frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 29 November 2023
DOI 10.3389/feduc.2023.1253926

OPEN ACCESS

EDITED BY
Marcel Mierwald,
Georg Eckert Institute (GEI), Germany

REVIEWED BY
Mario Carretero,
Autonomous University of Madrid, Spain
Arja Virta,
University of Turku, Finland
Martin Nitsche,
University of Applied Sciences and Arts Northwestern Switzerland, Switzerland

*CORRESPONDENCE
David Rosenlund
david.rosenlund@mau.se

RECEIVED 06 July 2023
ACCEPTED 14 November 2023
PUBLISHED 29 November 2023

CITATION
Rosenlund D (2023) The nature of historical knowledge in large-scale assessments – a study of the relationship between item formats and offerings of epistemic cognition in the Swedish national test in history.
Front. Educ. 8:1253926.
doi: 10.3389/feduc.2023.1253926

The nature of historical knowledge in large-scale assessments – a study of the relationship between item formats and offerings of epistemic cognition in the Swedish national test in history

David Rosenlund*

Faculty of Learning and Society, Malmö University, Malmö, Sweden

Issues of validity and reliability have an impact on the construction of tests. Since the 2010s, there has been increasing emphasis in Sweden on enhancing reliability in the large-scale test system to combat grade inflation. This study aims to examine how this increased focus on reliability has affected how the nature of historical knowledge is presented in the national test in history. Accordingly, it addresses the following research question: what kinds of epistemic cognition does the test communicate to students? The concept of *epistemic cognition* builds on Kuhn et al.'s discussion on epistemic understanding, regarding the balance between the objective and subjective dimensions of knowledge. Furthermore, the concept of *companion meanings* is used to establish a connection between the items in the test and students' epistemic cognition. The findings show that the selected-response tasks predominantly communicate an objective dimension of historical knowledge, while the constructed-response tasks communicate both subjective and objective dimensions of historical knowledge. The findings regarding the offerings of epistemic cognition are discussed in relation to validity, reliability, item formats and classroom practices.

KEYWORDS

epistemic cognition, epistemic beliefs, epistemic understanding, reliability, validity, history education, item formats

1 Introduction

In test construction, the balance between construct validity and assessment reliability is important. There are two aspects of validity that is addressed in the study presented here. First, the concept *construct validity* refers to the degree to which an assessment construct captures the knowledge dimensions in the intended construct. Second, *cognitive validity* addresses the degree to which an assessment instrument can be argued to elicit the intended cognitive processes (Kaliski et al., 2015). Reliability addresses the degree to which test items enable assessors to make equal evaluations of student responses with similar quality. Reliability concerns both the agreement between different teachers (inter-rater reliability) as well as

agreement between individual teachers' assessments of different student responses (intra-rater reliability). There is an ongoing debate in history education on the relation between validity, reliability, and item formats in test construction and assessment. One central aspect in this discussion is how appropriate the item formats used in the tests are to address the construct that is to be assessed. When it comes to reliability in assessment, there is a tendency to promote selected-response (SR) items, namely items in which select the correct answer from several predefined alternatives. These items are suitable for ensuring reliability because they reduce the room for different interpretations of student responses, so assessors are more likely to make equal evaluations of responses of the same quality, resulting in higher degrees of inter-rater reliability (Rodriguez, 2015). On the other hand, the literature tends to ascribe the ability to address more complex knowledge to open-ended questions, also known as constructed-response (CR) items, where pupils construct their answer on an item. This is because complex knowledge is often characterized by variety and issues of interpretation, which are more appropriate to express in open-ended contexts (Koretz, 2008). Another relevant aspect in the debate about validity, reliability, and item formats is that large-scale tests tend to influence what teachers include in their teaching, a phenomenon labeled as a 'washback effect' (Au, 2007; Hardy, 2015).

Existing studies on item formats in history tests have investigated challenges in the use of both SR and CR items in assessing historical knowledge. When constructing a test to tap students' ability to handle three historical thinking concepts— evidence, historical perspectives, and the ethical dimension— Seixas et al. (2015) chose to use a combination of both SR and CR items. The CR items were considered necessary to elicit information about the students' perceptions of the intended construct, which are fundamental to the discipline of history (Seixas, 2015). Similarly, US history tests in the National Assessment of Educational Progress (Lazer, 2015) and the Advanced Placement (Charap, 2015) programs use a combination of the two item formats. In these studies, the challenge that comes with CR items regarding reliability was also raised, stressing the need for assessment criteria. Such criteria are needed because students have to formulate their own answers to CR questions, the assessment of which leaves room for examiners' interpretations and risks lowering inter-rater reliability (Shemilt, 2018). In the three examples above, SR items are complemented with CR items because of the assumption that the latter are better suited to address more complex types of knowledge. This implies that SR items are seen as more appropriate for assessing the less complex knowledge type, factual, or content, knowledge. This position is problematized by Shemilt (2018), who argues that although using SR items to assess factual knowledge is practical, the format is still beset with challenges because its application is based on the questionable supposition that the items are interchangeable and thus of equal difficulty. Meanwhile, there are attempts to use SR items to address more complex knowledge of the history subject. Körber and Meyer-Hamme (2015) used SR items in a study to examine students' ability to handle historical accounts. The items included the content knowledge needed, and students were asked to apply more complex concepts and provide different answers to receive points. Other studies have problematized the use of SR items to assess complex knowledge in history because the format has proven to be difficult to use for addressing the interpretative nature of historical knowledge (Reich, 2009; Smith et al., 2019).

Based on the studies presented above, test developers face a challenge in the context of history, where the construct that is to be assessed consists of complex knowledge. In such a context, the need for reducing assessors' room for interpretations (to ensure reliability) has to be balanced with the need for measuring students' proficiency regarding knowledge with higher degrees of complexity. Such a situation is at hand in Sweden, where the balance between validity and reliability has become more prominent because of a persistent grade inflation. This inflation is characterized by an increasing tendency among teachers to hand out higher grades than are motivated by students' actual levels of knowledge. The main factor behind this process is the marketization of education in Sweden, initiated in the 1992, resulting in a competition between schools (Wennström, 2020). The Swedish National Agency for Education wants to handle this inflation by improving the national tests (Skolverket, 2021). This means that reliability is likely to be given more weight in the construction of national tests.

