



Inductive and deductive justification of knowledge: epistemological beliefs and critical thinking at the beginning of studying mathematics

Benjamin Rott¹ 

Accepted: 22 October 2020 / Published online: 14 November 2020

© The Author(s) 2020

Abstract

Epistemological beliefs are considered to play an important role in processes of learning and teaching. However, research on epistemological beliefs is confronted with methodological issues as for traditionally used self-report instruments with closed items, problems with social desirability, validity, and capturing domain-specific aspects of beliefs have been reported. Therefore, a new instrument with open items has been developed to capture mathematics-related epistemological beliefs, focusing on the justification of knowledge. Extending a previous study, the study at hand uses a larger and more diverse sample as well as a refined methodology. In total, 581 mathematics students (Bachelor of Science as well as pre-service teachers) completed the belief questionnaire and a test for mathematical critical thinking. The results confirm that beliefs can empirically be distinguished into belief position and belief argumentation, with only the latter being correlated to critical thinking.

Keywords Epistemological beliefs · Critical thinking · Mathematical competencies · Replication

1 Introduction

Studying beliefs has a long tradition in educational research generally (Hofer & Pintrich, 1997) and mathematics education specifically (Thompson, 1992; Philipp, 2007). A special emphasis has been put on *epistemological* (or *epistemic*) *beliefs* (EB), which are beliefs about the nature of knowledge and knowing (Hofer & Pintrich, 1997). Research indicates that reflected or sophisticated EB are correlated with greater learning success (e.g., Hofer & Pintrich, 1997;

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10649-020-10004-1>.

✉ Benjamin Rott
benjamin.rott@uni-koeln.de

¹ Faculty of Mathematics and Natural Sciences, University of Cologne, Gronewaldstr. 2, 50931 Cologne, Germany

Trautwein & Lüdtke, 2007). Students with sophisticated EB, that is, students who understand the way that scientific knowledge is created and founded, show better integrated and deeper knowledge than students with less sophisticated EB; in contrast, students seem to learn more superficially if they are convinced that knowledge is stable and secure and only needs to be passed on by authorities (Tsai, 1998a, 1998b). Additionally, sophisticated EB are considered not only as a prerequisite to successfully completing higher education but also for active participation in modern science- and technology-based societies (e.g., Bromme, 2005).

Much of the work on students' EB has been done under the assumption that such beliefs are domain-general (Muis, 2004, p. 346). There are, however, newer studies suggesting that EB can be domain-specific (Muis, 2004; Stahl & Bromme, 2007). Thus, domain-specific EB have been identified as an area in need of further study (Hofer & Pintrich, 1997, pp. 124 ff.).

The study described here deals with mathematics-related EB (MEB); it focuses on what students believe and answer to typical questions regarding epistemology discussed in the philosophy of mathematics (e.g., Barrow, 1992; Horsten, 2018): What does it mean to do mathematics? How do mathematicians work when they generate or find new mathematical knowledge? Is such knowledge created or discovered? And, most important for this study, do mathematicians justify their ideas and hypotheses either deductively or inductively or do they use both ways (cf. Lakatos, 1976; Russell, 1919)?

2 Background

In this section, conceptualizations of EB as well as ways of measuring EB are discussed. Thereafter, conceptualizations and ways of measuring critical thinking are introduced. Finally, research goals for the study at hand are presented.

2.1 Epistemological beliefs

In a literature review, Hofer and Pintrich (1997) identified two general areas of EB, *nature of knowledge* and *nature or process of knowing*, with two dimensions each. One of the dimensions of the latter area, which is important for the study at hand, is *justification of knowledge*, addressing assumptions about how to evaluate knowledge claims, how to use empirical evidence, how to refer to experts, and how to justify knowledge claims.

Beliefs generally and EB specifically can be conceptualized on a scale from subconscious beliefs over conscious beliefs to knowledge of concepts, rules, and so forth, or, in other words, from subjective (not necessarily justified) to objective knowledge (cf. Murphy & Mason, 2006; Pehkonen, 1999). Under this perspective, as knowledge is built, beliefs converge towards objective knowledge (i.e., “true belief” in a Platonic sense), which in its purest form is not accessible to humans (see Philipp, 2007, pp. 266 ff. for a further discussion). Therefore, all human knowledge is some kind of belief. In this line of thought, Stahl and Bromme (2007) proposed a distinction between (associative-)connotative and (explicit-)denotative aspects of EB. The former are based on feelings and often context-free, whereas the latter are reflected upon and context-specific. Especially denotative beliefs are dependent on a person's knowledge and experience. For example, mathematics is often identified as a deductive science that is characterized by well-formulated theorems and formal proofs. Without ever having worked like a research mathematician or knowing that famous mathematicians like Leonhard Euler collected countless examples and conducted “quasi experiments” before being able to

formulate and deductively prove a theorem (cf. Pólya, 1954), one cannot comprehend inductive aspects of mathematics. The study at hand focuses on denotative beliefs.

2.2 Measuring epistemological beliefs

At the beginning of research on EB, mainly interviews were conducted; later, the use of questionnaires became the norm, as they could be used for much larger samples (cf. Hofer & Pintrich, 1997). Questionnaires with closed items have proven to be most popular—especially the instrument developed by Schommer (1990).¹ In such instruments, the respondents are presented with statements that they should agree or disagree with on a Likert scale. Sample items from Schommer’s Epistemological Belief Questionnaire are “You can believe almost everything you read”; “Truth is unchanging”; and “Scientists can ultimately get to the truth.”

In such questionnaire studies, how “good” (i.e., reflective or sophisticated) test persons’ beliefs are is determined by their answers, that is, the *positions* they agree to, like “knowledge is certain or uncertain.” In a somewhat simplified way, “naïve” beliefs are attributed to respondents who consider knowledge as safe and stable, whereas more sophisticated beliefs are attributed to respondents who agree on the uncertainty and tentativeness of knowledge (Hofer & Pintrich, 1997). With regard to the history of science, this approach is reasonable—thinking of Newton’s theory of gravity, which has been superseded by Einstein’s theory of relativity.

