

Bias in teachers' assessments of students' historical narratives

Fredrik Leonard Alvén* – *Malmö University, Sweden*

Abstract

According to diverse research, historical thinking and historical accounts or narratives contain different dimensions. At least three such dimensions can be found: historical methods, rhetorical forms and ethical statements. The ethical dimension means that historical narratives contain ethical agendas; the rhetorical dimension implies that historical narratives consist of certain stylistic figures; the dimension of historical methods signifies that history is a science with certain methods that must be considered when constructing historical narratives. Although research on history teaching and assessment has made great progress in recent years, it almost exclusively deals with the dimension of historical methods. This is problematic as students' historical narratives, test responses or essays, contain all three dimensions, and all three dimensions seem to be taken into account when the students' narratives are assessed. This study problematizes what happens when teachers in Sweden are only required to assess the dimension of historical methods. The research is based on an empirical investigation where teachers, using the knowledge requirements from the syllabus in history, assessed four historical narratives with focuses on different dimensions of the three. The results suggest that teachers find it difficult to accept a historical narrative that, on the one hand, corresponds to the dimension of historical methods but, on the other hand, contains ethical statements that do not correspond with the assessor's own ethical understanding.

Keywords: Sweden; assessment; bias; curriculum; grading; historical narratives – students; history teaching

Introduction

According to philosophers of history and history education researchers, historical narratives comprise different dimensions (Alvén, 2017a; Rüsen, 2004; White, 1987; Ricœur, 1984). Researchers and thinkers base their conclusions on different ontological and epistemological starting points, but at least three important dimensions in historical narratives can be found in their works. There is a scientific approach to history, where the historical production is assessed by distinct criteria of how historical narratives should be interpreted and understood (Rüsen, 2004). There is also a political and moral dimension. This entails that history and historical mediation are included in political projects, such as nation building, the sanctioning of established power, or as effective tools to criticize contemporary society. It is not historical methods that are the primary tools to understand and convey history; rather, the value of history is embedded in the political goal (Alvén, 2017a, 2017b; Wertsch, 2002). When history is used to form or transform contemporary society, it also involves moral and ethical aspects (Gaddis, 2002; Rüsen, 2005). Finally, the aesthetic dimension determines that historical narratives

are clothed in a rhetorical form, which may be more or less appealing and persuasive. History as a subject has its own stylistic and rhetorical figures (Coffin, 2006; White, 1973, 1978). Indeed, the philosophers Ricœur (1984, 1985, 1988) and White (1973) agree that historical narratives are based on narrative forms.

In this article, these three dimensions in history narratives are termed: (1) the historical thinking dimension – corresponding to the scientific use of history; (2) the ethical dimension – corresponding to morality, and moral or political statements; and (3) the rhetorical dimension – corresponding to ideas of how history narratives should be linguistically formulated.

In the mid-nineteenth century, the subject of history became compulsory in many Western countries and was seen as an instrument to strengthen citizens' sense of national identity (Foster, 2011). Since then, national master narratives have played a central role in the subject of school history (Barton and Levstik, 2004; Carretero, 2011). These national master narratives give the public – in this case, students – insight into important cultural themes, heroes, values and expected ethical behaviour (Kessler and Wong-Mingji, 2009). This is a simplistic history without nuances or different perspectives, and this simplistic way of understanding the nation's history involves what Wertsch (1998) calls schematic narrative templates. Such schematic narrative templates may not only underpin destructive nationalist ideas, but also promote ideas about globalism, democracy, mutual international understanding and peace, but they are still often without nuances and immune to different perspectives (Alvén, 2017a; Foster, 2011; Pingel, 2010). Nietzsche (2010: 8–9) wrote about a monumental use of history, reminiscent of a master narrative:

History is necessary above all to the man of action and power who fights a great fight and needs examples, teachers and comforters; he cannot find them among his contemporaries ... And yet if we really wish to learn something from an example, how vague and elusive do we find the comparison! If it is to give us strength, many of the differences must be neglected, the individuality of the past forced into a general formula and all the sharp angles broken off for the sake of correspondence.

Traditional history teaching, where the teacher explains and applies a national master narrative, also includes the three dimensions in a natural way (Stoel *et al.*, 2015; VanSledright, 2011; Lee and Howson, 2009; Seixas, 2000). The national master narrative contains the rhetorical form by having an explicit beginning, a middle and an end, and distinct content that forces the story forward (Ricœur, 1984). The ethical dimension is relevant because the master narrative seeks to anchor patriotic feelings among students. Finally, the master narrative, although carefully selected, rests upon interpreted sources and historical events that underpin the historical thinking dimension. The use of history teaching as a means to build the nation through patriotic feelings, often characterized as a collective memory approach (Seixas, 2007; VanSledright, 2011), has recently come to be questioned. Many educationalists consider that this approach to teaching history belongs to the past – to a time when ethnically homogeneous nations were built with the help of a collective national master narrative (Carretero, 2017). Unfortunately, however, many political elites still see history as a nation-building exercise.

Instead of favouring memorizing a master narrative, recent research has emphasized that the teaching of history should consider how history is constructed, based on sources and causal beliefs. Research such as this was influential in Great Britain during the 1970s. Lee and Ashby (2000: 199) write that 'the changes in English history education can ... be described as a shift from the assumption that school history was

only a matter of acquiring substantive history to a concern with students' second-order ideas', that is, the disciplinary concepts of history. Investment in a new way of teaching history – the 'New History' – was extensive, headed by the inspirational Schools Council History Project (1972–4) (Berger, 2012). Influential researchers such as Ashby, Blow, Lee and Shemilt have documented results from the evaluation of the Schools Council History Project, and their further development of research over the following decades (Ashby and Edwards, 2010; Shemilt, 2000, 2009; Blow *et al.*, 2015; Lee, 2004; Lee and Howson, 2009). These researchers advocate a history teaching grounded in the discipline of history at the universities. Students should learn historical disciplinary methods. Much of this research is focused on second-order historical disciplinary concepts (Lee and Ashby, 2000), such as source criticism, historical causation, historical significance and historical empathy, which are supposed to help students understand the epistemology of history. Lee (1991: 48–9) summarizes the radically different approach originating from the New History, compared to views about history education that preceded it:

... it is absurd to say that schoolchildren know any history if they have no understanding of how historical knowledge is attained ... The ability to recall accounts without any understanding of the problems involved in constructing them or the criteria involved in evaluating them has nothing historical about it.

