

Article

How to Change Epistemological Beliefs? Effects of Scientific Controversies, Epistemological Sensitization, and Critical Thinking Instructions on Epistemological Change

Eric Klopp * and Robin Stark

Department of Education, Saarland University, 66123 Saarbrücken, Germany; r.stark@mx.uni-saarland.de

* Correspondence: e.klopp@mx.uni-saarland.de

Abstract: The present study investigates the combination of an epistemological sensitization and two different critical thinking instructions, i.e., the general and infusion approach, in the context of epistemological change induced by the presentation of resolvable scientific controversies. In a randomized study, we tested the hypothesis that the presentation of resolvable controversies generally reduces absolutism and multiplicism and increases evaluativism. We assume that these effects are strongest when the controversies are presented with an epistemological sensitization and the infusion approach. The results indicate an increase in absolutism when the general approach is employed without an epistemological sensitization. Combined with an epistemological sensitization, the increase in absolutism is only detected when the infusion approach is used. Concerning multiplicism, there is a reduction in all conditions, but the reduction is more effective without an epistemological sensitization. The general approach yields a larger increase in evaluativism without an epistemological sensitization, while the infusion approach fosters evaluativism only in combination with the sensitization. However, an argumentation task revealed that the desired level of an evaluativist argumentation only seems to emerge without an epistemological sensitization in combination with the infusion approach. In sum, the results show that there is no general way to reduce absolutism and multiplicism and increase evaluativism.

Keywords: epistemological beliefs; epistemological change; sensitization; critical thinking; scientific controversies; argumentation

Citation: Klopp, E.; Stark, R. How to Change Epistemological Beliefs? Effects of Scientific Controversies, Epistemological Sensitization, and Critical Thinking Instructions on Epistemological Change. *Educ. Sci.* **2022**, *12*, 499. <https://doi.org/10.3390/educsci12070499>

Academic Editor: Lisa Bendixen

Received: 29 April 2022

Accepted: 14 July 2022

Published: 20 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Several studies have provided evidence that beliefs about knowledge and knowing, i.e., epistemological beliefs, are an important prerequisite for scientific argumentation and thinking, e.g., [1] From a developmental perspective [2], epistemological beliefs encompass three stages: from absolutist to multiplicitist, and finally to evaluativist beliefs. Weinstock [3] characterizes these levels as follows: On the *absolutist* level, an individual believes that only one account of knowledge on the same issue is true. Other accounts are believed to fail due to erroneous or biased thinking. On the *multiplicitist* level, a person believes that there are many accounts of knowledge on the same issue without determining their correctness. Every one of these accounts may be true because these accounts are mere opinions. Finally, on the *evaluativist* level, a person constructs an account of knowledge on an issue based on *evidence*. To do that, pieces of evidence have to be weighed against each other to pave the way for a reasoned decision for one particular account. From a normative perspective, the *evaluativist* level is seen as the most sophisticated level of epistemological development. The well-known and extensive work of Kuhn [2,4] showed that evaluativist epistemological beliefs are, in particular, a prerequisite to achieving an advanced level of scientific *argumentation*. Proper scientific argumentation requires that evidence on

which the assertion rests is weighed appropriately and brought in to support the argument. Such an evaluation of supporting and contradicting evidence requires a proper understanding of how knowledge is generated. Evaluativist epistemological beliefs, i.e., the notion that pieces of evidence should be weighed against each other to come to a reasoned decision for the respective point of view, are particularly important for proper scientific argumentation. Aside from scientific argumentation, epistemological beliefs are also important when scientific information has to be evaluated. For instance, Feinkohl, Flemming, Cress, and Kimmerle [5] found that sophisticated epistemological beliefs are prerequisites for the critical evaluation of new scientific information provided in a journalistic article. Thus, epistemological beliefs are part of the concept of scientific competence [6] and have a central role in handling scientific evidence. This concerns students, scientists, and laypeople who have to rely on scientific knowledge. Teachers in particular can be regarded as laypersons who have to use scientific information regularly, e.g., when designing lessons or handling other school-related situations. In Germany, the Standing Conference of the Ministers of Education and Cultural Affairs of the States in the Federal Republic of Germany [7] requires that teachers and student teachers critically reflect on teaching and learning processes drawing on educational knowledge, a considerable part of which is knowledge from educational psychology. Thus, a large amount of teacher education consists of providing educational psychological knowledge.

However, in the domain of psychology in general and in educational psychology in particular, knowledge is *ill-structured*, i.e., it consists of many different theories, paradigms, concepts, and empirical results, which are partially contradictory. In other words, this content of psychological knowledge forms scientific controversies. According to Rosman et al. [8], such an ill-defined knowledge structure requires evaluativist epistemological beliefs to properly understand and address the contradictions.

But this ill-defined knowledge structure confronts students with more general problems. Klopp and Stark [9] argue that novice students in particular cannot cope with these kinds of inconsistencies due to their lack of knowledge of the philosophy of science and methodological concepts. Following Rosman et al. [8], these problems result in the development of multiplicitist epistemological beliefs, i.e., the belief that knowledge consists merely of opinions of which everyone is true in their own right. The same authors have also provided evidence that psychology students show, on average, a high level of multiplicity in the first three semesters, which decreases afterward. It can be assumed that psychology students have acquired enough philosophy of science and methodological knowledge after the first three semesters to develop evaluativist epistemological beliefs.

Nonetheless, this is a problem in other areas such as teacher education, where, primarily, content knowledge of educational psychology is provided, but the philosophy of science and methodological background is lacking. Therefore, student teachers can develop multiplicitist beliefs that interfere with the required competence to reflect critically on teaching and learning processes drawing on educational knowledge. However, this applies not only to student teachers but to all laypersons who have to use (educational) psychological knowledge to judge scientific controversies. Thus, interventions have to be developed and evaluated in this context to foster evaluativist epistemological beliefs.

Hereafter, we describe the theoretical background of an intervention designed for student teachers. The intervention draws on the presentation of psychological controversies to elicit epistemological change towards evaluativist beliefs [10]. We also describe a measure that aims to improve the intervention's effectiveness, i.e., an epistemological sensitization, cf., [9,11] and critical thinking instructions [12]. This intervention is evaluated with a sample of student teachers.

2. Theoretical Background

2.1. Approaches to Epistemological Beliefs

Epistemological beliefs can be conceptualized in two different ways [8,13]: the developmental approach, on the one hand, and the dimensional approach, on the other hand. In the developmental approach, a sequence of qualitatively different levels depicts the developmental trajectory of epistemological beliefs. Kuhn's [2] model, as described in the Introduction section, defines the three consecutive levels *absolutism*, *multiplism*, and *evaluativism*. In the dimensional approach, epistemological beliefs relate to interindividual differences, as it is known in personality research [14]. That is, epistemological beliefs are considered as dimensions of interindividual differences, an approach based on the work by Schommer [15]. The most widely used model in the dimensional approach is the one by Hofer and Pintrich [16]. The authors conceptualize four dimensions of knowledge: *Certainty of knowledge*, *Simplicity of knowledge*, *Source of knowledge*, and *Justification of knowledge*.

Weinstock [3] combined the developmental and dimensional approaches, resulting in the integrated approach, which is based on the idea that the developmental and dimensional approaches are not different but simply refer to two sides of the same coin. Weinstock [3] characterizes the developmental levels of epistemological beliefs as a specific profile of the dimensions in the model of Hofer and Pintrich. Barzali and Weinstock [17] presented an integration of both approaches based on a summary of recent models of interindividual differences in epistemological beliefs. In their model, each level of epistemological development is considered to be a specific profile 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 shows the characteristics of each dimension for each level of epistemological development.

Table 1. Description of the epistemological beliefs dimension for each developmental level according to Barzilai and Weinstock [17] (the table is adapted from [9]).

Dimension	Absolutism	Multiplism	Evaluativism
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 perspectives 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

The integrated approach allows the topic of epistemological change to be studied from either a dimensional or a developmental stage perspective. In this paper, we focus on the developmental stage perspective to fit this study in the previous research efforts of Rosman et al. [8] and Klopp and Stark [9]. In addition to this, epistemological change as the temporal development of epistemological beliefs results from an asynchronous change in individual profiles of epistemological belief dimensions [18]. The integrated approach thus provides a new perspective on the developmental approach. Instead of assuming discontinuous, qualitatively different levels over all kinds of knowledge, the integrated approach highlights development as a continuous process. The integrated approach proposes steady transitions of developmental levels based on different quantitative profiles instead of qualitatively different levels with abrupt transitions. In contrast to the classical developmental approach, which allows only the categorization of an individual as being an absolutist, multiplicitist, or evaluativist, e.g., [19]. The integrated approach allows the assessment of these three levels simultaneously while still building on the individual epistemological beliefs profile, e.g., [3,9,16,20]

2.2. Process Model of Epistemological Change

The *Process Model of Personal Epistemology Development* [21,22] describes the conditions that yield changes in epistemological beliefs. The model outlines necessary mechanisms to induce change from current beliefs to more adequate epistemological ones. These mechanisms are *epistemological doubt*, *epistemological volition*, and *resolution strategies*. Figure 1 depicts the process model. Firstly, epistemological doubt refers to the process that emerges when individuals recognize the dissonance of their current epistemological beliefs and have to reconcile these new experiences with their epistemological beliefs [22]. Beyond that, the new experience must be of personal relevance, i.e., the individual is interested in the outcome or the topic itself. Secondly, an essential factor for epistemological change is epistemological volition. This relates to focused effort as a consequence of experiencing epistemological doubt, to change the current beliefs with regard to the affordances and constraints of the new experience. Complementing these first two mechanisms, resolution strategies are the last mechanism for actual epistemological change. According to Bendixen [21], resolution strategies are both social interactions and reflection processes. In social interactions, individuals are confronted with different opinions and perspectives that lead to revising of their current beliefs. In reflection processes, a person first reviews past experiences and then considers current epistemological beliefs to analyze their implications, which can also lead to an epistemological change. Lastly, resolution strategies aim to reduce epistemological doubt introduced by the presentation of the controversies. Resolution strategies should push individuals further to evaluativism. According to Rosman and Kerwer [23], resolution strategies are elaborative processes that deal with epistemological doubt, e.g., the integration of the contradicting information from the controversies and one's own epistemological beliefs.

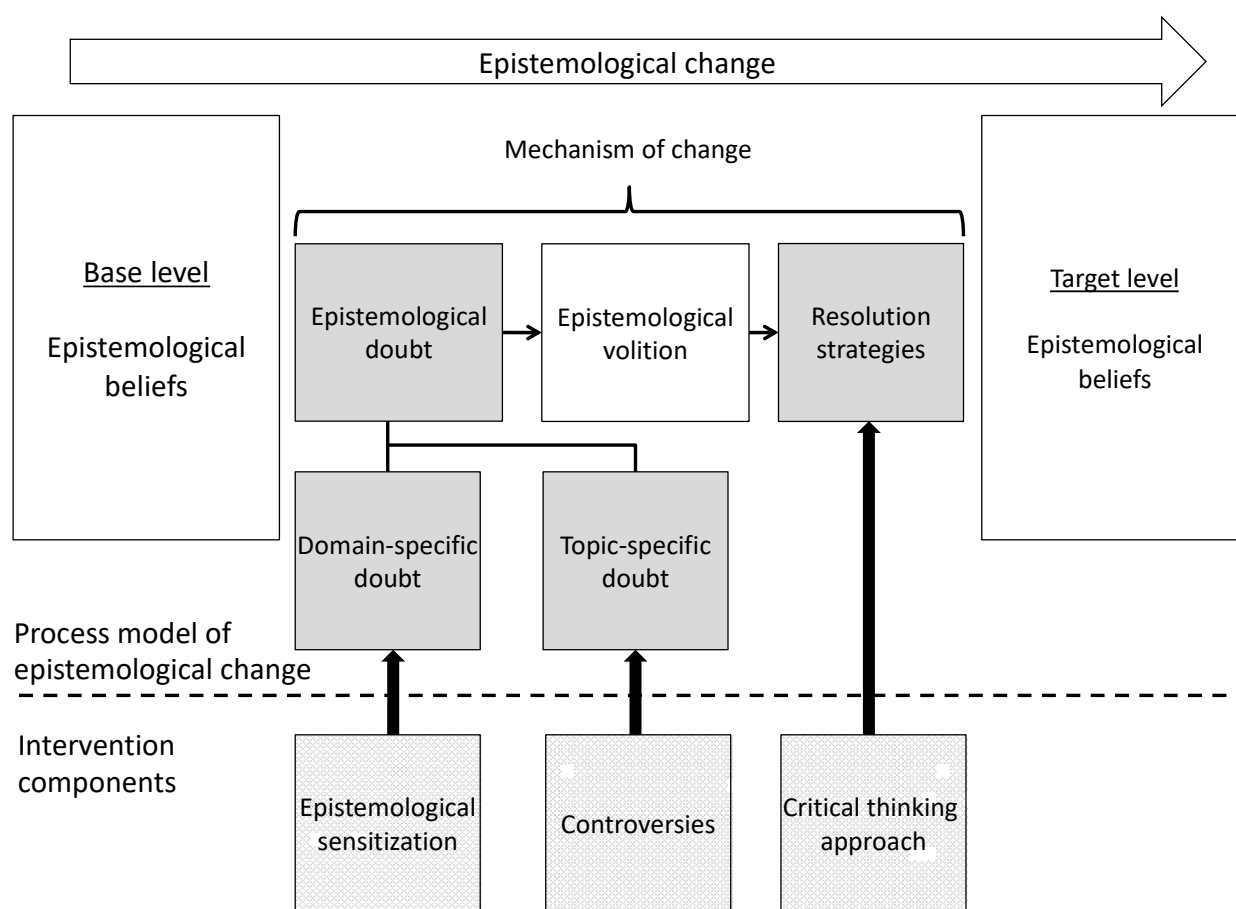


Figure 1. Components of the process model of epistemological change and the corresponding elements of the experimental variation, or intervention.