There are, however, problems with the use of closed self-report questionnaires such as the problem of social desirability; that is, test persons often mark answers that they think are expected of them (Di Martino & Sabena, 2010; Safrudiannur, 2020). Other problems of such instruments relate to their reliability and validity (Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002; Stahl, 2011) as well as to difficulties in their capability of measuring general or domain-specific EB (Muis, 2004; Stahl & Bromme, 2007).

Tying sophistication to the belief position may cause additional validity problems: For example, Clough (2007, p. 3) points out that belief positions “may be easily misinterpreted and abused” by being memorized as declarative knowledge. And Schommer herself argues for a “need for balance”; that is, extreme beliefs in positions like “knowledge is certain or tentative” are unfavorable as they might prevent persons from acting appropriately (Schommer-Aikins, 2004, p. 21). Especially for mathematics, in addition to arguments for uncertainty, there are also good arguments for knowledge not necessarily being regarded as uncertain and tentative (e.g., axioms and deductive reasoning) by persons with a sophisticated background. An interview study by Rott, Leuders, and Stahl (2014) has shown that research mathematicians partly argue for positions like certainty and stability of mathematical knowledge—and they do so in a very convincing way; that, is their *argumentation* shows reflected beliefs even though they argue for a position associated with “naïve” beliefs. Regarding the justification of knowledge, mathematics unlike other natural and educational sciences offers unique ways of working inductively as well as deductively (Horsten, 2018; Pólya, 1954). Again, either position (inductive vs. deductive) and its importance can be supported with convincing arguments (Rott & Leuders, 2016a). Therefore, at least for the domain of mathematics, it seems plausible to not tie the degree of sophistication to the belief *position* but instead to its

¹ Schommer’s study introduced not only a quantitative approach of assessing EB but also the idea of EB being a multi-dimensional construct (Hofer & Pintrich, 1997, p. 106).

argumentation (i.e., supporting the position inadequately or sophisticatedly). This, however, implies that it is not sufficient to just collect belief positions (via closed items).

These difficulties with traditional instruments led us to develop a questionnaire with open items (Rott & Leuders, 2016a). In this questionnaire, the participants are asked to give their written opinion on mathematical-philosophical questions as those given in Section 1 to assess their MEB. In a quantitative study using this questionnaire with 439 pre-service teachers (PST), we were able to show that belief position and argumentation (*inflexible* vs. *sophisticated*) can be differentiated (i.e., coded with high interrater agreement) and that a more convincing argumentation is not tied to one specific position (i.e., belief position and argumentation are statistically independent from each other).

However, the study by Rott and Leuders (2016a) was limited (a) to mathematics PST, whereby a larger and more diverse sample would be desirable. Additionally, (b) belief argumentation was only coded dichotomously, whereby investigating students' arguing in a more differentiated way would be preferable. Therefore, in order to replicate and expand the results from the previous study, the study presented here was conducted.

2.3 Critical thinking

To be able to locate beliefs within competence models for assessment in higher education (cf. Blömeke, Zlatkin-Troitschanskaia, Kuhn, & Fege, 2013; Baumert & Kunter, 2013), this study focuses not on only one cognitive dimension, that is, beliefs, but also on a second cognitive dimension, that is, an aspect of mathematical ability. We do not measure mathematical ability in terms of assessing success of attending university courses; instead, the *use* of students' knowledge is captured by drawing on the concept of critical thinking (CT) (Rott, Leuders, & Stahl, 2015).

Although CT has its roots in several fields of research (e.g., philosophy, psychology, and education) and there are many different conceptualizations of CT, it is commonly attributed to the following abilities (cf. Facione, 1990; Lai, 2011): analyzing arguments, claims, or evidence; making inferences using inductive or deductive reasoning; judging or evaluating and making decisions; or solving problems. Consistent with this list, Jablonka (2014, p. 121) defines CT as

a set of generic thinking and reasoning skills, including a disposition for using them, as well as a commitment to using the outcomes of CT as a basis for decision-making and problem solving.

All cited characterizations imply that CT is not a simple trait but a complex bundle of traits that cannot be located within one domain alone. However, as Facione (1990, p. 14) points out, there are domain-specific manifestations of CT. A mathematics-specific interpretation of CT, the importance of which is stressed by Jablonka (2014), is used in the study at hand.

To conceptualize CT, we refer to Stanovich and Stanovich (2010), who adapt and extend *dual process theory* (e.g., Kahneman, 2011). They distinguish fast, automatic, emotional, *subconscious* thinking, that is, the “autonomous mind,” from slow, effortful, logical, *conscious* thinking, the latter encompassing the “algorithmic” and the “reflective mind.” In this model, CT is identified within the reflective mind; persons think critically when they engage in hypothetical thinking or when autonomous and algorithmic processes are checked and, if necessary, overridden (Stanovich & Stanovich, 2010)—which fits Jablonka's definition of CT as thinking and reasoning skills, including a disposition for using them. Therefore, we define

CT in mathematics as those processes that consciously regulate the algorithmic use of mathematical procedures (Rott et al., 2015).

2.4 Measuring critical thinking

There have been several attempts at designing tests for CT, for example, the Ennis-Weir test in which participants are presented with a letter that contains complex arguments to which they are supposed to write a response (Ennis & Weir, 1985). Another example is the Watson-Glaser Critical Thinking Appraisal (Pearson Education, 2012), which consists of five subtests. In one of the subtests, for example, the probability of truth of inferences based on given information has to be rated on a 5-point Likert scale ranging from “true” to “false.” These tests, however, are domain-general and do not take into account mathematics-specific characteristics of CT (cf. Jablonka, 2014).