In the wake of this project, much research about history teaching has focused on students' own construction of history, that is, how they can be encouraged to think as historians and construct historical narratives on their own (Semmet, 2012; Laville, 2004). There are many examples of researchers participating in, and even leading, projects that directly strive to persuade history teachers to teach historical thinking skills to students, rather than to employ a nationalistic master narrative. In Canada, Peter Seixas, among others, developed the Historical Thinking Project from 2006 (<http://historicalthinking.ca>). In Germany, Andreas Körber is involved in the project Hitch: Historical Thinking: Competencies in History (Körber and Meyer-Hamme, 2015). Ercikan and Seixas (2015a: 255) write that 'history educators around the world have mobilized curricular reform movements toward including complex thinking in history education, advancing historical thinking, developing historical consciousness, and teaching competence in historical sense making'.

Other contemporary processes have also challenged the national master narrative as the obvious approach to the subject of history. Globalization and migration, for instance, challenge such narratives (Rüsen, 2011). These processes challenge master narratives by introducing new ways of interpreting and understanding historical events, in the light of globalization and through the perspective on history of different peoples. Today, these new perspectives on history have explicitly been introduced into many classrooms in the West by the inclusion of students from many parts of the world. According to many researchers, the answer to the challenge of the master narrative in the classroom is to look upon history as a subject that not only can provide different perspectives on the past, but actually draws its strength from this possibility (Nordgren and Johansson, 2015; Grever, 2012; Seixas, 2009).

Today, researchers in the field of school history ask for teaching that is characterized by the students' own thinking and their construction of historical accounts, as well as by allowing multiple perspectives on historical events. Indeed, the curricula in some countries even urge teachers to develop students' competences to make their own historical narratives, often via the concept of historical consciousness (Eliasson *et al.*, 2015; Kölbl and Konrad, 2015). VanSledright (2014: 6) interprets this to mean that the

students are supposed to *do history*. By extension, this means that history can no longer be taught as a single national master narrative that students are supposed to remember and retell in tests or in essays. Instead, students should learn how to make their own historical narratives. Consequently, the new ways of looking at history teaching and history learning require new ways of assessing students' knowledge in the subject of history: students' own historical narratives must now be assessed. Assessment researchers have taken this into account and have tried to find ways to assess students' historical thinking skills (Ercikan and Seixas, 2015b; Seixas and Morton, 2013). However, they have mostly focused on the historical thinking dimension, that is, the scientific approach to history (Eliasson *et al.*, 2015; Ercikan and Seixas, 2011). In fact, when they do history, students tend to utilize all three dimensions discussed above (Alvén, 2011, 2017b; Alvén and Zander, 2011).

Further complicating this approach, recent research indicates that how we understand and use history is strongly dependent on our identity (Ahonen, 2012; Grever, 2012) and the historical culture in which we live: historical culture as 'historical memory and historical consciousness working in its social context' (Rüsen, 2007: 22). Identity and historical culture affect both the moral dimension and the rhetorical dimension as we construct historical narratives. Several researchers have shown how students react and dissociate themselves from the historical narratives that schools provide, due to their identity, and stemming from other historical cultures outside the schools (Wertsch, 2000, 2002; Rosenzweig and Thelen, 1998; Barton and McCully, 2005). On the whole, it is reasonable to assume that, given the opportunity, some students will construct historical narratives different from those presented by their schools. Stobart (2005: 282) argues that 'There is no cultural neutrality in assessment or in the selection of what is to be assessed. This applies as much to mathematics as it does to history, and attempts to portray any assessment as acultural are a mistake.' He also concludes that assessment processes often strive to standardize knowledge, which encourages assimilation. Altogether there seem to be several things challenging a reliable, fair and neutral assessment of students' own historical narratives.

To summarize, in the classroom we ask for students' own construction of historical narratives, and these are assessed. Further, history in general, and historical narratives in particular, embrace at least three dimensions: historical thinking, ethics and rhetoric. Furthermore, the way one understands and uses history, the building blocks for one's historical narratives, is dependent on one's identity and the historical culture within which one lives. A fair approach to assessment must therefore be open about constructs and designs (Gipps, 1999). For example, if students are supposed only to show knowledge from the dimension of historical thinking, their historical narratives should not be assessed by the ethical or rhetorical dimensions.

This quantitative study problematizes how teachers in Sweden differentiate between the three dimensions when assessing historical narratives using knowledge requirements from the curriculum that only covers the historical thinking dimension. The purpose of this study was to examine the effects of the three dimensions in the subject of history in producing potential bias in scoring among assessors.

Research overview

Messick's (1989) concept of construct-irrelevant variance defines assessments where irrelevant aspects affect the outcome. Generally, it is used to find parts in tests that invalidate their reliability, that is, parts that require knowledge from students that is not supposed to be part of the assessment. In this case, however, the concept is used

to analyse teachers' grading. We can then understand irrelevance as a process where teachers take irrelevant aspects of students' narratives into account when grading. Therefore, grading is unfair because it is not true to the knowledge domain that is supposed to be assessed; further, it is not transparent to the students answering the questions. Rater bias could thus be considered as an example of a source of potential construct-irrelevant variance.

There is, however, research on bias in grading. It has shown that bias in grading can be both conscious (Malouff, 2008) and unconscious (Malouff *et al.*, 2014). Malouff and Thorsteinsson's (2016) meta-analysis (representing 20 studies and a total of 1,935 graders) concludes there are certain groups of students that are exposed to biased grading to a high degree. These are mostly students with negative educational labels, students belonging to specific ethnic groups, students who are already known to perform poorly, and students who are regarded as less attractive by the graders (*ibid.*). Students from minorities may get both higher and lower grades than other students due to bias (Fajardo, 1985). Gender also affects graders (Spear, 1984). Malouff and Thorsteinsson (2016) suggest that keeping students anonymous when possible during grading may be a way to counteract bias in grading (see also Brennan, 2008; Kahneman, 2011). Although this can help when bias stems from students' characteristics, it will not help if the construct-irrelevant variance assessment is located in the students' responses and not in the students themselves.