The process model of epistemological change targets epistemological change, regardless of the actual conceptualization of epistemological beliefs. In this study, we draw on the notion of epistemological change as the temporal development of epistemological beliefs resulting from an asynchronous change in individual profiles of epistemological belief dimensions.

2.3. Resolvable Controversies—An Intervention Concept

Rosman et al. [10] developed an intervention based on the process model. By presenting *resolvable controversies*, the intervention attempts to elicit epistemological doubt. Resolvable controversies are seemingly contradictory theories or empirical findings that can be resolved when other theories, variables, or paradigms are considered, cf., [24]. In this intervention, a resolution to this controversy is provided after the presentation of said controversy. The controversies aim to induce epistemological doubt, and the resolution strategies aim to induce epistemological change. The basic notion of this intervention concept is that the resolution of controversies is incompatible with absolutism and multiplicity [10]. As there is only one correct account for absolutists, they would deny the sheer possibility of resolving scientific controversies. Multiplicists, on the other hand, would not even take the existence of these controversies into account, as they claim that controversies are scientists' personal opinions that cannot be resolved. Thus, the intervention should minimize absolutist and multiplicitist epistemological beliefs because the resolution only fits the level of evaluativism. Rosman et al. [10] provided evidence for the effectiveness of this intervention. In a pre–post test design, the authors presented six controversies and their solution strategies. These controversies were fictional to prevent effects of prior

knowledge but referred to scientific issues. The authors were able to show a reduction in absolutist and multiplicitist epistemological beliefs due to the intervention. However, because the authors used the EBI-AM questionnaire [20], which only assesses absolutist and multiplicitist beliefs, they could not provide results regarding the change of evaluativist beliefs.

There is the question of if there are measures that alter the effectiveness of the controversy intervention. A possible starting point is the various components of the process model on which this intervention concept relies. Klopp and Stark [9] presented an epistemological sensitization that was aimed at epistemological doubt. In particular, they focused on the induction of domain-specific doubt as a facilitator for the topic-specific doubt that should result from the presentation of the controversies. That way, the epistemological sensitization is an add-on that is applied in combination with the controversy intervention. Another approach to fine-tune the effectiveness of the intervention consists of the resolution strategies. The resolution strategies can be reframed as an affordance to find the reasons for the controversy by using critical thinking. Critical thinking instructions have been found to alter epistemological beliefs [12]. In contrast to the epistemological sensitization, applying critical thinking instructions to the resolution strategy is not an add-on but an alteration of the method of how to resolve the controversy. In the following, we describe the concepts of epistemological sensitization and approaches to critical thinking as measures to potentially enhance the effectiveness of the controversy intervention.

2.4. Epistemological Sensitization

The process model of epistemological change consists of three necessary components to induce epistemological change. Rosman et al. [10] intervention concept aims at inducing epistemological doubt that should afterward be reduced by providing resolution strategies that aim to pave the road towards evaluativism. According to the TIDE model [25], epistemological beliefs range from *domain-general* to *topic-specific* beliefs. Accordingly, presenting a controversy referring to a specific topic may induce doubt about *topic-specific* epistemological beliefs. Thus, the question arises whether inducing *domain-general* epistemological doubt before presenting the controversies is beneficial for epistemological change or not [9].

According to Klopp and Stark [9], one possible method of inducing *domain-general* epistemological doubt is using an epistemological sensitization measure [11]. Porsch and Bromme [11] introduced the idea of epistemological sensitization in the context of research concerning source choices. The authors conducted an experiment in which they used two variations of epistemological sensitization measures. In the *naïve* sensitization, knowledge is characterized as structured and static, and the accurateness of scientific models was emphasized. In the *sophisticated* sensitization, selected facts and their interconnection are highlighted and, also, the epistemological features of knowledge were presented, e.g., the existence of scientific controversies, and thus targeting an evaluative view of knowledge. The sensitization's intention was to elicit either naïve or sophisticated epistemological beliefs. Porsch and Bromme [11] linked "sophisticated" epistemological beliefs to the use of more sources and investigated the number of used sources in the evaluation of a text dealing with tides. According to Porsch and Bromme [11], epistemological sensitization is a heuristic concept aiming to elicit specific epistemological features in a specific situation. The authors found that participants used more sources in the sophisticated sensitization condition than in the naïve condition.

Even though Porsch and Bromme's [11] epistemological sensitization is topic-specific, the principle can be generalized to any given level of domain-specificity [9]. Following the notion of the TIDE model, a topic-specific sensitization refers to the epistemological features of a given topic. In contrast to this, a domain-specific sensitization would refer to the features of the domain to which the topic belongs. Consequently, the features of a topic are always an instance of the epistemological features of its domain. For instance, if the effectiveness of an instruction method depends on a third moderator variable such as

the age of the children, then this is a topic-specific instantiation of the domain-specific epistemological feature of moderator variables. According to [9], presenting the controversies raises *topic-specific* epistemological doubt, whereas presenting an epistemological sensitization before the controversies introduces *domain-specific* epistemological doubt. The authors assume this to be beneficial to foster epistemological beliefs towards evaluativism, while, at the same time, absolutism and multiplicism are reduced. The authors assume that an epistemological sensitization yields a deeper elaboration and more elaborate processes, and spending more time on the task, which, in turn, should be beneficial for epistemological change, cf., [26].

A proper sensitization measure includes presenting basic epistemological features of the domain, such as the main reasons why controversies exist. Sensitization should only focus on the general epistemological aspects of domain knowledge rather than on topic-specific aspects provided in the controversies. However, it must be ensured that the controversies presented later reflect the epistemological features introduced in the sensitization measure.

Empirical support for the effectiveness of this kind of epistemological sensitization comes from Klopp and Stark [9]. In a pre–post test design, a computerized version of Rosman et al. [10] intervention was used. In the first condition, participants received only the intervention. In the second condition, participants received an epistemological sensitization combined with the intervention. In the control condition, the participants received a task not related to epistemological beliefs. Klopp and Stark [9] used the ETA questionnaire [17] that measures absolutism, multiplicism, and evaluativism as dependent variables, drawing on the integrated approach. The ETA is a scenario-based measure, i.e., it provides a controversy in combination with several individual items to measure the three levels of epistemological beliefs. In addition to the ETA scales, Klopp and Stark [9] had their participants write an essay about the controversy provided with the ETA scales. They rated whether the essay had an absolutist, multiplicitist, or evaluativist argumentation style.

Klopp and Stark [9] did not find any evidence of epistemological change in absolutism and multiplicism, i.e., their findings did not indicate a reduction or an increase in these beliefs. However, there was a significant increase for evaluativism in the condition with the epistemological sensitization, as opposed to the other conditions. Regarding the essay, while an increase in argumentation skills was found in both intervention conditions, an evaluativist argumentation level was only shown for participants in the condition with the epistemological sensitization.

2.5. Epistemological Beliefs and Critical Thinking Instruction

However, an epistemological sensitization is not the only method to induce epistemological change. Since one major component of the controversy intervention is the presentation of the resolution strategies, any method that allows an individual to come to a resolution should be beneficial for epistemological change. To elaborate on the resolution strategies, possible methods can be sourced from the research on fostering critical thinking. Critical thinking means purposeful reflection and reasoning about what to believe or how to act when confronted with complex issues. In doing so, relevant context features have to be considered [27]. This demonstrates the close relation between epistemological beliefs and critical thinking [6,28]. Greene and Yu [29] argue that epistemological beliefs can activate or deactivate critical thinking, and Valanides and Angeli as well as Muis and Duffy [12,30] have presented evidence that critical thinking alters epistemological beliefs. Critical thinking is essentially related to the developmental account of epistemological beliefs. According to Kuhn and Dean [31], only one side of a given controversy can be correct on the absolutist level. The other side of the issue is necessarily false due to biased thinking. In contrast, critical thinking is a tool to decide which one is correct depending on contextual circumstances. Thus, the use of critical thinking to resolve a controversy is incompatible with absolutist beliefs. On the multiplicitist level, a person believes

that there are many accounts of knowledge, each of which is a mere opinion, so that there is no need to determine their correctness, and critical thinking becomes irrelevant. However, on the evaluativist level, a person constructs an account of knowledge based on *evidence*, and pieces of evidence should be weighed against each other to pave the way to a reasoned decision. Thus, critical thinking is used to determine the validity of judgments about the presented evidence on the evaluativist level. Based on this direct relation, it should be possible to use critical thinking interventions to foster epistemological change.

According to Ennis [32] the general and the infusion approach are the two major approaches to teaching critical thinking. (Actually, Ennis [32] mentions four interventions approaches: the general, mixed, infusion and immersion approach. However, in this article, we consider only the general and infusion approach.) Abrami et al. [33] point out that, with regard to the general approach, critical thinking skills are separated from the content of a specific subject matter. In contrast, in the infusion approach, critical thinking skills are inseparable from the content of a specific subject matter. Valanides and Angeli [12] (p. 317) note that the description of the two approaches does not “provide any detailed guidance about the instructional design of each strategy”. Nevertheless, concerning controversies, the above description suggests that critical thinking principles should be applied to each side of the controversies in the infusion approach because of the combination of critical thinking and subject matter. Therefore, the resolution of the controversies is put into an argumentative context, in which a person has to consider the pros and cons of each side; cf., [34]. This perspective also highlights that the infusion approach relates directly to the resolution strategies. In the general approach, the critical thinking methods are set apart from the resolution of the controversies. The infusion approach should therefore be better suited to promoting epistemological change. This is supported by the results of Valanides and Angeli [12], who applied critical thinking instructions and analyzed their effects on epistemological change. The participants had to use critical thinking on controversial topics. Their results indicated that the infusion approach induced a significantly greater epistemological change than the general approach.

From these considerations, the question emerges whether critical thinking methods can be integrated into Rosman et al. [10] intervention concept. As the major component of this intervention concept consists of the resolution of the controversies, a purposeful reflection on the controversy and why it exists might drive epistemological change. Moreover, critical thinking may also be conceived as an additional measure to elicit epistemological doubt. As both the controversies and critical thinking are incompatible with absolutist and multiplicist epistemological beliefs, the instruction to think critically after reading a controversy could elicit epistemological doubt. From this perspective and in line with the findings of Valanides and Angeli [12], it seems plausible that the infusion approach is better suited to inducing epistemological change than the general approach.

