

- VanSledright, B. (2011). *The challenge of rethinking history education: On practices, theories, and policy*. New York, NY: Routledge.
- Wertsch, J. V. (1998). *Mind as action*. New York, NY: Oxford University Press.
- Wertsch, J. V. (2008). The narrative organization of collective memory. *ETHOS*, 36(1), 120–135.
- Wertsch, J. V. (2012). Texts of memory and texts of history. *L2 Journal*, 4(1), 9–20.
- Wilschut, A. (2009). Canonical standards or orientational frames of reference? The cultural and the educational approach to the debate about standards in history. In L. Symcox, & A. Wilschut (Eds.), *National history standards: The problem of the canon and the future of teaching history*, pp. 117–140. Charlotte, NC: Information Age.
- Wilschut, A. (2010). History at the mercy of politicians and ideologies: Germany, England, and the Netherlands in the 19th and 20th centuries. *Journal of Curriculum Studies*, 42(5), 693–723.
- Wineburg, S., Mosborg, S., Porat, D., & Duncan, A. (2007). Common belief and the cultural curriculum: An intergenerational study of historical consciousness. *American Educational Research Journal*, 44(1), 40–76.

CHAPTER 5

TESTING FRAME OF REFERENCE KNOWLEDGE IN NATIONAL EXAMINATIONS

Report on an Experiment in the Netherlands

Arie Wilschut

During the 1990s, national debates about “new history teaching” in the Netherlands—like those in many other western countries—led to political pressure demanding more “factual knowledge.” The commission that was given the task of developing a new national curriculum, however, did not produce the list of names and dates of important historical events and personalities that was widely expected, but put forward the idea of “frame of reference knowledge,” which could support historical orientation in time. Testing frame of reference knowledge meant developing a completely new paradigm for the national examinations. This chapter will report and evaluate aspects of this process of curriculum development.

HISTORY EXAMINATIONS IN THE NETHERLANDS

National written examinations for history at the end of secondary education have existed in the Netherlands since 1981. Secondary education in the Netherlands has three levels: a lower level, mainly preparing for vocational training; an intermediate level, preparing for professional education; and a higher level, preparing for academic studies. In 2011 the final examination was taken by 51% of Dutch students at the lower level, by 29% at the intermediate level and by 20% at the higher level. The number of students taking history at the lower level is comparatively limited. At the intermediate and higher levels, many more students take history for their final examinations. This chapter will only discuss the history examinations at the intermediate and higher levels (in the Netherlands known as HAVO and VWO respectively). This means that we will deal with about 77% of the total number of students taking history for their final examinations.

Since their introduction in 1981, the final examinations applied at the intermediate and higher levels have taken the form of two annually prescribed historical themes or topics, which are studied by all candidates. The themes or topics are selected two years prior to their actual implementation, so teachers will have enough time to prepare. The topics chosen are carefully outlined in an official description of subject matter in the form of a continuous text of several pages (up to 10 or more) with highlighted “important concepts” and references to relevant further reading. In the period between the publication of a newly selected topic and the actual examination about it, universities usually provide specialist in-service training courses for teachers, and educational publishers produce special study units for students preparing for the examination. Figure 5.1 shows a survey of the topics studied for the national examinations in the period 2006–2010, the period in which the pilot study reported in this chapter was conducted.

The selection of annually changing topics for national examinations, topics of a highly specific nature (Figure 5.1), suggests that it cannot be the content of the topics that is the most relevant subject matter. Otherwise, the selection of annually changing and very diverse content would not be justified for a national examination. So it has to be something beyond the content presented that is at stake, for example: the way a historical problem is approached, the historical meta-concepts applied and/or, the methods of historical thinking and historical reasoning required. After all, these are common to studying any historical topic. But an inspection of some of the questions asked at the examinations shows rather the opposite. The topic

Year	Topic 1	Topic 2
2006	Decolonization and Cold War in Vietnam.	From child to citizen. Educating people through schooling in the Netherlands 1780–1920.
2007	From child to citizen. Educating people through schooling in the Netherlands 1780–1920.	The colonial relationship between the Netherlands and the Netherlands East Indies.
2008	The colonial relationship between the Netherlands and the Netherlands East Indies.	“To war!” European wars 1789–1919. War as a societal phenomenon.
2009	“To war!” European wars 1789–1919. War as a societal phenomenon.	Dynamics and stagnation in the Dutch Republic, 15th to 18th centuries.
2010	Dynamics and stagnation in the Dutch Republic, 15th to 18th centuries.	Decolonization and Cold War in Vietnam.