Limited research investigates grading bias regarding the content of responses. Some researchers note that graders sometimes take into account aspects that are not part of the construct to be measured (Davidson *et al.*, 2000; Howell *et al.*, 1993; Bennett *et al.*, 1993; Diederich, 1974; Hogarth, 1987). For example, Davidson *et al.* (2000) have presented statistically significant results indicating that raters graded non-violent content in essays higher than violent content, although neither were aspects of the construct to be measured. Researchers using verbal protocol methodology and anonymous and coded student responses identify that graders show strong emotions while assessing students' responses (Crisp, 2008, 2012; Alvé, 2017b). In such cases, graders expressed feelings of pleasure, dislike and sympathy. In addition, they were found to have talked to or imagined a stereotype student congruent to the response while assessing (Crisp, 2012; Alvé, 2017b). A small study indicates specific problems for graders assessing a certain knowledge domain, namely knowledge in history as it is described in the Swedish curriculum for the compulsory school (Alvé, 2017b). The results show that graders used not only the knowledge requirements in the curriculum when grading students' responses, but also ethical statements and rhetorical qualities that were not described therein. As the study was built upon a small sample of graders, it contains no statistical significance for construct-irrelevant variance.

As a consequence, this study problematizes grading bias in history in the compulsory school in Sweden in a quantitative manner.

Research design

An authentic, well-developed student response, graded high by several teachers, was used as an original answer to be assessed. With this answer as a point of departure, three other answers were constructed: one without any historical evidence, one with a poorer form of rhetoric, and one with an ethical standpoint not in line with the schools' fundamental values (however, these were aspects that were not supposed to be assessed or graded). Except for the manipulated dimension, each constructed answer was true to the original answer. According to the knowledge requirement, the

only answer that was required to receive a lower grade was the one without historical evidence. Neither the students' rhetorical skills nor the ethical aspects in their historical accounts were supposed to be assessed in history. The four answers were then graded by random samples of active history teachers, with different groups for each answer.

Research question

According to the Swedish curriculum for history in the compulsory school Years 7–9 (ages 13–15), students are required to develop their own historical thinking (in fact, historical consciousness). Four competences are expected to be developed. These are their ability to:

- use a historical frame of reference that incorporates different interpretations of time periods, events, notable figures, cultural meetings and development trends
- critically *examine, interpret* and *evaluate* sources as a basis for creating historical knowledge
- *reflect* on their own and others' use of history in different contexts and from different perspectives
- use historical concepts to *analyse* how historical knowledge is organized, created and used. (Swedish National Agency for Education, 2011: 163)

The italicized words emphasize that it is the students doing history that is to be at the centre of history teaching. These actions, formulated through the four competences, connect to the historical thinking dimension. However, the curriculum does not request any aesthetic or rhetorical expressions when the students do history. This is also evident from the 'knowledge requirements' that are to be used by the teachers when assessing students' knowledge. This can create difficulties when assessing students' historical accounts, as they contain all three dimensions (Alvén, 2011). To address this problem, the following question was formulated for the study:

How do different teachers, using the same knowledge requirement, assess similar responses, but with different emphasis on the three dimensions: historical methods, rhetorical forms and ethical statements?

Research method

An authentic test response from a student in the final grade in the compulsory school, age 15–16, was chosen as a point of departure (see Appendix for the original response). Three anonymous history teachers had already assessed the response, which they deemed a high-quality answer, and thus awarded it the highest grade, A. The response was from a test task completed in the autumn of 2009. The class had studied the history of migration to and from Sweden, after which they sat an examination paper. To solve the task, which was built as a scenario, the students had to act as the Swedish Minister for Migration at a future date, namely the year 2011. In Russia, a group of soldiers had seized power by force and had started to annex territories that had once belonged to the Soviet Union. In Afghanistan, jihadists had seized power. As a consequence, tens of thousands of refugees from both Finland and Afghanistan sought asylum in Sweden. As the Minister for Migration, the student had to: (1) consider how she wanted Sweden to act in this extreme situation; and (2) support her arguments with examples from history, mainly from the history of immigration (Alvén, 2011).

With the original response as a starting point, three fictitious responses were constructed (see Table 1). These responses were constructed to be inferior to the

original one. Therefore, the working hypothesis was that the original answer would get the highest grades. Thus, one response contained very few historical examples and facts to strengthen the arguments. This response is called *No evidence* in the study, and it lacks much of the historical thinking dimension. One response used less rhetorical and appealing language, and it is called *No rhetoric*. The last response constructed contained values that were not congruent with the values explained in the first chapter in the curriculum, 'Fundamental values and goals' (Swedish National Agency for Education, 2011), and it is called *Another ethic*. The moral content in this response opposes two learning goals in the curriculum: that students realize the value of a multicultural society and that they learn to stand up for vulnerable people. This response was more difficult to construct because the same history examples contained in the original response could not be used to strengthen the arguments. The solution became a variant, where the moral content and the history examples from another student response to the same examination paper were used. The response, however, still adhered to the form of the original response. All the responses are given in the Appendix, translated from Swedish.

All the responses were also tested in LIX (www.lix.se). The LIX readability index (LIX – Läsbarhetsindex) is used as a way to assess how easy or difficult a text is to read. LIX is based on the average number of words per sentence and the proportion of long words (words with more than six letters) expressed as a percentage. Both the original answer and the constructed answers received a readability index <30, which corresponds to a very easy to read text.

The number of words and sentences in the *No evidence* response was, for obvious reasons, not as high as in the other responses. The structure of the other three responses was similar to each other, except that the response *Another ethic* had more historical arguments and slightly fewer other relevant facts than the original response. All in all, the responses were similar to each other, except for the focuses on different dimensions.

Table 1: The form of the responses

Response	Word count	Number of sentences	LIX index	Number of historical examples	Number of other relevant facts	Causal reasoning or not
Original	332	28	28	3	7	Yes
Another ethic	296	28	29	6	5	Yes
No evidence	184	14	26	1	0	No
No rhetoric	217	28	29	3	7	Yes

The knowledge requirement from the curriculum that the teachers had to use to assess the responses are shown below:

- **Grade E:** Pupils can study some trends where cultures interact, and in migration, politics and living conditions, and describe simple relationships between different time periods. Pupils also give some possible extrapolations of these trends, and justify their reasoning with simple and, to some extent, informed references to the past and the present.
- **Grade C:** Pupils can study some trends where cultures interact, and in migration, politics and living conditions, and describe relatively complex relationships

between different time periods. Pupils also give some possible extrapolations of these trends and justify their reasoning by developed and relatively well-informed references to the past and the present.

