6 (Guttman, 1945; Table 4) and an item analysis indicated to keep all items.

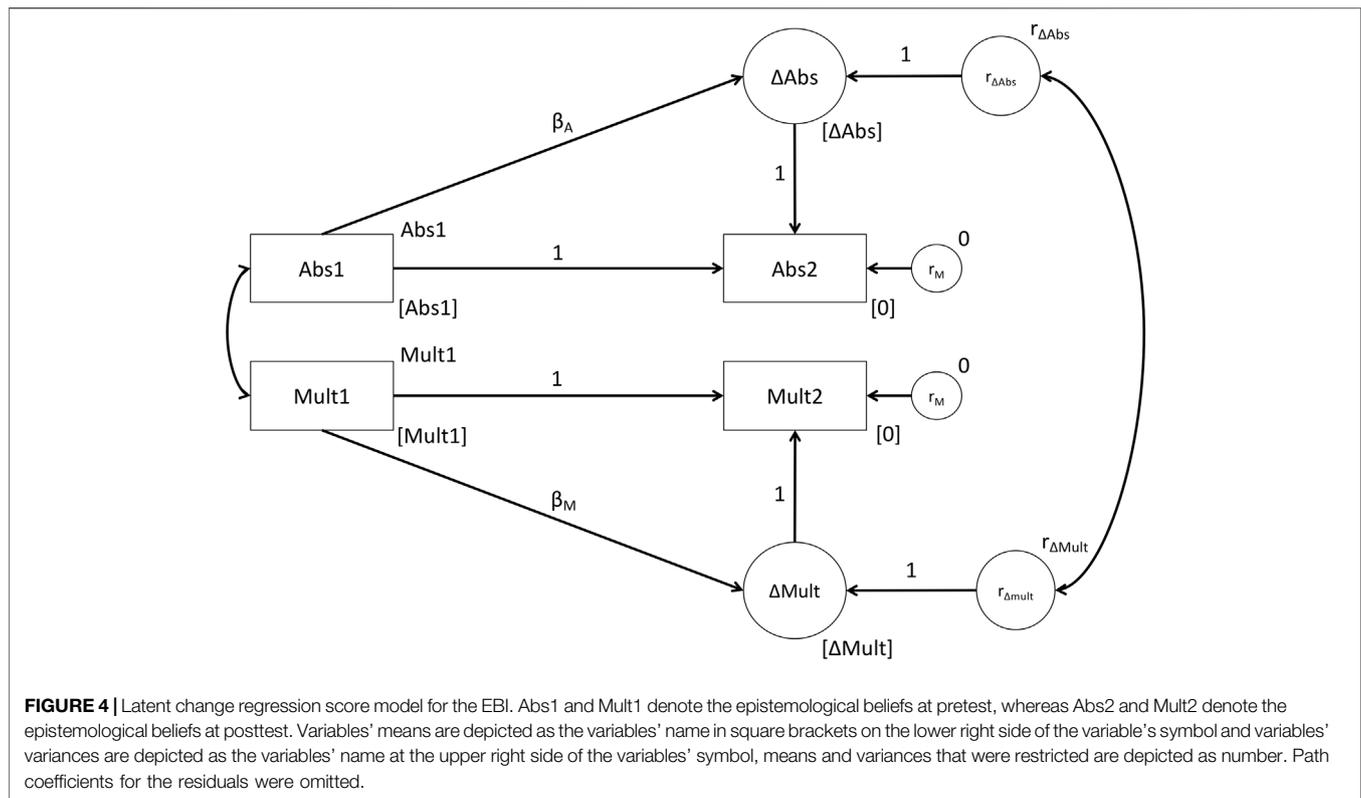
The ETA is a scenario-based questionnaire assessing domain-specific absolutism, multiplism, and evaluativism. The ETA draws on the integration of the developmental and the dimensional approach to epistemological beliefs. The three developmental levels of epistemological beliefs are considered as a multidimensional composition of the following nine dimensions: Right answer, Certainty of knowledge, Attainability of truth, Nature of knowledge, Source of knowledge, Multiple perspectives, Evaluate explanations, Judge accounts, and Reliable explanation³. For each of these dimensions, absolutism, multiplism, and evaluativism was assessed by one item. Thus, there were 27 items in total and were administered in conjunction with a six-point rating scale on which the subjects had to indicate their agreement with the item statement. The items themselves refer to a scenario, i.e., an epistemological dilemma. Such a dilemma is similar to the controversies given to the participants in the training intervention. The items are formulated in such a way that they prompt the participants to reason about a problem that belongs to a specific psychological topic provided in the scenario. Besides, the items are embedded in a specific domain, i.e., psychology. Consequently, the ETA items enable to measure domain-specific epistemological beliefs in conjunction with some topic-specific aspects of these epistemological beliefs (Barzilai and Weinstock,

2015, p. 144). We adapted the scenarios used in the ETA to the domain of psychology as the original scenarios provided by the authors referred to the domains of history and biology. The scenario of the pretest referred to the causes of depression and presented two conflicting statements about different origins. The pretest scenario consisted of 406 words and had a Flesch reading score of 52. The posttest scenario referred to the cause of schizophrenia and presented two conflicting statements about the etiology of schizophrenia. It consisted of 348 words and its Flesch reading score was 34. To generate scores for absolutism, multiplism, and evaluativism, we calculated the mean score for each scale. An item analysis indicated that some items belonging to the dimension Right answer correlated negatively with the corrected mean score for both the pre- and posttest. We, therefore, dropped these items from the analysis. Thus, the final set consisted of 24 items, eight items measuring absolutism, multiplism, and evaluativism, respectively. The final scales yielded good internal consistencies in terms of Guttman's λ_6 (Table 4).

2.3.2 Argumentation Task

As epistemological beliefs and argumentation skills are related, we expanded the ETA with an argumentation task. We prompted the participants to write an essay about the presented controversies in the ETA and to write an essay for the pre- and posttest, respectively. Particularly, students were not prompted to argue for one perspective in the controversies, because that may have prompted to argue for one side which either had been in the form of an absolutist or multiplist argumentation. Instead, the participants were asked to comment on the presented controversies and explain the conclusions. We developed a coding scheme reflecting the epistemological level of the argumentation in the essays

³These dimensions differ slightly from those described in the introduction section. In the questionnaire development process described in Barzilai and Weinstock (2015), the dimensions were slightly revised.



(Mason and Scirica, 2006). We coded if the essays reflect an absolutist, multiplist, or evaluativist argumentation, the coding is as follows: one to two points represent absolutist beliefs, three to four points represent multiplicitist beliefs, and five to six points represent epistemological beliefs. An overview of the task definition, the coding criteria, examples, and explanations for the criteria for each level and the points assigned to each level is provided in the (Supplementary Material SC).

The coding procedure was as follows: The first author and a student research assistant did the coding. The research assistant was not involved in the study but was introduced to the theoretical aspects of argumentation and epistemological beliefs. All essays were exported from the software and rated by each rater. Afterward, the ratings were compared. We used Cohen's κ as a measure of interrater agreement and we resolved the cases of disagreement by discussion. For the pretest, the interrater agreement is $\kappa = 0.54$ and for the posttest $\kappa = 0.76$ and indicates moderate or substantial agreement, respectively (Landis and Koch, 1997).