In the Swedish national test in history, administered annually to students in Grade 9 (15–16 years), the construct that is to be assessed is formulated in the official history curriculum. This curriculum is largely recontextualized from the academic discipline of history, and the complexity of this disciplinary knowledge is also transferred to the curricula (Samuelsson, 2014; Eliasson et al., 2015). Conversely, the share of SR items in the national history test increased between the years 2013 and 2015, a fact that is noteworthy, considering that these items address knowledge with a lower degree of complexity than the knowledge prescribed by the curriculum (Rosenlund, 2022). This observed discrepancy in complexity between the history curriculum and the SR items in the national history test raises a question regarding what kind of history subject the test offers to the students taking it. Is it a subject that aligns with the constructivist complexity of the historical discipline (Zeleňák, 2015), as recontextualized in the curriculum, or is it a subject where historical knowledge is communicated as objective statements?

The study presented here aims to further the understanding of how the balance between validity and reliability affects the subject of history that is communicated to students in the national test. To address this aim, the study pursued the following research question: What stances of epistemic cognition does the national test in history communicate to students?

2 Epistemic cognition and companion meanings

Epistemic cognition is a concept that is used to address individuals' perceptions of knowledge, how knowledge is constructed, validated and the limits it is beset with (Kitchner, 1983). In this study, the operationalization of the concept is rooted in a discussion about a closely related concept, epistemic understanding (Kuhn et al., 2000), where it is characterized by the coordination of the objective and subjective dimensions of knowledge. Since this division implies that there are elements in a subject that can be treated as objective entities, and that how individuals organize and make meaning of these objective elements can be described as a subjective dimension of knowledge, it provides elements suitable to establish an analytical framework for this study. The concept is relevant to address in an educational context, since a more qualified understanding of epistemic

issues affects how individuals can utilize the knowledge that they have (Kuhn et al., 2000). Further, when students have more nuanced epistemic stances toward historical knowledge, their proficiency in other aspects of the subject increases (Van Boxtel and van Drie, 2017). However, research has shown that many adolescents have more simplistic epistemic understandings of history (Lee and Shemilt, 2004; Miguel-Revilla, 2022), indicating that history education needs to address issues of epistemic cognition (Seixas, 2015). Research has provided examples of how that can be done (Marczyk et al., 2022).

In history education research, the first two decades of the 21st century saw increased attention to the concept of epistemic cognition. It was built on ideas that took a coherent form in the United Kingdom in the 1970s about school history as a subject where strategies from the academic historical discipline are prominent (Shemilt, 1983). Similar ideas about history education took form in both the United States (Wineburg, 1991) and Canada (Seixas, 2015), partially building on the ideas from the United Kingdom. The UK-based research on the disciplinary strategies resulted in progression models that formulated suggestions regarding how students' knowledge about such strategies develop. These disciplinary strategies are recontextualized into so-called second-order concepts. These concepts are *evidence*, which concerns how information in historical sources can be addressed (Lee and Shemilt, 2003); *accounts* about the construction of historical narratives (Lee and Shemilt, 2004); and *causation*, which emphasizes how historians establish relations between historical phenomena and their causes and consequences (Lee and Shemilt, 2009).

The definition of epistemic cognition as the coordination between the objective and subjective dimensions of knowledge is generic, thereby making it applicable to several knowledge domains. It has already been applied in studies addressing history education. For instance, Maggioni et al., 2009 used this definition as one source of inspiration in an influential study on epistemic cognition in history. Their second source of inspiration was the aforementioned progression model related to evidence. In this progression model, the coordination between the objective and subjective dimension is described differently on each level. Maggioni et al., 2009 used the model and the generic model of epistemic cognition as scaffolds to define three epistemic stances—the copier, the borrower, and the criterialist stance—and construct an instrument to map individuals' epistemic cognition. The three stances describe a development in epistemic cognition from a more naïve view of knowledge that is characterized by acknowledging only the objective dimension of knowledge, via a focus on the subjective dimension, to a more nuanced view that coordinates both the objective and subjective dimension. In history, this could mean moving away from an understanding that there only is one answer to the question of whether the consequences of the industrial revolution were positive or negative; individuals holding this understanding cannot use different perspectives or address different interpretations of the consequences. As epistemic cognition progresses, individuals move from this stance to one where the subjective dimension of knowledge replaces the objective, resulting in knowledge being seen as mere opinions. In the history subject, in line with Lee and Shemilt (2003, 2004, 2009), this would mean that students see historical knowledge as dependent only on the historian's viewpoint, bias, and other personal attributes. On this level, the methodological strategies that are used by historians to bridge the objective and subjective dimensions are not yet acknowledged. On the most advanced level, students form an understanding where both the objective and subjective dimensions are acknowledged and coordinated. Here, historical knowledge is seen as the result of a subjective arrangement of objectively observable phenomena.

In research on students' epistemic cognition, both models mentioned above have been used as analytical frameworks. Stoel et al. (2017) revised the instrument constructed by Maggioni and VanSledright and surveyed 922 secondary school students. They found that students who can coordinate the two dimensions of knowledge also find history as a subject more interesting. Basing their analytical framework on Kuhn et al.'s model, Ní Cassaithe et al. (2022) conducted an interview study with 17 primary school students and found that the students' view on the nature of history and the concept of evidence affect their possibilities for progression in epistemic cognition. Similar results were reported in a survey study with 62 undergraduate students; based on the framework by Stoel et al. (2017), Sendur et al. (2022) found a strong correlation between epistemic beliefs and the quality of source-based argumentation.

Teachers are important as educators of epistemic cognition, and in the context of this study, their assessment practices are of extra interest and there are studies that highlights epistemic cognition within history tests. Van Nieuwenhuyse et al. (2015) examined four tests by each of 70 Flemish upper-secondary school teachers and found that 3 % of the questions in the tests made the constructed nature of historical knowledge visible for the students. In a similar study, Rosenlund (2016) examined all tests used by 23 upper-secondary school teachers during one academic year and found that 3.5 percent of 893 tasks communicated what is labeled as an integrated epistemic cognition in this study.

In this study, the assumptions underlying the model established by Kuhn et al. (2000) are used to formulate an analytical framework consisting of three categories. Each category describes one particular way in which historical knowledge can be presented in the tests, and thus offered to the students. The three categories comprise historical knowledge presented from (a) an objective perspective, (b) a subjective perspective, and (c) both an objective and subjective perspective. Although Kuhn et al.'s model is developmental—that is, the authors explicitly stated that it is meant to describe a progression between levels of cognitive cognition—this study does not use the model to address issues of development. Rather, it aims to employ it to identify what dimensions of historical knowledge are presented to the students; accordingly, the approach can be described as more dimensional (Nitsche et al., 2022). Furthermore, how the balance between the objective and subjective dimensions of knowledge is communicated in the national test in history will be used to indicate the companion meanings regarding epistemic cognition that are offered to the students encountering the tests.