- **Grade A:** Pupils can study some trends where cultures interact, and in migration, politics and living conditions, and describe complex relationships between different time periods. Pupils also give some possible extrapolations of these trends and justify their reasoning by applying well-developed and well-informed references to the past and the present. (Swedish National Agency for Education, 2011: 170–1)

Grade E corresponds to a pass grade, while A is the highest grade. Grade F is a non-pass grade; therefore, it has no place in the knowledge requirements. If a response does not reach the requirements for the pass at level E, it is considered as an F. The knowledge requirement the teachers used to assess the answers calls for assessing the descriptions of relationships between time periods and references to historical events and the contemporary. However, there are no requirements regarding ethical or rhetoric dimensions.

Sampling

To achieve a random sample of 320 teachers that could assess the responses, a list of all Swedish municipal schools and a random number generator were used to randomize 320 schools. Thereafter, history teachers at the schools were contacted via email. The email asked if they wanted to assess the historical account that was enclosed, with the task the student had answered and the knowledge requirement they were supposed to use for assessment. Each teacher was only assigned one of the four responses. The teachers answered the email with a letter F if they thought that the response they had assessed did not reach the minimum pass level, and gave a grade of E, C or A if they thought the response corresponded to one of those levels in the knowledge requirement.

Data collection

In total, 320 emails were sent to 320 different teachers (80 for each of the four test responses). From these, 150 teachers replied (47 per cent). The reply rate for the different responses corresponded to approximately half of the emails sent, with the exception of the response to *No evidence*, which was lower (see Table 2).

Table 2: Reply rate

Response	<i>n</i>	Reply rate (%)
Original	42	53
Another ethic	39	49
No evidence	33	41
No rhetoric	36	45

Because the reply rate is low, this must be taken into account when interpreting the results. However, it is a randomized sample, and there are no signs of a skewed distribution of the sample, such as towards urban or rural schools, schools with a high

or a low proportion of students with a nationality other than Swedish, or high- or low-performing schools. The only possible reason for an uneven distribution might be responses coming only from interested and committed teachers. However, a sample skewed on that basis would not bias the study.

Findings

The assessments for the different responses differed in that the *Original* response received the highest grades, which means that the working hypothesis was confirmed. The response *Another ethic* was awarded the lowest grades. As Figure 1 shows, only the *Original* response and the *No rhetoric* response got grades that corresponded to the knowledge requirements at the level of A. The responses *Another ethic* and *No evidence* were the only responses that received grades that did not meet the knowledge requirement at all, that is, the grade F.

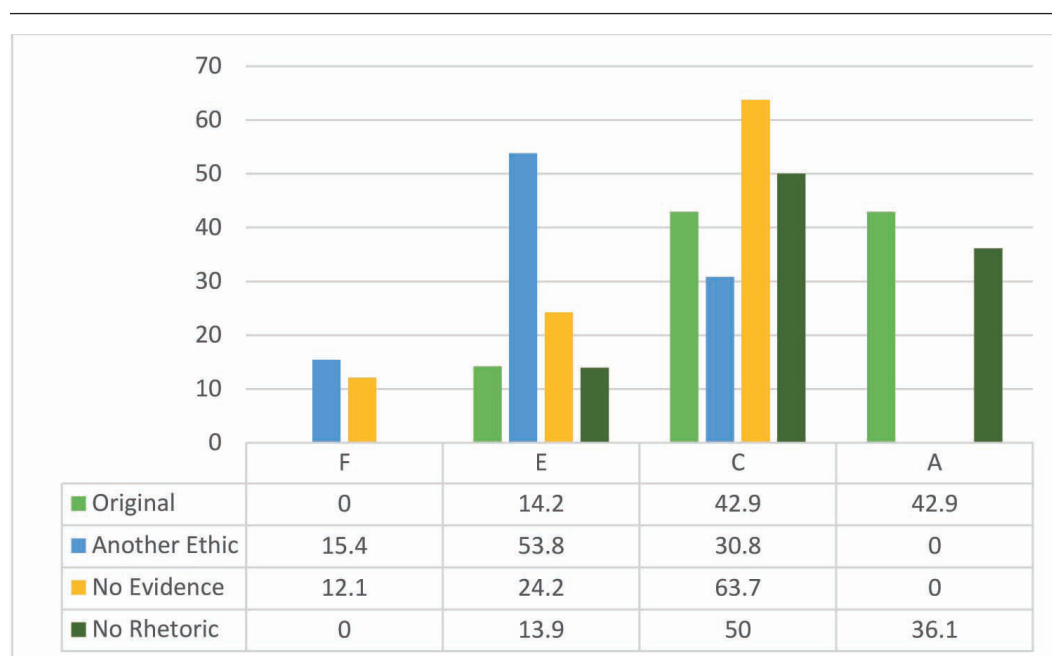


Figure 1: Grades awarded to each response (%)

Figure 1 reveals major differences between the assessments of the four responses. In order to determine whether there were statistically significant differences, a Kruskal-Wallis test was done. The result shows that there was a statistically significant difference between the assessments of the responses: $H(3) = 51.24, p < 0.000$.

Pairwise comparisons showed that there was no statistically significant difference between the assessments of the *Original* response and the response *No rhetoric*: $p = 0.729$. There were, however, statistically significant differences between the assessments for the *Original* response and the response *Another ethic* ($p < 0.000$), and between the assessments for the *Original* response and the response *No evidence* ($p < 0.000$). The assessments for the response *Another ethic* differed significantly from the response *No rhetoric* ($p < 0.000$) and from the response *No evidence* ($p = 0.015$). Assessments of the response *No evidence* and the response *No rhetoric* also differed significantly ($p = 0.002$).

