HISTORICAL ENCOUNTERS
A journal of historical consciousness, historical cultures, and history education

ISSN 2203-7543 | Double Blind Peer-Reviewed | Open Access | www.hej-hermes.net

Thinking aloud about epistemology in history: How do students understand the Beliefs about History Questionnaire?

Marcel Mierwald
Ruhr University Bochum, Germany

Maximilian Junius
Secondary School Gevelsberg, Germany

ABSTRACT

This study aims to explore the cognitive validity of a popular epistemological beliefs self-report instrument used in history education, namely the Beliefs about History Questionnaire (BHQ) developed by Maggioni (2010). The validity and reliability of this instrument were found to be problematic during the quantitative validation of both the original English questionnaire and its foreign language versions. Therefore, we conducted cognitive interviews with four students (all 17 years old) using a German version of the BHQ to gain a comprehensive insight into students' understanding of the questionnaire and the possible difficulties they experience in answering its items. The analysis of the interviews showed that the cognitive validity of the questionnaire was good. However, some items were found to be problematic because the students showed differences in understanding and difficulties in responding. Furthermore, four overarching problem areas were identified: the complexity of terms; epistemic ambiguity; length and comprehensibility; and irritating references to the school context. In this article, we address these and other difficulties in using the BHQ to assess students' thoughts about epistemology in history. Finally, possible improvements to the questionnaire and conclusions that can be applied to both research and practice are presented.

KEYWORDS

Epistemological beliefs, cognitive interviewing, questionnaire, cognitive validity, domain of history, secondary school education

CITATION

Mierwald, M., & Junius, M. (2022). Thinking aloud about epistemology in history: How do students understand the Beliefs about History Questionnaire? *Historical Encounters, 9*(1), 35-57. https://doi.org/10.52289/hej9.103

COPYRIGHT

© Copyright retained by Authors
Published 15 April 2022
Distributed under a CC BY-NC-ND 4.0 License

Introduction

While Sarah[1], a student who participated in our study, was thinking aloud about an item of the German version of the Beliefs about History Questionnaire (BHQ) originally developed by Maggioni (2010), she told us more about her understanding of the past and the role of historians:

> Well, I think that the past is actually what is not today, that is, it was before. And I do not know to what extent historians make the past. I think that the historian can represent the past in his way, but the past is still there. Thus, I do not think that the past depends on a historian.[2]

For researchers in the field of educational psychology and history education, the investigation of the concepts of knowledge and knowing (i.e., epistemological beliefs), such as those in this student's example, is important because they are assumed to influence both teaching and learning (e.g. Hofer & Bendixen, 2012; VanSledright & Limón, 2006). Regarding history education, some framework models exist that link epistemological beliefs and "knowing and doing" in history (e.g. Havekes, van Boxtel, Coppen, & Luttenberg, 2012; van Drie & van Boxtel, 2008; VanSledright, 2014). This argument is supported by observations from qualitative research, as students' difficulties in dealing with primary sources and historical accounts have been attributed to them having naïve beliefs about history (Lee & Shemilt, 2003; VanSledright, 2002; Wineburg, 2001). In addition, quantitative studies have shown that students' epistemological beliefs in history are related to second-order and strategic knowledge (Stoel, van Drie, & van Boxtel, 2017), the use of learning strategies and text comprehension (Ioannou & Iordanou, 2019), as well as argumentation skills (Mierwald, 2020; Mierwald, Seiffert, Lehmann, & Brauch, 2016).

As a result of the great importance of epistemological beliefs for historical learning and thinking, researchers have long been interested in how epistemological beliefs can be theoretically conceptualized and measured using valid and reliable instruments (e.g. Maggioni, Alexander, & VanSledright, 2004; Maggioni, VanSledright, & Alexander, 2009; Mierwald, Seiffert, Lehmann, & Brauch, 2017; Nitsche, 2019; Stoel, Logtenberg, Wansink, Huijgen, van Boxtel, & van Drie, 2017). In particular, the BHQ, developed by Maggioni (2010), became a key measurement tool used in research on epistemological beliefs in the domain of history. The theoretical model that the questionnaire is based on differentiates between three epistemic stances, including the perspectives of history as a "copy" of the past (*objectivist stance*), and history as just a matter of historians' interpretation (*subjectivist stance*), as well as the more developed view of history as a critical method of inquiry about the past (*criterialist stance*). Maggioni (2010) and many other researchers (e.g. Kidwai, 2015; Mierwald et al., 2017; Miguel-Revilla, Carril-Merino, & Sánchez-Agusti, 2020; Stoel, van Drie, & van Boxtel, 2015, 2016) have used the BHQ and its foreign language translations in their research with varying degrees of success. In particular, the validity and reliability of the instrument have been found to be problematic. For example, individual items did not load as expected in factor analyses, and certain scales showed low reliability. This could have arisen due to the theoretical background of the instrument itself, the individual formulations inserted in the instrument, the terminology of items, the translation or adaptation of items for non-English speaking participants, and the different educational and cultural contexts of respondents (Mierwald et al., 2017; Miguel-Revilla et al., 2020; Stoel et al., 2017).

To obtain more detailed insights into how students understand the questionnaire, we conducted cognitive interviews with a sample of secondary school students. The aim was to examine the cognitive validity of the existing BHQ using the German version as an example and to provide suggestions for future optimization of the instrument (e.g. Greene & Yu, 2014). Following this, it may be possible in the future to optimize the measurement of epistemological beliefs using the BHQ or its adaptions. Importantly, a better measure is essential for identifying educational implications at a diagnostic level and for verifying the effectiveness of interventions that aim to enhance epistemological beliefs about history (Mason, 2016).

Therefore, the research questions guiding this study are:

1. What are students' general impressions of the BHQ?
2. How do students understand the BHQ and what problems do they have with the individual items?
3. To what extent are the items that turned out to be particularly problematic in the exploratory factor analysis (EFA) perceived by the students?

Subsequently, we describe the modeling and assessment of epistemological beliefs in history education at the present time. Following this, we present the method and results of our study and, finally, the discussion of our findings.

Theoretical Framework

Theoretical conceptualization of epistemological beliefs in history education

Epistemological beliefs refer to the assumptions, views, and ideas of individuals in relation to the nature of knowledge and the process of knowing (Hofer & Pintrich, 1997; VanSledright & Limón, 2006). Therefore, they can be assigned to the psychological construct of epistemic cognition, which refers to research focusing on "what individuals believe about knowledge and knowing and how they think and reason about the epistemological aspects of knowing" (Hofer, 2016, p. 19). While empirical research on epistemic cognition began in the 1970s or even earlier (Hofer, 2016), the systematic study of the construct in history education is much more recent (VanSledright & Maggioni, 2016). However, various theoretical models of epistemological beliefs have emerged from psychological and educational studies that have influenced the work of history educators. The two main approaches can be characterized as developmental or stage models and models with independent dimensions (Buehl & Alexander, 2002; Hofer & Pintrich, 1997).

Developmental or stage models assume that epistemological beliefs can be divided into distinct stages ranging from naïve to developed beliefs about knowledge and knowing. These models include the three-stage reflective judgment model (King & Kitchener, 2002) and the four-stage model of epistemological understanding (Kuhn & Weinstock, 2002), which formed the basis for the development of the BHQ (Maggioni et al., 2004, 2009). Certain educational psychologists conducting research in the domain of history have also used stage models as their theoretical basis (e.g. Kuhn, Weinstock, & Flaton, 1994; Ioannou & Iordanou, 2019). Priemer (2006) stated that these different stages can be divided in terms of "(1) absolutist views of knowledge (in the sense of true or false), (2) relativist views of knowledge (with a subjectivity), and (3) a moderately relativist view of knowledge (plurality and a certain subjectivity)" (p. 163, translated). Further to this, other researchers (e.g. Schommer, 1990) propose the multidimensionality of the construct of epistemological beliefs. In a meta-analysis, Hofer and Pintrich (1997) identified two core areas that underpin most of the models, including beliefs about the nature of knowledge with the dimensions of knowledge certainty and simplicity, as well as beliefs about the nature of knowing with the dimensions of source of knowledge and justification for knowing. Recent research in the domain of history has utilized a combination of both approaches for modeling epistemological beliefs by conceptualizing the beliefs as separate stages comprising several dimensions (Barzilai & Weinstock, 2015; Stoel et al., 2017).