Figure 5.1 Survey of topics for national examinations, 2006–2010.

about the Dutch Republic (examined in 2009, 2010 and 2011) for instance, was tested with questions like:

- In 1483 a war broke out between a number of North-German Hanseatic cities and Holland. Give an economic cause for the tension between the Hanseatic cities and Holland.
- Agriculture in Holland contributed significantly to economic growth in the Republic until about 1670. Between 1670 and 1740, however, it hardly contributed anything any more. Give an explanation for this change by:

Explaining why agriculture contributed to economic growth of the Republic until 1670,

And naming two problems of agriculture in Holland between 1670 and 1740 which ended this contribution.

(CITO, 2011a, questions 1 and 3)

One could assume that the above questions were seen by their composers as dealing with historical concepts like cause and change, but in fact remembering the correct and specific factual information was essential. Other questions gave the impression that more historical reasoning, such as “using evidence,” was required; but in several cases they also predominantly focused on the reproduction of factual knowledge. For instance, a

17th century text (converted into modern Dutch language) was presented in which the author explained that even if all the land in Holland had been used for growing wheat, and even if many more people had been employed in agriculture, the country could not possibly have fed its inhabitants. The accompanying question to be answered was:

(Use source 1) What is described by the author has been characterized as Malthusian tension by later scientists. Show that the author describes this phenomenon and then explain how the Mother Trade could take away a part of this tension. (CITO, 2011b, question 2)

Content elements like the causes of the tension between Hanseatic cities and Holland; the reasons why agriculture contributed to economic growth until 1670; the problems that occurred in agriculture between 1670 and 1740; what is meant by Malthusian tension and Mother Trade (the trade between Holland and the Baltic area, which provided Holland with wheat); were all expounded upon in the official description of subject matter and therefore also explained in the special study units provided by commercial educational publishers, with highlighted concepts like Malthusian tension. In this way they were transferred to the students to be reproduced on the examination. So the impression could arise that it was in fact the special content of the diverse topics presented annually for the examinations that was at stake, and not something like historical methods or historical thinking; and it was safe to assume, at least, that many students must have had this idea.

CRITICISM AND CHANGES

In the 1990s, the content of the history curriculum became the object of political and societal debate (Wilschut, 2012). As in many other Western countries, the dispute concentrated on the lack of traditional historical knowledge, the kind of knowledge that children supposedly used to acquire previously in schools (Phillips, 1998; Granatstein, 1998; Symcox, 2002). In the Netherlands, the situation around the final examinations added fuel to the flames. After all, why did students have to study strange and exotic topics in detail, instead of acquiring survey knowledge about their “important historical heritage?”

The publishing of a cartoon like the one in Figure 5.2 on the front page of the national quality newspaper *NRC Handelsblad* could be the result in 2004. In that year, one of the examination topics was: *The Magnifying Glass on Lancashire: Cotton and Society 1750–1850*.

Like in many other countries, the general impression given by critics in the debate was that modern educationalists had destroyed traditional history teaching. A government installed commission (consisting only of



Figure 5.2 Final Examination. “What do you know about the English cotton industry in 1850?” “Nothing.” “In that case I cannot vouch for your future!” *Source:* Cartoon by Kamagurka in *NRC Handelsblad*, 2004.

outsiders, not of history teachers or history educators) had already recommended “a new balance between knowledge, understanding, and skills” (meaning more survey knowledge and fewer skills) (CG, 1998, p. 27). A subsequent commission was given the task of developing a new curriculum that would contain more general survey knowledge, instead of concentrating on exotic topics and themes.

This second commission, the Commission on Historical and Social Education (CHMV), was to report in 2001. The author of this chapter was the secretary of this commission. It did not produce the generally expected list of names and dates to be remembered. The purpose of survey knowledge, said the commission, was to enable students to “orient in historical time” and thus to acquire a historical consciousness. Simply remembering names and dates could not support this. The report presented a proposal to help develop students’ chronological understanding via a curriculum of ten “easily remembered” eras, bearing “simple, associative names” (CHMV, 2001, p. 20, my translation). The commission was of the opinion that the choice of particular historical persons or events would be arbitrary and that a nationally prescribed program had to include only the general characteristics of eras to enable orientation in time (CHMV, 2001, p. 23). The

chronological framework was developed to enable historical consciousness and historical thinking, it was not intended to be utilized for the memorization of facts about (national) history. Therefore there was no reference to names, dates, significant events, or heroes.

The framework provided by the commission consisted of ten eras, which were to be utilized throughout all history curricula, from initial history teaching in primary education to the last grades of secondary education, including the examination program. Each era had an “associative name” which was intended to evoke a “sense of period.” Furthermore, for the examination program, each era was characterized by specific features that described it in more detail. (Figure 5.3 shows a survey of the ten eras, each provided with a specific emblem. As an illustration, the specific features of two of the eras are specified.)