Table 3: Pairwise comparisons, Kruskal-Wallis test

Pairwise comparisons	Test statistic (H)	Adj. Sig.	Hierarchical conditions
Another ethic – No evidence	-23,413	,015	<i>No evidence</i> assessed as the better
No rhetoric – Original	3,190	,729	No significative difference
No evidence – No rhetoric	-30,152	,002	<i>No rhetoric</i> assessed as the better
No evidence – Original	33,342	,000	<i>Original</i> assessed as the better
Another ethic – No rhetoric	-53,564	,000	<i>No rhetoric</i> assessed as the better
Another ethic – Original	56,755	,000	<i>Original</i> assessed as the better

Seemingly, the *Original* response satisfied all the criteria from the three dimensions, and this was also the response that was awarded the highest grades by the teachers. This was in accordance with the working hypothesis. The assessment of the response that was constructed to have a less convincing aesthetic dimension, *No rhetoric*, did not differ significantly compared to the *Original* answer. Both the response constructed to have a less convincing historical thinking dimension and the one with *Another ethical* content were awarded significantly lower grades than the *Original* response. However, the most startling result is that the response *Another ethic* received significantly lower grades compared to the *No evidence* response, although the latter was constructed without any informed references to the past and the present, which is required by the knowledge requirement. According to the knowledge requirement in the curriculum, the *No evidence* response cannot be regarded as a valid solution to the task, while the response *Another ethic* should be. The teachers seem to penalize a morally dubious use of history with lower grades, as compared to a use of history without factual knowledge.

Discussion

The purpose of teaching history is changing in many countries, and many researchers call for education that allows students to use history to develop historical thinking skills. In Sweden, curriculum developers have heeded this request, and have thus created a curriculum that is supposed to develop four competences that require the students to do history in different ways.

However, the Swedish history curriculum for the compulsory school, and other models developed both for teaching and for assessing students' use of history, focus on cognitive skills reminiscent of the scientific use of history. On the whole, ethical and rhetorical dimensions are forgotten in the models for assessment. Nevertheless, this small study suggests that teachers also take into account aspects other than the historical thinking dimension when assessing student responses. This is problematic from the point of view of students' rights and a fair and reliable assessment. It is important for students to know what is to be assessed before they begin to study history in the classroom. It is also important that assessment is not made of criteria other than those outlined from the outset of a task. Otherwise, the subject of history suffers from what Samuel Messick (1995) calls construct-irrelevant variance, which means that

for students to get good grades, they must use competences that are not supposed to be assessed. These may be competences that the student does not have or is not aware of, which may therefore lead to the student failing (ibid.). In the case of history, we should not assess ethical positions or rhetorical forms if we do not inform students about them or teach students to handle them. The hypothesis that students outside the majority culture are discriminated against by such a non-transparent teaching model is not unreasonable. Indeed, research demonstrates that such students have difficulties agreeing with schools' history teaching (Harris and Reynolds, 2014; Traille, 2007; Grever *et al.*, 2008; Epstein, 2009), as well as experiencing discrimination against them in grading processes (Cumming, 2000; Stobart, 2005; Baker and O'Neil, 1994a, 1994b). This discrimination may, however, be due to different views of how history can be used, emanating from people coming from different historical cultures, and not from the fact that the students represent a minority. If so, research about transparency in assessment processes in the classroom is highly needed to ensure justice. Without such transparency, it is difficult to use assessment to provide 'powerful messages about what learning is valued' (Peck and Seixas, 2008: 1018).

More studies are needed to investigate whether teachers' assessments of students' historical accounts also contain aspects of ethical and rhetorical dimensions. Research must also address in what ways rhetorical and ethical dimensions in history teaching are compatible with the historical thinking dimension. In *Historical Discourse: The language of time, cause and evaluation*, Coffin (2006) argues for the inclusion of the rhetorical dimension, while in *Teaching History for the Common Good*, Barton and Levstik (2004) outline ideas for history teaching based on a political/moral dimension, grounded in human rights.

Whatever the answer to the dilemma, transparent criteria are a question of justice for the students. As Stobart (2005: 275) notes, 'Fair assessment is a sociocultural issue rather than a technical one' (see also Sutherland, 1996; Gipps, 1999). Not showing students all of the dimensions that are assessed in their historical narratives is unfair, and it does not meet the requirements that Gipps (1999: 385) proposes:

The best defense against inequitable assessment is openness. Openness about design, constructs and scoring, will bring out into the open the values and biases of the test design process, offer an opportunity for debate about cultural and social influences, and open up the relationship between assessor and learner. These developments are possible, but they do require political will.

Notes on the contributor

Fredrik Leonard Alvé is a lecturer in history and history didactics at Malmö University, Sweden. He is a member of the team constructing the national test in history for the last year of compulsory education in Sweden. His research interests include historical consciousness, historical culture and narratives, history teaching and assessment. He recently published, 'Making democrats while establishing their historical consciousness: A complex task' in the *Journal of Historical Consciousness, Historical Cultures and History Education*, 4 (1), 52–67 (2017).

Appendix: The original response and the three constructed responses

The responses are marked up as follows:

- underlined text contains facts
- **bold** text contains historical examples
- blue text contains causal reasoning.

The teachers in the study did not have access to the marked-up text.

The responses are translated from Swedish, which means that the number of words and sentences is not congruent with Table 1. The number of facts, historical examples and instances of causal reasoning are intact.

The original response

I definitely think we should accept all refugees coming to Sweden. It's not that they do not fit, right? Just take the UK as an example! A smaller country than Sweden with more than 9 times larger population! Yes, space here is. And so it was with this thing with jobs ... There will inevitably be built up several new communities when the refugees come here. More people and communities = many more new jobs. It weighs up itself. Of course the costs will be great to build houses etc but most of it will of course be taken back by gaining more taxpayers. Furthermore – Have we not been taught that a human life is priceless? Just think, how many lives can we save from an otherwise grim fate. In addition to this, we would create great contacts and future cooperation possibilities that can be very good for Sweden. Now that we welcome so many people, we must of course also take advantage of their skills and learn from them, things that will help us towards a better development. We should learn from the 1600s when we recruited **Walloons** to work for us. It was them who led Sweden to faster and more positive industrial development. Or the **Germans** in the 1100–1500s. Was it not those who helped us on the road when it comes to business? We should learn from this. But also learn from our mistakes. How many people were not killed in **World War II**? Many of these people we could have been able to save, but once we decided to act, it was too late. We do not want to make that mistake again, right? Moreover, if we refuse to take in refugees, it can also be interpreted as we place ourselves on the wrong side. For sure, we all agree on what side we are on? I think we must open the border for the Finnish population and use boats and planes like an expedition with the goal to bring all the Afghan refugees here. We must save them!