Stanovich and Stanovich (2010) suggest using tasks with an obvious but incorrect solution that needs to be checked thoroughly. Finding correct solutions of such tasks can help identify the disposition of thinking critically. We used this idea to construct a test for mathematical CT (MCT), arguing that such an assessment should reflect discipline-specific processes without requiring higher level mathematics (Rott et al., 2015). For this test, we adapted items from Stanovich, from Frederick (2005) who explicitly mentions the “mathematical content” (Frederick, 2005, p. 37) of his Cognitive Reflection Test, and others; details and examples of this test are presented in Section 3.

In previous studies with PST (Rott et al., 2015, Rott & Leuders, 2016a), this MCT test was able to differentiate between different programs of study: the more demanding the mathematics part of the program of studies, the greater the mean scores in the MCT test (PST for upper secondary school scored significantly higher than for lower secondary and primary schools). Additionally, this test supported the differentiation of MEB in *position* and *argumentation* by showing that belief position is not correlated to MCT, whereas belief argumentation is significantly positively correlated to MCT (i.e., students who argue sophisticatedly have greater scores in the MCT test; Rott et al., 2015; Rott & Leuders, 2016a).

2.5 Research goals

Building on the literature of the belief and knowledge dimensions and previous studies, with the goal to present an instrument for measuring MEB, the following *hypotheses* were tested:

- (H1) In students’ denotative MEB, belief position (whether they believe that mathematics is justified mostly deductively or inductively) and belief argumentation (how they reason for their position) can be distinguished. This was evaluated via an interrater agreement score.
- (H2) As shown in Rott and Leuders (2016a), we expect students’ belief positions and argumentations in this study to be statistically independent of each other, using a chi-square test.
- (H3) Students’ MCT scores will differ between programs of study: students aiming for “higher mathematics” (PST for upper secondary schools as well as Bachelor of Science who were attending “Analysis” and “Linear Algebra”) are more successful in MCT than students with less demanding mathematics lectures (PST for lower secondary,

primary, and special education schools who were attending “Elementary Mathematics”). This was evaluated with an ANOVA.

- (H4) As observed in previous studies, MCT scores will not be correlated with belief position but rather with belief argumentation. This was evaluated with another ANOVA.

3 Methodology

In this section, information on the study and its participants are given; thereafter, methods for measuring MEB and MCT are presented.

Several instruments were used in this study: a test for connotative EB (the CAEB by Stahl and Bromme (2007)), questionnaires for denotative MEB in two dimensions, and certainty and justification of knowledge, as well as the MCT test. In this article, only the latter two are reported.

3.1 Participants

At the beginning of the winter semester of 2017/2018, the instruments were distributed within mathematics lectures (or related tutorials) that specifically addressed students at the beginning of their studies at the University of Cologne, Germany. Choosing those lectures gave us access to all students with a focus on mathematics (see Table 3 for the students’ programs of study).

Participating in the study or refusing to participate did not affect the students’ outcome for those lectures. In total, more than 600 students (with individually generated pseudonyms) voluntarily participated, of whom 581 students (65% female) completed the instruments. Even though the study was conducted in lectures for first semester students, not all students were in their first semester: some repeated lectures or others had changed their program of study. The mean number of semesters in the participants’ current programs of study was 1.2 (standard deviation 0.8); the mean number of their total number of semesters at a university was 1.6 (standard deviation 1.6).

3.2 Measuring mathematics-related epistemological beliefs

The interview study by Rott et al. (2014) has shown that it is not easy (especially for students) to argue context-free (i.e., without given positions to refer to) in relation to challenging philosophical questions. Therefore, the interviewees were presented with two introductory quotations, with which they should agree or disagree at the beginning of the interviews. The corresponding statements are not necessarily mutually exclusive, but stimulate a discussion about the different positions. This approach has worked well insofar as the students were better able to discuss the topic. Therefore, we also used the introductory quotations in our questionnaire (Rott & Leuders, 2016a). The prompts regarding the dimension *justification of knowledge* are given in Table 1.²

The answers of the participants to the partial questions (a–d) were regarded as one text block and then coded in relation to two dimensions: (1) *position* and (2) *argumentation*. The first code records which *position* the participants have chosen, *deductive* or *inductive*. In this

² The participants were given German translations of the prompts; in mathematical contexts, the German word for justification, “Begründung,” captures the same meaning.

Table 1 Prompts and questions in the questionnaire regarding the justification of mathematical knowledge

Mathematical discovery is justified deductively	Mathematical discovery is justified inductively
<p>“Mathematics is a deductive science: starting from certain premises, it arrives, by a strict process of deduction, at the various theorems which constitute it. [...] No appeal to common sense, or ‘intuition,’ or anything except strict deductive logic, ought to be needed in mathematics after the premises have been laid down.” (Bertrand Russell) (1919, p. 144)</p>	<p>“Mathematics is an experimental science. The formulation and testing of hypothesis play in mathematics a part not other than in chemistry, physics, astronomy, or botany. [...] It matters little that the mathematician experiments with pencil and paper while the chemist uses testtube and retort, or the biologist stains and the microscope.” (Norbert Wiener) (1923, p. 237 ff.)</p>
<p>Which of the two positions can agree with? Please indicate this on the following scale: deductive ○ ○ ○ ○ ○ inductive</p>	
<p>a) Which of the two positions regarding the justification of mathematical knowledge can you identify yourself with?</p> <p>b) Please give reasons for your judgment regarding deductive/inductive justification of mathematical knowledge.</p> <p>c) Did you yourself make experiences that support one position or the other?</p> <p>d) Compare the justification of mathematical knowledge to that of knowledge from other domains. For example, is mathematical knowledge justified differently than knowledge from physics, language sciences, or educational sciences?</p>	

study, only five participants (less than 1%) were able to give good reasons, examples, and/or situations for both positions; arguing for both positions—not untypical for experts—can therefore be neglected in the present study with first-year students.