The BHQ was developed by Maggioni (2010) and colleagues (Maggioni et al., 2004, 2009), and to date represents the measurement instrument for epistemological beliefs about history that has the best theoretical foundation, is the most well-known, and has been most frequently adapted for other languages. Based on psychological development models (King & Kitchener, 2002; Kuhn & Winstock, 2002) and the progression model for the concept "evidence" in history established by Lee and Shelmit (2003), three epistemic stances were developed. Maggioni (2010) defines these stances as "a system of beliefs about the nature and justification of knowledge that people entertain at a certain moment in time" (p. 6). The first epistemic stance is the *objectivist stance,* formerly known as the copier stance, which is characterized by the objectivist view that history is

an image of the past and can be understood directly from historical sources. According to this stance, a comprehensive understanding of the past can be formulated by collecting all the facts from primary sources. The second epistemic stance is characterized by a subjectivist idea of history. In the *subjectivist stance,* formerly known as the borrower stance, history simply arises from the interpretation and is dependent on the personal views, preferences, and causal inferences of the historian. In the third epistemic stance, the *criterialist stance*, history is viewed as a research process in which the past is reconstructed based on a discipline-specific method. In this stance, history is the result of a subject encountering an object, such as the selection, interpretation, and evaluation of sources by the historian. Therefore, underpinning the epistemic stances is the progression of ideas about history that, for each level, determines what is considered historical knowledge for an individual and how this knowledge is both generated and applied (VanSledright & Reddy 2014; VanSledright & Maggioni, 2016). It should be noted that these three epistemic stances can exist simultaneously and in different forms within individuals. Maggioni (2010) describes this situation as epistemic inconsistency. Indeed, this epistemic inconsistency is typical for novices in a domain, as they may still have unestablished, different, and sometimes divergent conceptualizations of historical knowledge. However, the findings of Stoel and colleagues (2017) indicate that even among history experts, conflicting epistemic views are held, especially regarding the subjectivity of history. This complicates the measurement of epistemological beliefs in the domain of history.

Empirical assessment of epistemological beliefs in history education

In recent years, there has been increasing interest in the study of epistemological beliefs and their connection to historical thinking skills. Indeed, some studies have even examined the effectiveness of interventions that aim to promote both epistemic understanding and historical competencies (e.g. Mierwald, Lehmann, & Brauch, 2018; Stoel et al., 2017). In this domain, researchers have focused on secondary school and college students, as well as both training and practicing history teachers (VanSledright & Maggioni, 2016). In order to assess epistemological beliefs in history education, qualitative approaches are occasionally used, as well as the more common quantitative instruments.

For example, researchers in educational psychology often use a fictional historical scenario called the Livia problem to classify students as “absolutists”, “multiplists”, or “evaluativists” based on their responses (e.g. Ioannou & Iordanou, 2019; Iordanou, Muis, & Kendeou, 2019; Kuhn, Iordanou, Pease, Wirkala, 2008; Kuhn, Weinstock, & Flaton, 1994). Furthermore, Barzilai and Weinstock (2015) developed a closed-format Epistemic Thinking Assessment (ETA) based on Kuhns et al.’s (2000) developmental model of epistemic understanding integrated with the multidimensionality of epistemic thinking. In addition, researchers across different domains, such as history and biology, have used questionnaires to measure epistemological beliefs that are based on Schommers (1990) considerations about the dimensionality of the construct (Buehl, Alexander, & Murpy, 2002). Finally, questionnaires have also been developed focusing on sub-aspects of epistemic cognition, such as beliefs about the value of integrating information across multiple documents in history (Wiley, Griffin, Steffens, & Britt, 2020).

In history education itself, there is a wide range of possibilities for examining epistemological beliefs. Indeed, history educators have used interviews, class observations, and, more frequently, questionnaires for this purpose (e.g. Maggioni, 2010; Maggioni et al., 2004, 2009; Miguel-Revilla et al., 2020; Nitsche, 2019; Nokes, 2014; O’Neill, Guloy, & Sensoy, 2014; Stoel et al., 2017; Voet & De Wever, 2016). It should be mentioned that qualitative and quantitative methods can also be combined depending on the research question. In particular, the BHQ (Maggioni, 2010) has become important for research on epistemological beliefs in history. The BHQ represents a refinement of an earlier questionnaire that was used to assess the beliefs of primary school teachers about the teaching and learning of history (Maggioni et al., 2004, 2009). The questionnaire consists of 22 items based on the three epistemic stances, including objectivist, subjectivist, and criterialist stance, and is used to measure epistemic beliefs in history among secondary school students and history teachers, both prospective and practicing (e.g. Maggioni,

2010; VanSledright & Reddy, 2014; VanSledright, Burkholdt, & Montgomery, 2018). Previously, history educators have translated and adapted the questionnaire for different languages (Mierwald, 2020; Mierwald et al., 2016, 2017; Kidwai, 2015; Miguel-Revilla et al., 2017, 2020; Namamba & Roa, 2016; Stoel et al., 2015, 2016) and even developed new questionnaires from it (Nitsche, 2019; Stoel et al., 2017).

However, several problems have been identified in the use of the BHQ in research studies. Regarding construct validity, a factor analysis using the original questionnaire with a sample of 66 history teachers suggested a division into two factors; the first factor included the subjectivist and the objectivist items, and the second factor contained the criterialist items (Maggioni, 2010). Further validation of the German version of the questionnaire using a sample of 124 secondary school students also suggested such a two-factor solution, including both naïve and developed epistemic beliefs. However, an adaptation of the questionnaire for a sample of 272 prospective history teachers in Germany resulted in a three-factor solution and, thus, a subdivision including all three epistemic stances (Mierwald et al., 2016, 2017). Importantly, for this questionnaire, the contrast between the objectivist items and the subjectivist items was increased by changes to the item wording. Based on these studies, the contrast between the subjectivist items and the objectivist items was strengthened further by slight adaptations to the translation and the addition of three new items (i.e., two objectivist items and one criterialist item) in order to optimize the instrument for use with secondary school students. Following this, exploratory factor analysis in an intervention study with a sample of 161 students revealed a three-factor structure of the instrument with sufficient to good reliability of the factors (Mierwald, 2020). In a recent study with 92 students, a three-factor solution and similar internal consistency of the factors was also reported (Behrendt & Brauch, 2020). However, in studies using Dutch and Spanish translations of the questionnaire, it was observed that factor loading of single items was problematic, and there was low reliability, especially for the objectivist scale (Miguel-Revilla et al., 2017, 2020; Stoel et al. 2015, 2016). This result similarly applies to the new German-language student questionnaire (Mierwald, 2020).

Previously, researchers assumed that these problems may have arisen from shifts in the meaning as a result of translating the BHQ. Indeed, difficulties may be caused by differences in understanding of the items in the US context, for which the questionnaire was originally designed, and in other countries (Stoel et al., 2016), as well as by slight adaptation of the items (Mierwald et al., 2017). Additionally, the mix of items could also prove problematic, as they are mainly related to the domain of history, but some also focus on the educational context (Miguel-Revilla et al., 2020). Furthermore, the theoretical background of the questionnaire has been criticized. For example, Stoel et al. (2017) outlined the ambiguity of the subjectivist items, "…acknowledging the subjective character of historical knowledge can both reflect naïve beliefs (i.e., history as opinion) but also more nuanced beliefs (i.e., historical knowledge as interpretative and constructed)" (p. 123).

Based on these issues, history educators have emphasized that further research is required in this area (Miguel-Revilla et al., 2020; Stoel et al., 2017). Cognitive interviews have been shown to be effective in verifying the comprehensibility of questionnaires and identifying difficulties with items, which is important to understand and address the root of these problems (Barzilai & Weinstock, 2015; Greene & Yu, 2014).

Method

Design

We conducted cognitive interviews based on the problems with the validity and reliability of the BHQ outlined above.[3] In previous studies, cognitive interviewing or cognitive pretesting was found to be helpful for assessing the cognitive validity of questionnaires used to assess epistemological beliefs (Barzilai & Weinstock, 2015; Greene & Yu, 2014; Greene, Torney-Purta, Azevdeo, & Robertson, 2010). Cognitive interviewing is a method for examining the

comprehensibility of the individual items in a questionnaire and identifying the potential problems with them. In this way, it can be determined whether the items are interpreted as intended by the questionnaire respondents, in line with their theoretical background (Beatty & Willis, 2007; Willis, 2005; Woolley, Browen, & Browen, 2006). Therefore, the goal of our approach was to investigate and improve the cognitive validity of the self-report items of the German version of the BHQ (Karabenick et al., 2007). Specifically, we used a combination of a think-aloud procedure as the students filled out the questionnaire and semi-structured interviews or follow-up probes to obtain in-depth information regarding the students' understanding (Beatty & Willis, 2007).