2.6. The Current Study

The current study investigates how epistemological sensitization and critical thinking instructions affect epistemological change. We use resolvable controversies as the basic intervention to induce epistemological change. The aim of the current study is to explore in which way the epistemological sensitization and the critical thinking approach affect the controversy intervention. The epistemological sensitization aims to foster the induction of epistemological doubt, whereas the critical thinking approach aims to affect the resolution strategy. Drawing on the reasoning of Klopp and Stark [9] and Rosman et al. [10], we examine the following four hypotheses.

Hypothesis 1 (H1): The intervention reduces absolutism and multiplicism.

Hypothesis 2 (H2): The intervention fosters evaluativism.

Concerning the reasoning behind the epistemological sensitization and the critical thinking approach, we hypothesize that:

Hypothesis 3 (H3): The intervention is more effective when an epistemological sensitization is present.

Hypothesis 4 (H4): The intervention is more effective when using the infusion approach instead of the general approach.

Additionally, as a relation exists between epistemological beliefs and argumentation (Fischer et al., 2014), we did not only consider measures of epistemological beliefs but also a measure of argumentation as a form of enacted epistemological beliefs; cf., [6] and [35]. This is in line with Klopp and Stark [9], who also use a measure of argumentation. For this measure, we hypothesize the following three hypotheses.

Hypothesis 5 (H5): The intervention raises the epistemological level of the argumentation.

Hypothesis 6 (H6): The intervention is more effective when an epistemological sensitization is presented.

Hypothesis 7 (H7): The intervention is more effective when the infusion approach is used.

3. Methods

3.1. Sample, Design, and Procedure

The sample consisted of 106 student teachers (23 male) from a southwestern German university. The mean age was 22.50 ($sd = 4.30$), and the median semester was 2 (range 1). The participants were recruited in mandatory courses in psychology and took part voluntarily.

The study had a 2×2 pre–post test design; Figure 2 shows an overview of the procedure and the design. The participants were randomly allocated to four experimental conditions. The experimental conditions resulted from the combinations of the presence of an epistemological sensitization (Factor: SENS) and the critical thinking approach (Factor: APPR). The epistemological sensitization factor consisted of an epistemological sensitization condition (S), in which the participants received a sensitization text, and a neutral condition (N), in which the participants received a neutral text. The critical thinking approach factor consisted of a general condition (G) and an infusion condition (I), in which the epistemological intervention was designed according to the critical thinking approach. Due to the 2×2 design, there are four conditions to which we refer to hereafter as NG, NI, SG, and SI.

The participants took part in groups. In each group, a participant was assigned to one of the four conditions. The experiment spanned over four sessions and the sessions were conducted exactly one week apart. In the first session, participants completed the pretest consisting of demographic questions, the epistemological beliefs measures, and an argumentation task. The participants also received a unique and anonymous identifier code. In the second session, the participants were randomly allocated to one of the four experimental conditions and worked on two scientific controversies. In the third session, participants worked on three more scientific controversies. In the fourth and last class session, the participants received the posttest consisting of the epistemological beliefs measure and an argumentation task. The participants worked individually in a self-paced manner on the material. Each session lasted 90 min. All materials were presented as booklets, and students were allowed to write in them. The identifier code was used to ensure the assignment of the test and the experimental conditions.

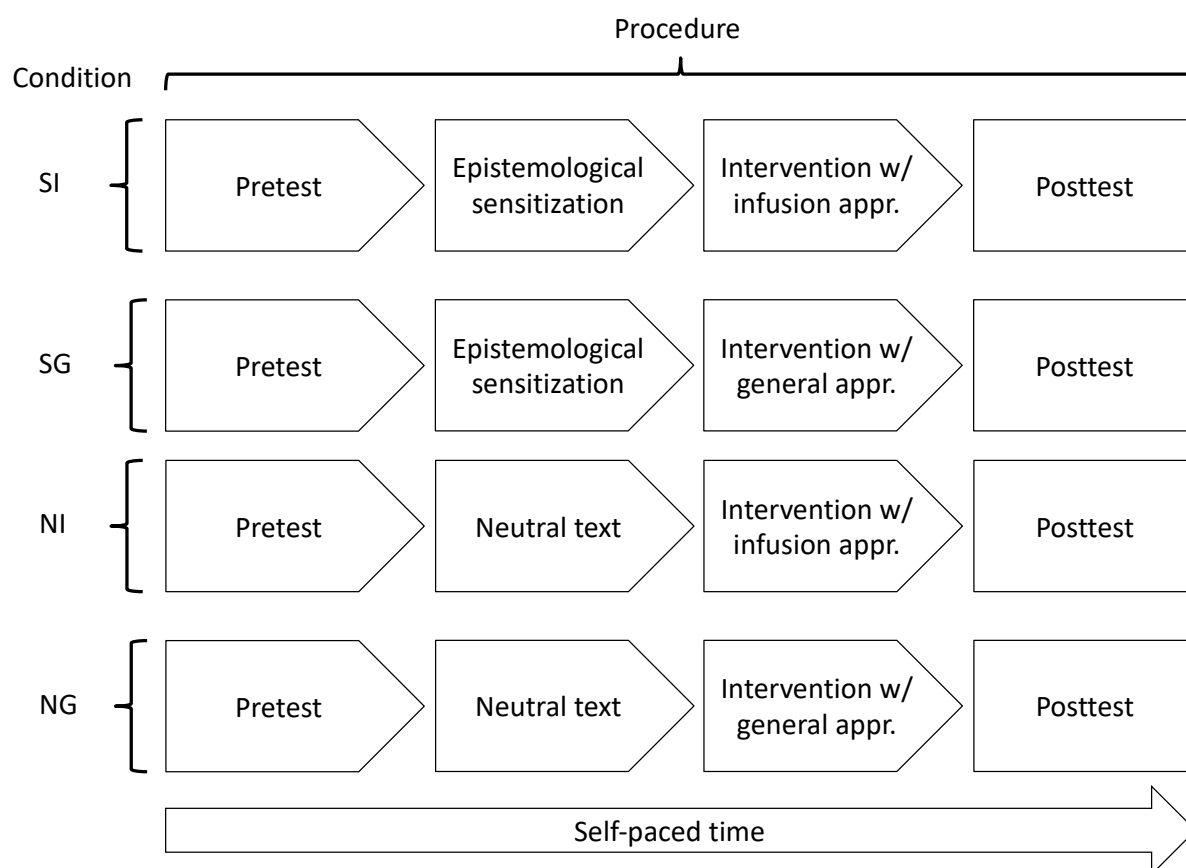


Figure 2. Design and procedure.

3.2. Epistemological Intervention and Experimental Variation

3.2.1. Epistemological Intervention

The epistemological intervention in the presentation consisted of scientific controversies. Our implementation of this intervention approach was based on authentic examples of scientific controversies, in contrast to Rosman et al. [10], who used fictive examples, cf., Klopp and Stark [9]. The intervention consisted of five controversies from the domain of educational psychology. The content of the examples was rated as relevant for teacher education students by two independent experts of teacher education. Additionally, the examples were selected to be understood without specialized educational psychology knowledge. The topics of the controversies and the reasons for them are presented in Table 2. Before each controversy, a short explanation outlined its relevancy for teacher education students. Each controversy consisted of two paragraphs. The first paragraph featured one side of the controversy, whereas the second paragraph presented the other side. After this, the participants received a task designed according to the infusion or general approach (see section Critical Thinking Approach).

Table 2. Overview of the topics in the intervention.

Topic	Cause for Conflicting Claims
Two types of dictation	Moderator effect moderating the effectiveness for different age types
Media use and aggression	Experimental, artificially induced effects of media violence with low ecological validity versus longitudinal effects with high ecological validity
Computer use in school	Effects of the use of computers on quality of text production quality in different languages (English vs. Chinese); incommensurable results due to different writing systems

Open learning	Different effects of open learning, results for academic achievement or affective learning outcomes
Development of performance anxiety	Different theoretical perspectives on the development of performance anxiety; different theoretical scopes

3.2.2. Epistemological Sensitization

The epistemological sensitization text consisted of an adaption of a textbook chapter covering the topic of conflicting scientific claims [36]. It was rewritten to address the reasons for conflicting claims in educational sciences and was designed to match the cause of the controversies presented to the participants. The text presented the issue either in the form of conflicting theories or conflicting empirical results. Before discussing the reasons for conflicting results, the text explained why conflicting results matter for teacher education students. Altogether, three reasons for conflicting results were provided. Firstly, the text stated that the interpretation of data always occurs on the basis of theoretical assumptions and that, consequently, different assumptions may yield different interpretations of the same data. Secondly, the text discussed the reversibility of educational theories or empirical results and whether the current state of knowledge depends on a consensus of the scientific community. It also emphasized that there may exist knowledge for which there is a high degree of consensus and that there exists knowledge, e.g., referring to new developments, for which there is only low consensus. Thirdly, different theoretical perspectives were provided as reasons for conflicting results. There was also a summary stating the importance of making one's own judgment when conflicting theories or results exist. The epistemological sensitization text had 569 words and a Flesch reading score of 35.

The neutral text featured a definition of educational sciences and a description of its various subdisciplines (educational psychology, pedagogy, education sociology, and educational economics). Each description contained a summary of the topics that the respective subdiscipline deals with, and care was taken not to mention anything concerning the topic of conflicting claims. The neutral text was shorter than the epistemological sensitization text; it had 409 words and a Flesch reading score of 12.

3.2.3. Critical Thinking Approach

In the infusion condition, the participants were instructed to use a questioning scheme after reading the first controversy. The scheme consisted of four questions, and a short explanation as well as an instruction accompanied each question. The first question asked students to locate the contradiction of the two presented claims. The participants were instructed to rephrase the contradiction in their own words. The second question had the students name possible reasons for the contradiction. Participants were asked to examine either the provided data, the variables involved, the stated assumptions, or the theoretical background of each claim. The third question asked participants to work out the weighting of each reason they had formulated in the second question. The fourth question drew on the answers to the second and third questions and asked the participants to determine the most plausible reason for the controversy. In the following four controversies, the same scheme of explanations and instructions was provided after each controversy. The participants were instructed to write down their answers.

In the general condition, after each controversy, the participants received a task to elaborate on the reasons for a contradiction between two presented claims. The participants were instructed to write down what they believed to be possible reasons. After each controversy, the same task was given again, and only the wording of the introduction sentence was adapted to the content of the controversy.

An elaborated sample resolution of the controversy was provided in each condition after the participants found their own resolution. The sample resolution contained a detailed description of the reasons why a controversy emerged. This was to ensure that the

participants had the necessary prerequisite for epistemological change to occur and to prevent adverse motivational effects in case the participants could not resolve the controversy. Following this, the participants received a short summary of the sample resolution in the form of a take-home message.

3.3. Dependent Variables

3.3.1. Epistemological Beliefs

To measure epistemological beliefs, we applied the ETA, a scenario-based questionnaire assessing domain-specific absolutism, multiplicity, and evaluativism according to the definitions provided above [17]. The ETA draws on the integrated approach to epistemological beliefs. The three developmental levels are considered a composition of the following nine dimensions (see Table 1): *Right answer*, *Certainty of knowledge*, *Attainability of truth*, *Nature of knowledge*, *Source of knowledge*, *Multiple perspectives*, *Evaluate explanations*, *Judge accounts*, and *Reliable explanation*. There is one item assessing each of the three levels of absolutism, multiplicity, and evaluativism for each of these dimensions, resulting in 27 items.

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. Higher values indicate a higher level of the respective belief. Scores for absolutism, multiplicity, and evaluativism were calculated as the mean score for each scale. Thus, higher mean scores indicate a higher level of absolutism, multiplicity, and evaluativism. An item analysis indicated that the multiplicity items belonging to the dimension *Right answer* correlated negatively with the corrected mean score for both the pretest and the posttest (a result also reported by Klopp and Stark [10]). We therefore dropped the *Right answer* items from the analysis. The final set consisted of 24 items, with 8 measuring absolutism, multiplicity, and evaluativism, respectively. The final scales had good internal consistencies in terms of Cronbach's α (see Table 3). In the following, we use the abbreviations ABS1, MULT1, and EVAL1 to refer to the absolutism, multiplicity, and evaluativism scores in the pretest as well as ABS2, MULT2, and EVAL2 to refer to the scores in the posttest.

Table 3. Means, standard deviations, correlations, and internal consistencies for the ETA scales and the median and interquartile ranges for the essay scores.