The another ethic response

I definitely think we should not accept the refugees coming to Sweden. Just take the UK as an example! A smaller country than Sweden with more than 9 times larger population! Yes, space here is. And so it was with this thing with jobs ... There will inevitably be built up several new communities, when refugees come here. More people = many more who need jobs. So, clearly there will be some new jobs when these areas will be built. Not everyone can get a job from this. We also already have a high unemployment rate and a financial crisis. We should learn from the 1990s when Sweden went into a **recession**. We then got **big problems with racism**. We had the **Lazer man, the party New Democracy and refugee camps, which were burnt down**. We should learn from this. How many refugees weren't ill treated then? Many people experienced terrible things in Sweden. We do not want to make that mistake again, right? Furthermore – Have we not been taught that a human life is priceless? People

should not have to escape from the atrocities only to face new ones here. It is cruel to let in refugees here who then will be badly treated. That would give Sweden a bad reputation internationally, and our contacts would be damaged. We should also learn from our mistakes. When Visby was dominated by the **Hanseatic League**, the Swedes felt sidelined and there were conflicts. So it can be again. Also, if racism increases in Sweden, it will give strange signals abroad. For sure, we all agree on what side we are on? I think we have to be very careful about how many we can welcome in to Sweden without suffering from severe racism again. Sweden can not rescue everyone!

The no evidence response

I definitely think we should accept all refugees coming to Sweden. It's not that they do not fit, right? Yes, space here is. Just think, how many can lives we save from an otherwise grim fate. Now that we welcome so many people, we must of course also take advantage of their skills and learn from them, things that will help us towards a better development. We should learn from the 1600s. But also learn from our mistakes. How many people were not killed in **World War II**? Many of these people we could have been able to save, but once we decided to act, it was too late. We do not want to make that mistake again, right? Moreover, if we refuse to take in refugees, it can also be interpreted as we place ourselves on the wrong side. For sure, we all agree on what side we are on? I think we must open the border for the Finnish population and use boats and planes like an expedition with the goal to bring all the Afghan refugees here. We must save them!

The no rhetoric response

I think Sweden should receive those who come here. They will fit. Britain is smaller than Sweden, with more than 9 times larger population. There will be no problem with jobs. There will be built new communities when people come here. More people and more communities provide more new jobs. The costs will be great to build houses etc. Most of it will be taken back while the country gets more taxpayers. Everyone has the right to life. Many can be saved. Sweden would create great contacts and future cooperation. It may be good for Sweden. When the country is bringing people here it must take advantage of their knowledge. Learn from them. Things that will help Sweden to a better development. Sweden should learn from the 1600s when it recruited **Walloons**. It was them who led Sweden to faster and more industrial development. Or the **Germans** in the 1100–1500s. They helped Sweden. Sweden should learn from this. But also learn from mistakes. Many of the people who died during the **Second World War**, Sweden had been able to save. When Sweden decided to act, it was too late. If the country refuses to take in refugees, it can be interpreted as Sweden places itself on the wrong side. I think that the country should open the border for the Finnish population. Using boats and flights to bring all the Afghan refugees.

References

- Ahonen, S. (2012) *Coming to Terms with a Dark Past: How post-conflict societies deal with history*. Frankfurt am Main: Peter Lang.
- Alvé, F. (2011) *Historiemedvetande på prov: En analys av elevers svar på uppgifter som prövar strävansmålen i kursplanen för historia*. Lund: Lunds universitet.
- Alvé, F. (2017a) 'Making democrats while developing their historical consciousness: A complex task'. *Historical Encounters*, 4 (1), 52–67.
- Alvé, F. (2017b) *Tänka rätt och tycka lämpligt: Historieämnet i skärningspunkten mellan att fostra kulturbärare och förbereda kulturbyggare*. Malmö: Malmö Högskola.