Participants

A sample of four students in eleventh grade (1 female, 3 male) from two different secondary schools participated in the cognitive interviews. These students were attending the highest-level educational track in the German secondary education system (i.e., Gymnasium) and were in the final years of their formal history education. The average age of the students was 17 years. This sample was chosen because the German-language version of the BHQ was designed for this target group. In addition, foreign-language versions of the BHQ are often conducted with students of this age (e.g. Stoel et al., 2016). We acknowledge that our sample size is very small. However, it should be noted that cognitive interviews are designed to elucidate the thoughts and provide a deeper understanding of a few participants rather than be representative (DeMaio et al., 1993). Therefore, smaller samples are often used, although these samples should cover a variety of characteristics relevant to the goal of the study, such as performance level and language skills (Beatty & Willis, 2007; Willis, 2005). In our study, the students who took part in the interviews were recruited by their history teachers at the request of the second author. The selection criteria were communicated to the teachers to appropriately evaluate the cognitive validity of the BHQ. Firstly, the study aimed to interview students with different levels of achievement in history, assessed by their history courses and grades. Furthermore, participating students had to be native German speakers so that any understanding difficulties with the BHQ could not be due to general language skills. Finally, the selected students had to be considered by their teachers as open-minded enough to think deeply about the questionnaire and verbally express their thoughts about it. Detailed information on the sample is presented in Table 1.

TABLE 1: Overview of the personal characteristics of the participants

Characteristics	**Ben**	**Paul**	**Sarah**	**Alex**
Age	17	17	17	17
Gender	male	male	female	male
Course type in the subject history	basic course	advanced course	basic course	advances course
Points in the subject history	13	4	14	7
Criterialist stance (*M*)	5.11	5.44	5.67	4.67
Subjectivist Stance (*M*)	3.56	3.11	3.56	2.78
Objectivist stance (*M*)	2.20	4.80	4.00	4.40

Notes: Advanced courses are taught five hours a week while the basic courses for only three hours a week. This means that the advanced course covers historical topics in greater depth than the basic course. Points are an equivalent of the grades in German secondary school system, ranging from 15 = best to 0 = worst. The mean values (*M*) of the three epistemic stances are on a scale ranging from 1 = strongly disagree to 6 = strongly agree.

Instrument

In this study, we used a German version of the original English BHQ by Maggioni (2010). We refined an earlier translated version of the instrument (Mierwald et al., 2017). In this process, small changes in the formulation of existing statements were made to enhance the applicability of the items to German secondary school students, and three new items were added. The entire questionnaire, including the changed and added items, can be found in Table 3 (see also the note below the Table). These changes were made to the German version of the BHQ because the first version indicated only two epistemic stances, and the two-factor idea was established based on this result. The first factor is related to differentiated beliefs, containing all criterialist items, and the second factor is related to naïve beliefs, containing the subjectivist and objectivist items. Although the internal consistency of the factors in scales that were created later was satisfactory, it was observed that the objectivist items did not contribute much to supporting the factor of naïve beliefs (Mierwald et al., 2017). Therefore, changes were enacted to improve the validity and reliability of the German language version of the questionnaire. Indeed, firstly, the changes enhance the contrast between the objectivist items and the subjectivist items on a linguistic level. Secondly, three new items (i.e. two objectivist items and one criterialist item) were added based on the theoretical background of the instrument to strengthen the reliability of the respective scales. Following modification, the final instrument contained a total of 25 items. When answering the questionnaire, students expressed their agreement or disagreement with pre-formulated statements on a 6-point Likert scale, ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). The statements were designed to exemplify the *objectivist*, *subjectivist*, and *criterialist stance* (see Table 3).

An exploratory factor analysis (EFA) was utilized to determine the factor structure of the instrument, as previous studies using foreign language adaptions of the questionnaire have identified different factor structures. Previously, an EFA with a sample of 161 students in a pre-post quasi-experimental intervention study (Mierwald, 2020, p. 324f.) revealed a three-factor structure of the instrument with sufficient to good reliability of the factors (objectivist scale [5 items]: $\alpha_{pre} = .53$, $\alpha_{post} = .59$; subjectivist scale [9 items]: $\alpha_{pre} = .78$, $\alpha_{post} = .83$; criterialist scale [9 items]: $\alpha_{pre} = .70$, $\alpha_{post} = .81$). Two items in the objectivist scale loaded against our theoretical assumption of three separate epistemic stances, so we decided to exclude these items. In addition, two criterialist items showed high cross-loadings on the objectivist factor. The objectivist scale did not reach an appropriate value for reliability recommended for standardized tests. However, it should be noted that studies measuring epistemological beliefs frequently work with lower reliabilities (i.e., less than .70) because of the complexity of the construct, so the reliability of the scale can be seen as still accepTable (e.g. Shraw 2013).

Procedure

The second author met with the history teachers, explained the study and was able to recruit suiTable participants for the cognitive interviews through them. The recruited students were asked about their willingness to participate in the interviews. In addition, written consent was obtained from parents for the students' participation in the study. Participants were interviewed during their regular class time in separate rooms. The second author, a trained student finalizing his Master of Education in History and German with experience in this field, interviewed each student. The individual interviews were conducted in the spring of 2018.

Each interview began with the students completing a questionnaire to obtain socio-demographic information, and an identification code was generated for each student to guarantee the anonymity of the collected data. In the first phase of the interview, the students were asked with written instructions (see Appendix A) to read each item of the BHQ aloud, mark their agreement or disagreement with the given statements, and verbalize the thoughts that came into their minds while they answered. The second phase was the semi-structured interview. For this purpose, an interview guide (see Appendix B) was used, which divided the interview into three parts. Firstly, in order to begin the interview in an easy way for the students, they were asked

about their general impressions of the whole questionnaire and the "thinking aloud" component. The main part of the semi-structured interview then included more specific questions about items that were either unexpected or not clearly assigned during factor analysis. Specifically, the questions focused on the two objectivist items that loaded on the factor for the subjectivist stance in the factor analysis (items 19 and 20), as well as the criterialist items that had high cross-loadings on the objectivist stance (items 18 and 21). Furthermore, the interviewer asked the students about important features that they noticed during the think-aloud component of the first phase. These questions related to the phrase "reconstruction of the past" that the students read in the questionnaire, as well as other such terms that may have influenced their assessment behavior. In order to end the interview process for the students, they were asked about their general understanding of history. With the consent of the students and their parents, the cognitive interviews were audio-recorded to enable accurate analysis of the verbal data afterward. The cognitive interviews lasted on average 43.21 minutes (*SD* = 10.08).

Analysis

The four recorded interviews were transcribed into text form with the software f4 (Dresing & Pehl, 2015). The data from the students thinking aloud in the first phase of the cognitive interviews were used as the basis for the thematic analysis. For the thematic analysis of the interviews (Kuckartz, 2014), we used a coding scheme comprising seven categories to identify the different types of problems associated with the items of the BHQ (research question 2). Our coding scheme was developed based on theoretical considerations from the literature, thus dividing the problems with the questionnaire into the four categories of comprehension, recall, judgment, and response selection (DeMaio & Landreth, 2004; Jabine, Straf, Tanur, & Tourangeau, 1984; Willis, 2005). In addition to these categories, we used an already existing coding scheme by Anschütz (2012). The first draft of the whole coding scheme was initially tested, discussed, and adapted using one of the interview protocols. The final coding scheme (see Table 2) was then independently applied by the first and second authors to the entire dataset.

TABLE 2: Coding scheme for questionnaire evaluation

Category	Subcategory	Description
No Problems	-	The students' statements indicate that the item was understood in the intended sense. Students can express the content of the item in their own words and give reasons for their opinion. They can add explanations, statements or examples from school or everyday life that reflect the meaning of the item presented. No significant problems can be identified in the statements of the person interviewed.
Understanding difficulties	**words and technical terms**	The students do not know a word *or* do not understand what is meant by this word in the context of the item. The students say that they find a word or phrase difficult to understand. They do not know a technical term in the item or have problems understanding the meaning of the technical term.
	misunder-standing	The students interpret an item statement differently from intended meaning of the item. They can add examples or explanations that show that their understanding does not correspond to the content or intention of the item. They need not be aware that they have misunderstood the item. This can also mean that they interpret the content or the epistemic beliefs (criterialist, subjectivist, or objectivist stance) of the item differently from the intended meaning (e.g., when an item aims at subjectivist stance and the students interpret the item statement as a criterialist stance).
	incomprehension	The students explicitly state that they do not understand all or part of the item's message. They do not know for sure what is meant by the item presented.
Reply difficulties	**reply format**	The students have difficulties in answering the item because of the answer's format (e.g., by double negation) *or* may say that they have misunderstood *or* do not fully understand the scale gradation (e.g. by labelling).
	uncertainty of response and explanation	The students are unsure of their answers to the item. They may think that the item cannot be answered in this way or may not know this. They may have an opinion on the statement presented, but do not really know how to answer it due to their level of knowledge, or they may state that they have no previous experience.
Assignment not possible	-	It is not clear from the students' statements whether the item was understood. They use the same words and phrases as it is present in the given items. The students say that they understood the meaning of the item, but do not explain how they understood the meaning of the item or why they arrived at the chosen answer.