	M	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ABS1	3.69	0.70	0.79						
MULT1	3.12	0.79	−0.58 ***	0.84					
EVAL1	4.39	0.49	0.39 ***	−0.32 **	0.67				
ABS2	3.89	0.75	0.63 ***	−0.41 ***	0.22 *	0.83			
MULT2	2.79	0.77	−0.44 ***	0.74 ***	−0.26 **	−0.56 ***	0.83		
EVAL2	4.43	0.58	0.13	−0.19	0.46 ***	0.21 *	−0.11	0.72	
ESSAY1	3.00	2.00	−0.11	0.06	−0.19	−0.11	0.10	0.01	−
ESSAY2	4.00	1.00	0.02	−0.06	−0.03	0.05	−0.01	0.12	1.00 ***

Note. The internal consistencies (Cronbach's α) are given on the main diagonal for the ETA scales, the correlation between the ETA scales and the essay scores; the two essay scores are Spearman correlations. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Before the participants work on the items, they read a controversy and the items themselves refer to this scenario. The items are formulated to prompt the participants to reason about the controversy. We adapted the scenarios used in the ETA to the domain of educational psychology, as the original scenarios provided by the authors referred to the domains of history and biology. Thus, the items are firstly embedded in the scenario's specific domain, and secondly, in the scenario's topic. Therefore, the ETA enables measurement of domain-specific epistemological beliefs combined with some topic-specific aspects of epistemological beliefs [17] (p. 144).

The scenarios were framed in the domain of educational psychology to match the presented controversies. The first scenario provided conflicting statements on computer-based learning and had 276 words. Its Flesch reading score was 34. The second scenario contained a controversy about grouping students according to their achievements. It had 288 words and a Flesch reading score of 35.

3.3.2. Argumentation Task

We followed Klopp and Stark [10] by expanding the ETA with an argumentation task to assess the participants' argumentation skills. In the pretest and posttest, the participants were instructed to write an essay about the scenarios presented in the ETA. We used the coding scheme from Klopp and Stark [10]. The coding scheme has a 6-point-rating scale and indicates if the essays reflect an absolutist (1–2 points), multiplicist (3–4 points), or evaluativist (5–6 points) argumentation. Table 4 presents an overview of the coding criteria. In the following, we use the labels ESSAY1 to refer to the pretest essay score and ESSAY2 to refer to the posttest essay score. The coding procedure was conducted as follows: A third person not involved in the study removed any information if the essays stem from either the pre- or posttest and to which condition they belong. Then, the first author and another student research assistant started coding. The research assistant was involved in the conducting of the study, but the coding research assistant was introduced to the theoretical aspects of argumentation and epistemological beliefs. Each rater rated all essays, and afterward, their ratings were compared. We used Cohen's κ to measure interrater agreement, and we resolved cases of disagreement through discussion. The interrater agreement is $\kappa = 0.75$ in the pretest and $\kappa = 0.85$ in the posttest, indicating moderate to substantial agreement [37].

Table 4. Coding criteria according to Klopp and Stark [10]

Points	Level	Criterion
1	ABS	The participant indicates one-sided, i.e., one side of the controversial topic is correct.
2		The participant indicates one-sided but indicated that there may be a second point of view which is equally right.
3	MULT	The participant indicates that both points of view are correct.
4		The participant indicates that both points of view are correct but there is a possibility that depending on the circumstances, one point of view may be more suitable than the other.
5	EVAL	The participant indicates that the available evidence has to be evaluated according to the given circumstances on which the point of view is correct.
6		The same as the criterion for five points but the participant indicates that both points of view may change according to new research.

Note. ABS: absolutism, MULT: multipicism, EVAL: evaluativism.

3.4. Statistical Analysis and Sample Size Considerations

All of the analyses were carried out with R [38] using the packages psych [39], lavaan [40], car [41], and simsem [42]. We used a latent change score regression model to analyze epistemological change within the ETA scores (LCSRM [43]). The LCSRM models between-person differences in within-person changes that are essential when changes occur after the initial measurement. This is the case in experiments when the manipulation occurs between the first and the second measurement. In LCSRM, the change score is regressed on the pretest measurement to provide a base-free measurement of change [43]. Figure 3 depicts a conceptual diagram of the LCSRM.

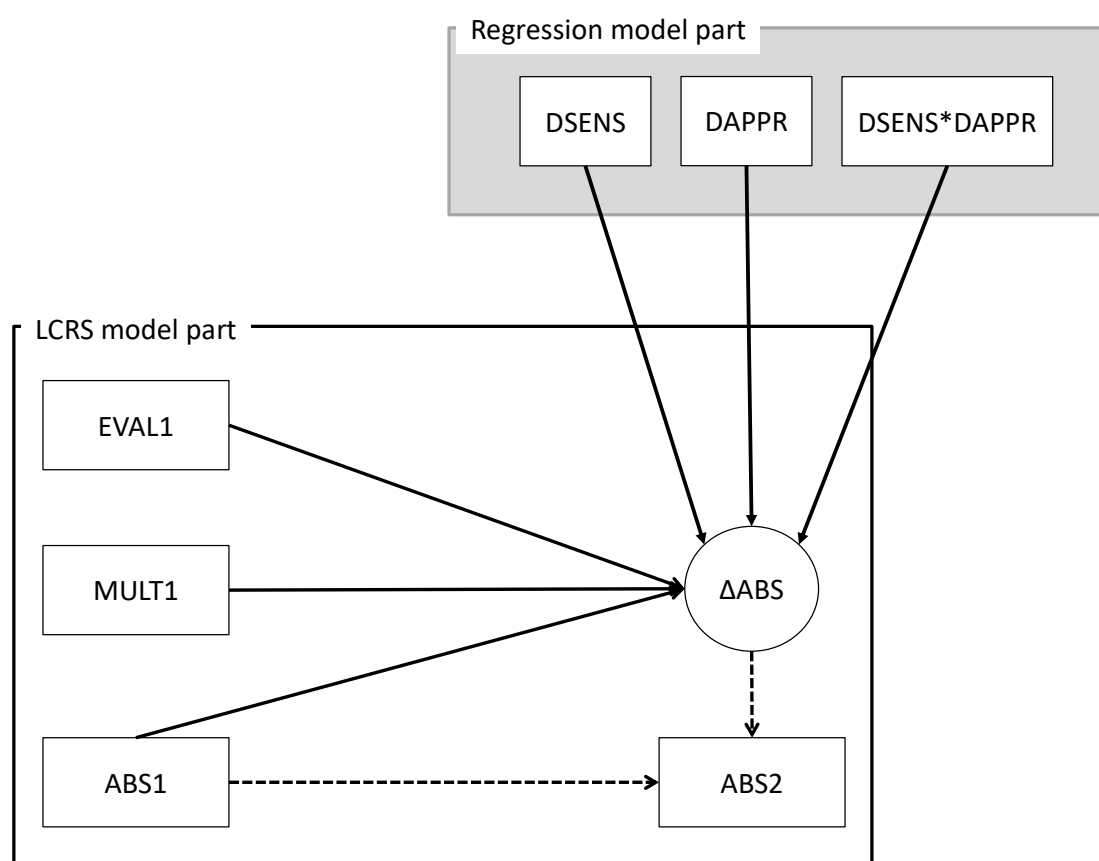


Figure 3. Conceptual diagram of LCSR model in combination with the ANOVA. Note. Δ ABS represents the latent change regression score for absolutism. The star symbol * denotes the statistical interaction between the dummy variables DAPPR and DSENS.

For all three ETA scales, we set up a combined LCSR model. Firstly, the respective epistemological belief was regressed on its pretest level. Secondly, the change of the respective epistemological beliefs was regressed on the pretest measurements of the two other epistemological beliefs. For instance, the change in absolutism was firstly regressed on the absolutism pretest measurement and, secondly, on the pretest measurements of multiplicism and evaluativism (see Figure 3). The reason to include all epistemological belief pretest measures into the regression is based on theoretical considerations. Not only can the respective epistemological belief that was investigated affect the change score, but also, the other two epistemological beliefs may have the potential to affect the change score. In that way, the latent change score is controlled for the effects of the other epistemological beliefs.

The experimental conditions were introduced using dummy variables in the LCSR model, i.e., the change scores were regressed on the dummy variables. For the SENS factor, the reference category of the dummy (DSENS) was the N condition. For the APPR factor, the reference category for the dummy (DAPPR) was the G condition. The interaction of both factors was modeled as the product term (DINT) of the dummy variables. Thus, we used the same approach as in a linear model to analyze the effects of the experimental condition on the latent change scores. The regression of the change scores on the dummy variables is depicted in the regression part in Figure 3. In total, the model is saturated, i.e., has zero degrees of freedom.

We computed the estimated marginal means (EMMs; [44]) of the latent change scores to compare the experimental conditions. EMMs are linear combinations of the intercepts, the regression coefficients of the dummy variables, and the regression coefficients of the

pretest epistemological beliefs on the change score, and thus represent the latent change score means in each experimental condition.

Every dependent variable was first checked for the presence of a statistically significant interaction between because the interpretation of the main effects depends on the presence of an interaction effect. We particularly scrutinized if the interaction was either ordinal, hybrid, or disordinal. Main effects can only be interpreted if there is no interaction or in the case of an ordinal interaction. A hybrid interaction allows some of the main effects to be interpreted. A disordinal interaction forbids the interpretation of the main effects. Therefore, the interaction effects are reported and interpreted first, cf., Fox, [45], Ch. 8.2] and [46]. In the case of a hybrid or disordinal interaction, we used the simple main effects for interpretation. The simple main effects were calculated as the difference in EMM between the non-reference category and the reference category of one factor, given a fixed category of the other factor. Regarding the interpretation of the simple main effects, one should keep in mind that due to the dummy variables, the regression coefficient for the SENS dummy corresponds to the simple main effect for the SENS factor in the G condition of the APPR factor. In other words, the regression coefficient for the SENS dummy represents the increase/decrease in the change score in the dummy variable's reference category of the APPR factor. Accordingly, the regression coefficient for the APPR dummy corresponds to the simple main effect for the APPR factor in the N condition of the SENS factor. We used η^2 as a measure of effect size, with 0.01 being a small, 0.058 being a medium, and 0.138 being a large effect [47].

We used the ML estimator for the model as all ETA scales had skewness and kurtosis values smaller than 2 or 7 in each condition, respectively [48]. We followed Klopp and Stark [9] concerning sample size and expected moderate effects, resulting in 30 participants per condition. To analyze the observed power, we performed a Monte Carlo study [49] with 10,000 replications of the fitted model results as a data-generating process. According to Beaujean [50], power is the proportion of repetitions for which the null hypothesis is rejected for a given parameter with $\alpha = 0.05$. Power should be at least 0.50, at best exceeding a value of 0.80 and above [51].

As the essay score is ordinal, we used Kruskal and Wilcoxon tests. As there is no dedicated method for 2×2 designs, we coded each of the four possible conditions differently. The general procedure was that Kruskal tests were performed for the pretest and the posttest essay score, using the generated code as a group indicator. Afterward, we conducted a paired Wilcoxon test for the whole sample, and finally, we conducted a Wilcoxon test in each condition. We adjusted the p -values with the Holm method for the last series of tests.

4. Results

4.1. Descriptive Statistics and Internal Validity

In the following, we first provide the descriptive statistics for all variables. Then, we present the results concerning the ETA scale and, afterward, the results on the argumentation task. The descriptive statistics for each ETA variable and the variables' correlations are given in Table 3, and Table 5 provides the correlations of the ETA scales per condition. Table 6 presents the means of the epistemological belief scales, as well as medians of the essay score in each experimental condition.

The correlations in Table 5 show that almost all ETA scales are strongly related. The only exception is the evaluativism scale in the posttest, which does not correlate with the absolutism and multiplicism scale in the pretest. In particular, the overall correlations show that absolutism and multiplicism, both in the pretest and in the posttest, correlate negatively, which is a result that is also reported in Klopp and Stark [9]. The same pattern of correlation results emerges in each condition; in particular, the negative correlation between the absolutism and multiplicism scale emerges again.

Table 5. Correlations of the ETA scales in each condition.