2.4 Statistical Analysis and Sample Size Considerations

For all statistical analyses, R version 3.2.1 (R Core Team, 2015) was used. Reliability analysis was done with the psych package Revelle (2016) in version 1.6.9 and we used lavaan (Rossee, 2012) in version 0.6.3. To analyze the effects of the experimental conditions on the EBI-AM and ETA scales, we used latent change regression score models (McArdle, 2009; McArdle and

Nesselroad, 2014). In these models, the difference between pre- and posttest is modeled as a latent variable and this latent difference variable is also regressed on the pretest. The mean of the latent difference variable thus represents the mean change in epistemological beliefs taking the epistemological belief level of the pretest into account. We used a multiple group setting in which for each experimental condition, a separate latent change regression score model was set up. Mean changes were calculated and compared between the three groups. For the EBI-AM and ETA, we set up separate multiple group models. We also allowed epistemological beliefs in the pretest to be correlated. The model for the EBI-AM is depicted in Figure 4. Differences between the epistemological beliefs at the pretest and between the means for the experimental conditions were evaluated using Wald tests. Additionally, we calculated Cohen's d for the pairwise group comparisons as effect size measure.

We used the ML estimator, and all relevant variables had skew and kurtosis values smaller than two or seven in each condition, respectively (Finney and DiStefano, 2013). Regarding the sample size, we followed Rosman et al. (2016) and expected moderate effects and used 30 participants per group as a target sample size to account for possible drop-outs. For the power analysis, we performed a Monte Carlo study (Zhang and Liu, 2019) with 10,000 replications using the analysis results for the EBI-AM and ETA as the data generating process. The power is defined as the proportion of repetitions for which the null hypothesis is rejected for a given parameter with $\alpha = 0.05$ (Beaujean, 2014). According to (Cohen, 1988; Kyriazos, 2018), power should at least be 0.50 with an ideal value of 0.80. Drawing on the values, we consider a

TABLE 3 | Descriptive statistics epistemological belief scales and the essays scores at pre- and posttest for each experimental condition.

	Condition	EBI absolutism		EBI multiplicism		ETA absolutism		ETA multiplicism		ETA evaluativism		Essay score	
		M	SD	M	SD	M	SD	M	SD	M	SD	Med	IQR
Pretest	EG1	2.66	0.61	3.76	0.61	4.02	0.67	2.59	0.62	4.71	0.40	3	1.00
	EG2	2.74	0.52	3.94	0.51	4.12	0.64	2.74	0.56	4.49	0.61	3	0.00
	CG	2.76	0.47	3.80	0.52	3.89	0.62	2.77	0.59	4.32	0.73	3	0.00
Posttest	EG1	2.35	0.49	4.00	0.60	4.04	0.81	2.48	0.57	4.97	0.52	5	1.00
	EG2	2.39	0.53	4.17	0.54	4.25	0.46	2.70	0.75	4.64	0.53	4	1.25
	CG	2.73	0.61	3.71	0.57	4.13	0.65	2.64	0.59	4.27	0.69	3	1.00

Note. For the essay score, the median and the interquartile range are reported, for all other variables, the mean and standard deviations are reported.

TABLE 4 | Descriptive statistics and correlations for the epistemological belief scales and the essays score at pre- and posttest for the whole sample.

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. EBI A1	2.71	0.53	(.70)	—	—	—	—	—	—	—	—	—	—	—
2. EBI M1	3.83	0.55	-0.45***	(.74)	—	—	—	—	—	—	—	—	—	—
3. ETA A1	4.02	0.64	0.35**	-0.18	(.79)	—	—	—	—	—	—	—	—	—
4. ETA M1	2.700	0.59	-0.08	0.32**	-0.54***	(.74)	—	—	—	—	—	—	—	—
5. ETA E1	4.51	0.60	-0.19	0.31*	0.08	0.01	(.79)	—	—	—	—	—	—	—
6. ESSAY 1	35.00	17.79	-0.15	-0.05	-0.04	-0.06	-0.12	(.54)	—	—	—	—	—	—
7. EBI A2	2.48	0.56	0.69***	-0.30*	0.07	0.05	-0.09	-0.11	(.75)	—	—	—	—	—
8. EBI M2	3.97	0.59	-0.31**	0.67***	-0.05	0.1	0.30*	0.04	-0.39**	(.69)	—	—	—	—
9. ETA A2	4.14	0.65	0.36**	-0.15	0.70***	-0.35**	0.14	0.02	0.24*	-0.16	(.82)	—	—	—
10. ETA M2	2.61	0.64	-0.10	0.39**	-0.39**	0.64***	-0.09	-0.10	0.04	0.24	-0.39**	(.79)	—	—
11. ETA E2	4.64	0.64	-0.19	0.26*	0.13	-0.09	0.7***	-0.10	-0.19	0.27*	0.12	-0.03	(.83)	—
12. ESSAY 2	35.00	19.30	-0.04	-0.11	0.05	-0.22	0.09	0.19	-0.19	0.09	-0.01	-0.23	0.34**	(.76)

Note. Descriptive statistics for the essay score are based on the rank transformed values. Correlations between epistemological belief scales are Pearson correlations between scale means scores, correlations between epistemological beliefs scales and essay score are Pearson correlations between the scales mean scores and rank transformed essay scores; correlation of pretest and posttest essay scores are based on Kendall's τ . Values on the diagonal in brackets indicate the internal consistency (Guttman's λ_6) of the epistemological belief scales and the interrater reliability (Cohen's κ) for the essay scores. Variables: EBI A1, A2: EBI pre- and posttest absolutism scale, EBI M1, M2: EBI pre- and posttest multiplicism scale, ETA A1, A2: ETA pre- and posttest absolutism scale, ETA M1, M2: ETA pre- and posttest multiplicism scale, ETA E1, E2: ETA pre- and posttest evaluativism scale, ESSAY 1, 2: pre- and posttest essay score.

power of 0.65 as sufficient. Due to the relatively small sample sizes in each condition, we used bootstrapped standard errors for significance testing using 10,000 bootstrap samples. To take care of possible a error inflation due to multiple comparisons, we adjusted the *p*-values for these using the Holm method.

Because the essay score is an ordinal measure, we used a Kruskal-Wallis-Test to analyze the effect of the epistemological sensitization. Although this test does not allow taking the pretest essay score into account, it has the advantage of dealing with ordinal data with a higher power than parametric procedures. For the pairwise comparisons of the experimental conditions, we used pairwise Wilcox tests and adjusted the *p*-value with the Holm method.

3 RESULTS

3.1 Descriptive Statistics, Correlational Analysis and Internal Validity

Table 3 shows the descriptive statistics for the EBI-AM and ETA scales, as well as the essay score at pre- and posttests for the three experimental conditions. The correlations of the dependent variables for the whole sample are shown in Table 4. The

correlational analysis of the pretest scores indicated a significant, negative correlation between the EBI-AM absolutism and multiplism as well as between the ETA absolutism and multiplism scales. Surprisingly, the ETA evaluativism scale also correlated positively with the EBI-AM multiplism scale. Furthermore, the EBI-AM and ETA absolutism scales as well as the EBI-AM and ETA multiplism scales correlated positively, providing evidence for the convergent validity of the scale. At the pretest, neither of the epistemological belief scales correlated with the rank transformed essay score.