For the content analysis of the students' answers from the semi-structured interviews, we used a technique of summarizing and inductive category formation from Mayring (2014). The authors independently prepared the summaries from the interviews. Additionally, the students' answers were categorized into positive and negative aspects depending on their impressions of the BHQ (research question 1) and their understanding of individual problematic items in yes/no form (research question 3). Following this, we compared the independent summaries and categorizations.

The inter-rater reliability (Cohen's κ) between the two raters for measurement of the entire dataset varied across all the categories from a minimum of $\kappa = .62$ to a maximum of $\kappa = 1.00$ (corresponding to "substantial" to "perfect" inter-rater reliability; Landis & Koch, 1977). Any disagreements concerning the coding between the raters were resolved through discussion, and a consensus was reached for all of them.

Results

General impressions of the BHQ

During the interviews following the thinking aloud section, the students discussed various positive and negative aspects of the questionnaire when asked to give their general impressions. Sarah and Paul reported that the questionnaire had a clear design and the items were mostly short and easy to understand. Furthermore, Paul emphasized that the answer options were easy to handle because they were based on a German grading scale. Alex even reported that he found using the questionnaire interesting in terms of thinking about what history is and how to study history appropriately, rather than just discussing content in history lessons.

However, Sarah, Ben and Alex expressed negative concerns about some items being very similar in their content. The students also presented criticisms relating to uncertainties in the understanding of certain terms in the questionnaire (i.e., Sarah: "history"; Paul: "historical facts"; Ben: "research methods"; Alex: "evidence"). Additionally, Paul stated that items 18 and 20 utilized very long sentences, meaning that he had to read the items several times to fully understand their meaning.

Understanding of the BHQ and problems with individual items

We used the coding scheme described in the method section for the analysis of the data from the thinking aloud section in the first phase of the cognitive interviews. Certain general observations were derived from the absolute frequency distribution of the categories per item (see Table 3). Firstly, in most cases (73) the four participating students had no problems with the items of the BHQ, meaning that the students understood them according to their intended meaning. Fourteen out of 25 items were found to be unproblematic based on the students' statements. In addition, it is of note that all of the items were understood by at least one participant, showing that there was always at least one student who understood each item. Secondly, the absolute frequencies demonstrated that there were some problems with individual items, which mainly arose due to difficulties in understanding (18 cases). These difficulties predominantly included misunderstandings and problems in relation to certain words and technical terms, but difficulties in responding to the items were less frequent. Indeed, there were only 5 cases where students were uncertain in their responses and explanations. However, in 4 cases, no assignment was possible due to the students' reports. Thirdly, in terms of the three scales (*criterialist*, *subjectivist*, and *objectivist*), we observed that the objectivist scale presented the main problem. For this scale, only 3 out of the 7 items could be classified as appropriate. For the criterialist scale and the subjectivist scale, 5 and 6 out of a total of 9 items were considered unproblematic, respectively. However, it is important to understand exactly which features were problematic regarding the remaining 11 items.

TABLE 3: Frequency and nature of problems experienced by students for individual items and subscales of the German Beliefs about History Questionnaire (BHQ_G)

Items	No problems	Understanding difficulties			Reply difficulties		Assignment not possible
		Words and technical terms	Misunder-standing	Incom-prehension	Reply format	Uncertainty of response and explanation	
Criterialist Scale							
1. It is fundamental that students are taught to support their reasoning with evidence.	4	0	0	0	0	0	0
3. A historical account is the product of a disciplined method of inquiry.	1	0	1	1	0	1	0
7. Students need to be taught to deal with conflicting evidence.	4	0	0	0	0	0	0
11. History is a critical inquiry about the past.	2	0	0	0	0	1	1
13. Comparing sources and understanding author perspective are essential components of the process of learning history.	4	0	0	0	0	0	0
15. Knowledge of historical method is fundamental for historians and students alike.	4	0	0	0	0	0	0
18. Reasonable accounts can be constructed even in the presence of conflicting evidence.	4	0	0	0	0	0	0
21. History is a reasonable reconstruction of past occurrences based on available evidence.	3	1	0	0	0	0	0
24. Historians reconstruct the past based on regulated research methods.*	1	0	1	0	0	2	0
Subjectivist Scale							
2. History is simply a matter of interpretation.	1	1	1	0	0	1	0
4. Students who read many history books learn that the past is what a historian makes it to be.	2	0	1	0	0	0	1
6. Good students know that history is basically a matter of opinion.	4	0	0	0	0	0	0
8. Historical claims cannot be *substantiated*, since they are simply a matter of interpretation.	1	2	1	0	0	0	0
10. Since there is no way to know what really happened in the past, *students can choose whatever story they believe.*	4	0	0	0	0	0	0
12. The past is what the historian makes it to be.	4	0	0	0	0	0	0
14. It is impossible to know anything *with certainty* about the past, since no one of us was there.	4	0	0	0	0	0	0
17. Students need to be aware that history is essentially a matter of interpretation.	4	0	0	0	0	0	0
22. There is no evidence in history.	4	0	0	0	0	0	0
Objectivist Scale							
5. Disagreement about the same event in the past is always due to lack of evidence.	4	0	0	0	0	0	0
9. Good general reading and comprehension skills are enough to learn history well.	2	0	2	0	0	0	0
16. The facts speak for themselves.	1	1	2	0	0	0	0
19. Even eyewitnesses do not always agree with each other, so there is no way *of knowing* what happened *in the past*.	4	0	0	0	0	0	0
20. Teachers should not question students' historical opinions, *but only check whether they know the historical facts.*	1	1	1	0	0	0	1
23. Differences in historical accounts result from absence or falsity of historical facts.*	4	0	0	0	0	0	0
25. History consists of the sum of collected historical facts.*	2	0	1	0	0	0	1
Total	73	6	11	1	0	5	4

Note: The formulations in italics indicate which changes were made in the German questionnaire compared to the original BHQ in order to strengthen the contrast between objectivist and subjectivist items. The marked items (*) were added to the corresponding scales.

Criterialist Scale: Regarding the items in the individual scales, items 3, 11 and 24 of the criterialist scale were observed to be problematic. Item 21 ("History is a reasonable reconstruction of past occurrences based on available evidence.") can be excluded here, as only Paul had slight difficulty explaining the meaning of "reconstruction (of past occurrences)". Although this term may be difficult for students to define, the other three students demonstrated that it could be understood through the context and wording of the item. However, item 3 ("A historical account is the product of a disciplined method of inquiry.") caused greater issues. For example, Sarah showed uncertainty about her response for item 3, as well as for items 11 and 24. In the case of item 3, her uncertainty resulted from the fact that she did not fully understand the term "historical accounts" nor know to what extent a historian should have done "subject-specific research" or "studied" to produce their texts. Ben reported that he did not understand the item because of the phrase "disciplined method of inquiry". However, he assumed that this was related to critical appraisal of sources. Alex made a statement suggesting that he misunderstood the item, as he confused historical accounts with primary sources and said that historians would write these sources.

Further to item 3, two students had difficulties with item 11 ("History is a critical inquiry about the past."). Indeed, Sarah was unsure of how to respond to the item. Although she agreed with the item and reported that she had been taught this in class, she experienced confusion regarding the term "history" and the phrase "a critical inquiry about the past". Paul reported that history is "always a critical inquiry about events". Moreover, he added, in general terms, that "historians or reporters should be critical of historical or political events". Therefore, his understanding of the items was unclear, making it difficult to assign his response to a coding category.

Finally, item 24 ("Historians reconstruct the past based on regulated research methods.") also caused problems for the students. Sarah reported that she could not assess the extent to which historians use "regulated research methods", and this view was shared by Ben. In relation to this, Sarah reported:

> I can't really say anything about this, because I'm only at school. I have little to do with historians, except with my history teacher who once studied history. I would now say that this is true. At least when you write a text in an exam, you make sure that a certain form is followed to meet criteria and so that everything seems more or less well-founded.

In this statement, Sarah compared the research methods of historians with the formal guidelines for writing texts in history exams. However, the research methods used to reconstruct the past are not clear from her statements. Alex misunderstood the item and thought that primary sources are written based on the research procedures of historians. In this way, he appeared unable to clearly distinguish between the concepts of "primary source" and "historical account", nor was he able to explain the meaning of regulated historical research methods.