The concept of *companion meanings* (Roberts, 1998) directs attention to the implicit learning that is present in educational contexts. Implicit learning is not just present in situations where learning is foregrounded; it is present in most educational situations, including testing. The national test in history, which is investigated in this study, carries with it companion meanings that offer students certain ways to understand historical knowledge. That the companion meanings are *offered* indicates that the meanings identified by the researcher do not have a one-to-one relationship with the meanings perceived by the students. Therefore, what this study examines is what stances to epistemic cognition that is *most likely* perceived by the

students taking the test. Since the concept companion meanings focuses on what is implicitly communicated, the intentions behind the items in the history test are of secondary importance in this study. Regardless of the complexity of the knowledge that the Swedish National Agency for Education is looking to assess, the items in the national tests communicate certain companion meanings regarding epistemic cognition to the students. Generally, tests communicate to students what aspects and dimensions of a subject are deemed as important (Black and Wiliam, 2009). Students are assumed to perceive the companion meanings in the national tests as important for three reasons: First, the tests are constructed externally from the schools; second, they are administered by an authoritative body, the National Agency of Education; and third, they are high-stake tests that affect the students' grades in history. These factors contribute to making the national tests appear authoritative to the students, which is why the tests' companion meanings regarding epistemic cognition are likely to have an impact on them.

In sum, three assumptions underlie the analytical framework (see Table 1) of this study: When an item in the test presents knowledge from an objective perspective, it offers students an *objectivist* epistemic stance concerning historical knowledge. In the same way, items presenting historical knowledge from a solely subjective perspective are assumed to offer students a *subjectivist* epistemic stance concerning historical knowledge. Finally, items containing both the objective and the subjective dimensions of historical knowledge are assumed to offer an *integrative* epistemic stance concerning historical knowledge.

3 Context of the national test and its role in teachers grading practice

As mentioned above, the construct that is assessed in the Swedish national test in history is the official history curriculum. The curriculum that was in effect when the tests investigated in this study were administered aimed at enabling students to develop four abilities after participating in history education: (1) to use a historical frame of reference that incorporates different interpretations of time periods, events, notable figures, cultural meetings, and development trends; (2) to critically examine, interpret, and evaluate sources as a basis for creating historical knowledge; (3) to reflect over their own and others' use of history in different contexts and from different perspectives; and (4) to use historical concepts to analyze how historical knowledge is organized, created, and used (Skolverket, 2011). As mentioned in the introduction, these four abilities share similarities with the academic subject of history (Samuelsson, 2014). A common feature of the abilities in the curriculum is that they communicate that interpretations have a central role in the subject, something that historians also express when describing the subject (McCullagh, 2004; Berkhofer, 2008). The level descriptors in the curriculum, characterizing the differences between the grades E, C and A, are not incorporated from the disciplines (Rosenlund, 2019), instead the Agency of Education (2011) have used generic descriptions of increasing complexity to describe them. In one line of progression, addressing how students handle relationships between time periods, the levels are described with *simple* (for grade A) – *relatively complex* (for grade B) and *complex* (for grade C).

TABLE 1 Description of coding categories.

Coding category	Description
Objectivist stance	Historical knowledge is communicated as consisting of objective elements and statements about history can be true or false.
Subjectivist stance	Historical knowledge is communicated as consisting of subjective elements and statements about history can be true or false
Integrative stance	Historical knowledge is communicated as consisting of a coordination between objective and subjective elements and statements about history can vary in quality.

In Sweden, teachers have a large degree of autonomy when it comes to grading their students. However, the national tests are an instrument that infringes on this autonomy. Namely, teachers are obliged to take the test results into account when they grade their students—a feature of the tests that was further strengthened in 2018 (Skolverket, 2018). The students receive a grade on each item in the tests, and these item grades are combined into a test grade: F (fail), E (pass), C (pass with distinction), and A (pass with special distinction). There are national tests in primary school and in both lower- and upper-secondary school. The tests are most common in the subjects Swedish (mother tongue), Mathematics, and English, which are administered at all three stages. In the natural science subjects, the tests are administered in lower- and upper-secondary school, whereas tests in social science subjects are used solely in lower-secondary school. Students have to take the national test in one of the four social science subjects (geography, history, religion, and social science) and one of the three natural science subjects (biology, chemistry, and physics). This study examines the history test.

4 Materials and methods

The empirical material examined in this study consists of the items in large-scale tests, the Swedish national history tests for Grade 9 conducted 2016–2019. The four tests included in this study contained 89 graded items categorized into two types: selected-response items and constructed response items. In many cases, the items comprised subitems that were graded individually, and these subitem-grades were combined to make up a total item grade (F, E, C, and A). This was the case for both SR items and CR items. In this study, each subitem is considered to offer a companion meaning to the students taking the test. This is based on the assumption that the companion meanings are inherent in each of the stems the students encounter and that this is valid both when the item in itself results in a grade and when its result is combined with other items. Accordingly, although the tests contained 89 graded items, 507 items were analyzed in this study.

The methodological approach used in the study is content analysis (Krippendorff, 2004), which is a useful method when looking for implicit meanings in texts. Each item was examined in relation to the analytical framework in a three-stage process (Bowen, 2009). The first two stages involved an initial, superficial reading followed by a thorough examination, and the third stage entailed a final categorization and coding based on the analytical framework (Bowen, p. 32–33). In order to enhance intra-rater reliability a reexamination

of the material was conducted. The aim was to ensure that the understanding of the analytical framework shifted during the first round of analysis. Following this process, each item was provided with two codes, one that categorizes the epistemic cognition that is communicated in the item and one that describes whether it is an SR or a CR item. The corresponding researcher was responsible for the coding procedure and in order to enhance transparency, the principles of the coding are presented in detail below.