The second code refers to the *argumentation*—in contrast to the previous study, in which two values (inflexible and sophisticated) were coded (Rott & Leuders, 2016a), four values (inflexible is now divided into “inadequate” and “simple”; “no argumentation” was added) can now be coded with high interrater agreement (Cohen’s $\kappa = 0.82$; see the Appendix for details). After calculating the interrater reliability, the differing scores have been re-coded by the two raters (the author and a research assistant) consensually.

The students’ texts were analyzed for the quality of their arguments. As it was not the goal of this research to analyze the structure of their reasoning, the texts were not coded with the Toulmin model (e.g., Knipping, 2008). However, the terms by Toulmin (2003) are used to operationalize the following categories with the claim for “deductive” or “inductive discovery.”

No Argumentation is coded if all text fields are empty. This category is also assigned if only individual words have been entered, e.g., “no” for the question whether one has gained experience with it, but the other fields have not been filled in.

Inadequate is coded if the argumentation does not fit the epistemological meaning of the question. This could be either a (more or less) consistent and elaborate argumentation that lacks reference to the question (e.g., arguing for the importance of “inductive methods like calculating an example in problem solving”) or an argumentation with reference to the question that lacks adequate warrants (e.g., an overemphasis of subjective experiences instead of arguing for mathematical discovery in general), maybe because they never thought about such a question.

Simple is coded if the argumentation contains warrants and/or data that fit the epistemological meaning of the question, but does not reveal an in-depth examination of the

Table 2 Coded extracts of sample statements for the discussion “deductive/inductive” (student codenames, group according to Table 3)

Position	Argumentation		
	Inadequate	Simple	Sophisticated
Deductive	“inductive [sic], because we learned the axioms in the first lectures. Mathematics is not tangible like e.g., physics.” (SHET-27, (1))	“Mathematics is made up from axioms and proven from there. In other areas, inductive work is more common.” (BBAB-11, (1))	“Since mathematics derives its theorems from definite axioms and does not prove them through experiments, it is deductive. Discoveries from other sciences must be statistically verified or derived from statistics. That is why mathematical discoveries are different.” (KVAP-12, (1))
Inductive	“An equation does not always apply directly to everything. Sometimes you have to try it out.” (ABTI-08, (6))	“Because mathematics is a science that is developed through experimentation. Yes, they are different, because e.g., in educational sciences they are based on experience and mathematical discoveries are based on experiments.” (MBEN-21, (5))	“You have to experiment to come up with new theorems, because without hypotheses you have nothing to prove, explain, or grasp. Every science is based on theses and experiments. Especially in the educational sciences.” (HMAI-26, (5))

topic. Sometimes, facts and opinions are not clearly separated. Usually, the contradictory position is not rebutted. Examples for warrants or data of this category would be “mathematical knowledge is deductively justified because one proceeds logically” or “mathematical reasoning starts with axioms,” respectively. This definition includes argumentations that simply reproduce warrants from the introductory quotations (without additional reflection).

Sophisticated is coded if the argumentation contains warrants and/or data and/or backings that show an in-depth understanding of the topic, that is if (at least two) arguments have been combined to reach a conclusive argument, for example “By great mathematicians like Leonhard Euler and Bernhard Riemann research diaries have been handed down. This shows that they have tried out examples page by page before setting up rules—so they have proceeded inductively.” A single (good) argument can be sufficient, if it is not taken from the introductory quotations. Usually, data are stated explicitly and the contradictory position is rebutted.

In this way, representatives were found for all possible combinations of position and argumentation; one example each is given in Table 2 (translated from German to English).

3.3 Measuring mathematical critical thinking

We have developed a mathematical test with 11 items similar to and including the well-known bat-and-ball task:

A bat and a ball cost \$ 1.10 in total. The bat costs \$ 1 more than the ball. How much does the ball cost? (cf. Frederick, 2005; Kahneman, 2011)

All items were designed to be solvable with simple mathematics procedures (i.e., no items were used that resemble written argumentations like in the Ennis & Weir, 1985 test), but suggest an intuitive but incorrect answer like \$ 0.10 in the bat-and-ball task.

The items allow measuring the disposition to reflect upon “obvious” algorithms and solutions, that is, critical thinking (see Section 2). All items have been validated both quantitatively and qualitatively (Rott & Leuders, 2016b), and were rated dichotomously with 1 point for a correct and 0 point for a wrong solution. Because of the non-linear nature of raw data³ (Boone, Staver, & Yale, 2014, p. 6 ff.), a Rasch model (software Winsteps Version 3.91.0; Linacre, 2005) was used to transform the students’ test scores into values on a one-dimensional competence scale (confirming that all items measure MCT instead of other traits like computational skills). In the previous study, two items had been eliminated because of underdiscrimination (Rott et al., 2015). Here, both items have been replaced and the Rasch analysis shows that all MCT items are within the reasonable mean-square (MNSQ) ranges (between 0.5 and 1.5; Boone et al., 2014, p. 166 f.). All MCT items, their empirical solution rates, and their Rasch values are presented in the [Appendix](#).

The students filled in the instruments on paper and were given as much time as they needed to do so. On average, they took 10–12 min for the MCT test and thereafter 12–15 min for the MEB test regarding justification of knowledge.

4 Results and discussion

In this section, results of the MEB and MCT tests are analyzed individually; thereafter, possible connections between both cognitive dimensions are explored.

4.1 Mathematics-related epistemological beliefs

Table 3 shows the distribution of students for the position code, sorted by their programs of study. Overall, in those groups, the numbers of students arguing for either position are surprisingly evenly distributed. Looking at the curricula for German Secondary School (KMK, 2003), we had expected a higher proportion for the position “inductive” in the study at hand. Oftentimes, algorithms and theorems are presented by teachers and then practiced with exercises, but not proven, meaning deductive reasoning plays a negligible role. Fittingly, in the previous study (Rott & Leuders, 2016a), almost two-thirds of the students argued for “inductive” and only one-third for “deductive.” Further research is needed to explain the distribution in the study at hand.