Subjectivist Scale: Regarding the subjectivist scale, items 2, 4 and 8 can be considered problematic based on the results. For item 2 ("History is simply a matter of interpretation."), both the terms "history" and "interpretation" caused difficulties. Sarah experienced particular problems with the term "history" due to her school experiences. Indeed, during the thinking aloud section she reported:

> This is difficult to answer because I wonder how history is defined. In our class, there are discussions about history or working with primary sources. So, I think that this is true. But especially in school, we are taught views. A lot of teachers tell us: "So and so it was in the past". Consequently, there is not so much room for interpretation at school.

Alex was unsure how to answer the same item because his views were equivocal. Indeed, he stated that history is based on evidence, but, contrastingly, he acknowledged that there are always opportunities for interpretation. Based on Pauls' statements, it appears that he misunderstood both terms:

> I understand the item to mean that history can be understood in several ways. In the subject history, there are actually rather few different interpretations. Of course, there are different areas of history. There is military history, there is the history of any country, especially of any people. These could be different interpretations of history. But I think in itself that history is always something historical, a certain point in time, a certain person, so the interpretation is always the same.

As is clear from the statement, this student's explanations of the concepts of "history" and "interpretation" were vague. For example, Paul reported that there are different "areas" for which "interpretations" occur, but he did not explain what these interpretations are. In contrast, he approximately equated history with the past. In the end, it was clear that this student understood the item more as an objectivist rather than a subjectivist item. Overall, Paul appears to conceptualize history like a fixed fact, meaning that the interpretations of the past should always be the same.

Two students experienced problems with item 4 ("Students who read many history books learn that the past is what a historian makes it to be."). Sarah deviated from the core meaning of the item and began thinking about her history lessons, so her statements could not be assigned to any category of the coding scheme. In Ben's case, whether he understood the item as an item from the subjectivist scale or the criterialist scale was unclear. He reported that "Depending on what you read and in which country you are hearing the history, historiography is already a little different. For example, when you look at the colonies, they have a different view of the past of the colonies compared to Great Britain." One can conclude from this statement that Ben was aware that historiography exists and that it arises from different interpretations that depend on individual perspectives.

Regarding item 8 ("Historical claims cannot be substantiated, since they are simply a matter of interpretation"), three students showed difficulties. Alex's interpretation of the item was ambiguous regarding whether historical claims originate directly from primary sources or represent contemporary claims about the past. Although the latter is in line with the intended meaning of the item, Alex did not decide on a meaning in his further interpretation of the item. Sarah expressed criticisms regarding the word "historical" in the context of the item. She commented:

> Well, claims are always something that is very subjective. Therefore, they are of course a matter of interpretation. But theoretically, "historical" always means that something should have been researched. But if there is a lack of information or there is little source evidence, then it is definitely a matter of interpretation.

Sarah justified her agreement with the item by referring to the general subjectivity of all claims. Concurrently, she referred to the fact that history is based on research. Ultimately, she believed that gaps in understanding can be filled using interpretations. Therefore, Sarah was also unsure what "historical claims" meant in the context of the item. Paul also misunderstood the same, assuming that historical claims arise from the individuals involved in the history themselves.

Objectivist scale: For items 9, 16, 20 and 25 in the objectivist scale, it was found that the students had some problems giving answers. Sarah and Paul misunderstood item 9 ("Good general reading and comprehension skills are enough to learn history well."), as they did not understand the item as intended in terms of directly extracting knowledge from primary sources to learn about the past. Instead, it was clear from their answers that they understood "general reading and comprehension skills" as referring to the critical-analytical competencies used when working with historical media. This suggests an interpretation of item 9 that is more typical of the criterialist stance than the objectivist stance. In this regard, Sarah reported that "In history lessons, we have to read a lot and understand these texts. I think that if you can't get along with these texts, you can't acquire knowledge. On the other hand, we also have caricatures, for which you need good comprehension skills." Similarly to Sarah, Paul did not consider the word "enough" but emphasized that good reading and comprehension skills are necessary to reconstruct historical

processes, understand the political and economic aims of wars, and to be able to understand historical contexts.

Item 16 ("The facts speak for themselves.") was also found to present problems. Firstly, Sarah had issues with the term "historical facts". Indeed, more precisely, she was not sure to what extent historical facts relate to source evidence. Even for Paul and Alex, the term "historical facts" remained unclear, thus leading to vague interpretations. For example, Paul described his understanding of the item as follows:

> My understanding is that historical facts are largely the core statement of historical events. Even if historians or persons talk about historical facts and give their own interpretation of them, the historical proof is actually the most fundamental. All opinions or reports are based on it. That is why I agree with it if the own interpretation is present.

From this statement, it is not clear to what extent the item is interpreted as an objectivist item or expresses beliefs more typical of the subjectivist or criterialist stances. Overall, Paul modified the content of the item in his understanding and agreed with the item for reasons that were not apparent from the item.

The interpretation of item 20 ("Teachers should not question students' historical opinions, but only check whether they know the historical facts.") was also challenging for the students. Ben related the content of the item to his own history lessons and discussed his ideas about what methods would be better for studying history. Therefore, it was not possible to assignment his response to a category of the coding scheme. Sarah also related the item to her own history lessons and said that agreement with this statement depends on the teacher. As with item 16, she emphasized that she would not know what "historical facts really are now". Finally, Paul misunderstood the item because he read over the phrase "should not", stopped following the first part of the sentence, and then related the item to his own ideas about what constitutes good history teaching. In general, the item appears too long in its wording and, due to its relation to the school context, it may encourage the students to express personal opinions instead of considering the item's content.

In contrast to item 20, item 25 ("History consists of the sum of collected historical facts.") appeared less problematic. Paul and Ben had difficulties with this item, as they struggled with the term "historical facts" again. When thinking aloud, Paul repeated and reformulated the content of the item, thus making it difficult to assign his response to a category of the coding scheme. Ben agreed with the item but interpreted it more as a subjectivist or criterialist item rather than objectivist. He explained that there are no facts in history, only opinions that must be justified and interpreted carefully. Therefore, for this student, history is derived from "what can be deduced from the collection of sources".

Perception of problematic items from the exploratory factor analysis (EFA)

In an exploratory factor analysis from an earlier study, which used data from the BHQ and a sample of 161 students, four items showed unusual loadings (Mierwald, 2020). Specifically, items 19 and 20 of the objectivist scale loaded unexpectedly on the factor uniting all subjectivist items. Additionally, items 18 and 21 of the criterialist scale showed high cross-loadings on the factor that includes all objectivist items. Based on these findings, our students were questioned in-depth about these items during the semi-structured interviews.

During the thinking aloud section, item 19 ("Even eyewitnesses do not always agree with each other, so there is no way *of knowing* what happened *in the past*.") was correctly understood and explained by all four students. Furthermore, this did not change when they were asked again about their understanding in the interview section. Theoretically, the students may have interpreted the item to mean that, if you cannot know what happened in the past, only your own views of the past matter. Indeed, this type of interpretation could explain why the item loaded on the subjectivist factor. In our interviews, the students did not express such strong subjectivist

views of the item. However, a few statements from Paul and Alex may indicate that they believe there are ways for historians to find out what happened in the past, suggesting some subjectivism in their understanding of the item.

As mentioned previously, Sarah, Paul and Ben had significant difficulties understanding item 20 ("Teachers should not question students' historical opinions, *but only check whether they know the historical facts."*). Firstly, the reference to the school context appeared problematic for this item, as the students tended to express their experiences with history lessons and teachers rather than addressing the actual message of the item. For example, Paul said: "I ticked 'I strongly disagree' because I don't like to learn things by heart and of course I want the lessons to be a bit interesting." Secondly, the students were unsure of the meaning of the term "historical facts". Their statements regarding this term ranged from "what is already certain from the past" (Sarah), "probable truth" (Ben) to "dates", "treaties like the Treaty of Versailles", "events like the discovery of America" and "quotes by politicians and kings" (Paul). These statements suggest that the students had diverse and varied understanding of the term "historical facts". It is also possible that the contrast between "historical opinions" and "historical facts" in the item led to it being associated with subjectivist views, thus potentially explaining the incorrect loading of the item on the subjectivist factor.

Item 18 ("Reasonable accounts can be constructed even in the presence of conflicting evidence.") showed a high cross-loading on the factor which all objectivist items loaded onto in the factor analysis. In this item, the focus on "evidence" may attract the students' attention rather than the term "reasonable accounts". This, in turn, may have led the students to associate the item more with objectivist views. Although this item did not cause many difficulties during the thinking aloud section, it should be noted that Sarah and Paul considered the term "evidence" for a long time during the semi-structured interviews. When explicitly asked about their understanding of the term, they appeared to find it difficult to explain. For example, Sarah reported:

> Evidence is something that is already definite in my eyes. [...] Instead of evidence, one can also say proof synonymously. Thus, already something, for which there are several facts, which speak for itself. Now, I actually find it difficult to say what evidence is. Particularly since this is to be conflicting.