5 The balance between the objective and subjective dimensions of historical knowledge

The 507 items were categorized according to the analytical framework presented above. The analysis of the empirical material will be presented in the following order: first, two items coded as offering an objectivist epistemic stance concerning historical knowledge; second, two items coded as offering a subjectivist epistemic stance; and lastly, two items coded as offering an integrative stance, combining the subjective and the objective dimensions of historical knowledge. The items that serve as examples in each of the following subsections are all from the 2017 test, since it at the time of publication is the most recent test that is not confidential. These items were chosen as representative examples of the item formats in the tests, and they will also be used to explain how the analytical framework has been applied in relation to the empirical material. This is because they share important characteristics with items in the national tests in history of the same format. This goes both for the other items in test from 2017, but also for the items in the tests conducted in 2016, 2018 and 2019.

5.1 Items representing an objectivist stance concerning historical knowledge

As mentioned in the theoretical section, knowledge is understood to contain both an objective and a subjective dimension. This section addresses the 432 items in the national test that solely present the objective dimension of historical knowledge. In the four analyzed tests, only SR items present knowledge from solely an objective perspective.

The first example is item number 4, which has been selected as a representative SR example due to its large number of subitems; it consists of 12 subitems, four of which are presented in Figure 1. These subitems have been categorized as communicating an objectivist epistemic stance because for each of them, there is only one correct alternative. Consequently, the subjective dimension of historical knowledge (for example, sampling, interpretation, and representation) is not offered to the students. This is not an argument for a view of historical knowledge where there are no objective aspects; the tricolor was indeed an important artifact during the French Revolution, and the Soviet Communist Party did use the symbol of the hammer and sickle. However, there are historical themes in these subitems where the subjective dimension is present but not made visible in the test. For example, two subjective aspects can be argued to nuance the objectivist stance communicated in these two items. First, what criteria can be used to substantiate that the tricolor was introduced during the French Revolution, not in the Netherlands in the 15th century? Second, the use of a hammer and a sickle to address the unity between workers and peasants can be found in other contexts before the Russian Revolution, so linking its origin to the Russian Revolution can also be nuanced by adding on a subjective perspective.

Another example of an item consisting of multiple subitems is number 18 (see Figure 2). In this item, students are asked to fill in a table with letters provided in four lists, where each letter represents (a) dates, (b) people, (c) events, and (d) countries/regions. The information from the lists is to be matched with the periods that are found in the table's leftmost column. In this item, there are 19 pieces of information that the students have to match to one of the historical periods. This type of item is present in several of the tests. It has been categorized as offering an objectivist epistemic stance for two reasons. First, each of the 19 pieces of information and the five historical periods are presented as objective entities; namely, there are no indications of a subjective aspect that can be related to them. Second, there can be only one correct response in each cell, meaning that students cannot take their subjective perspectives on history into account when responding to the item, which is also a factor for coding this kind of item as objective.

The four tests administered between 2016 and 2019 contained 432 items that have been coded as presenting historical knowledge from an objective perspective. That is, students responding to these items are being offered an objectivist epistemic stance concerning historical knowledge.

Regarding the issue of reliability and validity, the items categorized into the objective category have one important thing in common: they leave no room for interpretation in the assessment process. This feature increases both the inter- and the intra-rater reliability of the tests containing such items. On the issue of validity, since the items categorized to the objective dimension are void of the interpretative aspects of history that are present in the history curriculum, the construct validity is compromised. Also, as the number of correct results on subitems are combined to an item grade without any differentiation between the subitems regarding difficulty, the cognitive validity is reduced. This is because the level descriptors in the curriculum for the grades E, C and A are characterized by increasing complexity and not by an accumulation of content knowledge – which is what is rewarded in this kind of tasks in the national test in history.

5.2 Items representing a subjectivist stance concerning historical knowledge

Six items in the tests were coded as showing only the subjective dimension of history. These items address two different aspects of the history curriculum, namely, the uses of history and the handling of historical sources.

An example where the subjective dimension is likely to be prominent for students taking the test is item number 21 (see Figure 3). In this item, the students are supposed to provide a reason why a company in an advertisement is using a reference to its origin in 1880. This item aims to tap into student knowledge regarding the use of history, an aspect that was introduced in the history curriculum in 2011 and that history teachers in Swedish lower-secondary schools tend to neglect in their teaching (Skolinspektionen, 2015; Eliasson and Nordgren, 2016). Due to this neglect, a large share of students have likely not met this aspect as

Connections between the present day and the American, French and Russian Revolutions

4 Complete each sentence by putting a cross against the American, French or Russian Revolutions. Put **one** cross for each task (1-12).

1. Many flags of the world consist of three stripes in different colours. Such flags are called tricolour. The tricolour comes from ...

The American Revolution	
The French Revolution	
The Russian Revolution	

An example of tricolor: the Belgian flag.

2. One country is governed by a Senate and a House of Representatives. The country decided on this form of government after ...

The American Revolution	
The French Revolution	
The Russian Revolution	

3. A symbol for a particular political ideology is "the hammer and sickle". It comes from ...

The American Revolution	
The French Revolution	
The Russian Revolution	

The hammer and sickle.

4. The Marseillaise is a famous national anthem. The Marseillaise was sung for the first time during ...

The American Revolution	
The French Revolution	
The Russian Revolution	

8

FIGURE 1
Item 4 in the Swedish national test in history, 2017.

a part of their history education. The item lacks scaffolds that could inform students about how to respond to it, which increases the likelihood that students would not be aware of the objective aspects in this item. Students who lack knowledge of strategies that are productive for analyzing the uses of history and who are confronted with these tasks without any scaffolds are likely to perceive this kind

Periods of history

18 Complete tasks 1-4 by writing the correct **letters** in the table.
You should only write **one letter** in each square.
We have already written in the letter for Great Britain in Column 4: Important areas/states.

Period of History	1: Dates	2: People	3: Events	4: Important regions/ states
Ancient Egypt				
European Colonialism				
The Industrial Revolution in Great Britain				F
The World Wars				
The Post-War Period				

1. Dates

Write **the letter** for the date the period of history **begins** in column 1.
Three letters will remain over.

- **A.** 3000 BC
- **B.** 13th century
- **C.** 14th century
- **D.** 15th century
- **E.** 18th century
- **F.** 1914
- **G.** 1946
- **H.** 1989

12

FIGURE 2
Item 18 in the Swedish national test in history, 2017.

of item as presenting history solely from a subjective perspective. This means that these students are likely to perceive their response to this item as merely a personal standpoint.