Condition SENS = N and APPR = G						
	(1)	(2)	(3)	(4)	(5)	(6)
ABS1	-					
MULT1	-0.61 **	-				
EVAL1	0.50 *	-0.09	-			
ABS2	0.52 *	-0.37	0.35	-		
MULT2	-0.24	0.71 ***	-0.1	-0.46 *	-	
EVAL2	0.14	0.07	0.38	-0.05	0.44 *	-
Condition SENS = N and APPR = I						
	(1)	(2)	(3)	(4)	(5)	(6)
ABS1	-					
MULT1	-0.61 **	-				
EVAL1	0.40 *	-0.26	-			
ABS2	0.73 ***	-0.57 **	0.57 **	-		
MULT2	-0.49 *	0.82 ***	-0.42 *	-0.64 **	-	
EVAL2	0.10	-0.28	0.54 **	0.50 *	-0.31	-
Condition SENS = S and APPR = G						
	(1)	(2)	(3)	(4)	(5)	(6)
ABS1	-					
MULT1	-0.60 **	-				
EVAL1	0.63 ***	-0.63 ***	-			
ABS2	0.53 **	-0.41 *	0.30	-		
MULT2	-0.36	0.70 ***	-0.42 *	-0.51 **	-	
EVAL2	0.35	-0.42 *	0.49 **	0.22	-0.22	-
Condition SENS = S and APPR = I						
	(1)	(2)	(3)	(4)	(5)	(6)
ABS1	-					
MULT1	-0.52 **	-				
EVAL1	0.20	-0.57 **	-			
ABS2	0.79 ***	-0.41 *	0.08	-		
MULT2	-0.62 ***	0.73 ***	-0.44 *	-0.61 ***	-	
EVAL2	-0.07	-0.23	0.51 **	0.04	-0.27	-

Note. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table 6. Means and standard deviation of the ETA scales and the medians and interquartile ranges for the essay scores per experimental condition.

SENS: Neutral					SENS: Sensitization			
	APPR: General		APPR: Infusion		APPR: General		APPR: Infusion	
	N = 23		N = 25		N = 29		N = 29	
Variable	M	SD	M	SD	M	SD	M	SD
ABS1	3.87	0.78	3.62	0.66	3.61	0.66	3.68	0.70
ABS2	4.32	0.58	3.82	0.75	3.59	0.81	3.90	0.68
MULT1	2.98	0.90	3.06	0.78	3.08	0.68	3.34	0.79
MULT2	2.52	0.78	2.69	0.87	2.87	0.62	3.01	0.76
EVAL1	4.20	0.47	4.28	0.51	4.50	0.50	4.52	0.41
EVAL2	4.44	0.48	4.38	0.57	4.30	0.62	4.60	0.62
	Median	IQR	Median	IQR	Median	IQR	Median	IQR
ESSAY1	4.00	1.00	4.00	2.00	3.00	1.75	3.00	2.00
ESSAY2	4.00	1.75	5.00	1.00	4.00	0.25	4.00	1.00

Concerning internal validity, we used linear models to regress the ETA pretest scales on the experimental condition dummies. For absolutism, the linear model is not statistically significant ($F(2, 103) = 0.390, p = 0.536, R^2 = 0.008$), and there is neither an effect of the SENS factor ($\beta_{\text{SENS}} = -0.093, p = 0.499$) nor the APPR factor ($\beta_{\text{APPR}} = -0.079, p = 0.536$). The same holds for multiplicism ($F(2, 103) = 1.452, p = 0.239, R^2 = 0.027, \beta_{\text{SENS}} = 0.194, p = 0.207, \beta_{\text{APPR}} = 0.179, p = 0.247$). However, for evaluativism, the linear model is statistically significant ($F(2, 103) = 4.336, p = 0.004, R^2 = 0.081$), and there is an effect of the SENS factor ($\beta_{\text{SENS}} = 0.273, p = 0.004$) but not an effect of the APPR factor ($\beta_{\text{APPR}} = 0.051, p = 0.577$). A Welsch test indicates a difference in evaluativism between the two sensitization conditions, regardless of the critical thinking approach ($t(97.28) = -2.949, p = 0.004, d = 0.579$). For the pretest essay task, a Kruskal test does not indicate any difference between the conditions ($\chi^2(3) = 3.721, p = 0.293$). Additionally, a χ^2 test does not indicate any deviation of the gender distribution between the conditions ($\chi^2(4) = 3.074, p = 0.546$). Apart from slight evidence of the ETA evaluativism scale, the randomization procedure worked properly.

4.2. Results for the ETA Scales

Concerning absolutism (H1, H3, H4), Table 7 shows the EMMs that indicate significant increases in the NG and SI conditions. Table 8 reveals a statistically significant disordinal interaction effect of the experimental conditions, which can also be seen in Figure 4, and which has a medium effect and a relatively high power. Table 9 shows the simple main effects. The simple main effect of the APPR factor in the N condition indicates that the change score is greater in the G condition than in the I condition. Furthermore, the EMM in the G condition differs significantly from zero, whereas the EMM in the I condition does not. The simple main effects of the APPR factor in the S condition indicate that the change score is greater in the I condition than in the G condition. However, in this case, the EMM differs significantly from zero in the I condition, whereas the EMM in the G condition does not. The simple main effect of the SENS factor in the G condition indicates a significant difference, with the change score being greater in the N condition than in the S condition. The EMM indicates a significant positive change in the N condition, whereas the EMM is not significant in the S condition. All significant simple main effects reach at least the minimum threshold. These results suggest that without an epistemological sensitization, there is an increase in absolutism when using the general approach. However, when an epistemological sensitization is presented, there is only an increase in absolutism when the infusion approach is used. Other combinations do not yield epistemological change.

Table 7. Estimated marginal means of the latent change scores.

Variable	SENS	APPR	Estimate	SE	<i>p</i>	Power
Absolutism	N	G	0.533	0.121	0.000	0.988
	N	I	0.181	0.112	0.107	0.378
	S	G	−0.068	0.105	0.521	0.105
	S	I	0.227	0.107	0.034	0.565
Multiplicism	N	G	−0.720	0.131	0.000	0.999
	N	I	−0.584	0.116	0.000	0.998
	S	G	−0.397	0.104	0.000	0.967
	S	I	−0.432	0.103	0.000	0.948
Evaluativism	N	G	0.412	0.138	0.003	0.827
	N	I	0.270	0.137	0.049	0.491
	S	G	0.055	0.141	0.697	0.071
	S	I	0.369	0.134	0.006	0.781

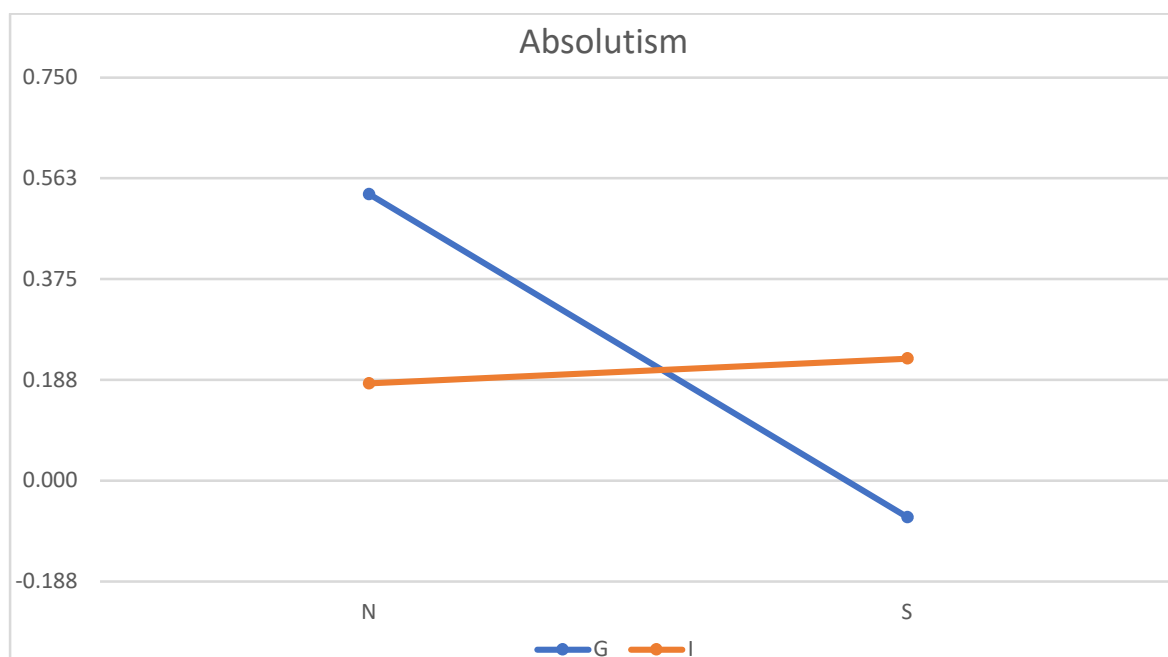


Figure 4. Estimated marginal means for absolutism change score in each condition.

For the multiplicism scales (H1, H3, H4), all EMMs are negative and significant (see Table 7), which indicates a general decline in multiplicism. Table 8 shows no interaction effect, and only the effect of the SENS dummy is significant. This indicates that the decline in the change score is smaller in the presence of an epistemological sensitization (see Figure 5). The effect size of the effect of SENS is small and, correspondingly, the power is almost at the lowest level. However, as the power for the EMM is relatively high, the results are reliable. Thus, providing an epistemological sensitization slightly affects the reduction of multiplicism. Stated otherwise, the reduction in multiplicism is more effective without an epistemological sensitization. In addition to this, as the absence of a significant interaction effect does not preclude an analysis of the simple main effects, there is a significant simple main effect for SENS in the G condition of the APPR factor. Therefore, it has to be supplemented that the reduction in multiplicism is most effective in the absence of an epistemological sensitization and when the general approach is used.

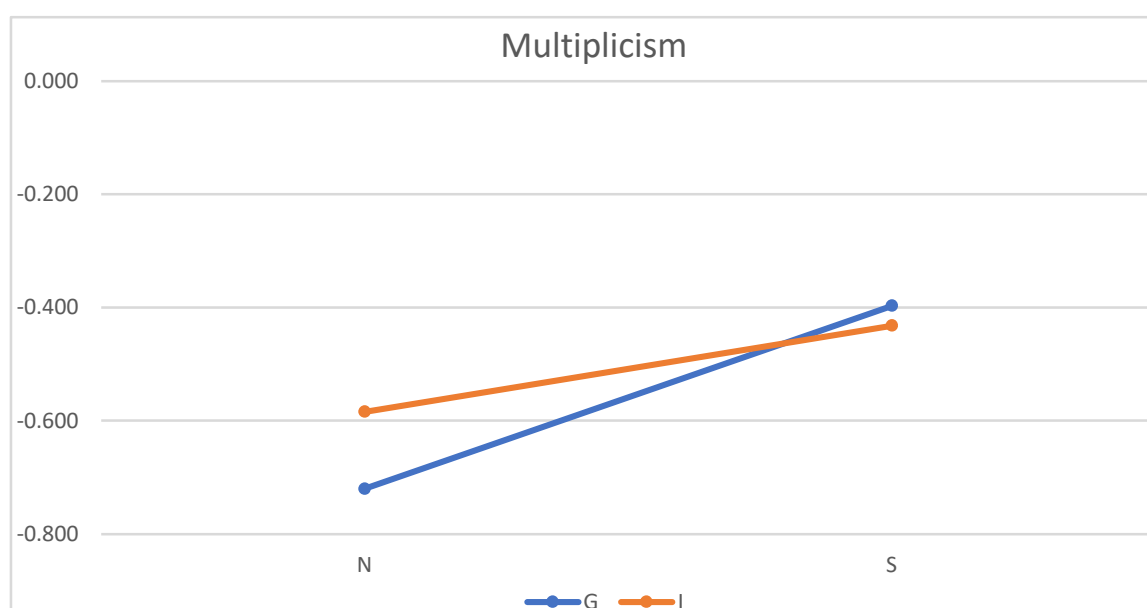


Figure 5. Estimated marginal means for multiplicism change score in each condition.

Table 8. Effects of the dummy coded experimental conditions on the latent change regression scores.