As stated above, the argumentation code could be applied with high interrater agreement (confirming H1). The distribution of the coded statements is given in Table 4. Additionally, the mean number of words for each category is given (for comparison, the introductory prompts in the questionnaire each contain 50 words). There is a general trend of “more words go along with a higher argumentation category.” However, “inadequate” has a higher average word count than “simple;” together with the minima⁴ and maxima, this shows that a large

³ Empirical data shows that a lot of students manage to gain 5 instead of 4 points, but only a handful of students gain 10 instead of 9 points, i.e., the latter difference of 1 is more difficult to gap than the former one.

⁴ An example for a “simple” argumentation with four words is “axioms and deductive reasoning” given by the student with the codename SBEN-23.

Table 3 Number of coded belief positions, sorted by programs of study

Programs of study	No position	Deductive	Inductive	Total
(1) Bachelor of Science	12 (14.8%)	25 (30.8%)	44 (54.3%)	81
(2) Business Informatics	8 (13.6%)	31 (52.5%)	20 (33.9%)	59
(3) PST Upper Secondary	8 (9.6%)	39 (47.0%)	36 (43.4%)	83
(4) PST Lower Secondary	9 (32.1%)	8 (28.6%)	11 (39.3%)	28
(5) PST Primary Education	1 (0.9%)	55 (50.0%)	54 (49.1%)	110
(6) PST Special Education	9 (5.9%)	68 (44.7%)	75 (49.3%)	152
Other	9 (13.2%)	28 (41.2%)	31 (45.6%)	68
Sum	56 (9.6%)	254 (43.7%)	271 (46.6%)	581

Table 4 Average number of words per argumentation category

	No argumentation	Inadequate	Simple	Sophisticated	Total
Number of codes	178	27	356	20	581
Words mean (SD)	0.3 (1.3)	24.0 (14.4)	21.9 (13.1)	47.6 (19.7)	15.9 (16.1)
Words median	0	20	19	41	12
Words min/max	0/12	7/75	4/82	21/103	0/103

number of words is not sufficient for a high category. (See the [Appendix](#) for a further analysis.)

In [Table 5](#), the distribution of the argumentation code is broken down by the participants' programs of study. The bottom line in the "No argumentation" column reveals that 30.6% of the participants did not write arguments in this part of the questionnaire. The low percentages in the column "Inadequate" suggest that almost all students who were willing to write down an answer were at least able to use warrants from the given prompts (i.e., being coded with "simple"). Finally, the column "Sophisticated" shows that, with regard to their argumentation, the Bachelor of Science (Mathematics and Business Mathematics combined) students show higher absolute and relative numbers than the PSTs (for different school types). Some of the students (11.1%) who have decided to start studying mathematics as a major thus seem to have thought intensively about the subject and its epistemology.

The number of students with sophisticated codes is low (3.4%; see [Table 5](#)). This is in line with the previous study which had shown that arguing sophisticatedly is difficult for students (7.3% in [Rott & Leuders, 2016a](#)). The even lower proportion in this study was expected, as most participants were in their first semester and therefore inexperienced in university-level mathematics and its philosophy (a semester effect was also observed in [Rott et al., 2015](#)).

Table 5 Number of coded belief argumentations, sorted by programs of study

Programs of study	No argumentation	Inadequate	Simple	Sophisticated	Total
(1) Bachelor of Science	27 (33.3%)	3 (3.7%)	42 (51.9%)	9 (11.1%)	81
(2) Business Informatics	32 (54.2%)	0 (0.0%)	25 (42.4%)	2 (3.4%)	59
(3) PST Upper Secondary	29 (34.9%)	4 (4.8%)	49 (58.3%)	1 (1.2%)	83
(4) PST Lower Secondary	14 (50.0%)	1 (3.6%)	12 (42.9%)	1 (3.6%)	28
(5) PST Primary Education	8 (7.3%)	7 (6.4%)	92 (83.6%)	3 (2.7%)	110
(6) PST Special Education	29 (19.1%)	10 (6.6%)	110 (72.4%)	3 (2.0%)	152
Other	39 (57.4%)	2 (2.9%)	26 (38.2%)	1 (1.5%)	68
Sum	178 (30.6%)	27 (4.6%)	356 (61.3%)	20 (3.4%)	581

Table 6 Comparison of both dimensions of the questionnaire for all students; the numbers in brackets indicate expected frequency under the assumption of statistical independence

Position	Argumentation				Sum
	No argumentation	Inadequate	Simple	Sophisticated	
Deductively	62 (57.8)	13 (13.0)	163 (171.5)	14 (9.6)	252
Inductively	58 (62.2)	14 (14.0)	193 (184.5)	6 (10.4)	271
Sum	120	27	356	20	523
No position	58	–	–	–	581

$$\chi^2 = 5.22; df = 3; p = 0.16$$

A chi-square test (Table 6) comparing belief position and argumentation (expected frequencies assuming statistical independence are given in parentheses) gives no evidence for both dimensions being statistically dependent on each other. This confirms the corresponding result from Rott and Leuders (2016a) using the finer grained argumentation coding (H2). This finding supports the validity concerns regarding traditional questionnaires that measure belief argumentation by collecting belief positions (see Section 2).