The second area where the subjective dimension is likely to speak most clearly to students concerns the historicity of individuals' actions and ideas. This area is addressed in item number 11, which references

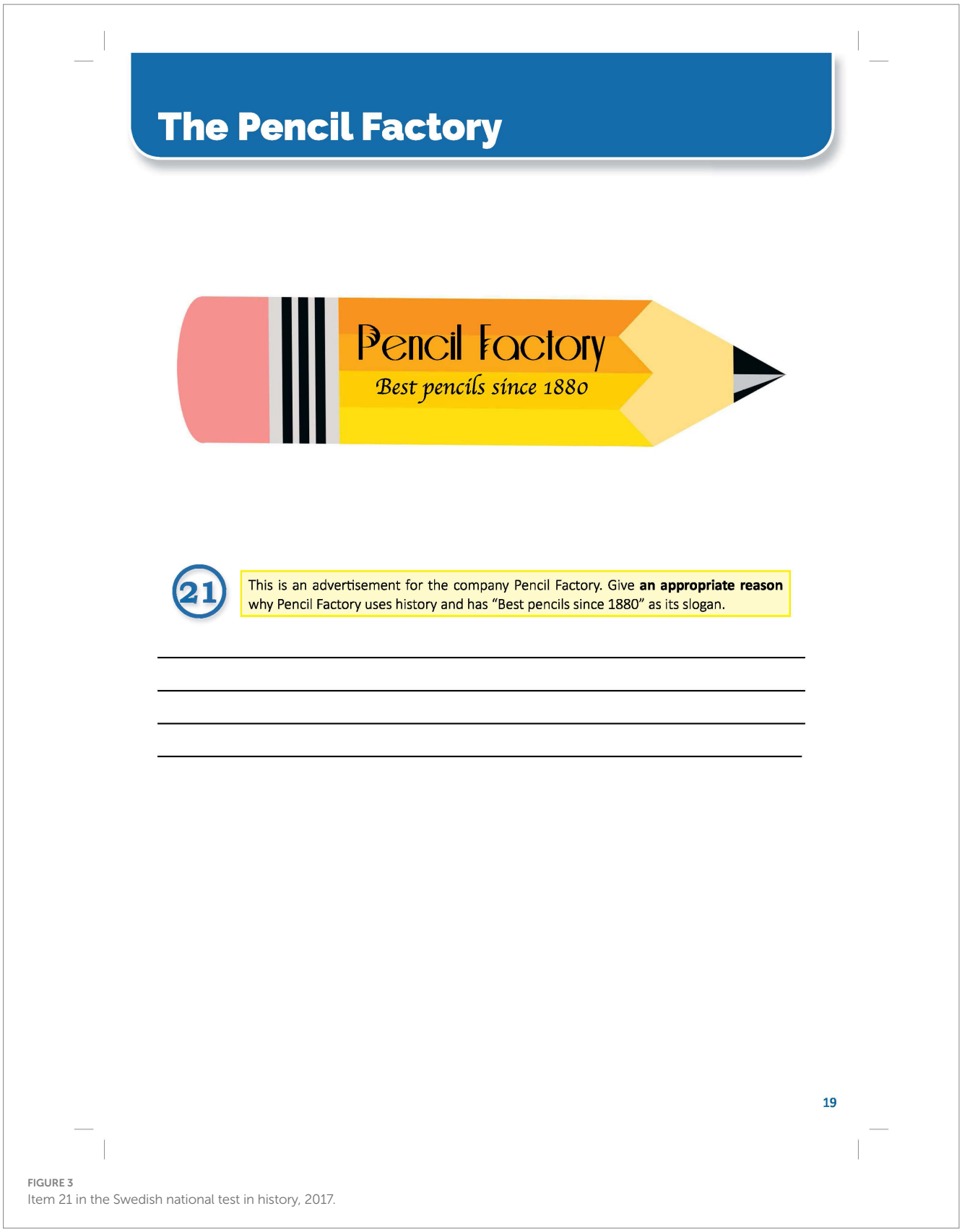


FIGURE 3
Item 21 in the Swedish national test in history, 2017.

a commercial for men's clothing (see Figure 4). In this item, students are supposed to provide reasons why the person in the commercial is smoking. To be able to do this, students need to have met the idea of historicity and learned how to apply it. Considering that historicity is a rather complex concept and a procedural aspect of history, an aspect that is seldom addressed in Swedish history education

Advertising then and now

Advertisement from 1953. The picture is advertising suits for men. The men in the picture are smoking on board an aeroplane.

11 Why did people smoke in advertising pictures in 1953? Give **one** reason.

22

FIGURE 4
Item 11 in the Swedish national test in history, 2017.

(Skolinspektionen, 2015; Eliasson and Nordgren, 2016), a large share of students are likely unfamiliar with the concept. As with item 21 above, the lack of scaffolds also contributes to the possibility that students fail to identify the objective aspects in the item and, thus, perceive the historical knowledge addressed in this kind of item as solely a subjective enterprise.

There are six items in the examined tests that have been categorized as addressing historical knowledge from a subjective perspective, all being in the CR format. The conclusions regarding these items should be considered as tentative since they are based on indications found in previous research. Nonetheless, these items likely offer students a subjectivist epistemic stance concerning historical knowledge, meaning that knowledge in the subject is more akin to opinions than based on criteria and related to objective reality.

The fact that these items lack scaffolds and address aspects that teachers tend to neglect in their teaching is likely to reduce inter-rater reliability. Regarding construct validity, the items tap into content of the history curriculum, providing students with the opportunity to show their proficiency in relation to a relevant historical content. The lack of scaffolds, however and the indications of scarce teaching in these areas risks compromising the cognitive validity. Reasons for this is that many students are likely to be unaware of what to include in their responses, something that might result in a disconnect between the level descriptors and the responses.

5.3 Items representing an integrative stance concerning historical knowledge

In the analyzed tests, 69 items have been characterized as addressing historical knowledge from a perspective where both the subjective and objective dimensions are visible. The example in Figure 5 concerns one of the CR items that have been coded as part of this integrative category. In this item, students are supposed to respond by providing examples of consequences of the industrial revolution, and they have one and a half pages at their disposal for this task. The grading rubric contains descriptions of the criteria that a response has to meet for each of the three passing grades (E, C and A). The objective dimension is present in this item in that it directs students' attention to a historical phenomenon and asks for its consequences. This is likely to communicate to the students that there are events that have happened in the past and that these have real consequences. The subjective dimension is first communicated to the students through the term *discuss* in the stem, indicating that there is not one correct way to address this item. *Discuss* is a frequently used term in the Swedish school system; it tells students that there are several aspects to be considered and that an evaluation of some sort can be included in the response. Thus, both the objective and subjective dimensions of historical knowledge are present in this item, demonstrating that it offers students an integrative epistemic stance concerning historical knowledge.