Variable	Effect	Estimate	SE	<i>p</i>	η^2	Power
Absolutism	DSSENS	−0.601	0.160	0.000	0.102	0.952
	DAPPR	−0.353	0.159	0.027	0.036	0.593
	DINT	0.647	0.216	0.003	0.065	0.840
Multiplicism	DSSENS	0.323	0.147	0.028	0.037	0.590
	DAPPR	0.136	0.146	0.353	0.007	0.167
	DINT	−0.170	0.198	0.390	0.006	0.150
Evalutativism	DSSENS	−0.357	0.145	0.014	0.045	0.678
	DAPPR	−0.142	0.144	0.325	0.007	0.191
	DINT	0.456	0.196	0.020	0.040	0.642

Table 9. Simple main effects of the latent change regression scores.

Variable	Constant Factor	Difference	Estimate	SE	<i>p</i>	Power
Absolutism	N	I-G	−0.353	0.159	0.027	0.593
	S	I-G	0.294	0.145	0.042	0.532
	G	S-N	−0.601	0.160	0.000	0.952
	I	S-N	0.046	0.154	0.764	0.069
Multiplicism	N	I-G	0.136	0.146	0.353	0.167
	S	I-G	−0.035	0.133	0.794	0.071
	G	S-N	0.323	0.147	0.028	0.590
	I	S-N	0.152	0.141	0.281	0.212
Evalutisvism	N	I-G	−0.142	0.144	0.325	0.191
	S	I-G	0.314	0.131	0.017	0.658
	G	S-N	−0.357	0.145	0.014	0.678
	I	S-N	0.099	0.140	0.479	0.133

Concerning evaluativism (H2, H3, H4), there is a significant disordinal interaction effect with medium effect size and an adequate power that suggests the simple main effects need to be interpreted (see Table 8). The significant interaction is also revealed in Figure 6. All EMMs are positive, i.e., there is no decrease in evaluativism (see Table 7). However, the EMM in the condition with epistemological sensitization and combined with the general approach is not significantly different from zero, indicating that no change occurred. Furthermore, in the condition without sensitization in combination with the infusion approach, the EMM is just below the nominal significance level. Thus, the findings warrant only a cautious interpretation, namely that change has occurred in this condition. The only significant simple main effect of the APPR factor is in the S condition of the SENS factor, with the I condition having the higher EMM. For the SENS factor, the only significant main effect is in the G condition of the APPR factor. As this effect is negative, it indicates that the increase in evaluativism is greater in the N condition than in the S condition. These results imply that without an epistemological sensitization, the general approach yields a larger increase in evaluativism. This is most notable since the evidence for an increase in evaluativism without an epistemological sensitization in combination with the infusion approach is rather fragile, which supports the notion that, in combination with a sensitization, only the infusion approach increases evaluativism.

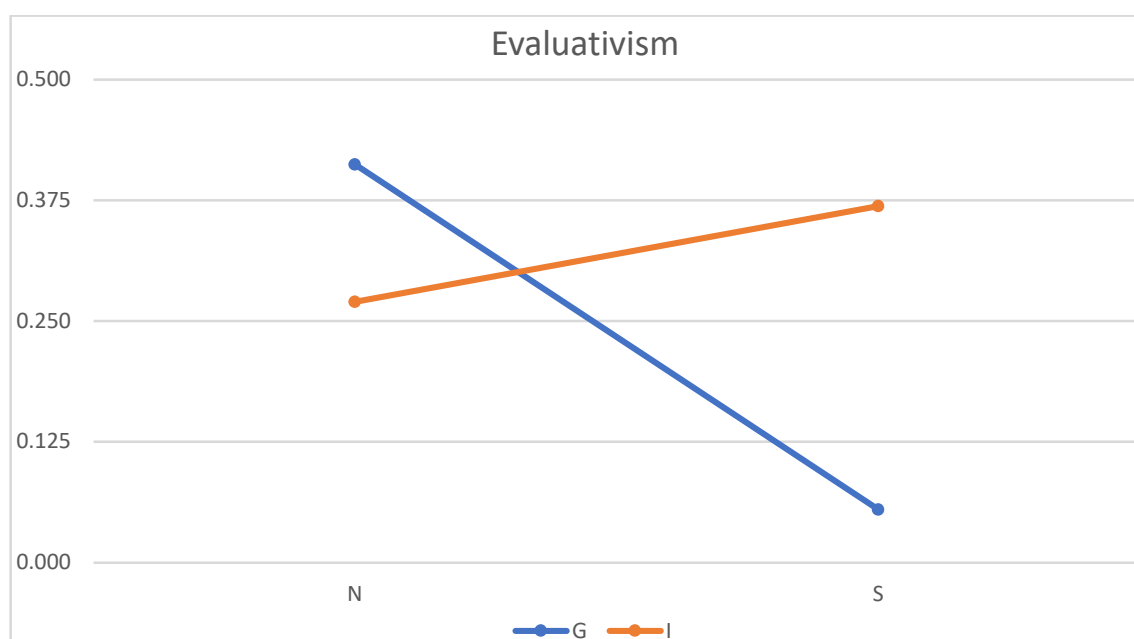


Figure 6. Estimated marginal means for evaluativism change score in each condition.

4.3. Results for the Argumentation Task

Table 6 shows the medians and interquartile ranges for the essay score (H5, H6, H7; see Figure 7). There is no difference between the conditions in the pretest (see above section on internal validity). There is also no difference between these conditions in the posttest ($\chi^2(3) = 3.906, p = 0.272$). Testing all possible pairwise comparisons between the single conditions in the posttest does not yield significant results. However, a paired Wilcoxon test for the whole sample indicates a difference between the pretest and posttest ($V = 627, p < 0.001$). The descriptive statistics indicate that the median increases in all conditions but not in the neutral condition and the general approach.

There was no significant change in a paired Wilcoxon test for the neutral condition without the sensitization and general approach ($V = 28, p = 0.059$; please note that this and the following p -values are adjusted). There is a significant change in the neutral condition in combination with the infusion approach ($V = 38, p = 0.036$). Next to this, there was no change when the general approach was combined with an epistemological sensitization ($V = 91.5, p = 0.153$). However, there was a significant increase when the infusion approach was used in combination with an epistemological sensitization ($V = 13.5, p < 0.001$). The results for the essay score are inconclusive, and there seems to be a general increase in the essay score. Above all, this increase does, with one exception, not depend on the condition. However, as revealed by investigating the increase in each condition separately, the essay score only seems to increase when the infusion approach is used, regardless of whether there is an epistemological sensitization. In particular, the desired level of an evaluativist argumentation only seems to emerge without an epistemological sensitization in combination with the infusion approach. Thus, no firm conclusion should be drawn on this result.

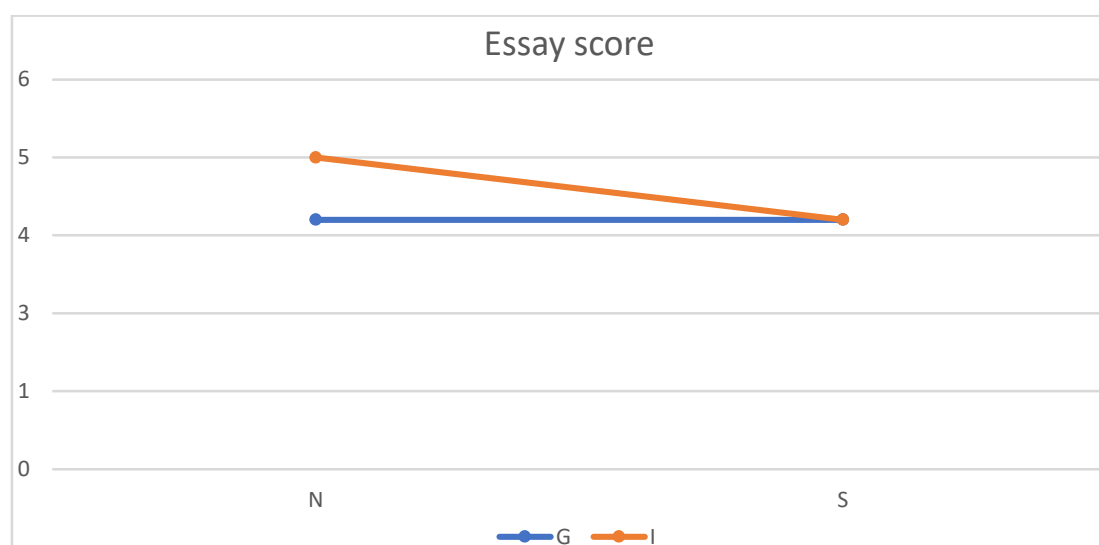


Figure 7. Plot of medians for the posttest essay score for each condition.

5. Discussion

In general, the results are not in line with the hypotheses developed from theoretical reasoning. However, they provide valuable insight into the mechanism that underlies epistemological change. The first conclusion is that a “one-in-all” effect, i.e., reducing absolutism and multiplicity, and enhancing evaluativism, is impossible. Instead, interventions should be tailored explicitly according to the goal that must be achieved.

For instance, when the goal is to reduce absolutism (see hypotheses H1, H3, and H4), there should be no epistemological sensitization when using the infusion approach because, in this condition, there was an increase in absolutism. As the epistemological sensitization presents the reasons for possible controversies, the instruction to elaborate on the resolution strategy as well as to figure out the causes for the controversy and to provide the insight into an existing side of the controversy that *fits best in the given context*—a feature of evaluativism—could have provided the wrong impression that only one side in the controversy is correct—a feature of absolutism. The following presentation of the resolution strategy may have strengthened this impression. On the other hand, the general approach should not be used when there is no epistemological sensitization. A possible explanation for this finding is that a general assignment to resolve the controversy indicates one true account of the controversy, which is a feature of absolutism as well.

Concerning multiplicity (see hypotheses H1, H3, and H4), there was generally no effect of the critical thinking approach, but only an effect of the epistemological sensitization, indicating that a reduction in multiplicity is hampered by the presence of an epistemological sensitization. On a more fine-grained level, there is a simple main effect for the epistemological sensitization for the general approach, but not for the infusion approach, indicating that the mitigation of the multiplicity reduction is more severe with the general approach. A possible reason for this finding is that epistemological sensitization has fostered the notion that science is arbitrary in its very nature, cf., [9]. Furthermore, this notion is enhanced by the kind of question used in the general approach.

Concerning evaluativism (see hypotheses H2, H3, and H4), there was an increase in all conditions for evaluativism except for the one with the epistemological sensitization and the general approach. The missing increase in evaluativism in this particular condition is surprising and not easy to explain. A possible cautious interpretation is that the combination of epistemological sensitization and broad question to reflect on the controversies overstrains the participants. In future research, a possible remedy is to assess cognitive load to gain insight into the participant’s cognitive strain.

The essay score (see hypotheses H5, H6, and H7) is generally higher in conditions with the infusion approach than in the conditions with the general approach. A change

towards a more evaluativist argumentation occurred only in the conditions with the infusion approach, regardless of the presence of an epistemological sensitization. Theoretically, this result is not unexpected and may be ascribed to the nature of the infusion approach. The infusion approach consisted of a scheme that led the participants to work on the pros and cons of each side of the controversy, among other tasks. However, the question scheme did not ask the participants to take the context into account to argue for one side. The critical thinking instruction only had the goal of identifying the reason for the controversy. Additionally, the increase in evaluativist argumentation level does not depend on the presence of the epistemological sensitization. On the one hand, this is not an unexpected result as the sensitization only provides reasons why scientific controversies exist and neither provides an explanation nor an example of how to cope with this knowledge in an argumentation. On the other side, this result is remarkable because there was an increase towards an evaluativist argumentation level—in the neutral condition. In the sensitization condition, by contrast, only an increase towards a multiplicist argumentation sharply at the edge of an evaluativist argumentation level was detected (see the definition of the level of argumentation in Table 3 and the medians per condition in Table 4). A possible explanation for this is that the participants in the neutral condition had better argumentation skills. Therefore, as a critical thinking instruction, the infusion approach yielded an increase towards the evaluativist level. In contrast to the sensitization condition, the highly multiplicist argumentation level improved only slightly toward a precursor of an evaluativist argumentation level. However, this explanation is highly speculative as there were no statistically significant differences in argumentation level for the pretest score.