The text of the proposed examination program, specifying what students were supposed to be able to do with their knowledge of eras, was as follows:

The candidate can, for each of the eras:

- Name the specific features indicated for the era
- Provide, for each of the specific features, an adequate example of an event, or phenomenon, or development, or of the acts or ways of thinking of a person, and use this example to explain the specific feature concerned
- Explain how knowledge of the era concerned can influence views of the present
- Explain how the significance attributed to eras depends partly on the time, position and circumstances in which people deal with the past. (Wilschut, 2007, p. 51)

This was to be the only way in which the historical content of the examination program was specified. Apart from this, students had to study “historical thinking,” but that was nothing new in this program: it had been part of the requirements for the examinations before. So the new program implied that, instead of studying specific content about two topics, with the help of a careful and extensive official description of subject matter with highlighted “important concepts,” students had to deal with a general outline of (mostly European) history described in general terms. This was meant as a frame of reference to contextualize specific events, phenomena, persons, or developments. As such, frame of reference knowledge is not the same as “big history” as it has been described in recent American and British publications. Big history is meant to be an overall view of, or narrative about, human existence (or sometimes even the existence of the universe) through the centuries (Dunn, 2000; Dunn, 2009; Howson & Shemilt, 2011). On the one hand, frame of reference knowledge is less comprehensive, but on the other hand it is more practical; it is a tool for

The Ten Eras	Examples of Specifications of Eras
 Era of Hunters and Farmers	<p><i>Specific features of Era 4—Cities and States:</i></p> <ul style="list-style-type: none"> • The rise of trade and crafts, providing the base for a revival of urban society. • The emergence of an urban citizenry and a growing autonomy of cities. • The conflict in the Christian world about the question whether a spiritual or a secular authority should have primacy. • The expansion of the Christian world, such as manifested in the Crusades. • The beginnings of national and centralized states.
 Era of Greeks and Romans	
 Era of Monks and Knights	
 Era of Cities and States	
 Era of Discoverers and Reformers	<p><i>Specific features of Era 8—Citizens and Steam Engines:</i></p> <ul style="list-style-type: none"> • The industrial revolution in the western world, providing the base for an industrial society. • Discussions about the “social issue”. • The modern kind of imperialism resulting from industrialization. • The emergence of emancipation movements. • Ongoing democratization, with more and more men and women taking part in the political process. • The emergence of sociopolitical movements: liberalism, nationalism, socialism, confessionalism and feminism.
 Era of Princes and Regents	
 Era of Wigs and Revolutions	
 Era of Citizens and Steam Engines	
 Era of World Wars	
 Era of Television and Computers	

Figure 5.3 The 10 era curriculum as proposed by the Commission on Historical and Social Education (CHMV) in 2001.

orientation such as European students need while working on the topics of history that they usually study.

In an examination program like this, the reproduction of specific factual knowledge would be out of the question. Students could expect to

be confronted on the examination with historical examples that they had never seen or heard of before, but which they would need to be able to contextualize with the help of their general frame of reference. It was obvious that this implied a considerable change compared to the examination practices that had developed since 1981. The annual selection of topics would be abolished. Instead, students' ability to apply historical thinking would be tested, part of which would involve using a general frame of reference described in ten eras.

QUESTIONS

Although there has not been extensive research on this point to date, the literature available suggests that students struggle to build up a usable and comprehensive framework of history (Lee, 2005; Shemilt, 2009; Lee & Howson, 2009; Kropman, Van Boxtel, & Van Drie, Chapter 4, this volume). Usually their general surveys of the past are fragmented and do not lead to any usable outline they can refer to. Furthermore, they seem to see the past as an "event-space" (Shemilt, 2009, p. 178) rather than viewing knowledge of the past as coherent narratives for which there may be alternatives. The first and most important question in the experiment with the new ten-era curriculum was, therefore: Is it possible for students to create a comprehensive framework that they can use to contextualize new and unknown cases of historical events, developments, and phenomena? And how does this influence their conception of comprehensive narratives of the past?

The other questions were of a more practical nature. How would teachers, who were used to carefully described examination topics and the official description of subject matter as their "guide books," be able to cope with an examination program formulated only in very general outlines? Could they do without the study units that were provided by commercial publishers each year, and that supported the studying of the specified examination topics? How were they going to deal with a kind of history teaching that concentrated on "orientational" framework knowledge instead of very specific factual knowledge?

To have an answer to these questions was essential to the successful introduction of the new curriculum, and even more so because national examinations were involved. Therefore it was decided to conduct a pilot study in a number of test schools with experimental examinations. This pilot study took place before a final decision about the national introduction of the new examination program, and was intended to inform the final decision.