In item number 22, the SR-subitems are meant to address students' understanding of the uses of history (see Figure 6). The students are provided with two contrasting standpoints on Swedish membership in NATO through fictive persons who use historical arguments to substantiate the two standpoints. Subitems one, two, and six (shown in Figure 6) have been coded as addressing both the objective and the subjective dimension of historical knowledge because they apply to both the positive and the negative standpoints. They communicate that how the objective dimension (i.e., the historical examples referenced in the item) is used can be dependent on the subjective dimension. In this case, the subjective dimension relates to attitude toward NATO membership.

The tests contain 69 items where both the objective and the subjective dimensions have been identified. These items offer students an integrative epistemic stance, where both dimensions of knowledge are present. Among these items, 61 are in the CR format and 9 are subitems in the SR format. The SR-items present in the tests from 2016 and 2017 have high inter rater reliability since there are no room for interpretation in the assessment of student responses, and that both the objective and subjective dimensions are present increases construct validity. However, as the case is with the items addressing only the objective dimension discussed above, the cognitive validity of these SR-items is compromised because there is no differentiation regarding difficulty between them.

One aspect included in the CR-items that can help strengthen both inter- and intra-rater reliability is assessment rubrics, reducing the room for interpretations in assessments of the student responses. These rubrics also have the potential to strengthen the cognitive validity because they make it transparent what differences there are between the three grading levels. Finally, that the two dimensions of knowledge are combined strengthens the construct validity of these items.

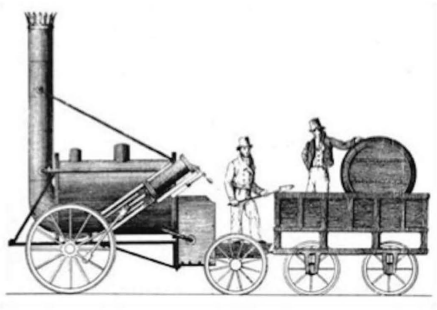
6 Conclusions and discussion

There is one consistent pattern in the four tests examined in this study: a vast majority of the items address the knowledge dimensions separately. In each of the four test years, between 84 and 90% of the items address one of the dimensions in isolation. Among this majority of items, only a small share addresses the subjective dimension, whereas between 82 and 87% of the items address the objective dimension. Moreover, only 10 to 16% of the items present historical knowledge as consisting of both the objective and the subjective dimensions. These numbers are presented in Table 2. For each year, the gray rows show the number of items that correspond with each of the three categories of epistemic cognition. The non-shaded rows show the number and percentage of SR and CR items each year.

A majority of the 507 items offer the students in Grade 9 a one-dimensional and objectivist epistemic stance concerning historical knowledge. I will in the following discuss this finding from two perspectives: first, in relation to students' proficiency in history and, second, in relation to test construction and issues of reliability and validity. Regarding students' proficiency in history, there is a risk that an objectivist approach in large-scale testing strengthens the objectivist preconceptions of historical knowledge already held by many adolescents (Lee and Shemilt, 2004; Miguel-Revilla, 2022). This is likely to impede the endeavors of history teachers to educate their students so that they can acquire an integrated epistemic cognition of historical knowledge. Such a nuanced epistemic cognition is necessary for developing other relevant aspects of the history subject, such as historical reasoning (Van Boxtel and van Drie, 2017) and source based argumentation (Sendur et al., 2022). Importantly, a more nuanced understanding of historical knowledge also increases individuals' ability to handle conflicting accounts (Nokes, 2014), an ability of great importance for active participation in society (Lee, 2011).

Regarding test construction, the analysis of the relation between the item formats and the epistemic cognition that is offered to students shows two significant correlations. The first is that the objectivist epistemic stance is exclusively offered in SR items and the second is

Consequences of the industrial revolution



Drawing of Stephenson's steam engine Rocket in an American technical journal from 1829.

7 Write about and discuss the **consequences** of the industrial revolution using **historical examples**.

Your discussion should include		
E	C	A
Two consequences	**Three** consequences. **At least two different types** of consequence, for example social, economic, political etc.	The same as for C, but also one consequence that **occurred then** and a consequence that **had a long-term affect**. Discuss this by **describing in what way** it was a consequence that lasted for a long time.

14

FIGURE 5
Item 7 in the Swedish national test in history, 2017.

that the integrative epistemic stance is offered mostly in CR items. This is in line with previous research where the use of SR-items for assessing complex constructs have been questioned (Reich, 2009; Smith et al., 2019). This study adds two aspects to this this research. First, that history assessments communicate perceptions about the nature of historical knowledge. Second, that SR-items, the way they

FIGURE 6
Item 22 in the Swedish national test in history, 2017.

are used in the context of this study, communicate an objectivist stance concerning historical knowledge. This finding strongly indicates that increasing the share of SR items in tests is likely to result in a more persistent offering of an objectivist epistemic stance. Since SR items are efficient in achieving high inter-rater reliability (Rodriguez, 2015), the share of SR items in large-scale tests is likely to

TABLE 2 Number of items for each category, divided by item type.

	Epistemic stances							
Year	Objectivist		Subjectivist		Integrative		Total	
2016	109 (87%)		0 (0%)		16 (13%)		125	
	SR: 109 (100%)	CR: 0	SR: 0	CR: 0	SR: 4 (25%)	CR: 12 (75%)	SR: 109 (87%)	CR: 16 (13%)
2017	95 (82%)		2 (2%)		19 (16%)		116 (100%)	
	SR: 95 (100%)	CR: 0	SR: 0	CR: 2 (100%)	SR: 4 (21%)	CR: 15 (79%)	SR: 99 (85%)	CR: 17 (15%)
2018	122 (87%)		4 (3%)		14 (10%)		140 (100%)	
	SR: 122 (100%)	CR: 0	SR: 0	CR: 4 (100%)	SR: 0	CR: 14 (100%)	SR: 122 (87%)	CR: 18 (13%)
2019	106 (84%)		0 (0%)		20 (16%)		126 (100%)	
	SR: 106 (100%)	CR: 0	SR: 0	CR: 0	SR: 0	CR: 20 (100%)	SR: 106 (84%)	CR: 20 (16%)

For each year, the gray rows show the number of items that correspond with each of the three categories of epistemic cognition. The non-shaded rows show the number and percentage of SR and CR items each year.

be substantial when reliability is a prioritized objective, as is the case in Sweden. Also, such a prioritization would risk compromising both the construct- and the cognitive validity of the national test in history. Regarding construct validity, the results presented in this study show that the objective and subjective dimensions of knowledge rarely are combined in SR-items. The risk for cognitive validity comes from the fact that the SR-items in the test are interchangeable and do thus not mirror the progression of complexity described in the curriculum.