- Alvén, F. and Zander, U. (2011) "'Och deras skott långt fjärran eko väckt": Historiebruk och The Minute Man'. *Historielärarnas Förenings Årsskrift*, 85–101.
- Ashby, R. and Edwards, C. (2010) 'Challenges facing the disciplinary tradition: Reflections on the history curriculum in England'. In Nakou, I. and Barca, I (eds) *Contemporary Public Debates Over History Education*. Charlotte, NC: IAP, 27–46.
- Baker, E.L. and O'Neil, H.F. (1994a) *Diversity, Assessment, and Equity in Education Reform*. Los Angeles: University of California.
- Baker, E.L. and O'Neil, H.F. (1994b) 'Performance assessment and equity: A view from the USA'. *Assessment in Education: Principles, Policy and Practice*, 1 (1), 11–26.
- Barton, K.C. and Levstik, L.S. (2004) *Teaching History for the Common Good*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Barton, K.C. and McCully, A.W. (2005) 'History, identity, and the school curriculum in Northern Ireland: An empirical study of secondary students' ideas and perspectives'. *Journal of Curriculum Studies*, 37 (1), 85–116.
- Bennett, R.E., Gottesman, R.L., Rock, D.A. and Cerullo, F. (1993) 'Influence of behavior perceptions and gender on teachers' judgments of students' academic skill'. *Journal of Educational Psychology*, 85 (2), 347–56.
- Berger, S. (2012) 'De-nationalizing history teaching and nationalizing it differently! Some reflections on how to defuse the negative potential of national(ist) history teaching'. In Carretero, M., Asensio, M. and Rodríguez-Moneo, M. (eds) *History Education and the Construction of National Identities*. Charlotte, NC: Information Age Publishing, 33–47.
- Blow, F., Rogers, R., Shemilt, D. and Smith, C. (2015) 'Only connect: How students form connections within and between historical narratives'. In Chapman, A. and Wilschut, A. (eds) *Joined-Up History: New directions in history education research*. Charlotte, NC: Information Age Publishing, 279–315.
- Brennan, D.J. (2008) 'University student anonymity in the summative assessment of written work'. *Higher Education Research and Development*, 27 (1), 43–54.
- Carretero, M. (2011) *Constructing Patriotism: Teaching history and memories in global worlds*. Charlotte, NC: Information Age Publishing.
- Carretero, M. (2017) 'Teaching history master narratives: Fostering Imagi-Nations'. In Carretero, M., Berger, S. and Grever, M. (eds) *Palgrave Handbook of Research in Historical Culture and Education*. London: Palgrave Macmillan, 511–28.
- Coffin, C. (2006) *Historical Discourse: The language of time, cause and evaluation*. London: Continuum.
- Crisp, V. (2008) 'Exploring the nature of examiner thinking during the process of examination marking'. *Cambridge Journal of Education*, 38 (2), 247–64.
- Crisp, V. (2012) 'An investigation of rater cognition in the assessment of projects'. *Educational Measurement: Issues and Practice*, 31 (3), 10–20.
- Cumming, J. (2000) 'After DIF: What culture remains?'. Paper presented at the 26th International Association for Educational Assessment (IAEA) Annual Conference, Jerusalem, 14–19 May 2000.
- Davidson, M., Howell, K.W. and Hoekema, P. (2000) 'Effects of ethnicity and violent content on rubric scores in writing samples'. *Journal of Educational Research*, 93 (6), 367–73.
- Diederich, P.B. (1974) *Measuring Growth in English*. Urbana, IL: National Council of Teachers of English.
- Eliasson, P., Alvén, F., Axelsson Yngvéus, C. and Rosenlund, D. (2015) 'Historical consciousness and historical thinking reflected in large-scale assessment in Sweden'. In Ercikan, K. and Seixas, P. (eds) *New Directions in Assessing Historical Thinking*. New York: Routledge, 171–82.
- Epstein, T. (2009) *Interpreting National History: Race, identity, and pedagogy in classrooms and communities*. New York: Routledge.
- Ercikan, K. and Seixas, P. (2011) 'Assessment of higher order thinking: The case of historical thinking'. In Schraw, G. and Robinson, D.R. (eds) *Assessment of Higher Order Thinking Skills*. Charlotte, NC: Information Age Publishing, 245–61.
- Ercikan, K. and Seixas, P. (2015a) 'Issues in designing assessments of historical thinking'. *Theory into Practice*, 54 (3), 255–62.
- Ercikan, K. and Seixas, P. (eds) (2015b) *New Directions in Assessing Historical Thinking*. New York: Routledge.
- Fajardo, D.M. (1985) 'Author race, essay quality, and reverse discrimination'. *Journal of Applied Social Psychology*, 15 (3), 255–68.
- Foster, S. (2011) 'Dominant traditions in international textbook research and revision'. *Education Inquiry*, 2 (1), 5–20.

- Gaddis, J.L. (2002) *The Landscape of History: How historians map the past*. Oxford: Oxford University Press.
- Gipps, C. (1999) 'Socio-cultural aspects of assessment'. *Review of Research in Education*, 24, 355–92.
- Grever, M. (2012) 'Dilemmas of common and plural history: Reflections on history education and heritage in a globalizing world'. In Carretero, M., Asensio, M. and Rodríguez-Moneo, M. (eds) *History Education and the Construction of National Identities*. Charlotte, NC: Information Age Publishing, 75–91.
- Grever, M., Haydn, T. and Ribbens, K. (2008) 'Identity and school history: The perspective of young people from the Netherlands and England'. *British Journal of Educational Studies*, 56 (1), 76–94.
- Harris, R. and Reynolds, R. (2014) 'The history curriculum and its personal connection to students from minority ethnic backgrounds'. *Journal of Curriculum Studies*, 46 (4), 464–86.
- Hogarth, R.M. (1987) *Judgement and Choice: The psychology of decision*. 2nd ed. Chichester: Wiley.
- Howell, K.W., Moore, E., Bigelow, S. and Evoy, A.M. (1993) 'Validity of authentic assessment'. *Diagnostique*, 19 (1), 387–400.
- Kahneman, D. (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kessler, E.H. and Wong-Mingji, D.J. (eds) (2009) *Cultural Mythology and Global Leadership*. Cheltenham: Edward Elgar Publishing.
- Kölbl, C. and Konrad, L. (2015) 'Historical consciousness in Germany: Concept, implementation, assessment'. In Ercikan, K. and Seixas, P. (eds) *New Directions in Assessing Historical Thinking*. New York: Routledge, 17–28.
- Körper, A. and Meyer-Hamme, J. (2015) 'Historical thinking, competencies, and their measurement: Challenges and approaches'. In Ercikan, K. and Seixas, P. (eds) *New Directions in Assessing Historical Thinking*. New York: Routledge, 89–101.
- Laville, C. (2004) 'Historical consciousness and historical education: What to expect from the first for the second'. In Seixas, P. (ed.) *Theorizing Historical Consciousness*. Toronto: University of Toronto Press, 165–82.
- Lee, P. (1991) 'Historical knowledge and the national curriculum'. In Aldrich, R. (ed.) *History in the National Curriculum*. London: Kogan Page, 39–65.
- Lee, P. (2004) 'Understanding history'. In Seixas, P. (ed.) *Theorizing Historical Consciousness*. Toronto: University of Toronto Press, 129–64.
- Lee, P. (2014) 'Fused horizons? UK research into students' second-order ideas in history: A perspective from London'. In Köster, M., Thünemann, H. and Zülsdorf-Kersting, M. (eds) *Researching History Education: International perspectives and disciplinary traditions*. Schwalbach am Taunus: Wochenschau Verlag, 170–94.
- Lee, P. and Ashby, R. (2000) 'Progression in historical understanding among students ages 7–14'. In Stearns, P.N., Seixas, P. and Wineburg, S. (eds) *Knowing, Teaching, and Learning History: National and international perspectives*. New York: New York University Press, 199–222.
- Lee, P. and Howson, J. (2009) "'Two out of five did not know that Henry VIII had six wives": History education, historical literacy, and historical consciousness'. In Symcox, L. and Wilschut, A. (eds) *National History Standards: The problem of the canon and the future of teaching history*. Charlotte, NC: Information Age Publishing, 211–61.
- Malouff, J. (2008) 'Bias in grading'. *College Teaching*, 56 (3), 191–2.
- Malouff, J.M., Stein, S.J., Bothma, L.N., Coulter, K. and Emmerton, A.J. (2014) 'Preventing halo bias in grading the work of university students'. *Cogent Psychology*, 1 (1), Article 988937, 1–9.
- Malouff, J.M. and Thorsteinsson, E.B. (2016) 'Bias in grading: A meta-analysis of experimental research findings'. *Australian Journal of Education*, 60 (3), 245–56.
- Messick, S. (1989) 'Validity'. In Linn, R.L. (ed.) *Educational Measurement*. 3rd ed. New York: American Council on Education, 13–104.
- Messick, S. (1995) 'Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into scoring meaning'. *American Psychologist*, 50 (9), 741–9.
- Nietzsche, F. (2010) *The Use and Abuse of History*. Charleston, SC: CreateSpace.
- Nordgren, K. and Johansson, M. (2015) 'Intercultural historical learning: A conceptual framework'. *Journal of Curriculum Studies*, 47 (1), 1–25.
- Peck, C. and Seixas, P. (2008) 'Benchmarks of historical thinking: First steps'. *Canadian Journal of Education*, 31 (4), 1015–38.
- Pingel, F. (2010) *UNESCO Guidebook on Textbook Research and Textbook Revision*. 2nd ed. Paris: United Nations Educational, Scientific and Cultural Organization.