Item 21 ("History is a reasonable reconstruction of past occurrences based on available evidence.") could be explained correctly by almost all students. Only Paul, as noted above, showed ambiguities in his understanding. When asked what a "reconstruction" is during the interview section, his answer fluctuated between "reproduction of any event based on available source evidence" and "self-interpretation of sources". Overall, a high cross-loading of the item onto the objectivist factor could be due to the item's strong emphasis on the importance of evidence, which could have led to an implicit objectivistic interpretation of the item.

Interestingly, all the items showing unusual loadings (items 18, 19, 20 and 21) were presented next to each other and at the end of the questionnaire. Therefore, it is possible that the order of the items influenced their interpretations. In addition, the large number of items in the questionnaire may have caused a reduction in concentration and, thus, inaccuracies in the students' interpretations.

Discussion and Conclusions

Summary and interpretation of the findings

In summary, the translated and adapted BHQ was designed effectively for our sample of German-speaking students, the Likert response scale was easy to use, and most items were clearly formulated and understandable. Furthermore, the problematic items were not equally difficult to understand across all the students in our cognitive interviews. Regarding the cognitive validity of the scales, coherent connections were identified between the students' interpretation and the intended meaning of most BHQ items.

However, some students had problems with individual items based on the results of the cognitive interviews. These problems were mainly due to difficulties in understanding some items and, to a lesser extent, difficulties in responding to them. It was found that items from the objectivist scale were more problematic than items from the criterialist and subjectivist scale. This may explain why the objectivist scale, or the factor which most objectivist items loaded onto, had relatively low reliability in the factor analysis. Furthermore, the sequential arrangement of the problematic items (18, 19, 20 and 21) at the end of the questionnaire could have influenced the students' responding behavior. Based on the results of this study, four overarching general and potential problem areas were identified, and corresponding items were assigned to them. Specifically, these problem areas represent potential issues that some students may experience difficulties with when interpreting the items. Some of the interviewed students struggled with these issues when they thought aloud about the items of the BHQ or were explicitly asked for their understanding. Additionally, in some cases, one item caused several potential problems.

The first problem area relates to the *complexity of terms*. This refers to the fact that some items (2, 3, 8, 11, 16, 18, 20, 21, 24 and 25) of the BHQ utilized terms that were abstract and ambiguous to the interviewed students. Such semantic problems often occur with questionnaire items (e.g. Willis, 2005), and they are also common for questionnaires that assess epistemological beliefs, as their items often relate to complex concepts such as "truth" or "expert" (e.g. Hyytinen, Postareff, & Lindblom-Ylänne, 2020). In the case of the BHQ, students are given items that relate to complex meta-concepts, such as "disciplined method of inquiry", "evidence", "interpretation", and "historical facts". As in other studies (e.g. Greene et. al., 2010; Muis, Duffy, Trevors, Ranellucci, & Foy, 2014), the students often refer to previous experiences in class or everyday life in order to understand these terms. Furthermore, students' prior knowledge of these meta-concepts in history is often implicit and incomplete, thus making them difficult to understand (van Drie & van Boxtel, 2008). Additionally, it is difficult for students to understand because the same terms occur in various items with different connotations (e.g. "History is a critical inquiry about the past." vs. "History is simply a matter of interpretation.").

Besides the complexity of terms, the *epistemic ambiguity* of items presented a challenge to the interviewed students. Epistemic ambiguity means that the wording of individual items (2, 4, 9, 16, 18, 19, 20, and 21) causes the item to be unintentionally associated with one or more of the other epistemic stances. Indeed, items can be formulated in such a way that the students do not focus on their intended core statement, as certain words or phrases may allow a different interpretation. For example, as in item 21 ("History is a reasonable reconstruction of the past occurrences based on available evidence."), the emphasis on the term "evidence" may cause the students to interpret the item in a more objectivistic way. This could also be related to a reinterpretation of items, which would be most consistently related to the students' prior knowledge or current epistemological beliefs (Barzilai & Weinstock, 2015). For some items, the students' difficulty in differentiating between the subjective and objective components of historical knowledge and knowing was also evident (Stoel et al., 2017).

Further to this, another problem relates to the *length and comprehensibility* of items (18 and 20), which caused problems for some of the interviewed students. For example, with item 20 ("Teachers should not question students' historical opinions, but only check whether they know the historical facts."), students may misunderstand the content because they may not understand the statement when they first read it, overlook certain words, or focus on only one part of the sentence.

Finally, some of the items (4, 9, 20) distracted the students from the item's message and made them think about school experiences or their opinions about history teaching. As mentioned in the literature, the issue of *irritating references to the school context* may lead to inconsistent response behavior (Miguel-Revilla et al., 2020). In our study, these sporadic references to the school context were found to be problematic for some respondents. This could also be related to the fact that these items address less epistemological beliefs and more beliefs about teaching and learning history (Nitsche, 2019).

Limitations of the study

The findings of this study were obtained using a small sample of four students from two secondary schools in Germany. Furthermore, we use a German-language version of the BHQ as the basis for our cognitive interviews. These factors limit the generalizability of the findings to other German students and students from other countries. For example, students from other educational tracks in Germany may experience problems completing the questionnaire that are different from the difficulties of the students interviewed in this study. Moreover, students from other countries may experience other types of difficulties with translated versions of the BHQ due to their socio-cultural and educational contexts. For example, it is conceivable that some of the complex historical meta-concepts, such as "historical facts", "evidence", or "account", may be more comprehensible to students from North America due to differences in history instruction (Seixas, 2016). However, the goal of this study was not specifically to generalize to other groups but to identify the potential strengths and weaknesses of the BHQ. For this purpose, we used a sample of secondary school students, as the German-language BHQ questionnaire was designed for this group (Mierwald, 2020), and this group is commonly used in studies assessing foreign-language adaptations of the BHQ (e.g. Stoel et al., 2016). An additional limitation is that this study did not consider the effects of instructional practices or teachers on the epistemological views of the students in our study sample (VanSledright & Maggioni, 2016). Overall, future research in this area may benefit from larger, more diverse samples that have been recruited systematically (Greene & Yu, 2014).

It should also be mentioned that the method employed in this study of asking the students to think aloud may have influenced the results (Funke & Spering, 2006). Two students reported no difficulties with this method. However, the other two students said that they may have answered the items a little differently in writing without giving verbal comments. Indeed, they reported that thinking aloud encouraged them to examine the questionnaire more intensively. Furthermore, the thinking aloud method depends on the participants' abilities to verbalize their cognitions. Regarding some items, the students' comments were very brief and, in some cases, it appeared that comments given on a previous item influenced their response to a later item. All of these factors complicated the researchers' interpretations of the student statements. Although we aimed to achieve objectivity through the coding scheme and the use of two raters, the experience of the raters with epistemic cognition may have influenced their interpretation of the thinking aloud protocols (Greene & Yu, 2014). Despite these issues, a range of potential problem areas was identified that may be helpful for improving the BHQ in the future.

Implications for research and practice

The problems with the items of the BHQ identified in this article are noteworthy in general. Firstly, they represent potential problems that may arise in the development of other self-report questionnaires. Secondly, questionnaires that assess epistemological beliefs may be particularly affected by these problems because they measure very complex and sometimes vague constructs (Mason, 2016), which inherently makes it difficult to formulate items that are clear and easily comprehensible.

In terms of future optimization of the BHQ items, certain important solutions have been identified for the potential problem areas. Firstly, researchers should reduce the *complexity of terms*, and we recommend rewording the BHQ to improve its clarity and interpretation (Barzilai & Weinstock, 2015). For example, terms could be reworded, such as texts by historians vs. historical accounts and claim about the past vs. historical claim. If possible, the definitions of relevant terms should be included in the instructions (Muis et al., 2014). The *epistemic ambiguity* identified in the BHQ could be addressed by increasing the emphasis on the intended epistemic stance for each item (Barzilai & Weinstock, 2015). This could be done by omitting terms that are strongly identified with other epistemic stances (e.g. avoid using the term "historical opinions" in objectivist items) or by utilizing qualifiers (e.g. emphasize subjectivist items by using phrases such as "not with certainty" or "you can believe what you want"). The issue with the *length and*

comprehensibility of some items could be solved by shortening or reformulating the items. Furthermore, *irritating references to the school context* should be removed from the affected items. In addition, due to the low reliability of the objectivist scale, we suggest the development and addition of further items in this stance. Indeed, the statements of the interviewed students could help with this task. Further to this, a new arrangement of the items at the end of the questionnaire could also be beneficial for the instrument and such changes should, in turn, be studied by cognitive interviews. These changes would also require further factor analyses with larger sample sizes to validate the newer versions of the BHQ.