At the posttest, the correlation pattern between ETA and EBI-AM absolutism and multiplism scales remained approximately the same as at the pretest. Also, the EBI-AM and ETA multiplism scales correlated positively, providing evidence for their convergent validity. However, in contrast to the pretest, the EBI-AM and ETA absolutism scales did not significantly correlate. Furthermore, the ETA absolutism scale also correlated positively with EBI-AM multiplism scale. In general, the correlations of epistemological beliefs in the pretest were numerically smaller than the correlation in the posttest. In contrast to the pretest, there was a significant correlation between the rank transformed essay score and the ETA

TABLE 5 | Results of the latent change regression score models.

	EBI absolutism			EBI multiplicism			ETA absolutism			ETA multiplicism			ETA evaluativism						
	M	SE	p	M	SE	p	M	SE	p	M	SE	p	M	SE	p	M	SE	p	
Means																			
EG1	-0.31	0.08	0.001	0.24	0.08	0.003	0.86	0.10	0.831	0.08	0.831	0.08	-0.11	0.11	0.332	0.20	0.09	0.003	0.83
EG2	-0.35	0.09	0.001	0.23	0.10	0.027	0.63	0.11	0.226	0.69	0.226	0.69	-0.04	0.12	0.732	0.08	0.12	0.182	0.31
CG	-0.03	0.08	0.685	-0.09	0.09	0.301	0.20	0.09	0.013	0.71	0.013	0.71	-0.12	0.09	0.150	0.32	0.07	0.490	0.51
Pairwise comparisons																			
EG1-CG	-0.27	0.11	0.032	0.33	0.12	0.014	0.80	0.14	0.339	0.36	0.339	0.36	0.02	0.14	1.00	0.36	0.12	0.018	0.91
EG2-CG	-0.31	0.12	0.027	0.32	0.14	0.024	0.68	0.15	0.900	0.07	0.900	0.07	0.08	0.15	1.00	0.07	0.13	0.266	0.44
EG1-EG2	0.04	0.12	0.754	0.01	0.13	0.870	0.06	0.15	0.900	0.39	0.900	0.39	-0.07	0.16	1.00	0.39	0.15	0.459	0.13
Effect sizes																			
EG1-CG	0.81	-	-	0.88	-	-	-	-	-	-	-	-	0.03	-	-	-	-	-	-
EG2-CG	0.80	-	-	0.75	-	-	-	-	-	-	-	-	0.17	-	-	-	-	-	-
EG1-EG2	0.10	-	-	0.03	-	-	-	-	-	-	-	-	0.12	-	-	-	-	-	-

Note. Pow: Power.

evaluativism scale providing evidence for the validity of our scoring. Especially, there were significant correlations between the pre- and posttest measures for all EBI-AM and ETA scales, thus, justifying their inclusion in the form of the latent change regression score models. The sole exception to this was the correlation between the essay scores at the two measurements, which was not significant.

To check for the internal validity, we tested if there were systematic differences in age, the number of semesters, or in the distribution of the sexes between the experimental conditions. There were neither significant differences for the age, $F(2, 66) = 0.21, p = 0.809$, nor the number of semesters, $F(2, 66) = 0.13, p = 0.883$, nor the distribution of the age, $\chi^2(2) = 2.56, p = 0.278$. None of the epistemological belief pretest measures differed significantly between the experimental conditions (all Wald test, n. s., $0.44 \leq \chi^2(2) \leq 1.68, 0.803 \leq p \leq 0.433$) expect the ETA evaluativism scale [$\chi^2(2) = 6.02, p = 0.049$], which was slightly below the significance threshold. Thus, the randomization procedure was effective.

3.2 Results for the EBI-AM

The results for the EBI-AM as well as the ETA scales are shown in **Table 5**. For absolutist beliefs, the Wald test indicates that the mean changes are significantly different between the groups [$\chi^2(2) = 8.01, p = 0.018$]. The mean changes indicate that there is a significant reduction of absolutist beliefs in both experimental conditions but not in the control condition. The mean comparisons revealed that both the means of EG1 and EG2 significantly differed from the mean of the CG with a large effect size. However, the means of EG1 and EG2 did not show any significant difference. Thus, it seems the intervention effectively reduces absolutism, regardless of the presence of an epistemological sensitization.

Regarding multiplist beliefs, the Wald test indicates differences between the experimental conditions [$\chi^2(2) = 7.32, p = 0.026$]. In detail, the results indicate an effect that was in contrast to our expectations. In both experimental conditions EG1 and EG2, there was a significant increase in multiplist beliefs, whereas there was no change in the CG. Furthermore, both the means of EG1 and EG2 significantly differed from the mean of the CG; in case of the comparison between EG1 and CG with a large effect and in case of the comparison of EG2 with CG with a medium effect size. The difference in the means between EG1 and EG2 was, however, not significant.

In summary, these results suggest a possible backfire effect, i.e., the reduction in absolutism causes an increase in multiplism (cf., Rosman et al., 2016). The descriptive statistics in **Table 3** also supports a possible backfire effect. As can be seen, absolutism decreases from pre- to posttest in both experimental conditions but remains approximately the same in the control condition. For multiplism, there is an increase in both experimental conditions, whereas the multiplism level remains approximately the same in the control condition. The correlations between absolutism and multiplism within each condition may also support a possible backfire effect: The correlation in the EG1 is $r = -0.49 (p = 0.014)$, in the EG2 $r = -0.60 (p = 0.002)$, and $r = -0.04 (p = 0.861)$ in the CG. As the correlation within the conditions suggests, the

backfire effect occurs only in the experimental condition and is possibly stronger in the condition without epistemological sensitization.

Thus, these results only lend partial support to hypothesis H1. Especially, the intervention only reduces absolutism but possibly produces a backfire effect for multiplism. Regarding hypothesis H3, there is no evidence that the epistemological sensitization strengthens the intervention's effectiveness in the reduction of absolutism. Moreover, there is weak evidence that the epistemological sensitization may lessen the possible backfire effect, but a definite conclusion regarding these findings cannot be drawn.

3.3 Results for the ETA

For the absolutism scale, there are no differences between the experimental conditions [$\chi^2(2) = 2.34, p = 0.310$]. The mean changes in EG1 and EG2 did not indicate a significant change. However, for the CG, the mean change indicated a significant increase. Besides, neither of the pairwise comparisons was significant. For the multiplism scale, there are no differences for the mean changes [$\chi^2(2) = 0.29, p = 0.863$] and all other results are also not statistically significant.

Regarding the evaluativism scale, the Wald test indicates a significant difference between the groups [$\chi^2(2) = 7.12, p = 0.028$]. The mean change indicated a significant increase in evaluativism in EG1 but no significant change in EG2 and CG. The pairwise comparisons indicated a significant and strong difference between the mean changes in EG1 and CG. All other comparisons were not significant.