THE PILOT PROJECT

The Netherlands Ministry of Education and Sciences supported the set up of a pilot in nine pilot schools. An experimental examination took place in these schools at the intermediate level in the year 2006. This meant that the students of these schools, instead of being examined on the two scheduled topics "Vietnam" and "From child to citizen" (see Figure 5.1), were questioned on their frame of reference knowledge about ten eras. Their examinations were valid and official, replacing the national examinations. For this reason, the project was led by a trio consisting of a test specialist from CITO (the Dutch National Testing Institute, which produces all the national examinations for all school subjects), a curriculum specialist from SLO (the National Curriculum Institute) and an academic expert on teaching and learning history from the Amsterdam University of Applied Sciences, who was also the general manager of the project (author of this chapter). The project was to start in 2004, so over the course of two school years students of the experiment schools were prepared for their examination in 2006.

The involvement of the CITO-specialist meant that the production of the experimental examinations was conducted in exactly the same way as the production of traditional examinations. The appearance of the experimental examinations was very similar to the appearance of traditional examinations: a booklet with (open) questions, and an accompanying booklet with source material. The key difference was in the content and nature of the questions. Examples of questions used in the experimental examinations were as follows:

- In 1792 the Marquis de Condorcet wrote: "We must assume that the progress of reason will keep pace with the progress of arts and sciences. That, for instance, the progress of medicine will eventually result in the annihilation of all contagious and transferable diseases and also of diseases which are the consequence of climate, nutrition and the burdens of daily labor. Is it therefore strange to suppose that the improbability of the fate of man knows no limits?"

The views of the Marquis de Condorcet are an example of a new way of thinking in the eighteenth century. Name a characteristic of this new way of thinking and explain how this is manifest in the views of the Marquis. (CITO, experimental examination, intermediate level, 2006, question 16)

- Congressman Robert Winthrop from Massachusetts delivered a speech to the American Congress (= parliament) in 1846 about a conflict with the British on the Oregon Territory: "There is one element in our title, however, which I confess that I have not named, and to which I may not have done entire justice. I mean that new revelation of right which has been designated as the right of our manifest destiny to spread over this

whole continent. It has been openly avowed in a leading administration journal that this, after all, is our best and strongest title; one so clear, so pre-eminent, and so indisputable, that if Great Britain had all our other titles in addition to her own, they would weigh nothing against it. The right of our manifest destiny! There is a right for a new chapter in the law of nations; or rather in the special laws of our own country; for I suppose the right of a manifest destiny to spread, will not be admitted to exist in any nation except the universal Yankee nation!"

Robert Winthrop's plea, based on the idea of a manifest destiny, can be seen as nationalistic. Explain why. (CITO, experimental examination, higher level, 2010, question 17)

These questions were explicitly designed to test the application of conceptual or framework knowledge. In the case of these two examples the corresponding texts in the examination program were: "rational optimism and enlightened thinking applied at all levels of society: religion, politics, economy and social relations" and "the rise of the sociopolitical movements of liberalism, nationalism, socialism, confessionism, and feminism." This meant that students confronted with these questions were not supposed to reproduce specific knowledge about the views of the Marquis De Condorcet about the progress of medicine or 19th-century American ideas about a manifest destiny; they had to apply their general knowledge about the Enlightenment and 19th-century nationalism to these examples. During their preparation for the examination, they had practiced this application of knowledge using concrete examples, but these differed between the schools depending on the teachers' or students' choices. So they probably would not have heard about the Marquis de Condorcet, and definitely not about Congressman Winthrop nor the idea of a manifest destiny for Americans. The examination makers did not even assume that they knew that the American Congress was composed of representatives of the American people, and added an elucidation (Congress = parliament) to make sure that this information was available. To be able to answer these questions, students had to rely on their general outline knowledge of all of the ten eras ranging from prehistory to the present, because they did not know which topics would be presented to them in the examination questions.

The schools involved in the study had been acquired via an open call in the Dutch journal for history teachers. This meant that the teachers who agreed to take part were positively inclined toward the new examination program. Of course this was a setback for the objectivity of the inquiry, but no other option was available. After all, it would have been a hazard for the students involved to have been exposed to an experimental examination with an unwilling teacher who had been forced to do this. Whole schools took part however, which meant that complete teams of history teachers

had to agree to take part, which may imply some difference of opinion and differing degrees of enthusiasm for the new approach.

Some 700 students in nine schools took part in the experiment. The schools were spread geographically from north to south and east to west of the Netherlands; some schools were in rural districts and some were in cities; however, the intake of the schools was largely middle class and white. The first stage of the pilot took place only at the intermediate level, because the Ministry of Education and Sciences took the view that a successful experiment at that level would guarantee that the new examination would also be possible at the higher level. By limiting the experiment, they hoped to save some money. After the report on the first experimental examination in 2006 was published, however, the Dutch Association of Teachers of History insisted that an experiment also should be held at the higher level, because it had some doubts whether more demanding examination questions (requiring a higher level of thinking) would be possible if they were based only on an outline description of subject matter. This higher-level experimental examination took place in 2010. In the end, experimental examinations were held at nine schools at the intermediate level in 2006, 2007, 2008 and 2009; and at both intermediate and higher levels in 2010. The research reported here is based on the evaluations of the intermediate level examination in 2006 and both the intermediate and higher level examinations in 2010.