Considering the possibility of a washback effect (Au, 2007), it is not a farfetched concern that if the objectivist epistemic stance will predominate in the large-scale tests, it is likely that will also be so in the history classrooms. If students are to be acquainted with an integrative epistemic stance, teachers have to promote such an understanding in their own teaching despite what is emphasized in the tests. History education researchers have suggested concept such as evidence, causation, and significance (Lee and Shemilt, 2004; Seixas, 2015) as disciplinary strategies that an education promoting an integrative epistemic stance can focus on. That said, research on teachers' assessment practices indicates that these findings from history education research have yet to be applied in history classrooms (Van Nieuwenhuyse et al., 2015; Rosenlund, 2016).

Bearing in mind that the four social science subjects examined through national tests in Sweden (i.e., geography, history, civics, and religious education) share many similarities, both regarding the complexity of curriculum content (Samuelsson, 2014) and the construction of the national tests, endeavors to ensure reliability in assessment are likely to result in similar consequences regarding the epistemic cognition offered in all four subjects. Furthermore, an approach to ascribe items the same value with no differentiations based on difficulty is likely to strengthen the image of the subjects as consisting of non-related pieces of information (Shemilt, 2018), increasing the obstacles for students to construct coherent images of the subject at hand (Shemilt, 2009).

Returning to the balance between reliability and validity, this study cannot provide an answer regarding how a reasonable balance can be achieved. However, the study can offer one recommendation that may be useful when seeking such a balance. Test developers are encouraged to consider how the balance between validity and reliability impacts how the subject is perceived by the test takers. If, for example, the share of SR items are increased in an attempt to increase inter- and intra-rater reliability, they should be mindful of the effects this strategy can have on the epistemic cognition offered to students. This is a crucial point because companion meanings in the test items impact students' perception of what is important in a subject (Black and Wiliam, 2009). In addition, offering to students a one-dimensional cognition of knowledge hinders their ability to use their knowledge in constructive ways (Kuhn et al., 2000) when participating in society.

Finally, it is important to mention two limitations of his study. First, this study is not a reception study, meaning that the relation between the epistemic stances communicated in the items and the statements made here about how students perceive the nature of historical knowledge is theoretical. Second, this study does not take into account other aspects of history education where epistemic stances are communicated, history teaching and teacher-made tests, being two examples. This means that this study addresses one of several aspects that influence students' epistemic cognition. However, the consistency of results presented above, regarding the stances that are communicated in the tests, calls for further research on the washback effects of high-stakes testing on students' epistemic cognition.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

DR: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Agency of Education. (2011). *Curriculum for the compulsory school system, the pre-school class and the leisure-time Centre 2011*. Stockholm: Swedish National Agency for Education Skolverket: Fritze distributör.

Au, W. (2007). High-stakes testing and curricular control: a qualitative metasynthesis. *Educ. Res.* 36, 258–267. doi: 10.3102/0013189X07306523

Berkhofer, R. F. (2008). *Fashioning history*. New York: Palgrave, Macmillan.

Black, P., and Wiliam, D. (2009). Developing the theory of formative assessment. *Educ. Assess. Eval. Account.* 21, 5–31. doi: 10.1007/s11092-008-9068-5

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qual. Res. J.* 9, 27–40. doi: 10.3316/QRJ0902027

Charap, L. G. (2015). "Assessing historical thinking in the redesigned advanced placement United States history course and exam" in *New directions in assessing historical thinking*. eds. K. Ercikan and P. Seixas (New York: Routledge), 159–170. doi: 10.4324/9781315779539

Eliasson, P., Alvén, F., Yngvéus, C.A. and Rosenlund, D. (2015). "Historical consciousness and historical thinking reflected in large-scale assessment in Sweden," in *New directions in assessing historical thinking*. eds. K. Ercikan and P. Seixas. Routledge, 171.

Eliasson, P., and Nordgren, K. (2016). Vilka är förutsättningarna i svensk grundskola för en interkulturell historieundervisning? [what are the conditions in Swedish lower secondary school for intercultural history teaching?]. *Nordidicatica-J Hum Soc Sci Educ* 2, 47–68.

Hardy, I. (2015). A logic of enumeration: the nature and effects of national literacy and numeracy testing in Australia. *J. Educ. Policy* 30, 335–362. doi: 10.1080/02680939.2014.945964

Kaliski, P., Smith, K., and Huff, H. (2015). "The importance of construct validity evidence in history assessment: what is often overlooked or misunderstood?" in *New directions in assessing historical thinking*. eds. K. Ercikan and P. Seixas (New York: Routledge)

Kitchner, K. S. (1983). Cognition, metacognition, and epistemic cognition: a three-level model of cognitive processing. *Hum. Dev.* 26, 222–232. doi: 10.1159/000272885

Körber, A., and Meyer-Hamme, J. (2015). "Historical thinking, competencies, and their measurement: challenges and approaches" in *New directions in assessing historical thinking*. eds. K. Ercikan and P. Seixas (New York: Routledge), 89–101.

Koretz, D. M. (2008). *Measuring up*. Cambridge, Mass: Harvard University Press.

Krippendorff, K.H. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks: Sage Publications.

Kuhn, D., Cheney, R., and Weinstock, M. (2000). The development of epistemological understanding. *Cogn. Dev.* 15, 309–328. doi: 10.1016/S0885-2014(00)00030-7

Lazer, S. (2015). "A large-scale assessment of historical knowledge and reasoning: NAEP US history assessment" in *New directions in assessing historical thinking*. eds. K. Ercikan and P. Seixas (New York: Routledge), 145–158.

Lee, P. (2011). "History education and historical literacy," in *Debates in history teaching*, ed. I. Davies. Routledge London).

Lee, P., and Shemilt, D. (2003). A scaffold, not a cage: progression and progression models in history. *Teach. Hist.* 113, 13–23.

Lee, P., and Shemilt, D. (2004). 'I just wish we could go back in the past and find out what really happened': progression in understanding about historical accounts. *Teach. Hist.* 117, 25–31.