Thus, the results do not support the hypothesis H1 because there are no changes in absolutism and multiplism, but partially support the hypothesis H2 and fully support the hypothesis H3 because there was only an increase of evaluativism in the condition with the epistemological sensitization. The results are not clear-cut regarding absolutism. As the pairwise comparisons between the three conditions and the Wald test for mean changes are all not statistically significant, it seems to be a rather random fluctuation than a systematic change. Nevertheless, the results warrant no conclusion regarding the effects of the intervention and the epistemological sensitization on the ETA absolutism scale.

3.4 Results for the Essay Score

Regarding the essay score, a Kruskal-Wallis-test indicated that there were no differences between the conditions of the pretest, $\chi^2(2) = 2.79, p = 0.248$. During the posttest, there was a significant effect between the conditions, $\chi^2(2) = 26.25, p < .001$. Comparing the EG1 with the CG yielded a significant difference, $W = 442, p < 0.001$, and comparing the EG2 with the CG also yielded a significant difference $W = 392, p < .001$. The comparison between the two experimental conditions was also significant, $W = 374, p = 0.030$. The training intervention seems to foster the participants' argumentation skills. Still, as indicated by the median scores (Table 3) in both experimental conditions, the epistemological sensitization is necessary to foster an evaluativist argumentation as the median in the experimental condition without an epistemological sensitization only reflects a

multiplism argumentation. Thus, these results support our hypothesis H4 that the intervention fosters psychology students' argumentation skills, and the hypothesis H5 that the epistemological sensitization strengthens the effects of the intervention. Moreover, it seems to be the case that an evaluative argumentation skill is only obtained when the epistemological sensitization is present.

4 DISCUSSION

The results of our study only partially support our hypotheses. We will now discuss the results in the light of the hypotheses for each epistemological belief measure. Regarding the EBI-AM, there was a reduction of absolutist epistemological beliefs, but in contrast to our expectation, this was the case in both experimental conditions. In contrast to our expectation, there also was an increase in multiplism in both experimental conditions. Thus, the intervention seems to be effective for reducing absolutism regardless of the epistemological sensitization.

The reduction of the EBI-AM absolutism is consistent with the prediction from the Bendixen and Rule (2004) epistemological change model. Regarding multiplism, there was an unexpected increase. The increase of multiplism is in contrast to Rosman et al. (2016) findings that reported a reduction of absolutism and multiplism. One reason for this may be the computerized format of our intervention. Rosman et al. (2016) cautioned that a backfire effect might occur when an instructor, who serves as a role model for the participants, does not carefully moderate the phases of intervention. A similar situation may have been present in our computerized intervention. On the other hand, it is an open question if these findings indicate a backfire effect. Per definition, a backfire effect consists in a decrease in one measure and an increase in another measure at the same time, which is indicated by a negative correlation between the two measures. The same is the case in both experimental conditions. Still, it should be taken into account that the sample size in each condition is relatively small, so the correlation coefficients may be unstable (Bonett and Wright, 2000). Moreover, regarding the correlation between the EBI-AM's absolutism and multiplism scales, there is the general problem that, according to the correlations reported in Table 4, the EBI-AM scales for absolutism and multiplism correlate negatively at the pretest as well as at the posttest in the whole sample. Peter et al. (2016) also reported in the two studies, in which the EBI-AM was developed, a negative correlation between absolutism and multiplism. Thus, if absolutism and multiplism are negatively related before any intervention takes place, the question arises if absolutism and multiplism can be reduced at the same time. Moreover, such a constellation would indicate that any decrease in absolutism would be automatically accompanied by an increase in multiplism and yield a backfire effect. The aforementioned considerations are based on an empirical argument, i.e., the negative correlation between absolutism and multiplism. From a more theoretical point of view, the conceptualization of epistemological development as a stage

model implies that one has to first move from the absolutist stage to the multiplist stage. Substantively, an individual has to move from the belief that there is only one correct account of knowledge in favor of the belief that there are many possible accounts of knowledge either of which is seen as correct. In this way, it seems plausible that reducing absolutism can only occur if multiplism increases. Thus, taking the stage model for granted, the expected pattern of results should correspond to the pattern found in this study but not the pattern reported in Rosman et al. (2016).

Additionally, we have to consider the format of our intervention. As mentioned in the introduction, in Rosman et al. (2016) original intervention, the resolution strategies were presented by an instructor. The instructor should carefully moderate the intervention process to avoid a possible backfire effect. Thus, trying to avoid the possible dependency of the interventions' effectiveness on the instructor, which was the reason to introduce the computer-based intervention in this study, may have the possible drawback to generate a backfire effect. Therefore, it seems necessary to scrutinize this question further using comparisons of computer-based and instructor-based interventions. In particular, the role of the instructor should be investigated experimentally, e.g., by using more than one instructor and by handling the associated instructor variable as a random factor.

For the ETA scales, we found no effects for absolutism and multiplism. Only evaluativism increased in the condition with the epistemological sensitization. Thus, for evaluativism, the change mechanism is exactly as hypothesized. Inducing domain-specific epistemological doubt before inducing topic-specific epistemological doubt increases the effect of the intervention. Nevertheless, the results do neither provide a clue why there was no effect in the experimental condition without sensitization for evaluativism nor do they indicate why there is no effect for absolutism and multiplism.

Another noteworthy outcome of this study is the different pattern of results for the EBI-AM and ETA. A possible explanation may be that the EBI-AM only captures to domain-specific epistemological beliefs, whereas the ETA captures a mixture of domain-specific and topic-specific epistemological beliefs. Thus, the participants may comprehend the intervention in terms of domain-specific epistemological beliefs when working on the EBI-AM but have a stronger focus on the topic-specific side when working on the ETA. The fact that the ETA items explicitly refer to the scenario can bring the topic-specific epistemological aspects into the foreground. A consequence of such an explanation is that epistemological beliefs captured by the various questionnaires may not be directly comparable. However, they provide evidence for convergent validity, at least, for absolutism and multiplism. Regarding our results and the results of epistemological beliefs interventions in general, one may, therefore, ask if the results generalize to other measurements of epistemological beliefs as well.

Regarding the essay task, the pattern of results corresponds to our expectations. There was an increase in argumentation skills in both experimental conditions, but only in the condition with

epistemological sensitization, we found an evaluativist argumentation. Here, the epistemological sensitization worked as hypothesized, but the question why this is the case remains. A possible explanation for this may be that the sensitization emphasizes the fragility of scientific knowledge and highlights the context-sensitivity of scientific claims. In other words, the epistemological sensitization strengthens the participants' awareness of the ill-defined structure of psychological knowledge, which in turn is considered by the participants when writing an argumentative essay.

5 LIMITATIONS

There are several limitations to this study. One limitation refers to the process model of epistemological belief change, which was the theoretical foundation for the presented intervention. According to Braten (2016), there is a lack of empirical evidence backing the process model of epistemological belief change even though its basic assumptions are common in domains like conceptual change research. Thus, this research is rather an empirical investigation of the process model than a theoretically sound derived intervention whose effectiveness is evaluated.