During the course of the project, the project team consisting of the three specialists mentioned above and at least one representative teacher from each school met regularly to discuss the development of the teaching and the experiences of teachers and students. (In addition, a number of experiments were conducted that cannot be reported here, due to constraints of space.¹) In this way, a lot of information was collected that could be used for the project's reports, which were published in 2006 and 2010 (Wilschut, Boom, Van Boxtel & Van Drie, 2006; Wilschut, Boom & Van der Kaap, 2010). The main parts of the reports were based on a formal collection of data, quantitative as well as qualitative. A synopsis of these results will be presented in the next sections.

METHOD OF QUANTITATIVE RESEARCH

The quantitative part of the inquiry consisted of a questionnaire of 27 items to which students could react via a 5-point Likert-scale from (strongly) disagree to (strongly) agree (see Appendix). The questionnaire was composed in 2006 by colleagues from the University of Amsterdam who were not involved in the examination project.² The same questionnaire was used again in 2010. The questionnaire was distributed among all the students taking part in the experiment,

but not all of them responded. They were to answer the questions immediately after they had finished their written examination work. In order to compare the results with results from regular, non-pilot students, a comparison group was created of five schools in 2006 and three different schools in 2010. The same questionnaire was distributed among students of the comparison group in 2006 (students doing their examination about the topics Vietnam and from child to citizen, see Figure 5.1) and among students of the comparison group in 2010 (students doing their examination about the topics dynamics and stagnation in the Dutch Republic and Vietnam, see Figure 5.1). These students were also asked to complete the questionnaire immediately after they had finished the examination. Table 5.1 shows the number of questionnaires that were included in the research in 2006 and 2010. The Appendix shows the 27 items of the questionnaire and the mean results of the experimental groups and the comparison groups. By means of shading, it has been indicated which differences were statistically significant between the experimental groups and the comparison groups.

A factor analysis was executed, which exposed five factor components, explaining 46% of variance in total. See Table 5.2.

TABLE 5.1 Number of Questionnaires in 2006 and 2010

	2006	2010	
	Intermediate Level	Intermediate Level	Higher Level
Experimental Group	561	531	316
Comparison Group	299	111	100

TABLE 5.2 Factor Loadings of Questionnaire Items

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Explained Variance	22.7%	9.2%	5.2%	4.8%	4.1%
Item	26 0.677	10 0.630	08 0.793	22 0.750	18 0.759
Numbers and Factor Loadings	27 0.664	03 0.610	09 0.749	21 0.745	19 0.664
	14 0.628	02 0.597	07 0.697	23 0.545	20 0.635
	25 0.527	12 0.583	01 0.392	24 0.538	15 0.418
	10 0.373	11 0.531		27 0.363	
	24 0.332	04 0.499			
	11 0.315	01 0.384			
		06 0.358			
		07 0.305			

If only the items with factor loadings larger than 0.5 are included, the following five factors emerge from the data:

- Factor 1 (items 14, 25, 26 and 27): History considered an important and interesting subject.
- Factor 2 (items 2, 3, 10, 11 and 12): Examination considered suitable and adequate as a test.
- Factor 3 (items 7, 8 and 9): Examination considered the right kind of challenge to show abilities.
- Factor 4: (items 21, 22, 23 and 24): Subject matter considered useful and applicable.
- Factor 5: (items 18, 19 and 20): The extent to which a usable framework of contextual knowledge of history had been acquired.

Variables for these five factors were computed by adding up the scores of the questionnaire items mentioned for each factor, so factor 1 could assume values between 4 and 20, factor 2 between 5 and 25, factor 3 between 3 and 15, factor 4 between 4 and 20, and factor 5 between 3 and 15.

QUANTITATIVE RESULTS

The mean factor values for the experimental groups and comparison groups were compared, and by means of *t*-tests it was made clear which differences were statistically significant (see Table 5.3). One factor displayed statistically different values in all of the three instances (factor 5: students' ideas about the extent to which a usable framework of contextual knowledge of history had been acquired). This implies that there was a consistent opinion among students that the experimental program provided a usable framework of contextual knowledge to a greater extent than the traditional program. Considering the fact that the new program dealt with the framework of 10 eras, and the traditional program with two selected themes, this result is not surprising, but it can be interpreted as a confirmation that one of the important aims of the new program had been achieved.