Based on the results of the cognitive interviews, we suggest that historical meta-concepts should be thoroughly discussed with the students (e.g. Limón, 2002). As Stoel and colleagues (2017, p. 131) suggest, a stronger emphasis on the "epistemological dimension of historical inquiry" and the development of "pedagogies that incorporate classroom reflection on the nature and sources of historical knowledge" is also advisable in the domain of German history teaching.

In the future, the development of questionnaires to assess domain-specific epistemological beliefs may pose significant challenges for history education, both theoretically and methodologically. Importantly, we believe that the present study contributes to the improvement of the BHQ and other questionnaires based on this measurement tool.

References

Anschütz, A. (2012). *Epistemologische Überzeugungen von Schülerinnen und Schüler: Entwicklung eines Erfassungsinstrumentes für die Jahrgänge 3 bis 6*. Berlin: Logos Verlag.

Barzilai, S., & Weinstock, M. (2015). Measuring epistemic thinking within and across topis: A scenario-based approach. *Contemporary Educational Psychology, 42,* 141–158. http://dx.doi.org/10.1016/j.cedpsych.2015.06.006

Beatty, Paul C., & Willis, Gordon, B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 1–25. https://doi.org/10.1093/poq/nfm006

Behrendt, L., & Brauch, N. (2020). Auf die Frage kommt es an! Zur Veränderung von epistemologischen Überzeugungen durch Förderung historischer Fragestellungskompetenz im Schülerlabor. In K. Sommer, J. Wirth, & M. Vanderbeke (Eds.), *Handbuch Forschen im Schülerlabor – Theoretische Grundlagen, empirische Forschungsmethoden und aktuelle Anwendungsgebiete* (pp. 91–98). Münster: Waxmann-Verlag.

Buehl, M. M., Alexander, P. A., & Murphy, P.K. (2002). Beliefs about schooled knowledge: Domain specific or domain general? *Contemporary Educational Psychology, 27*(3), 415–449. https://doi.org/10.1006/ceps.2001.1103

DeMaio, Th. J., Mathiowetz N., Rothgeb, J., Beach, M. E., & Durant, S. (1993). *Protocol for pretesting demographic surveys at the Census Bureau*. Washington, DC: U.S. Bureau of the Census.

DeMaio, T., & Landreth, A. (2004). Do different cognitive interview methods produce different results. In S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Questionnaire development and testing methods* (pp.89–108). Hoboken, NJ: Wiley.

Dresing, T., & Pehl, T. (2015). *Praxisbuch Interview, Transkription & Analyse. Anleitung und Regelsysteme für qualitativ Forschende*. Marburg: audiotrankription.de.

Funke, J., & Spering, M. (2006). Methoden der Denk- und Problemlöseforschung. In J. Funke (Ed.), *Denken und Problemlösen. Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Band 8: Denken und Problemlösen* (pp. 675–744). Göttingen: Hogrefe.

Greene, J. A., & Yu, S. B. (2014). Modeling and measuring epistemic cognition: A qualitative re-investigation. *Contemporary Educational Psychology, 39*(1), 12–28. https://doi.org/10.1016/j.cedpsych.2013.10.002

Greene, J. A., Torney-Purta, J., Azevedo, R., & Robertson, J. (2010). Using cognitive interviewing to explore elementary and secondary school students' epistemic and ontological cognition. In L. D. Bendixen & F. C. Feucht (Eds.), *Personal epistemology in the classroom: Theory, research, and implications for practice* (pp. 368–406). Cambridge University Press. https://doi.org/10.1017/CBO9780511691904.012

Havekes, H., Van Boxtel, C., Coppen, P.-A., & Luttenberg, J. (2012). Knowing and doing history: A conceptual framework and pedagogy for teaching historical contextualisation. *International Journal of Historical Learning, Teaching and Research, 11*(1), 72–93.

Hofer, B. K. (2016). Epistemic cognition as a psychological construct: Advancements and challenges. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 19–38). New York, London: Routledge.

Hofer, B. K., & Bendixen, L. D. (2012). Personal epistemology: Theory, research, and future directions. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Bd. 1: Theories, constructs, and critical issues* (pp. 227–256). Washington: American Psychological Association. https://doi.org/10.1037/13273-009

Hyytinen, H., Postareff, L., & Lindblom-Ylänne, S. (2020). Challenges in exploring individual's conceptions of knowledge and knowing: Examples of research on university students. In E. K. Kallio (Ed.), *Development of adult thinking: Interdisciplinary perspectives on cognitive development and adult learning* (pp. 177–190). Routledge. https://doi.org/10.4324/9781315187464-11

Ioannou, K., & Iordanou K. (2020). Elementary school students' epistemic perspective and learning strategies in history. *Learning: Research and Practice, 6*(2), 150–166. https://doi.org/10.1080/23735082.2019.1591492

Iordanou, K., Muis, K. R., & Kendeou, P. (2019). Epistemic perspective and online epistemic processing of evidence: Developmental and domain differences. *The Journal of Experimental Education, 87*(4), S. 531–551. https://doi.org/10.1080/00220973.2018.1482857

Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines. Report of the advanced research seminar on cognitive aspects of survey methodology.* Washington, D.C. NAP.

Karabenick, S. A., Wooley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., De Groot, E., Gilbert, M. C., Musu, L., Kempler, T. M., & Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42*(3), 139–151. https://doi.org/10.1080/00461520701416231

Kidwai, L. K. (2015). *Investigating a relationship of a perception of agency in task-based discourse and change in concepts: A practitioner research in education*. Dissertation. Bradford: University of Bradford.

King, P. M., & Kitchener, K. S. (2002). The reflective judgment model: Twenty years of research on epistemic cognition. In B. K. Hofer, & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 37-61). Mahway, NJ: Lawrence Erlbaum, Publisher.

Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice and using software*. London: SAGE

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

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development, 23*(4), 435–451. https://doi.org/10.1016/j.cogdev.2008.09.006

Kuhn, D., Winestock, M., & Flaton, R. (1994). Historical reasoning as theory-evidence coordination. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 377–402). Hillsdale, N.J: L. Erlbaum Associates.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159. https://doi.org/10.2307/2529310

Lee, P., & Shemilt, D. (2003). A scaffold, not a cage: Progression and progression models in history. *Teaching History, 13*(113), 13–24.

Limón, M. (2002). Conceptual change in history. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (S. 259–289). Dordrecht: Kluwer Academic Publishers.

Maggioni, L. (2010). *Studying epistemic cognition in the history classroom: Cases of teaching and learning to think historically.* [PhD Dissertation]. University of Maryland. http://hdl.handle.net/1903/10797

Maggioni, L, Alexander, P., & VanSledright, B. (2004). At a crossroads? the development of epistemological beliefs and historical thinking. *European Journal of School Psychology, 2*(1-2), 169–197.

Maggioni, L., VanSledright, B., & Alexander, P. A. (2009). Walking on the borders: A measure of epistemic cognition in history. *The Journal of Experimental Education, 77*(3), 187–214. https://doi.org/10.3200/JEXE.77.3.187- 214

Mason, L. (2016). Psychological perspectives on measuring epistemic cognition. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 375–392). New York, London: Routledge.

Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt. Retrieved from: https://www.ssoar.info/ssoar/handle/document/39517

Mierwald, M. (2020). *Historisches Argumentieren und epistemologische Überzeugungen. Eine Interventionsstudie zur Wirkung von Lernmaterialien im Schülerlabor*. Wiesbaden: Springer VS.

Mierwald, M., Lehmann, T., & Brauch, N. (2018). Zur Veränderung epistemologischer Überzeugungen im Schülerlabor: Authentizität von Lernmaterialien als Chance der Entwicklung einer wissenschaftlich angemessenen Überzeugungshaltung im Fach Geschichte? *Unterrichtswissenschaft, 46*(3), 279–297. https://link.springer.com/article/10.1007/s42010-018-0019-7

Mierwald, M., Seiffert, J., Lehmann, T. & Brauch, N. (2016). *"Do they affect it all?" - Measuring epistemological beliefs in history education and their relationship to students' argumentation skills and future history teachers' planning skills*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), 11.04.2016, Washington, DC, USA.

Mierwald, M., Seiffert, J., Lehmann, T. & Brauch, N. (2017). Fragebögen auf dem Prüfstand. Ein Beitrag zur Erforschung und Weiterentwicklung eines bestehenden Instruments zur Erfassung epistemologischer Überzeugungen in der Domäne Geschichte. In M. Waldis & B. Ziegler (Eds.), *Geschichtsdidaktik heute: Band 8. Forschungswerkstatt Geschichtsdidaktik 15: Beiträge zur Tagung "Geschichtsdidaktik empirisch 15"* (pp. 177–190). Bern: hep.