A potential issue is the stability of the results or, in other words, the replicability. A possible source to partly judge the replicability is the comparison with the results of Klopp and Stark [9]. In this study, the authors varied the presence of an epistemological sensitization. Their instruction to resolve the controversies was very similar to the infusion approach in this study. They used a questioning scheme in which the participants had to work on resolving the presented controversies. It is thus possible to compare the effects of the epistemological sensitization in the infusion approach condition. The current study is a partial conceptual replication of [9] from a methodological perspective. These authors' study showed epistemological change in the form of an increase in evaluativism only in the condition with the sensitization measure. This was also the case in the present research, but epistemological change was observed in other conditions, too. Additionally, the comparison shows that neither in [9] nor in the current study, group differences between the neutral condition and the condition with the epistemological sensitization appeared. In summary, in the particular condition of the infusion approach, our results resemble those of Klopp and Stark [9] and extend their results. However, as only one condition could be compared between the two studies, the comparison gives only a small picture of the replicability of the results.

The small effect sizes provide another minor limitation of the results. In contrast to Klopp and Stark [9] or Rosman et al. [10], who reported medium to large effect sizes, the effect sizes in the present study may be due to its long-term duration. Both aforementioned studies were limited to one session in which the pretest, the intervention, and the posttest took place. However, the present study lasted four weeks, with the pretest in the first week, followed by two weeks of interventions, and the posttest in the fourth week. It is likely that the effects of the intervention have decreased over time during this period. Another possible explanation for the smaller effect sizes is that the intervention did not induce an epistemological change as large as in [9] and [10]. The participants of both former studies were psychology students who can be considered intermediates, i.e., stand somewhere in the continuum between novices and experts, who have sufficient prior knowledge in methodological and philosophy of science issues. The resolution strategies draw on such knowledge. However, student teachers typically do not attend corresponding classes. They usually do not have the required prior knowledge, and when they have

it, the knowledge base is not as profound as the knowledge base of psychology students. Consequently, student teachers might not fully understand the resolution strategies which results in negative impact on epistemological change. Indeed, it is possible that both explanations combined contributed to the small effect sizes.

Another aspect refers to the power of the results. In general, the power of the results is in the lower part of the acceptable range. This relates to the issues of relatively small effect sizes mentioned above. This study was planned with the expectation of at least medium effect sizes as in [9] and [10]. Therefore, the current research is possibly underpowered. As a result, the power of the estimate is smaller, too. Further studies should draw on smaller effect sizes to consider the necessary sample size.

We hypothesized from a substantial theoretical base, referring to Klopp and Stark [9] and Rosman et al. [10], that the intervention should reduce absolutism and multiplicity. However, as also found in [9], this general pattern was not observed. Instead, we saw a general decline in multiplicity combined with an increase in absolutism in the conditions with the general approach and without the epistemological sensitization, and in the condition with the infusion approach and the epistemological sensitization. In the current study and in Klopp and Stark [9] there were negative correlations between absolutism and multiplicity in the pretest and the posttest. This pattern was observed in the whole sample as well as in each condition. Klopp and Stark [9] discuss these results in the context of a backfire effect, i.e., the decline in multiplicity yields the unwanted effect of increasing absolutism. The authors argue that, from the perspective of a stage model, to leave the absolutist stage, a person has to abandon the belief that there is only one correct account of knowledge and has to come to the belief that there are many possible accounts of knowledge, which are all equally correct. Therefore, reducing absolutism may increase multiplicity. In this sense, the current study results firstly replicate the results from [9] and, secondly, provide a constellation of conditions in which a decrease in multiplicity accompanies an increase in absolutism. However, as the negative correlation between absolutism was observed in all conditions, it is unclear whether the absence of an increase in multiplicity is random or not.

More importantly, this explanation does not provide an answer to the question of why the reduction in multiplicity is related to an increase in absolutism and not with an increase in evaluativism, the latter being a consequence that would follow from the stage theory of epistemological development. A possibility would be motivational aspects. Besides the cognitive factors, the participants must have the motivation to further develop their epistemological beliefs. This aspect is partially covered within the epistemological volition component in Bendixen's model [21]. However, from a motivational perspective, the participants must be determined to develop their epistemological beliefs towards evaluativism and not to fall back to absolutism.

Therefore, questions arise as to whether specific combinations of circumstances hinder the backfire effect and, if this is the case, why this result emerges. A theoretical underpinning for this result seems to be of vital interest and very important for a tailored development of epistemological belief interventions and further enhancing the theory.

To sum up, the overall results show that the intervention reduces, in general, multiplicity and enhances evaluativism in almost all conditions. This is a desirable result because multiplicity implies a notion that science consists merely of a collection of opinions, and the intervention acts to reduce this notion and foster the correct notion of science in the same vein. However, there is also an increase in absolutism in two conditions. On the one hand, this is not a desirable result, and on the other hand—and in contrast to multiplicity—absolutism implies a positive conceptualization of science. Both absolutism and evaluativism have in common that a person positively evaluates the scientific endeavor. The only difference is that absolutism endorses the inadequate belief that only one side of a controversy is correct regardless of contextual circumstances. From the perspective of the integrated approach in the sense of Barzilai and Weinstock [17], the profiles of absolutism and evaluativism are almost identical, apart from the dimensions *Nature of*

knowledge and *The role of multiple perspectives*. The direction of the epistemological change in these two conditions is thus appropriate but lacks an appropriate change in these two dimensions. Therefore, a noteworthy extension for further studies should include a more detailed view of epistemic beliefs regarding the dimensions that compose absolutism, multiplicity, and evaluativism, and the conditions that drive epistemological change in each of these dimensions.

6. Limitations

Although the current study provides viable results concerning the development of epistemological belief interventions, the study also has its limitations. The first limitation refers to the absence of manipulation checks. As discussed in detail in Klopp and Stark [9], the presence of manipulation checks is a possible source of confounding effects (cf., [52]). For instance, introducing a measure asking if the participants knew of the controversy may be a demand characteristic and cause epistemological doubt. For that reason, following these authors, we did not include a manipulation check in this study. However, the lack of a measure for the effectiveness of the controversy intervention narrows its theoretical explanatory power. The same considerations apply certainly to the parts of the interventions, i.e., the epistemological sensitization measure and the resolution strategy.

From the perspective of limitations, the lack of an untreated control group may also limit the interpretation of the results. As the experiment lasted over four weeks, an occurring fluctuation of epistemological beliefs is not impossible. Thus, such fluctuations would confound the epistemological change resulting from the intervention. Possible causes of fluctuations are either random changes or systematic influences. Random fluctuations occurred, e.g., for the absolutism scale in [9]. Systematic influences, e.g., may be caused by instruction. However, for the four weeks of the study, a review of the local curriculum showed no kind of instruction that could systematically alter psychology-specific epistemological beliefs apart from the teaching of educational psychology content knowledge. Therefore, the lack of a control group can be considered tolerable, but future studies should include one.

In the context of possible limitations, the sample solution itself may pose a limitation. Providing a sample solution may indeed avoid negative motivational consequences. Nonetheless, the sample solution also relieves the participants from finding their own solution, which impedes the necessary elaborative processes in the resolution strategies component in the process model of epistemological change. From a methodological perspective, providing no sample solution would be preferable. To control for negative motivational consequences, further studies could apply motivational measures taken into account in the statistical modeling. In addition to this, providing a sample solution and impeding the participants' cognitive elaborative processes on the resolution may also cause small effect sizes.

A more general limitation refers to the process model of epistemological change itself. As Braten [53] argues, the model of Bendixen [21,22] and its mechanism of change lacks empirical support, although it is a reasonable way to conceptualize the process of epistemological change. Furthermore, citing a study by Ferguson, Braten, and Stromso [54], Braten argues that there is evidence for the postulated sequence of epistemological doubt, epistemological volition, and resolution strategies. Thus, the process model of epistemological change is more a framework to describe epistemological change rather than a theoretical model to draw on.

The same consideration applies to the notion of epistemological sensitization. As introduced by Porsch and Bromme [11], the epistemological sensitization is more a heuristic than a worked-out theoretical account [9]. Klopp and Stark [9] relate the epistemological sensitization to the activation of task-relevant prior-knowledge that, in turn, yields a greater attention to the relevant epistemological features of the controversy. However, they discuss this cognitive mechanism of the epistemological sensitization only as a possible account. Up to now, there is no theoretical account with empirical support for the

mechanism behind the epistemological sensitization. However, such an account is needed to design further sensitization measures to be used in instructional settings.

In addition, the explanation by means of the activation of task-relevant prior-knowledge is not the only feasible theoretical account. An epistemological sensitization could also induce epistemological doubt on a domain-specific level in case the individual epistemological beliefs are incompatible with the epistemological feature presented in the sensitization. This, in turn, could facilitate the development of epistemological doubt during the presentation of the controversies. As mentioned by Bendixen and Rule [22], epistemological doubt is related to cognitive dissonance (cf., [21]). Thus, an alternative explanation is the induction of epistemological doubt in the form of cognitive dissonance. This dissonance state is later increased by the presentation of the controversies. Further research into the cognitive mechanisms of epistemological sensitization and epistemological change is needed.

The last limitation refers to the essay score. The essay score was introduced to provide an alternative assessment of epistemological beliefs. It draws on coding the epistemological level of the argumentation to infer the participant's epistemological beliefs. Current theoretical reasoning and empirical results suggest that epistemological beliefs and the level of argumentation are related (e.g., [2,55,56]). However, a theory must causally connect epistemological beliefs and argumentation to use a scored essay as a proxy for epistemological beliefs (e.g., [57,58]). This requirement states that the variations in the property to be measured should cause variations of the indicators and not vice versa. Nonetheless, current theoretical reasoning supports a correlational relation between epistemological beliefs and argumentation. For instance, Iordanou [59] showed that argumentation could change epistemological beliefs. The relation of the essay score and epistemological beliefs is thus not straightforward. Additionally, the writing of the essay always occurred after working on the ETA items. From a methodological perspective, this may have distorted the essay score in addition to the effects of the intervention.

7. Educational Implications

The main goal of this paper was to evaluate the epistemological sensitization and the approach to critical thinking to enhance the controversy intervention to alter epistemological beliefs. The research was conducted in the realm of teacher education. At the end of our paper, we want to elaborate on some educational implications that may follow from this study. However, to avoid an overstrain of the particular results of our study, we want to stress that we do not intend to give general recommendations. Instead, we want to highlight some critical points in the sense of a take-home message that could be of interest to educators in general and teacher educators in particular.

Firstly, as stated at the beginning of the discussion, educators should be aware that there is no general way to “enhance” students' epistemological beliefs, i.e., there is no all-in-one concept to reduce absolutism and multiplicism and to foster evaluativism. Thus, educators should determine the epistemological beliefs of their students and apply measures that fit the level of their beliefs. Secondly, drawing on our finding that the epistemological sensitization increases multiplicism, educators should be conscious of the problem that confronting students with the epistemological features of domain-knowledge may have the potentially unwanted effect of eliciting multiplicitist epistemological beliefs. Thus, educators should be careful when they introduce students to issues such as moderators, different paradigms, etc., in particular when they are followed by controversies arising from these issues. In a more general sense, educators should be aware that epistemological beliefs develop during the socialization of students in a domain (cf., [9,60]), which may be problematic when the domain's knowledge is ill-structured (cf., [9]).

Thirdly, as we explicated at the end of our discussion, educators should foster epistemological beliefs that go along with a positive evaluation of science, i.e., educators should think positively of absolutism in particular when considering the background of

students. Put differently, educators should have an evaluativist attitude. For instance, a student teacher aiming to become a mathematics and physics teacher is likely to have absolutist epistemological beliefs. Educators should take into consideration that this student potentially applies the epistemological beliefs of their primary domains to educational psychology. Although this is not appropriate considering the different knowledge structures of the domains under consideration, the student would value educational psychology and not disregard this knowledge as mere opinions. Against the background of problems concerning normative assumptions provided by stage models (cf., [9] for a summary), the educator should try to support evaluativism and avoid fostering multipicism first.