The other three statistically significant differences showed up in one of the instances in 2010. At the intermediate level, factor 1 (history considered an important and interesting subject) was significantly stronger in the experimental group, while factor 2 (examination considered suitable and adequate as a test) was significantly stronger in the comparison group. At the higher level, factor 3 (examination considered the right kind of challenge to show abilities) was stronger in the experimental group, though only at the significance level of $p < 0.05$. So there seems to be some indication that students in the experimental groups considered history more

TABLE 5.3 Mean Factor Values of Experimental Groups and Comparison Groups, 2006 and 2010

	Experimental Group	Comparison Group	Significance Level (p)
2006: Intermediate Level			
Factor 1	13.0	13.4	0.230
Factor 2	15.8	16.2	0.051
Factor 3	9.6	9.5	0.476
Factor 4	11.4	11.8	0.055
Factor 5	10.8**	9.7	0.000**
2010: Intermediate Level			
Factor 1	13.3**	12.5	0.005**
Factor 2	15.4	16.2**	0.009**
Factor 3	9.2	9.4	0.308
Factor 4	11.8	11.8	0.867
Factor 5	11.1**	10.1	0.000**
2010: Higher Level			
Factor 1	14.3	14.8	0.166
Factor 2	17.2	17.1	0.673
Factor 3	10.2*	9.7	0.020*
Factor 4	13.0	13.1	0.813
Factor 5	11.7**	10.7	0.000**

** Means: difference is statistically significant at level $p < 0.005$

* Means: difference is statistically significant at level $p < 0.05$

Note: In cases of significant differences, the larger values are printed in bold.

important and interesting than members of the comparison groups; and that they thought their exam was a challenge to show abilities. On the other hand, the experimental groups may have tended to see their examinations as a more adequate test, which may be ascribed to the fact that they studied specific subject matter they were questioned on directly in a recognizable manner; while the experimental students were confronted with unknown examples, usually not referring to the specific subject matter they had dealt with in their history lessons.

If we look at the individual questions (see Appendix), two questions stand out with relatively large and statistically significant differences between the experimental group and the comparison group in all three instances. Question 18 clearly shows that the experimental group students felt that they had acquired a better survey from antiquity to the present than the other students. A result like this was to be expected, given the composition of the curricula, and is also consistent with the results of the analysis of factor 5 described above. Question 5 indicates that questions in the experimental

examination looked more similar to each other than the questions in the traditional examination. Perhaps this means that there were fewer “surprises” and unexpected irregularities for the experimental group students.

Questions with statistically significant differences in two of the three instances were 17, 1, 12, 14, and 16. Question 17 is the most important one in this category, because it meets the higher level of significance ($p < 0.005$) in two instances. The results of this question may imply that experimental group students considered the subject matter they studied more useful and relevant, because they were less inclined to forget it. Question 12 and 16 produced significantly higher results in two instances in the comparison groups ($p < 0.005$ and $p < 0.05$). The implication of question 12 results could be that the congruence between the content of history lessons and the examination questions was greater in the traditional program, because it dealt with two narrowly described themes.

Question 16 shows that in two cases, the comparison groups thought that history is a difficult subject to a greater extent than the experimental groups. This result is somewhat difficult to interpret, because studying two precisely outlined themes may also be considered easier than studying a general frame of reference. Anyhow, the experimental groups' results shows that studying frame-of-reference knowledge does not make history more difficult.

Questions 1 shows contradictory results. In 2010, the higher-level, experimental-group students found the examination easier to do than the comparison group ($p < 0.005$), but the intermediate group found the opposite ($p < 0.05$), while no significant difference was indicated in 2006. The explanation may be that intermediate-level students found application of knowledge to new situations comparatively difficult, while higher-level students were more confident in applying knowledge. Some teachers, however, thought that this result indicated that questions in the experimental examination were “too easy” for the higher level. If this had been the case, the general level of performance of students doing the experimental examinations would also have to be considerably higher, but the results did not show this effect. (Table 5.4 shows the average performance of the experimental-group students and of the students doing the traditional examination). In

TABLE 5.4 Average Performance in 2006 and 2010^a

	2006	2010	
	Intermediate Level	Intermediate Level	Higher Level
Experimental Group	6.1	6.0	7.1
All Other Students	6.7	6.6	7.0

^a In the Netherlands student performances are measured on a 10-point scale with one decimal. Performances in the range 1–5.4 are considered unsatisfactory (fail) and performances in the range 5.5–10 are considered satisfactory (pass).

this case the table shows the average results of *all* experimental group students compared to the average results of *all* other students, nationwide.

In 2010 the levels of performance at the higher level were largely comparable in the two groups of students. The levels of performance at the intermediate level were considerably lower in the experimental groups. Based on these data and the results of question 1 in the questionnaire, we can conclude that the intermediate-level students found the experimental examination more difficult, while the higher-level students of the experimental group found it comparatively easier to perform at a level which was adequate for their school type.