Miguel-Revilla, D., Carril-Merino, M., & Sánchez-Agustí, M. (2017). Accediendo al pasado: creencias epistémicas acerca de la Historia en futuros profesores de Ciencias Sociales. *REIDICS, 1,* 86–101.

Miguel-Revilla, D., Carril-Merino, T., & Sánchez-Agustí, M. (2020). An examination of epistemic beliefs about history in initial teacher training: a comparative analysis between primary and secondary education prospective teachers. *The Journal of Experimental Education, 89*(1), 54-73. https://doi.org/10.1080/00220973.2020.1718059

Muis, K. R., Duffy, M. C., Trevors, G., Ranellucci, J., & Foy, M. (2014). What were they thinking? Using cognitive interviewing to examine the validity of self-reported epistemic beliefs. *International Education Research, 2*(1), 17–32.

Namamba, A., & Rao, C. (2016). Teachers' beliefs about history and instructional approaches: A survey of secondary school teachers in Kigoma Municipality, Tanzania. *International Journal of Education and Research, 4*(8), 203–220.

Nitsche, M. (2019). *Beliefs von Geschichtslehrpersonen: Eine Triangulationsstudie.* Bern: hep-Verlag.

Nokes, J. D. (2014). Elementary students' roles and epistemic stances during document-based history lessons. *Theory & Research in Social Education, 42*(3), 375–413. https://doi.org/10.1080/00933104.2014.937546

O'Neill, D. K., Guloy, S., & Sensoy, Ö. (2014). Strengthening methods for assessing students' metahistorical conceptions: Initial development of the historical account differences survey. *The Social Studies, 105*(1), 1–14. https://doi.org/10.1080/00377996.2013.763524

Priemer, B. (2006). Deutschsprachige Verfahren der Erfassung von epistemologischen Überzeugungen. *Zeitschrift für Didaktik der Naturwissenschaften, 12,* 159–175.

Sandmann, A. (2011). Lautes Denken: Die Analyse von Denk-, Lern- und Problemlösungsprozessen. In D. Krüger, I. Parchmann, & H. Schecker (Eds.), *Methoden der naturwissenschaftsdidaktischen Forschung* (pp. 179-188). Berlin/Heidelberg: Springer.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82*(3), 498–504. https://doi.org/10.1037/0022-0663.82.3.498

Schraw, G. (2013). Conceptual integration and measurement of epistemological and ontological beliefs in educational research. *ISRN Education, 1,* 1–19. https://doi.org/10.1155/2013/327680

Seixas, P. (2016). Translation and its discontents. Key concepts in English and German history education. *Journal of Curriculum Studies, 48*(4), 427-439. https://doi.org/10.1080/00220272.2015.1101618

Stoel, G. L., van Drie, J. P., & van Boxtel, C.A.M. (2015). Teaching towards historical expertise. Developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies, 47*(1), 49–76. https://doi.org/10.1080/00220272.2014.968212

Stoel, G. L., van Drie, J. P., & van Boxtel, C. A. M. (2017). The effects of explicit teaching of strategies, second-order concepts, and epistemological underpinnings on students' ability to reason causally in history. *Journal of Educational Psychology, 109*(3), 321–337. https://doi.org/10.1037/edu0000143

Stoel, G., Logtenberg, A., Wansink, B., Huijgen, T., van Boxtel, C., & van Drie, J. (2017). Measuring epistemological beliefs in history education: An exploration of naïve and nuanced beliefs. *International Journal of Educational Research, 83,* 120–134. https://doi.org/10.1016/j.ijer.2017.03.003

Van Drie, J., & van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review, 20*(2), 87–110. https://doi.org/10.1007/s10648-007-9056-1

VanSledright, B. (2002). *In search of America's past: Learning to read history in elementary school.* New York: Teachers College Columbia Univ.

VanSledright, B. (2014). *Assessing historical thinking and understanding: Innovative designs for new standards.* New York, NY: Routledge.

VanSledright, B. & Limón, M. (2006). Learning and teaching social studies: A review of cognitive research in history and geography. In Alexander, P.A. and Winne, P.H. (Eds.), *Handbook of educational psychology* (pp. 545–570). Hillsdale NJ: Erlbaum.

VanSledright, B., & Maggioni, L. (2016). Epistemic cognition in history. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 128–146). NY: Routledge.

VanSledright, B., & Reddy, K. (2014). Changing epistemic beliefs? An exploratory study of cognition among prospective history teacher. *Revista Tempo E Argumento, 6*(11), 28–68. https://doi.org/10.5965/2175180306112014028

VanSledright, B., Burkholdt, S., & Montgomery, S. (2018). Exploring change in epistemic beliefs among history teachers. *Social Studies Education Review, 7*(1), 1–24.

Voet, M. & De Wever, B. (2016). History teachers' conceptions of inquiry-based learning, beliefs about the nature of history, and their relation to the classroom context. *Teaching and Teacher Education, 55,* 57-67. https://doi.org/10.1016/j.tate.2015.12.008

Willis, G. B. (2005). *Cognitive interviewing. A tool for improving questionnaire design*. Thousand Oaks, California: Sage.

Wiley, J., Griffin, T. D., Steffens, B., & Britt, A. M. (2020). Epistemic beliefs about the value of integrating information across multiple documents in history. *Learning and Instruction, 65*. https://doi.org/10.1016/j.learninstruc.2019.101266

Wineburg, S. S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past. Critical perspectives on the past*. Philadelphia: Temple University Press.

Woolley, M. E., Browen, G. L., & Browen, N. K. (2006). The development and evaluation of procedures to assess child self-report item validity. *Educational and Psychological Measurement, 66*(4), 687-700. https://doi.org/10.1177/0013164405282467

Appendix A

Before the students started to think aloud, they were given the following instructions, which were adapted from Sandmann (2011, p. 184):

> PLEASE THINK LOUD!
> Please tell us everything you think when you answer the questionnaire. Continue until you have assessed the last statement in the questionnaire.
> To do this, please READ ALOUD.
> Ideally, you should speak out your thoughts about the statements you have just assessed without interruption, i.e. as much as possible without pauses.
> However, you should not order your statements in any particular order before speaking, nor should you express your thoughts in a particularly comprehensible way, nor should you explain your thoughts to the interviewer.
> Imagine that you are ALONE IN THE ROOM and speaking only to themselves. It is important that you are ALWAYS SPEAKING.

Appendix B

1. **Warm-Up**
   - Now you have filled out the questionnaire and thought about it aloud. Thank you very much for your effort! Would you like to start with a general comment on the questionnaire?
   - Did you notice anything that you have not yet "thought aloud"?
   - What is the first thing you think of when you think back to filling out the questionnaire?

2. **General questions about the items**
   - In retrospect, were there any item in the questionnaire that were difficult for you to understand? (Can you find them again? Why were the item difficult to understand?)

3. **More detailed questions about problematic items**
   - Let us take a closer look at item 19 again. Please describe in your words what should be assessed. How do you understand this item?
   - Let's go straight on to item 20: Why did you check the box like you did? What are "historical facts" for you?
   - Please explain why you agreed/disagreed with item 18. What did you think source evidence is? Can sources be contradictory?
   - To the last item: How do you understand item 21? There the "reconstruction of past events" appears. What do you understand by this?

4. **Closing**
   - We come to the end of the interview: Imagine your little brother coming into the fifth grade and getting history as a new subject. How would you explain what the subject history is about?

About the Authors

Marcel Mierwald, PhD is a postdoctoral researcher and history teacher educator at the Department of Didactics of History at the Ruhr University Bochum (Germany). His research interests include learning with (digital) media, measuring historical competencies and beliefs, language education in history, and anti-Semitism research. He has published essays and journal articles on, among other things, the assessment and fostering of historical argumentation skills and epistemological beliefs in the domain of history. Currently he is working in the project EMPATHIA[3] ("Empowering Police Officers and Teachers in Arguing Against Antisemitism"), which is funded by the German Federal Ministry of Education and Research.

Email: marcel.mierwald@rub.de ORCID: 0000-0003-3530-4728

Maximilian Junius, M.Ed. is a secondary school history teacher based in Germany. He holds a Master of Education from the Ruhr University Bochum and currently teaches German and History at a secondary school in North Rhine-Westphalia. He wrote his master's thesis in history education on the qualitative assessment of school students' epistemological beliefs.

Email: maximilian.junius@rub.de

Endnotes

[1] The names of students who participated in our study were anonymized.

[2] The student quotes in this article were translated from German into English.

[3] The interviews from this study are taken from the thesis of the second author, which was written under the supervision of the first author. The present study is based on a re-analysis of the data.