Another limitation refers to the concept of personal relevance that is a major ingredient in the process of epistemological change. In the presented intervention, the implicit assumption was that the presented fictitious examples are of personal relevance to the participants. A severe objection is that not all examples were of personal relevance. Although the examples referred to the domain of psychology, they may have been of different personal relevance for the participants because of different interests. For instance, the fourth example referred to the theory of work satisfaction. It may be the case that this example is of low relevance for a participant who has a strong interest in clinical psychology but not in work psychology. In contrast, the second example that referred to the effectiveness of a new psychotherapy may be of high personal relevance for this participant. As a consequence, this participant may have worked the second example in-depth, whereas he/she worked the fourth example only superficially. In short, different interests may moderate the personal relevance of the examples. A measure to control for such effects would be to assess topic-specific interests before the intervention takes place. A possible drawback of such a procedure may be that the assessment triggers the activation of situational interest. Therefore, the examples should not follow the assessment immediately.

A further limiting factor consists in the relation of the fictitious examples to prior knowledge. The reason for Rosman et al. (2016) to use fictitious examples is to prevent effects of prior knowledge. Although the examples were fictitious, they were embedded in different domains and or topics like the ones mentioned before. Reading a fictitious example that is embedded in the domain of clinical psychology and the topic of psychotherapy could trigger the appertaining knowledge a participant has in this domain and or this topic. Moreover, the participants were not aware that the examples are fictitious before they were debriefed after the posttest. Thus, although designed to prevent effects of prior

knowledge, it remains unclear if the fictitious examples perform well in this task.

Along with this, the concept of epistemological sensitization also has severe limitations because it lacks a comprehensive theoretical account. In its current form, the epistemological sensitization concept is a rather heuristic approach than a theory-based intervention strategy. However, a plausible theoretical explanation may draw on the activation of task-relevant prior knowledge. Task-relevant prior knowledge determines which information attention is paid to and which aspects of a task are considered as important (e.g., Alexander, 1996). Thus, activating the relevant prior knowledge about the epistemological features of psychology may help to pay attention to the correct epistemological features of a controversy and judge their importance correctly. Moreover, the relevant epistemological features were prompted through questions, epistemological sensitization may help to judge the question content accordingly and to evaluate its importance. Notwithstanding these considerations, epistemological sensitization remains a heuristic concept until a theoretical account is developed.

The scenarios presented with the ETA scales in the pre- and posttest also constitute a limitation. Although both scenarios came from the domain of clinical psychology, the topic of the pretest scenario was depression, whereas the topic of the posttest was schizophrenia. Thus, there is the implicit assumption that both measurements constitute a form of “parallel” measurements that are comparable. Because the ETA measures domain-specific epistemological beliefs in conjunction with some topic-specific aspects of epistemological beliefs, this assumption is at least questionable. It may be the case that the change in evaluativism is not due to the intervention, but due to *per se* higher topic-specific evaluativist epistemological beliefs about schizophrenia as compared to the topic-specific evaluativist epistemological beliefs about depression. In such a case, the change in epistemological beliefs is confounded with a different level of topic specific-epistemological beliefs. An alternative would be to use the same scenario in the pre- and posttest. But such a procedure carries the risk of memory effects even though the participants are instructed not to consider their previous answers to the items. Another alternative would be to use two scenarios in the posttest: the same scenario as in the pretest and a new scenario. Such a design allows for repeated measurement but has the disadvantage of posing high demands on the motivation of the participants. Memory effects may occur when the same measure is used during the pre- and posttest; this limitation also applies to the assessment of domain-specific epistemological beliefs with the EBI-AM. The significant correlations between pre- and posttest of the EBI-AM scales can be interpreted as evidence for this.

A further limitation of this study is the absence of a manipulation check. A manipulation check consists of one or more questions that capture whether the participants got aware of relevant features of the intervention, i.e., a question that captures if the participants figured out the reason for the controversy. Thus, there is no evidence that the intervention worked as intended. But a manipulation check also bears the risk that it

amplifies the intervention or even triggers the effect of the intervention (cf., Hauser et al., 2018). In the context of this study, asking the participants if they recognized the controversy or the respective resolution strategy could act as a prompting-intervention itself and thus, results in confounding the effects of the intervention. In consequence, the absence of a manipulation check leaves open the question if the participants were aware of the critical features of the intervention. In turn, the absence of an effect, like in the case of ETA scales for absolutism and multiplism, could not be unambiguously attributed to deficits in the intervention or the lack of the awareness of the controversy and the resolution strategies.

From a methodological perspective, a possible limiting factor may be undetected epistemological belief differences between the experimental conditions in the pretest. Although we used a randomization procedure and the control variables age and semester did not show any statistically significant differences between the experimental conditions. But this absence of statistical evidence for possible inequivalence of the experimental conditions is not a sufficient evidence for the successes of the randomization. In fact, there were significant differences between the experimental condition for the ETA evaluativism pretest scores. According to the descriptive statistics, the highest level of evaluativism was in the EG1 which may be potentially bias the results. However, we used a latent change score regression model that provide a base-free level of change, i.e., the part of the individual change that is related to the initial level is removed from the change score (McArdle, 2009, pp. 583–584). Thus, even when the randomization procedure may have failed for evaluativism, the latent change score and its comparison between the groups should be free from base-level effects.

A final limitation is the theory of epistemological development itself. The stage theory, in general, has been criticized (e.g., Driver, 1978). Also, in the domain of epistemological beliefs research, many researchers noted that such an assumption is not consistent with empirical findings. For instance, Bromme (2005) states that the development stages can reoccur and this is diametrical to the assumptions of the stage model. Moreover, Muis et al. (2006) explicitly account for reoccurring stages in their TIDE model. Another potential interpretation of the stages may be that there is a growing skill about how to deal with scientific knowledge in general and, in particular, with psychology-specific or topic-specific knowledge. The theory of hierarchical skill development (Fischer, 1980) provides a theoretical basis for such a stance.

In this perspective, being on the absolutist level means that a person has basic scientific and argumentation skills with the inherent belief that only one account is true. Moving from absolutism to multiplism means that a person acquires additional scientific and argumentation skills. Especially, a person recognizes that different methods may yield different assertions about the same issue but lacks the skills to consider contextual factors and, therefore, the inherent belief emerges, that the scientific claims are arbitrary. Finally, during the transition from multiplism to evaluativism, a person gains the skills to consider contextual factors that induce the belief that, depending

on the context, a certain view may be more appropriate. Skill theory (Fischer, 1980) assumes that skills are hierarchical, i.e., the skills from the previous levels are the basis for the skills of the next level. Moreover, it is important to note, that in this view and in contrast to the assumption of the developmental stage theory, the skills and beliefs of the previous level are not replaced by those from the next level. Instead, the skills and beliefs become enriched; a perspective which is also common in the conceptual change research.

This perspective is also in line with the backfire effect found in this study, but the interpretation differs. Instructing the students that there may be two possible accounts on one issue may prompt the necessary skills to handle scientific controversies with the associated multiplist beliefs. Thus, after the intervention, these skills with inherent beliefs are more salient, which, in turn, would explain why the backfire effect was observed. However, the question remains, why the backfire effect was only observed for the epistemological beliefs measured with the EBI-AM and not for those measured with the ETA.