Question 14 shows that in 2006 the comparison group at the intermediate level found history lessons more interesting than the experimental group ($p < 0.05$); but in 2010 the results at the intermediate level were opposite ($p < 0.005$). This may have been due to the different examination themes selected in these two years in the comparison groups (while the program in the experimental groups remained the same).

A question *not* showing significant differences in 2010 (both levels) is number 3. This is noteworthy, because in 2010 both groups thought that the most important aspects of subject matter were covered, while one would have expected this to be more the case with specified topics. Apparently, in 2010, students thought that generally-described subject matter could be covered adequately. In 2006 the experimental group had been neutral on this point, while the comparison group had scored more positively.

Question 15 shows that students in both groups in 2010 thought that they had learned quite a bit of new subject matter in senior secondary education, while in 2006 this was not the case, but more so in the experimental group than in the comparison group. This is noteworthy, because in junior secondary education the same 10 eras had been studied, as in senior secondary education new style. So it could be expected that the experimental group students would think that they learned "a lot more of the same stuff," while the other students studied specific and not very usual themes. The results of this question are not unambiguous, but at least we cannot conclude that the experimental group students felt that they only learned "a lot more of the same stuff."

Other differences between the experimental groups and the comparison groups were either too small or too insignificant to draw any conclusions, or the differences showed up only in one of the three instances. The conclusion based on these aspects of the questionnaire could be that the experimental examination was as satisfactory as the traditional examination in many respects, and that some of the main objectives of the new style examination had been achieved.

METHOD OF QUALITATIVE RESEARCH

In all but one of the project schools, semi-structured interviews were conducted with two of the teachers involved in the project; and in one case, the interviews were conducted with three of the teachers. A total of 17 teachers were interviewed in group sessions of two or three interviewees. The interviews were held after the teachers had completed grading of the students' work.³

Further qualitative data were collected through interviews with students who had taken part in the examination. We selected two to three students in each school who had taken part at the intermediate level in 2008, and then continued at the higher level in 2010. These students had a double experience with the new style examinations and could assess the difference between the levels. One of the aims of the interviews was to check whether the new style examination was adequate for the higher level, as was explicitly asked by the Dutch Association of Teachers of History, which had insisted on extending the pilot project to the higher level. A total of 20 students were interviewed, in groups of two to three students.

The interviews were recorded, transcribed, and then analyzed, after which conclusions were summarized.

QUALITATIVE RESULTS

Teacher Interviews

The first questions in the interviews with teachers were about the level of the examinations, especially the higher level. Did the teachers think that the examinations met the standard normally expected in higher-level examinations?

At two of the schools, teachers thought the higher level was clearly visible in the examination. "One can see that they have tried to really make it a higher level examination." To substantiate this view, these teachers referred to the length and difficulty of the source material employed in the examination, as well as to questions that were clearly aiming at students' understanding of history. At two other schools the teachers thought that the high-level examination was completely at the correct standard. At the remaining four schools however, there was some doubt about this; the main point of criticism was that too little reproduction of knowledge was required.

For the intermediate level, this is all right, but students at the higher level can be challenged more as it comes to their knowledge. I would have expected

more questions in which students were asked to compare historical situations with each other. That is what we have practiced in the lessons.

Some teachers at these four schools thought that more factual knowledge should have been specified in the examination program. "That would take away my main point of criticism." At three of the schools, teachers thought that the difference between the intermediate and the higher level was too small. They thought that the higher-level students should have met higher demands, especially in the way their answers were formulated. This should have been made visible especially in the scoring instruction, they thought. Similar questions should have been answered differently at the higher level. The scoring instruction as it presently was, did not provide for this.

Asked whether the higher-level examination did justice to the students concerned, there were some doubts at three of the schools. These teachers thought there was insufficient opportunity for high-level students to excel. According to them, the examination resulted in a narrowing of the range of achievement shown in student responses. At six schools however, teachers thought that the examination did justice to the level of these students.

They have learned something and that became manifest. Students were also of this opinion: a good examination, very doable, but certainly not too easy. One has to have the skills, one has to be able to think critically, to combine and to analyze things. So yes, I think justice has been done to the higher level students. It hasn't been all that simple at all.

If many students give a correct answer, does it mean that the question was too easy, or just that students had been prepared well for the examination, one teacher wondered. The assessment of a level of difficulty is sometimes very different from the perspectives of teachers and students. "For us as professionals, certain things are completely self-evident, but not for students. They may sometimes mix up complete periods."

The next question was what, according to the teachers, had been tested in the examinations: knowledge, skills, understanding, and/or historical thinking? The answers were rather diverse. Teachers in two schools thought that the complete program had been dealt with in a balanced way. "The whole range of eras and their specific features was dealt with very well. Also the part of historical thinking was adequate." At four other schools, teachers thought the main part of the examination was understanding and application of framework knowledge. Teachers of the remaining two schools thought the questions were only about the skills of historical thinking. They thought "everything except knowledge" played a role. But a teacher of one of the schools where the main opinion was that the examination was well balanced, played down this kind of opinion. "They are not always aware how much knowledge is needed to formulate a well substantiated answer."