Lee, P., and Shemilt, D. (2009). Is any explanation better than none? Over-determined narratives, senseless agencies and one-way streets in student's learning about cause and consequence in history. *Teach. Hist.* 137, 42–49.

Maggioni, L., VanSledright, B., and Alexander, P. A. (2009). Walking on the borders: a measure of epistemic cognition in history. *J. Exp. Educ.* 77, 187–214. doi: 10.3200/JEXE.77.3.187-214

Marczyk, A. A., Jay, L., and Reisman, A. (2022). Entering the historiographic problem space: scaffolding student analysis and evaluation of historical interpretations in secondary source material. *Cogn. Instr.* 40, 517–539. doi: 10.1080/07370008.2022.2042301

McCullagh, C.B. (2004). *The logic of history: Putting postmodernism in perspective*. Routledge, New York

Miguel-Revilla, D. (2022). What makes a testimony believable? Spanish students' conceptions about historical interpretation and the aims of history in secondary education. *Historical Encounters* 9, 101–115. doi: 10.52289/hej9.106

Ní Cassaithe, C., Waldron, F., and Dooley, T. (2022). We can't really know cos we weren't really there: identifying Irish primary children's bottleneck beliefs about history. *Historical Encounters* 9, 78–100. doi: 10.52289/hej9.105

Nitsche, M., Mathis, C., and O'Neill, D. (2022). Epistemic cognition in history education, historical. *Encounters* 9, 1–10. doi: 10.52289/hej9.101

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

Reich, G. A. (2009). Testing historical knowledge: standards, multiple-choice questions and student reasoning. *Theory & Research in Social Educ* 37, 325–360. doi: 10.1080/00933104.2009.10473401

Roberts, D. A. (1998). "Analyzing school science courses: the concept of companion meaning" in *Problems of meaning in science curriculum*. eds. D. A. Roberts and L. Ostman (New York: Teachers College Press), 5–12.

Rodriguez, M. C. (2015). "Selected-response item development" in *Handbook of test development*. eds. S. Lane, M. R. Raymond and T. M. Haladyna (New York: Routledge).

Rosenlund, D. (2016). "*History education as content, methods or orientation? A study of curriculum prescriptions, teacher-made tasks and student strategies*". (Frankfurt: Peter Lang).

Rosenlund, D. (2019). Internal consistency in a Swedish history curriculum—a study of vertical knowledge discourses in aims, content and level descriptors. *J. Curric. Stud.*, 51, 1, 84–99. doi: 10.1080/00220272.2018.151356

Rosenlund, D. (2022). Två olika berättelser om historieämnets komplexitet?: En jämförelse mellan grundskolans kursplan och det nationella provet i historia, [two different narratives about the complexity of the subject of history?: a comparison between the primary school curriculum and the national exam in history], in *Cross-sections: Historical Perspectives from Malmö University* ed. J.H. Glaser, Julia; Lund, Martin; Lundin, Emma (Malmö: Malmö university).

Samuelsson, J. (2014). Ämnesintegrering och ämnesspecialisering: SO-undervisning i Sverige 1980-2014. [subject integration and subject specialization: social-studies teaching in Sweden 1980-2014]. *Nordidactica: J Human Soc Sci Educ* 2014, 85–118.

Seixas, P. (2015). *"looking for history," in joined up history: New directions in history education research, eds. A. Chapman & a. Wilschut*. (Charlotte: Information Age Publishing).

Seixas, P., Gibson, L., and Ercikan, K. (2015). "A design process for assessing historical thinking" in *New directions in assessing historical thinking*. eds. K. Ercikan and P. Seixas (New York: Routledge), 102.

Sendur, K., van Drie, J., and van Boxtel, C. (2022). Epistemic beliefs and written historical reasoning: exploring their relationship. *Historical Encounters* 9, 141–158. doi: 10.52289/hej9.108

Shemilt, D. (1983). The devil's locomotive. *History and theory* 22, 1–18. doi: 10.2307/2505213

Shemilt, D. (2009). "Drinking an ocean and pissing a cupful" in *National history standards: The problem of the canon and the future of teaching history*. eds. L. Symcox and A. Wilschut (Charlotte, NC: Information Age Pub)

Shemilt, D. (2018). "Assessment of learning in history education: past, present, and possible futures" in *The Wiley international handbook of history teaching and learning*. eds. S. A. Metzger and L. M. Harris (New York: Wiley), 449–471.

Skolinspektionen (2015). *Undervisningen i historia*.

Skolverket (2011). *Curriculum for the compulsory school system, the pre-school class and the leisure-time centre 2011*. Swedish National Agency for Education.

Skolverket. (2018). *Genomföra och bedöma nationella prov i grundskolan[conduct and assess national tests in primary school]*.

Skolverket (2021). *Betygen måste bli mer likvärdiga [The grades must become more equal]*. Stockholm: Skolverket.

Smith, M., Breakstone, J., and Wineburg, S. (2019). History assessments of thinking: a validity study. *Cogn. Instr.* 37, 118–144. doi: 10.1080/07370008.2018.1499646

Stoel, G., Logtenberg, A., Wansink, B., Huijgen, T., van Boxtel, C., and van Drie, J. (2017). Measuring epistemological beliefs in history education: an exploration of naïve and nuanced beliefs. *Int. J. Educ. Res.* 83, 120–134. doi: 10.1016/j.ijer.2017.03.003

Van Boxtel, C., and van Drie, J. (2017). "Engaging students in historical reasoning: the need for dialogic history education" in *Palgrave handbook of research in historical culture*

and education. eds. M. Carretero, S. Berger and M. Grever (London: Palgrave), 573–589. doi: 10.1057/978-1-137-52908-4_30

Van Nieuwenhuyse, K., Wils, K., Clarebout, G., Draye, G., and Verschaffel, L. (2015). "Making the constructed nature of history visible. Flemish secondary history education through the lens of written exams" in *Joined-up history: New directions in history education research*. eds. A. Chapman and A. Wilschut (Charlotte, NC: Information Age Publishing), 231–253.

Wennström, J. (2020). Marketized education: how regulatory failure undermined the Swedish school system. *J. Educ. Policy* 35, 665–691. doi: 10.1080/02680939.2019.1608472

Wineburg, S. (1991). Historical problem solving: a study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *J. Educ. Psychol.* 83, 73–87. doi: 10.1037/0022-0663.83.1.73

Zeleňák, E. (2015). Two versions of a constructivist view of historical work. *History and Theory* 54, 209–225. doi: 10.1111/hith.10754