Author Contributions: Conceptualization, E.K. and R.S.; methodology, E.K.; software, E.K.; validation, E.K. and R.S.; formal analysis, E.K.; investigation, E.K.; resources, R.S.; data curation, E.K.; writing—original draft preparation, E.K.; writing—review and editing, E.K. and R.S.; visualization, E.K.; supervision, R.S.; project administration, E.K.; funding acquisition, not applicable. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and Saarland University within the funding program Open Access Publishing.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, no institutional approval necessary in accordance with local law.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgements: The authors thank Martin Klein for his valuable comments and his support in the preparation of this work. We also thank Hannah Lass and Sina Wilhelm for their help in the preparation of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fischer, F.; Kollar, I.; Ufer, S.; Sodian, B.; Hussmann, H.; Pekrun, R.; Neuhaus, B.; Dorner, B.; Pankofer, S.; Fischer, M.; et al. Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learn. Res.* **2014**, *4*, 28–45.
2. Kuhn, D. *The Skills of Argument*; Cambridge University Press: Cambridge, UK, 1991.
3. Weinstock, M.P. Psychological research and the epistemological approach to argumentation. *Informal Log.* **2006**, *26*, 103.
4. Kuhn, D. How do people know? *Psychol. Sci.* **2001**, *12*, 1–8.
5. Feinkohl, I.; Flemming, D.; Cress, U.; Kimmerle, J. The impact of epistemological beliefs and cognitive ability on recall and critical evaluation of scientific information. *Cogn. Process.* **2016**, *17*, 213–223.
6. Klopp, E.; Stark, R. Persönliche Epistemologien—Elemente wissenschaftlicher Kompetenz. In *Denken über Wissen und Wissenschaft. Epistemologische Überzeugungen als Gegenstand Psychologischer Forschung*; Mayer, A.-K., Rosman, T., Eds.; Pabst Science Publishers: Lengerich, Germany, 2016; pp. 39–69.
7. Standing Conference of the Ministers of Education and Cultural Affairs of the States in the Federal Republic of Germany. Standards for Teacher Education: Educational Sciences. Available online: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf (accessed on 14 March 2021).
8. Rosman, T.; Mayer, A.-K.; Kerwer, M.; Krampen, G. The differential development of epistemic beliefs in psychology and computer science students: A four-wave longitudinal study. *Learn. Instr.* **2017**, *49*, 166–177.
9. Klopp, E.; Stark, R. Scientific controversies and epistemological sensitization—Effects of an intervention on psychology students' epistemological beliefs and argumentation skills. *Front. Educ.* **2022**, *6*, 785241. <https://doi.org/10.3389/feduc.2021.785241>.
10. Rosman, T.; Mayer, A.-K.; Peter, J.; Krampen, G. Need for cognitive closure may impede the effectiveness of epistemic belief instruction. *Learn. Individ. Differ.* **2016**, *49*, 406–413.
11. Porsch, T.; Bromme, R. Effects of epistemological sensitization on source choices. *Instr. Sci.* **2010**, *39*, 805–819.
12. Valanides, N.; Angeli, C. Effects of instruction on changes in epistemological beliefs. *Contemp. Educ. Psycho.* **2016**, *30*, 314–330.
13. Hofer, B.K.; Bendixen, L.D. Personal epistemology: Theory, research, and future directions. In *Handbook of Educational Psychology*; Harris, K., Ed.; American Psychological Association: Washington, DC, USA, 2012; pp. 225–254.
14. Hammer, D.; Elby, A. On the form of a personal epistemology. In *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*; Hofer, B.K., Pintrich, P.R., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA, 2002; pp. 169–190.

15. Schommer, M. Effects of beliefs about the nature of knowledge on comprehension. *J. Educ. Psychol.* **1990**, *82*, 498–504.
16. Hofer, B.K.; Pintrich, P.R. The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Rev. Educ. Res.* **1997**, *67*, 88–140.
17. Barzilai, S.; Weinstock, M. Measuring epistemic thinking within and across topics: A scenario-based approach. *Contemp. Educ. Psychol.* **2015**, *42*, 141–158.
18. Schommer-Aikins, M. Explaining the epistemological belief system: Introducing the embedded systemic model and coordinated research approach. *Educ. Psychol.* **2004**, *39*, 19–29.
19. Kuhn, D.; Cheney, S.; Weinstock, M. The development of epistemological understanding. *Cogn. Dev.* **2000**, *15*, 309–328.
20. Peter, J.; Rosman, T.; Mayer, A.-K.; Lechner, N.; Krampen, G. Assessing epistemic sophistication by considering domain-specific absolute and multiplicitic beliefs separately. *Br. J. Educ. Psychol.* **2016**, *86*, 204–221. <https://doi.org/10.1111/bjep.12098>.
21. Bendixen, L.D. A process model of epistemic belief change; In *Personal Epistemology: The psychology of Beliefs about Knowledge and Knowing*; Hofer, B.K., Pintrich, P.R., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA, 2002; pp. 191–208.
22. Bendixen, L.D.; Rule, D.C. An integrative approach to personal epistemology: A guiding model. *Educ. Psychol.* **2004**, *39*, 69–80.
23. Rosman, T.; Kerwer, M. Mechanisms of Epistemic Change: The Roles of Reflection and Social Interaction. *Psychol. Learn. Teaching*. **2022**. <https://doi.org/10.1177/14757257221098860>.
24. Kienhues, D.; Ferguson, L.E.; Stahl, E. Diverging information and epistemic change. In *Handbook of Epistemic Cognition*; Greene, J., Sandoval, W., Braten, I., Eds.; Routledge: New York, NY, USA, 2016; pp. 318–330.
25. Muis, K.R.; Bendixen, L.D.; Haerle, F.C. Domain-general and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educ. Psychol. Rev.* **2006**, *18*, 3–54.
26. Pieschl, S.; Bromme, R.; Porsch, T.; Stahl, E. 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, Utrecht, The Netherlands, 23–28 June 2008; Lulu Enterprises: London, UK, 2008; pp. 213220–213222.
27. Ennis, R. A taxonomy of critical thinking dispositions and abilities. In *Teaching Thinking Skills: Theory and Practice*; Baron, J.; Sternberg, R., Eds.; Freeman: New York, NY, USA, 1987; pp. 9–26.
28. Chan, N.-M.; Ho, I.T.; Ku, K.Y.L. Epistemic beliefs and critical thinking of Chinese students. *Learn. Individ. Differ.* **2011**, *21*, 67–77.
29. Greene, J.A.; Yu, S.B. Educating Critical Thinkers: The Role of Epistemic Cognition. *Policy Insights Behav. Brain Sci.* **2016**, *3*, 45–53.
30. Muis, K.R.; Duffy, M.C. Epistemic climate and epistemic change: Instruction designed to change students' beliefs and learning strategies and improve achievement. *J. Educ. Psychol.* **2013**, *105*, 213–225.
31. Kuhn, D.; Dean, D. Metacognition: A bridge between cognitive psychology and educational practice. *Theory Pract.* **2004**, *43*, 268–273.
32. Ennis, R.H. The degree to which critical thinking is subject specific: Clarification and needed research. In *The Generalizability of Critical Thinking: Multiple Perspectives on an Educational Ideal*; Norris, S.P., Ed.; Teachers College: New York, NY, USA, 1992; pp. 21–37.
33. Abrami, P.C.; Bernard, R.M.; Borokhovski, E.; Waddington, D.I.; Wade, A.C.; Persson, T. Strategies for Teaching Students to Think Critically: A Meta-Analysis. *Rev. Educ. Res.* **2014**, *20*, 1–40.
34. Driver, R.; Newton, P.; Osborne, J. Establishing the norms of scientific argumentation in classrooms. *Sci. Educ.* **2000**, *84*, 287–312.
35. Lang, F.; Kammerer, Y.; Stürmer, K.; Gerjets, P. Investigating professed and enacted epistemic beliefs about the uncertainty of scientific knowledge when students evaluate scientific controversies. *Eur. J. Psychol. Educ.* **2021**, *36*, 125–146.
36. Bromme, R.; Kienhues, D. Wissenschaftsverständnis und Wissenschaftskommunikation. In *Pädagogische Psychologie (6. Auflage)*; Seidel, T., Krapp, A., Eds.; Beltz: Weinheim, Germany, 2014; pp. 55–81.
37. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174.
38. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019. Available online: <http://www.R-project.org/> (accessed on 1 April 2022).
39. Revelle, W. *Psych: Procedures for Personality and Psychological Research*; Northwestern University: Evanston, IL, USA, 2020.
40. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. Available online: <http://www.jstatsoft.org/v48/i02/> (accessed on 1 April 2022).
41. Fox, J.; Weisberg, S. *An R Companion to Applied Regression*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2019.
42. Pornprasertmanit, S.; Miller, P.; Schoemann, A.; Jorgensen, T.D. Simsem: SIMulated Structural Equation Modeling. R Package Version 0.5–16. 2021. Available online: <https://CRAN.R-project.org/package=simsem> (accessed on 1 April 2022).
43. McArdle, J.J. Latent variable modeling of differences and changes with longitudinal data. *Annu. Rev. Psychol.* **2009**, *60*, 577–605.
44. Searle, S.; Speed, F.; Milliken, G. Population marginal means in the linear model: An alternative to least squares means. *Am. Stat.* **1980**, *34*, 216–221.
45. Fox, J. *Applied Linear Regression and Generalized Linear Regression*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2016.
46. Sawilowsky, S. ANOVA: Effect sizes, simulation interaction vs. main effects, and a modified ANOVA table. In *Real Data Analysis (Ch. 14)*; Sawilowsky, S., Ed.; Information Age Publishing: Charlotte, NC, USA, 2007.
47. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Lawrence Erlbaum: Hillsdale, Australia, 1988.

48. Finney, C.; DiStefan, C. Nonnormal and categorical data in structural equation modeling. In *Structural Equation Modeling—A Second Course*; Hancock, G.R., Mueller, R.O., Eds.; Information Age Publishing: Charlotte, NC, USA, 2013; pp. 439–492.
49. Zhang, Z.; Liu, H. Sample Size Planning for Latent Change Score Models through Monte Carlo Simulation. In *Longitudinal Multivariate Psychology*; Ferrer, E., Boker, S.M., Grimm, K.J., Eds.; Taylor & Francis: New York, NY, USA, 2019; pp. 189–211.
50. Beaujean, A.A. *Latent Variable Modeling Using R A Step-by-Step Guide*; Routledge: New York, NY, USA, 2014.
51. Kyriazos, T.A. Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology* **2018**, *9*, 2207–2230.
52. Hauser, D.J.; Ellsworth, P.C.; Gonzalez, R. Are manipulation checks necessary? *Front. Psychol.* **2018**, *9*, 998.
53. Braten, I. Epistemic cognition interventions. Issues, challenges, and directions. In *Handbook of Epistemic Cognition*; Greene, J., Sandoval, W., Braten, I., Eds.; Routledge: New York, NY, USA, 2016; pp. 360–371.
54. Ferguson, L.E.; Bråten, I.; Strømsø, H.I. Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learn. Instr.* **2012**, *22*, 103–120.
55. Iordanou, K.; Panayioti, K.; Beker, K. Argumentative Reasoning. In *Handbook of Epistemic Cognition*; Greene, J., Sandoval, W., Braten, I., Eds.; Routledge: New York, NY, USA, 2016; pp. 39–53.
56. Mason, L.; Scirica, F. Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learn. Instr.* **2006**, *16*, 492–509.
57. Bornstein, R.F. Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychol. Assess.* **2011**, *23*, 532–544.
58. Borsboom, D.; Cramer, A.O.J.; Kievit, R.A.; Zand Scholten, A.; Franic, S. The two disciplines of scientific psychology, or: The disunity of psychology as a working hypothesis. In *Developmental Process Methodology in the Social and Developmental Sciences*, Valsiner, J., Molenaar, P.C.M., Lyra, M.C.D.P., Chaudary, N., Eds.; Springer: New York, NY, USA; pp. 67–97.
59. Iordanou, K. Developing epistemological understanding in scientific and social domains through argumentation. *Z. Für Pädagogische Psychol.* **2016**, *30*, 109–119.
60. Palmer, B.; Marra, R. Individual domain-specific epistemologies: Implications for educational practice. In *Knowing, Knowledge, and Beliefs: Epistemological Studies Across Diverse Cultures*; Khine, M.S., Ed.; Springer: New York, NY, USA, 2008; pp. 423–441.