- Ricoeur, P. (1984) *Time and Narrative: Volume 1*. Chicago: University of Chicago Press.
- Ricoeur, P. (1985) *Time and Narrative: Volume 2*. Chicago: University of Chicago Press.
- Ricoeur, P. (1988) *Time and Narrative: Volume 3*. Chicago: University of Chicago Press.
- Rosenzweig, R. and Thelen, D. (1998) *The Presence of the Past: Popular uses of history in American life*. New York: Columbia University Press.
- Rüsen, J. (2004) *Berättande och förnuft: Historieteoretiska texter*. Göteborg: Daidalos.
- Rüsen, J. (2005) *History: Narration, interpretation, orientation*. New York: Berghahn Books.
- Rüsen, J. (2007) 'Memory, history and the quest for the future'. In Cajani, L. and Ross, A. (eds) *History Teaching, Identities, Citizenship*. Stoke-on-Trent: Trentham Books, 13–34.
- Rüsen, J. (2011) 'Forming historical consciousness: Towards a humanistic history didactics'. In Nordgren, K., Eliasson, P. and Rönqvist, C. (eds) *The Processes of History Teaching*. Karlstad: Karlstads universitet.
- Seixas, P. (2000) 'Schweigen! die Kinder! or, does postmodern history have a place in the schools?'. In Stearns, P.N., Seixas, P. and Wineburg, S. (eds) *Knowing, Teaching, and Learning History: National and international perspectives*. New York: New York University Press, 19–37.
- Seixas, P. (2007) 'Who needs a canon?'. In Grever, M. and Sturman, S. (eds) *Beyond the Canon: History for the twenty-first century*. Basingstoke: Palgrave Macmillan, 19–30.
- Seixas, P. (2009) 'A modest proposal for change in Canadian history education'. *Teaching History*, 137, 26–32.
- Seixas, P. and Morton, T. (2013) *The Big Six Historical Thinking Concepts*. Toronto: Nelson Education.
- Semmet, S. (2012) 'Controversiality and consciousness: Contemporary history education in Germany'. In Taylor, T. and Guyver, R. (eds) *History Wars and the Classroom: Global perspectives*. Charlotte, NC: Information Age Publishing, 77–88.
- Shemilt, D. (2000) 'The caliph's coin: The currency of narrative frameworks in history teaching'. In Stearns, P.N., Seixas, P. and Wineburg, S. (eds) *Knowing, Teaching, and Learning History: National and international perspectives*. New York: New York University Press, 83–101.
- Shemilt, D. (2009) 'Drinking an ocean and pissing a cupful: How adolescents make sense of history'. In Symcox, L. and Wilschut, A. (eds) *National History Standards: The problem of the canon and the future of teaching history*. Charlotte, NC: Information Age Publishing, 141–209.
- Spear, M.G. (1984) 'The biasing influence of pupil sex in a science marking exercise'. *Research in Science and Technological Education*, 2 (1), 55–60.
- Stobart, G. (2005) 'Fairness in multicultural assessment systems'. *Assessment in Education: Principles, Policy and Practice*, 12 (3), 275–87.
- Stoel, G.L., Van Drie, J.P. and 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.
- Sutherland, G. (1996) 'Assessment: Some historical perspectives'. In Goldstein, H. and Lewis, T. (eds) *Assessment: Problems, developments and statistical issues*. Chichester: Wiley, 9–20.
- Swedish National Agency for Education (2011) *Curriculum for the Compulsory School, Preschool Class and the Recreation Centre 2011*. Stockholm: Swedish National Agency for Education.
- Traille, K. (2007) '"You should be proud about your history. They made me feel ashamed": Teaching history hurts'. *Teaching History*, 127, 31–7.
- VanSledright, B.A. (2011) *The Challenge of Rethinking History Education: On practices, theories, and policy*. New York: Routledge.
- VanSledright, B.A. (2014) *Assessing Historical Thinking and Understanding: Innovative designs for new standards*. New York: Routledge.
- Wertsch, J.V. (1998) *Mind as Action*. New York: Oxford University Press.
- Wertsch, J.V. (2000) 'Is it possible to teach beliefs, as well as knowledge about history?'. In Stearns, P.N., Seixas, P. and Wineburg, S. (eds) *Knowing, Teaching, and Learning History: National and international perspectives*. New York: New York University Press, 38–50.
- Wertsch, J.V. (2002) *Voices of Collective Remembering*. Cambridge: Cambridge University Press.
- White, H. (1973) *Metahistory: The historical imagination in nineteenth-century Europe*. Baltimore: Johns Hopkins University Press.
- White, H. (1978) *Tropics of Discourse: Essays in cultural criticism*. Baltimore: Johns Hopkins University Press.
- White, H. (1987) *The Content of the Form: Narrative discourse and historical representation*. Baltimore: Johns Hopkins University Press.