4.2 Mathematical critical thinking

The Rasch model of the participants' MCT ability provides metrical latent variables ranging from -3 to 3 with low values indicating a low ability (Table 7). The MCT results (column "Rasch-M") reveal significant differences between the groups of different programs of study (one-way ANOVA: $F = 19.2$; $df = 5$; $p < 0.0001$). Tukey post hoc tests (all either non-significant or $p < 0.01$) show that there are three groups: PSTs for Upper Secondary schools, Business Informatics students, and Bachelor of Science students form the top group. They do not differ significantly from each other, but achieve significantly higher scores than all other groups. PSTs for Lower Secondary Schools and Special Education Schools do not differ from each other, but score significantly higher than PST for Primary Schools (confirming H3). In Cologne, students of the latter group are obliged to attend mathematics lectures (as all primary school teachers in Germany have to teach mathematics); all other students chose mathematics (as their major or as a school subject to teach) voluntarily. As most of the students were at the beginning of their first semester, this result cannot be caused by their university education. Instead, it implies a selection effect for the different programs of study. Students with higher MCT abilities seem to be more likely to choose study programs with higher mathematical demands.

Table 7 Means (M) and standard deviations (SD) of "mathematical critical thinking" (MCT) sorted by the programs of study

Programs of study	Number	Min	Max	Median	Rasch-M	Rasch-SD
(1) Bachelor of Science	81	1	11	5	0.03	1.05
(2) Business Informatics	59	2	9	6	0.05	0.80
(3) PST Upper Secondary	83	2	9	6	0.18	0.93
(4) PST Lower Secondary	28	2	7	5	-0.44	0.72
(5) PST Primary	110	1	8	4	-0.74	0.78
(6) PST Special Education	152	0	9	4	-0.50	0.93
Sum (including Other)	581	0	11	5	-0.30	0.96

4.3 Connections between mathematics-related epistemological beliefs and mathematical critical thinking

An interesting result from the previous study was a correlative relationship between the MEB codes and the results of the MCT test (Rott & Leuders, 2016a). There was no correlation with regard to belief position (*deductive* vs. *inductive*), but a significant correlation with regard to belief argumentation (the latter was coded in two stages in the previous study: *inflexible* vs. *sophisticated*).

Table 8 shows the mean values (and standard deviations) of the MCT Rasch scores, sorted according to the two belief dimensions. A two-way ANOVA (after excluding the data for “No position”) shows that there are no significant differences in the MCT scores for students holding different belief *positions* (totals in the rows “Deductive” and “Inductive”; $F = 0.26$, $df = 1$, $p = 0.61$), but significant differences for belief *argumentation* (totals in the columns “No argumentation,” “Inadequate,” “Simple,” and “Sophisticated”; $F = 3.49$, $df = 3$, $p = 0.02$) and no interaction effects ($F = 2.14$, $df = 3$, $p = 0.09$).

Tukey post hoc tests regarding argumentation show that students arguing sophisticatedly score higher in the MCT test than students with no, inadequate, or simple argumentations ($p < 0.01$); other differences are not significant. This confirms the results from the previous study (H4). An explanation for this finding might be the fact that both MCT and belief argumentation tap into a kind of reflectiveness of the students, which belief position does not. (Versions of this table for sub-groups of different programs of study are given in the Appendix.)

Final note The results in this section rely on statistical methods to differentiate patterns that occur by chance from patterns with meaning. The use of significance testing should be reflected upon as this might lead to non-reproducible results—discussed as the “replication crisis” (cf. Inglis, Schukajlow, Van Dooren, & Hannula, 2018)—or false conclusions—discussed as the “ p value problem” (cf. Matthews, 2017). Regarding the former, researchers call for more replication studies. There is, however, no easy solution to fix the latter; no substitute method fixes all problems. Instead, we have to be careful regarding conclusions drawn from significance testing. Triangulating and replicating results—which is one of the goals of this study—might be one way of dealing with such issues.

Table 8 Means and standard deviations (SD) of “mathematical critical thinking” sorted by the two belief dimensions (position and argumentation)

		Argumentation				Total
		No argumentation	Inadequate	Simple	Sophisticated	
Position	No position	−0.04 (0.93) <i>n</i> = 58	--- (---) <i>n</i> = 0	--- (---) <i>n</i> = 0	--- (---) <i>n</i> = 0	−0.04 (0.93) <i>n</i> = 58
	Deductive	−0.47 (0.83) <i>n</i> = 62	−0.45 (1.34) <i>n</i> = 13	−0.32 (0.94) <i>n</i> = 163	0.59 (1.34) <i>n</i> = 14	−0.31 (0.98) <i>n</i> = 252
	Inductive	−0.26 (1.12) <i>n</i> = 58	−0.36 (1.00) <i>n</i> = 14	−0.40 (0.87) <i>n</i> = 193	−0.28 (0.97) <i>n</i> = 6	−0.35 (0.94) <i>n</i> = 271
	Total (excluding “No position”)	−0.37 (0.98) <i>n</i> = 120	−0.27 (1.17) <i>n</i> = 27	−0.36 (0.90) <i>n</i> = 356	0.33 (1.29) <i>n</i> = 20	−0.33 (0.96) <i>n</i> = 523

5 General discussion

Research on beliefs, especially on domain-specific EB, is still an open field for theoretical and methodological developments (Safrudiannur, 2020). Studies on EB, especially studies using a quantitative approach (e.g., Schommer, 1990), often only address belief positions, deriving a degree of sophistication from the positions. Theoretical considerations, however, suggest that such an approach might not be sufficient to adequately capture belief sophistication. At least for the domain of mathematics, there are convincing arguments supporting different belief positions, even those that might generally be associated with “naïve” beliefs (e.g., regarding the certainty of knowledge). Therefore, the study at hand presents and reflects upon an instrument for measuring denotative MEB in which university students’ belief positions and argumentations are differentiated. Both aspects of beliefs can be coded separately and the results suggest that they are statistically independent of each other. This result has clear implications for research: Many studies (like COACTIV; Krauss et al., 2008) use beliefs (including EB) as covariates; generally, beliefs are collected with self-report Likert scale items measuring belief positions rather than argumentations. In the study at hand, however, correlations of both aspects of beliefs with another cognitive dimension, that is, mathematical critical thinking (measured with an MCT test), suggest that belief argumentations seem to be more meaningful than belief positions (as the former show significant correlations to MCT scores whereas the latter do not). It could be inferred that measuring beliefs should not only address belief positions.