Moreover, another debatable point is that the normative sequence from absolutist to evaluativist epistemological beliefs applies to all domains. The sequence from absolutism to multiplism may be appropriate in domains with an ill-defined knowledge structure like psychology. But in the case of domains with a well-defined knowledge structure, this developmental sequence may be inappropriate. Consequently, in mathematics, assertions are true once they are proven, which has the consequence that these assertions apply regardless of any contextual factors. For instance, once it is proven that the sum of the first n natural number equals $n(n+1)/2$, this assertion is always true regardless of any contextual factors. Thus, regarding this assertion, neither multiplist nor evaluativist beliefs are normatively appropriate; only absolutist beliefs are appropriate. Another example for this purpose may be the domain of computer science as part of applied mathematics. Empirically, Rosman et al. (2017) report that computer science students develop more absolutist beliefs during the first four semesters. The authors argue that in this domain, in contrast to psychology, the structure of knowledge is well-defined because most of the knowledge rests on mathematical facts. Therefore, the question arises if a general sequence of absolutist, multiplist and evaluativist epistemological beliefs is productive in the domain of computer science. Thus, the theory of epistemological development, on which the intervention presented in this paper draws, should itself be critically scrutinized.

A related issue is that even in domains with ill-structured knowledge, like in the domain of psychology, evaluativist beliefs may not always be appropriate. Although evaluativist beliefs are beneficial in evaluating evidence, the rules for interpreting evidence are given by the scientific community. They are usually part of fixed paradigms that contain a basic set of theories, rules, laws, paradigms, traditions, etc., that are taken for granted (Kuhn, 1962). Applying evaluativist thinking to these fixed paradigms would neither be correct nor productive. The concepts of correctness and productivity were introduced by Elby and Hammer (2001). Correctness means that an epistemological belief is consistent with the prescriptions of the scientific

community. Productivity means that a specific epistemological belief leads to learning success. Applied to controversies, this distinction means that the methods used to resolve the controversy are not subject to evaluativist thinking. In particular, these methods are to be viewed from an absolutist perspective. For instance, in the controversy about a learning method shown in **Figure 3**, the resolution is the existence of a moderator. In another controversy, the resolution draws on the presence of two different paradigms. Thus, the participants have to apply evaluativist reasoning towards the resolution of the controversies. However, concerning the resolution of the controversies, absolutist beliefs are required. For instance, the concept of moderating variables and the existence of different paradigms are basic contents of the knowledge base accepted by the scientific community. Thus, as in the case of the sum of the first n natural numbers mentioned above, absolutist beliefs towards the concept of moderating variables and the existence of different paradigms. Applying evaluativist beliefs would neither be correct nor productive. It would not be correct because the two issues are taken for granted, so there is no need to question the applicability of the two concepts. Indeed, challenging these concepts would not fit into normal science. It would also be not productive because applying evaluativist beliefs towards these accepted concepts would not provide a learning gain for students. Another enlightening factor is the requirement outlined above that some disciplines like mathematics require absolutist beliefs. In the example, the concept of a moderating variable refers to a methodological notion. Methodology, in turn, is an auxiliary science that is usually taken for granted and that therefore requires absolutist thinking (except the methodological researcher themselves). Taken a step further, similar reasoning applies to almost any teaching situation. When teaching new content in which evaluativist beliefs towards the new content is desirable, there is a set of prior knowledge for which absolutist beliefs are best. Developing the contents of this prior knowledge within the new content would simply not be productive. To sum up, we have the antinomy that fostering evaluativist beliefs requires absolutist beliefs towards the knowledge base accepted. In the context of both the correctness and the productivity of epistemological beliefs, the antinomy vanishes.

However, that distinction calls for further refinement of epistemological beliefs theories, especially for refining the concept of sophisticated epistemological beliefs. Elby and Hammer (2001; see also²) denote sophistication as a flexible adaptation to contextual factors. In our case, such an

²In this paper, we consciously opted not to use the common term sophisticated epistemological beliefs. Elby and Hammer (2001) criticized the normative concept of sophisticated epistemological beliefs. Epistemological beliefs that are commonly termed sophisticated may either not be productive concerning a specific learning goal or may not be correct with regard to the body of knowledge considered by a scientific community. Given this critique, we therefore use the term adequate, i.e., something is satisfactory in quantity and quality (cf., Lexico, 2021c). With regard to evaluativist epistemological beliefs, as mentioned in the main text, they are satisfactory to cope with the knowledge structure in psychology. However, evaluativist beliefs may not always be adequate, please see our consideration in the last paragraph of the Limitations section.

adaptation means to apply absolutist epistemic beliefs concerning the resolution strategy and to open to change one's domain- and topic epistemological beliefs towards evaluativism. In the ideal case, a learner emerges that can flexibly adapt their beliefs depending on the context (cf., Stahl, 2011). Although the issue is not a problem on a theoretical level, it may cause problems from an instructional perspective. Without any instructional measure to cope with the required flexibility, it may be a limiting factor to the effectiveness of the intervention. Further research in this regard is needed.

To conclude, this study revealed that the intervention based on scientific controversies induces an epistemological change in boundaries discussed above. In summary, the results seemingly depend on the measure used to assess epistemological beliefs. For the EBI-AM, there is a decrease in absolutism regardless of the presence of an epistemological sensitization. However, for this measure, there seems to be a backfire effect as there is a raise of multiplism in each condition. Concerning the ETA scales, there is profound evidence for a change in absolutism and multiplism. Still, there is evidence for an increase of evaluativism in the presence of the epistemological sensitization. For the argumentation task, there also seems an increase in evaluativist argumentation when an epistemological sensitization is present. Drawing on this pattern of results and the boundaries discussed in the limitations, the most settled results are the effects of the epistemological sensitization and the intervention of the increase of ETA evaluativism and argumentation skills. These results provide some implications for educational practice in tertiary psychology education. The effects of the EBA-AM absolutism and the backfire support should be interpreted with caution, as they are only supported by this study and need further investigation. However, all results of this study, in combination with the discussed limitations, provide some interesting directions for future research.

A first direction refers to the role of prior knowledge. In the Limitations section, we related the epistemological sensitization to the activation of relevant domain-specific prior knowledge. To explore this theoretical account, research should explore the effects of prior knowledge about a domain's epistemological features on the effects of the sensitization measure in experimental studies. Additionally, a useful addition for a research agenda about the epistemological sensitization refers to the question if the sensitization measure works as an intervention on its own.

A second direction of future research refers to the examples in the controversies. Fictitious examples also constitute a further direction for future research. To achieve a higher ecological validity, real-world examples should be used. This would be the case especially when the goal of future research goes beyond basic questions like uncovering the psychological mechanism of epistemological change.

Also, further research should concentrate on the pre- and posttest measurements with the scenario-bases approach. It should be investigated if the results of interventions depend

on the measurement of epistemological belief using scenarios. For instance, using the controversies intervention, and varying the pre- and posttest measures systematically, the effects of different scenario combinations, i.e., either the same scenario in pre- and posttest or different scenario in pre- and posttest. In the case of different scenarios in the pre- and posttest, sequence effects could also be examined, e.g., one half of the participants would receive the depression scenario first and then the schizophrenia scenario and the other half would receive the reversed sequence.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available because of legal reasons. Participants gave their consent to store, process and analyze the data as well as to the publication of the analysis results. Participants were assured that the data will not be distributed but are in an anonymized version available to researchers on request. Requests for the data set should be directed to Eric Klopp, e.klopp@mx.uni-saarland.de.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EK made the conceptualization and planned the study, developed the epistemological sensitization measure as well as the ETA scenarios, analysed the data the data and wrote up the manuscript. RS supervised the conceptualization, read and commented the draft of the manuscript, and reviewed the analysis results.