Further implications, especially for teaching mathematics (at schools or universities), need to be studied in the future. It seems plausible that teachers should know arguments both speaking for and against certain epistemological positions if they want to teach their students about mathematics, that is, what it means to think and act like mathematicians, instead of only about mathematical procedures.

For the sub-group of the PST in the sample, the study at hand confirms previous results on MEB (Rott & Leuders, 2016a). In a wide sense of the term, this study could be interpreted as a replication, using a refined methodology (as claimed by Maxwell, Lau, & Howard, 2015) to address limitations of the original study, extending the results to a more diverse sample (mathematics, business mathematics, and informatics majors).

Additionally, using open items instead of Likert scale items could help in decreasing the social desirability problem (cf. Di Martino & Sabena, 2010) as the students cannot simply choose statements that they think are desirable; instead, participants have to argue for themselves.

However, the need to write down answers instead of checking boxes is not only a feature but also a *limitation* of this study. The proportion of participants (30.6%) not arguing indicate that the questionnaire also measures the willingness to write down an argumentation. Additionally, the choice of specific prompts starting the discussion and the fact that they do not represent a real dichotomy might influence participants’ answers. The latter is not regarded as problematic because the prompts succeeded in initiating discussions; nevertheless, different prompts will be tested.

An important factor might be the epistemological content, as only *justification of knowledge* has been addressed in this article. However, previous studies (Rott et al., 2015) have shown similar results for *certainty of knowledge*, indicating that independence of belief position and argumentation as well as correlations of the latter to MCT are not tied to a single EB dimension.

From the way the MEB questionnaire is designed, it follows that specific kinds of beliefs are measured. On the scale from subjective to objective knowledge (see Section 2), compared to most other studies (especially those in the tradition of Schommer, 1990), the beliefs that are measured in this study are more on the “objective” side, drawing heavily on knowledge about epistemology.

Another limitation regards the selection of the sample, as all participants are enrolled in one university. This limitation is mitigated because this specific university is very large and not limited to attracting a special group of students.

Familiarity of the participants with MCT items like the bat-and-ball task may present another limitation. However, this kind of research is not part of school or university curricula in Germany. Very similar solution rates of the MCT items in this and previous studies that were conducted in two different cities also suggest that knowledge about the items does not distort the results: for example, the bat-and-ball task had a solution rate of 55% in Rott et al. (2015), compared to a solution rate of 59% in this study; the dice task (red and green sides) had solution rates of 22% and 23%, respectively.

It will be very interesting to explore the influence of knowledge on the MEB test and the development of students’ MEB, respectively. As Bromme, Kienhues, and Stahl (2008) point out, knowledge gain does not necessarily lead to more sophisticated EB. The results presented in this article set a clear starting point, as exposure to knowledge that could distort the measurement previously to joining the university is very unlikely. In Germany, teaching epistemology is not part of any school curricula. Students with more knowledge have already participated in a continuation of this study: At the beginning of the winter terms 2018/2019 and 2019/2020, we collected data regarding MEB and MCT in mathematics lectures, mainly addressing students in respectively their third and fifth semester, in a pseudo-longitudinal and a real longitudinal study (with ca. 450 students in 2018, 850 students in 2019, and 100 who participated on all three survey dates). As the students have been experiencing university mathematics generally and proving specifically, it will be interesting to see how the distribution of belief position and argumentation have changed.

Acknowledgments I would like to thank the editors Vilma Mesa and Wim Van Dooren as well as the reviewers for their constructive feedback on previous versions of this manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Barrow, J. D. (1992). *Pi in the sky. Counting, thinking, and being*. Oxford, UK: Oxford University Press.

- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (pp. 25–48). New York, NY: Springer.
- Blömeke, S., Zlatkin-Troitschanskaia, O., Kuhn, C., & Fege, J. (Eds.). (2013). *Modeling and measuring competencies in higher education*. Rotterdam, the Netherlands: Sense.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht, the Netherlands: Springer.
- Bromme, R. (2005). Thinking and knowing about knowledge – A plea for critical remarks on psychological research programs on epistemological beliefs. In M. Hoffmann, J. Lenhard, & F. Seeger (Eds.), *Activity and sign – Grounding mathematics education* (pp. 191–201). New York, NY: Springer.
- Bromme, R., Kienhues, D., & Stahl, E. (2008). Knowledge and epistemological beliefs: An intimate but complicate relationship. In M. S. Khine (Ed.), *Knowing, knowledge and beliefs: Epistemological studies across diverse cultures* (pp. 423–441). Dordrecht, the Netherlands: Springer.
- Clough, M. P. (2007). Teaching the nature of science to secondary and post-secondary students: Questions rather than tenets, The Pantaneto Forum, Issue 25, <http://www.pantaneto.co.uk/issue25/front25.htm>, January. Republished (2008) in the *California Journal of Science Education*, 8(2), 31–40.
- Di Martino, P., & Sabena, C. (2010). Teachers' beliefs: The problem of inconsistency with practice. In M. Pinto & T. Kawasaki (Eds.), *Proceedings of the 34th Conference of the International Group for the Psychology of Mathematics Education* (vol. 2, pp. 313–320). Belo Horizonte, Brazil: PME.
- Ennis, R. H., & Weir, E. (1985). *The Ennis-Weir critical thinking essay test*. Pacific Grove, CA: Midwest Publications.
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Executive summary “The Delphi Report”. Millbrae, CA: The California Academic Press.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140.
- Horsten, L. (2018). Philosophy of mathematics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2018 Edition). Retrieved from <https://plato.stanford.edu/archives/spr2018/philosophy-mathemat/philosophy-mathematics/>. Accessed 1 Oct 2020.
- Inglis, M., Schukajlow, S., Van Dooren, W., & Hannula, M. (2018). Replication in mathematics education. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.), *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (vol. 1, pp. 195–196). Umeå, Sweden: PME.
- Jablonka, E. (2014). Critical thinking in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 121–125). Dordrecht, the Netherlands: Springer.
- Kahneman, D. (2011). *Thinking, fast and slow*. London, UK: Penguin Books Ltd..
- Knipping, C. (2008). A method for revealing structures of argumentations in classroom proving processes. *ZDM-Mathematics Education*, 40, 427–441.
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Kunter, M., & Jordan, A. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematiklehrerinnen und -lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29(3/4), 223–258.
- Kultusministerkonferenz (KMK). (2003). *Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss* [Resolutions by the Standing Conference of Ministers of Education and Cultural Affairs: Educational standards in mathematics for the intermediate school leaving certificate], http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-Bildungsstandards-Mathe-Mittleren-SA.pdf. Accessed 1 Oct 2020.
- Lai, E. R. (2011). *Critical thinking: A literature review*. Upper Saddle River, NJ: Pearson www.pearsonassessments.com/hai/images/tmrs/criticalthinkingreviewfinal.pdf. Accessed 1 Oct 2020.
- Lakatos, I. (1976). *Proofs and refutations*. New York, NY: Cambridge University Press.
- Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, 39(6), 497–521.
- Linacre, J. M. (2005). *Winsteps Rasch analysis software*. PO Box 811322, Chicago IL 60681-1322, USA. <http://www.winsteps.com/index.htm>
- Matthews, R. (2017). The ASA's *p*-value statement, one year on. *Significance*, 14(2), 38–41.
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, 70(6), 487–498.
- Muis, K. R. (2004). Personal epistemology and mathematics: A critical review and synthesis of research. *Review of Educational Research*, 74(3), 317–377.