ACKNOWLEDGMENTS

We thank Stephanie Knoll and Martin Klein for editing the last version of the manuscript as well as Julia Stöferle for her help with the first draft.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2021.785241/full#supplementary-material>

REFERENCES

- Alexander, P. A. (1996). The Past, Present, and Future of Knowledge Research: A Reexamination of the Role of Knowledge in Learning and Instruction. *Educ. Psychol.* 31, 89–92. doi:10.1080/00461520.1996.10524941
- Astleitner, H., Brünken, R., and Leutner, D. (2003). The Quality of Instructional Materials for Argumentative Knowledge Construction. *J. Instr. Psychol.* 30, 3–11.
- Barzilai, S., and Weinstock, M. (2015). Measuring Epistemic Thinking within and across Topics: a Scenario-Based Approach. *Contemp. Educ. Psychol.* 42, 141–158. doi:10.1016/j.cedpsych.2015.06.006
- Beaujean, A. A. (2014). *Latent Variable Modeling Using R A Step-by-step Guide*. New York: Routledge.
- Bendixen, L. D., and Rule, D. C. (2004). An Integrative Approach to Personal Epistemology: A Guiding Model. *Educ. Psychol.* 39, 69–80. doi:10.1207/s15326985ep3901_7
- Bendixen, L. D. (2002). “A Process Model of Epistemic Belief Change,” in *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*. Editors B. K. Hofer and P. R. Pintrich (Mahwah: Lawrence Erlbaum), 191–208.
- Bonett, D. G., and Wright, T. A. (2000). Sample Size Requirements for Estimating Pearson, Kendall and Spearman Correlations. *Psychometrika* 65, 23–28. doi:10.1007/bf02294183
- Braten, I. (2016). “Epistemic Cognition Interventions. Issues, Challenges, and Directions,” in *Handbook of Epistemic Cognition*. Editors J. Greene, W. Sandoval, and I. Braten (New York: Routledge), 360–371.
- Bromme, R., and Kienhues, D. (2014). “Wissenschaftsverständnis und Wissenschaftskommunikation [Understanding of Science and Scientific Communication],” in *Pädagogische Psychologie [Educational Psychology]*. Editors T. Seidel and A. Krapp (Weinheim: Beltz), 55–80.
- Bromme, R. (2005). “Thinking and Knowing about Knowledge: A Plea for and Critical Remarks on Psychological Research Programs on Epistemological Beliefs,” in *Activity and Sign - Grounding Mathematics Education*. Editors M. Hoffmann, J. Lenhard, and F. Seeger (New York: Springer), 191–201.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum.
- Dietrich, H., Zhang, Y., Klopp, E., Brünken, R., Krause, U.-M., Spinath, F. M., et al. (2015). Scientific Competencies in the Social Sciences. *Psychol. Learn. Teach.* 14, 115–130. doi:10.1177/1475725715592287
- Driver, R. (1978). When Is a Stage Not a Stage? A Critique of Piaget’s Theory of Cognitive Development and its Application to Science Education. *Educ. Res.* 21, 54–61. doi:10.1080/0013188780210108
- Elby, A., and Hammer, D. (2001). On the Substance of a Sophisticated Epistemology. *Sci. Ed.* 85, 554–567. doi:10.1002/sce.1023
- Finney, C., and DiStefan, C. (2013). “Nonnormal and Categorical Data in Structural Equation Modeling,” in *Structural Equation Modeling – A Second Course*. Editors G. R. Hancock and R. O. Mueller (Charlotte: Information Age Publishing), 439–492.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learn. Res.* 4, 28–45. doi:10.14786/flr.v2i2.96
- Fischer, K. W. (1980). A Theory of Cognitive Development: The Control and Construction of Hierarchies of Skills. *Psychol. Rev.* 87, 477–531. doi:10.1037/0033-295x.87.6.477
- Green, H. J., and Hood, M. (2013). Significance of Epistemological Beliefs for Teaching and Learning Psychology: A Review. *Psychol. Learn. Teach.* 12, 168e178. doi:10.2304/plat.2013.12.2.168
- Greene, J. A., Azevedo, R., and Torney-Purta, J. (2008). Modeling Epistemic and Ontological Cognition: Philosophical Perspectives and Methodological Directions. *Educ. Psychol.* 43, 142–160. doi:10.1080/00461520802178458
- Guttman, L. (1945). A Basis for Analyzing Test-Retest Reliability. *Psychometrika* 10, 255–282. doi:10.1007/BF02288892
- Hammer, D., and Elby, A. (2002). “On the Form of a Personal Epistemology,” in *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*. Editors B. K. Hofer and P. R. Pintrich (Mahwah: Lawrence Erlbaum), 169–190.
- Hauser, D. J., Ellsworth, P. C., and Gonzalez, R. (2018). Are Manipulation Checks Necessary. *Front. Psychol.* 9, 998. doi:10.3389/fpsyg.2018.00998
- Hayslip, B., Neumann, C., Louden, L., and Chapman, B. (2006). “Developmental Stage Theories,” in *Comprehensive Handbook of Personality and Psychopathology*. Editors M. Hersen and J. C. Thomas (Hoboken: John Wiley & Sons).
- Heftner, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., and Berthold, K. (2015). Effects of a Training Intervention to Foster Precursors of Evaluativist Epistemological Understanding and Intellectual Values. *Learn. Instruction* 39, 11–22. doi:10.1016/j.learninstruc.2015.05.002
- Hofer, B. K., and Bendixen, L. D. (2012). “Personal Epistemology: Theory, Research, and Future Directions,” in *Handbook of Educational Psychology*. Editor K. Harris (Washington, DC: American Psychological Association), 225–254. doi:10.1037/13273-009
- Hofer, B. K., and Pintrich, P. R. (1997). The Development of Epistemological Theories: Beliefs about Knowledge and Knowing and Their Relation to Learning. *Rev. Educ. Res.* 67, 88–140. doi:10.3102/00346543067001088
- Iordanou, K., Panayiotou, K., and Beker, K. (2016). “Argumentative Reasoning,” in *Handbook of Epistemic Cognition*. Editors J. Greene, W. Sandoval, and I. Braten (New York: Routledge), 39–53.
- Iordanou, K. (2016). Developing Epistemological Understanding in Scientific and Social Domains through Argumentation. *Z. für Pädagogische Psychol.* 30, 109–119. doi:10.1024/1010-0652/a000172
- Kienhues, D., Ferguson, L. E., and Stahl, E. (2016). “Diverging Information and Epistemic Change,” in *Handbook of Epistemic Cognition*. Editors J. Greene, W. Sandoval, and I. Braten (New York: Routledge), 318–330.
- Klopp, E., and Stark, R. (2016). “Persönliche Epistemologien - Elemente Wissenschaftlicher Kompetenz. [Personal Epistemologies - Elements of Scientific Competence],” in *Denken über Wissen und Wissenschaft. Epistemologische Überzeugungen als Gegenstand psychologischer Forschung [Thinking about knowledge and science. Epistemological beliefs as a subject of psychological research]*. Editors A. K. Mayer and T. Rosman (Lengerich: Pabst Science Publishers), 39–69.
- Klopp, E., and Stark, R. (2020). Learning to Argue from Others’ Erroneous Arguments - Fostering Argumentation Competence through Learning from Advocacy Errors. *Front. Educ.* 5, 126. doi:10.3389/educ.2020.00126
- Kuhn, D., Cheney, R., and Weinstock, M. (2000). The Development of Epistemological Understanding. *Cogn. Dev.* 15, 309–328. doi:10.1016/s0885-2014(00)00030-7
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kuhn, D. (1991). *The Skills of Argument*. Cambridge: Cambridge University Press.
- Kuhn, D. (2001). How Do People Know. *Psychol. Sci.* 12 (1), 1–8. doi:10.1111/1467-9280.00302
- Kyriazos, T. A. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psych* 09, 2207–2230. doi:10.4236/psych.2018.98126
- Landis, J. R., and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159–174. doi:10.2307/2529310
- Lexico (2021a). Meaning of Epistemological in English. Available at: <https://www.lexico.com/definition/epistemological> (Accessed December 30, 2021).
- Lexico (2021b). Meaning of Epistemic in English. Available at: <https://www.lexico.com/definition/epistemic> (Accessed December 30, 2021).
- Lexico (2021c). Meaning of Adequate in English. Available at: <https://www.lexico.com/definition/adequate> (Accessed December 30, 2021).
- Mason, L., and Scirica, F. (2006). Prediction of Students’ Argumentation Skills about Controversial Topics by Epistemological Understanding. *Learn. Instruction* 16 (5), 492–509. doi:10.1016/j.learninstruc.2006.09.007
- McArdle, J. J., and Nesselroad, J. R. (2014). *Longitudinal Data Analysis Using Structural Equation Models*. Washington: American Psychological Association.
- McArdle, J. J. (2009). Latent Variable Modeling of Differences and Changes with Longitudinal Data. *Annu. Rev. Psychol.* 60, 577–605. doi:10.1146/annurev.psych.60.110707.163612
- Merk, S., Rosman, T., Muis, K. R., Kelava, A., and Bohl, T. (2018). Topic Specific Epistemic Beliefs: Extending the Theory of Integrated Domains in Personal Epistemology. *Learn. Instruction* 56, 84–97. doi:10.1016/j.learninstruc.2018.04.008
- Muis, K. R., Bendixen, L. D., and Haerle, F. C. (2006). Domain-Generality and Domain-Specificity in Personal Epistemology Research: Philosophical and

- Empirical Reflections in the Development of a Theoretical Framework. *Educ. Psychol. Rev.* 18, 3–54. doi:10.1007/s10648-006-9003-6
- Palmer, B., and Marra, R. (2008). “Individual Domain-specific Epistemologies: Implications for Educational Practice,” in *Knowing, Knowledge, and Beliefs: Epistemological Studies across Diverse Cultures*. Editor M. S. Khine (New York: Springer), 423–441.
- Perry, W. G. (1970). *Forms of Intellectual and Ethical Development in the College Years: A Scheme*. New York: Holt Rinehart & Winston. doi:10.1007/978-1-4020-6596-5_16
- Peter, J., Rosman, T., Mayer, A. K., Leichner, N., and Krampen, G. (2016). Assessing Epistemic Sophistication by Considering Domain-specific Absolute and Multiplicistic Beliefs Separately. *Br. J. Educ. Psychol.* 86 (2), 204–221. doi:10.1111/bjep.12098
- Pieschl, S., Bromme, R., Porsch, T., and Stahl, E. (2008). “Epistemological Sensitisation Causes Deeper Elaboration during Self-Regulated Learning. International Perspectives in the Learning Sciences: Creating a Learning World,” in Proceedings of the Eighth International Conference for the Learning Sciences - ICLS, Netherlands, June 23–28, 2008 (London: Lulu Enterprises), 213220–213222.
- Porsch, T., and Bromme, R. (2011). Effects of Epistemological Sensitization on Source Choices. *Instr. Sci.* 39 (6), 805–819. doi:10.1007/s11251-010-9155-0
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <http://www.R-project.org/>.
- Revelle, W. (2016). *Psych: Procedures for Personality and Psychological Research*. Evanston, Illinois, USA: Northwestern University. Available at: <https://CRAN.R-project.org/package=psych>.
- Richter, T., and Schmid, S. (2010). Epistemological Beliefs and Epistemic Strategies in Self-Regulated Learning. *Metacogn. Learn.* 5, 47–65. doi:10.1007/s11409-009-9038-4
- Rosman, T., Mayer, A.-K., Peter, J., and Krampen, G. (2016). Need for Cognitive Closure May Impede the Effectiveness of Epistemic Belief Instruction. *Learn. Individual Diff.* 49, 406–413. doi:10.1016/j.lindif.2016.05.017
- Rosman, T., Mayer, A.-K., Kerwer, M., and Krampen, G. (2017). The Differential Development of Epistemic Beliefs in Psychology and Computer Science Students: A Four-Wave Longitudinal Study. *Learn. Instruction* 49, 166–177. doi:10.1016/j.learninstruc.2017.01.006
- Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* 48 (2), 1–36. doi:10.18637/jss.v048.i02
- Sadler, T. D. (2004). Informal Reasoning Regarding Socioscientific Issues: A Critical Review of Research. *J. Res. Sci. Teach.* 41, 513–536. doi:10.1002/tea.20009
- Schommer, M. (1990). Effects of Beliefs about the Nature of Knowledge on Comprehension. *J. Educ. Psychol.* 82, 498–504. doi:10.1037/0022-0663.82.3.498
- Schommer-Aikins, M. (2002). “An Evolving Theoretical Framework for an Epistemological Belief System,” in *Personal Epistemology. The Psychology of Beliefs about Knowledge and Knowing*. Editors B. K. Hofer and P. R. Pintrich (Mahwah: Lawrence Erlbaum), 103–118.
- Stahl, E. (2011). “The Generative Nature of Epistemological Judgments: Focusing on Interactions Instead of Elements to Understand the Relationship between Epistemological Beliefs and Cognitive Flexibility,” in *Links between Beliefs and Cognitive Flexibility*. Editors J. Elen, E. Stahl, R. Bromme, and G. Clarebout (Dodrecht: Springer). doi:10.1007/978-94-007-1793-0_3
- Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., and Winne, P. H. (2016). Identity and Epistemic Emotions during Knowledge Revision: A Potential Account for the Backfire Effect. *Discourse Process.* 53 (5-6), 339–370. doi:10.1080/0163853x.2015.1136507
- Weinstock, M. (2006). Psychological Research and the Epistemological Approach to Argumentation. *Informal Logic* 26, 103–120. doi:10.22329/il.v26i1.435
- Zhang, Z., and Liu, H. (2019). “Sample Size Planning for Latent Change Score Models through Monte Carlo Simulation,” in *Longitudinal Multivariate Psychology*. Editors E. Ferrer, S. M. Boker, and K. J. Grimm (New York: Taylor & Francis), 189–211.

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.

Copyright © 2022 Klopp and Stark. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.