- Murphy, P. K., & Mason, L. (2006). Changing knowledge and beliefs. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology – Second edition* (pp. 305–324). London, UK: Lawrence Erlbaum Publishers.
- Pearson Education. (2012). *Watson-Glaser critical thinking appraisal user-guide and technical manual*. <http://www.talentlens.co.uk/assets/news-and-events/watson-glaser-user-guide-and-technical-manual.pdf>. Accessed 14 Mar 2019.
- Pehkonen, E. (1999). Conceptions and images of mathematics professors on teaching mathematics in school. *International Journal of Mathematical Education in Science and Technology*, 30(3), 389–397.
- Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257–315). Charlotte, NC: Information Age.
- Pólya, G. (1954). Mathematics and plausible reasoning. In *Volume 1: Induction and analogy in mathematics*. Princeton, MA: Princeton University Press.
- Rott, B., & Leuders, T. (2016a). Inductive and deductive justification of knowledge: Flexible judgments underneath stable beliefs in teacher education. *Mathematical Thinking and Learning*, 18(4), 271–286.
- Rott, B., & Leuders, T. (2016b). Mathematical critical thinking: The construction and validation of a test. In C. Csikos, A. Rausch, & J. Szitányi (Eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* (vol. 4, pp. 139–146). Szeged, Hungary: PME.
- Rott, B., Leuders, T., & Stahl, E. (2014). “Is mathematical knowledge certain? – Are you sure?” An interview study to investigate epistemic beliefs. *Mathematica Didactica*, 37, 118–132.
- Rott, B., Leuders, T., & Stahl, E. (2015). Assessment of mathematical competencies and epistemic cognition of pre-service teachers. *Zeitschrift für Psychologie*, 223(1), 39–46.
- Russell, B. (1919). *Introduction to mathematical philosophy*. London, UK: Allen & Unwin.
- Safrudianur. (2020). *Measuring beliefs quantitatively. Criticizing the use of Likert scale and offering a new approach*. Wiesbaden, Germany: Springer.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82, 498–504.
- Schommer-Aikins, M. (2004). Explaining the epistemological belief system: Introducing the embedded systemic model and coordinated research approach. *Educational Psychologist*, 39(1), 19–29. https://doi.org/10.1207/s15326985ep3901_3
- Stahl, E. (2011). The generative nature of epistemological judgments: Focusing on interactions instead of elements to understand the relationship between epistemological beliefs and cognitive flexibility. In J. Elen, E. Stahl, R. Bromme, & G. Clarebout (Eds.), *Links between beliefs and cognitive flexibility – lessons learned* (pp. 37–60). Dordrecht, the Netherlands: Springer.
- Stahl, E., & Bromme, R. (2007). The CAEB: An instrument for measuring connotative aspects of epistemological beliefs. *Learning and Instruction*, 17, 773–785.
- Stanovich, K. E., & Stanovich, P. J. (2010). A framework for critical thinking, rational thinking, and intelligence. In D. Preiss & R. J. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching and human development* (pp. 195–237). New York, NY: Springer.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematic learning and teaching* (pp. 127–146). New York, NY: Macmillan.
- Toulmin, S. (2003). *The uses of argument, updated edition*. Cambridge, UK: Cambridge University Press.
- Trautwein, U., & Lüdtke, O. (2007). Epistemological beliefs, school achievement, and college major: A large-scale longitudinal study on the impact of certainty beliefs. *Contemporary Educational Psychology*, 32, 348–366.
- Tsai, C.-C. (1998a). An analysis of scientific epistemological beliefs and learning orientations of Taiwanese eight graders. *Science Education*, 82(4), 473–489.
- Tsai, C.-C. (1998b). An analysis of Taiwanese eighth graders' science achievement, scientific epistemological beliefs and cognitive structure outcomes after learning basic atomic theory. *International Journal of Science Education*, 20(4), 413–425.
- Wiener, N. (1923). *Collected works: With commentaries*. Cambridge, MA: The MIT Press.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.