The teachers were then asked about the way they had prepared their students for the examination. Did they know how to deal with the novelty of the curriculum? One teacher answered that he had been completely puzzled for a long time, especially about what kind of historical knowledge he should discuss. What could be expected of the examination? But all the other teachers reacted positively to this question. Their students were not taken by surprise in the examination, they said. "I have spoken to a few of them right after the examination session. They said they expected something like this more or less." Another teacher was "pleasantly surprised" when he saw the examination questions. "I thought: I have discussed all of this exactly in my teaching. This cannot go wrong." There was enough time in the lessons to deal with the extensive range of contexts of the 10 eras, especially at the higher level, where students had more history lessons than at the intermediate level. "It was wonderful to have the time to discuss all this, and my students thought it was interesting. All these histories, they tuck into them." All in all, the preparation to the examination appeared to be well adapted to the examination itself, according to the large majority of the teachers.

Another question was whether the grading of the examinations caused problems. After all, the lack of precisely described examination topics and the liberty to deal with very diverse concrete examples at the diverse schools, could result in student answers that would vary more than in the traditional examinations, especially where students were asked to substantiate something with "an example." In those cases, the CITO scoring instruction could tell the teachers no more than that "a correct example" should be scored at a particular level.

In four of the schools, teachers replied that there had not been any problems worthy of mentioning, and that the grading of the new style examination was even easier than the grading of traditional examinations. In two schools, teachers replied that more time could have been involved, because now and then an answer given by students would have to be checked in authoritative history books. But this was by no means insurmountable. In one school, teachers took the view that the variety in students' answers was such that problems in grading in the first instance and second grading had increased. The general tendency in the answers was that grading of the examinations was not more complicated than usual.

The final question was whether teachers thought that the new curriculum should be introduced nationally. In one school, there were serious doubts about this. Teachers in this school would have preferred to have a diachronic theme annually selected for the examination. One of the main reasons for this was that the performances of students at the pilot examination at this school were worse than expected. Especially in cases of re-examinations, students had less chance to quickly review the subject matter, as

had been the case when they had studied two topics based on special study units, with lists of important dates, persons and important concepts. The repair work to a bad result was harder with this new style history examination, and this caused the school management to criticize the history teachers.

There was one more school where one teacher favored a selected diachronic theme only for the higher level; for the intermediate level he thought the new program was adequate. The other 13 teachers were in favor of implementing the new program nationally, even if four of them had some criticism.

It has to go on anyhow. It is no option to go back. The program is more interesting for us and also for the students. Not all the expectations have been realized, but this is no reason to step back. We must concentrate on optimizing the new style examinations.

In the other schools, the nine remaining teachers were more unambiguous:

I am very enthusiastic about these changes. There are enough handles for students to hold on to. And students think this way of studying history is very useful.

I would deplore a return to the annual themes. Because with this program, one is really dealing with history. Otherwise one only has two topics which are just singled out from history and which have no relationship between them whatsoever.

Introduce the new program tomorrow! I consider the fact of starting in Pre-history and continuing until the present an enrichment for the education of students. They can do far more with this in their futures than with the examination programs such as they have been since 1981. So I am all for it.

They understand a lot more if they encounter something. Even if they do not know the exact details, they can contextualize things. I find it inconceivable that some people want to retain the old situation. I don't understand that at all. Unknown, unloved, that's what it is.

Student Interviews

The first question to the students was, whether they thought the examination was at the right level, too easy, or too difficult. None of the students thought that the examination was too difficult. We have to consider here that only students of the higher level were interviewed. Fourteen students thought that the examination had the right level. For instance, Marije replied:

With this examination, one really had to think. For instance there was one question about the Enlightenment and those articles from America. The question was to name the Enlightenment ideas and relate them to the texts. One had to think really deeply for that.

Even if they thought the level adequate, six of the students were disappointed that they could not display more of their knowledge of all the history that they had been studying. They thought it was strange that topics that according to them were "clearly important" had not been examined, such as the French Revolution and the World Wars. A general point of criticism, shared by 16 of the students, was the number and length of texts and sources that were employed in the examination. Much time was needed to read and comprehend all these. Eight of these students thought that the use of sources implied that someone clever could easily distil the right answers from the texts, which implied that there was too little distinction between those students who had studied well and those who had not. For example, Suzanne said:

Just a few more questions without texts or sources would make more sense to me. Otherwise there are always those people who are good at history, but not good at applying their knowledge to texts and that sort of thing.

Six students thought that the examination was actually too easy. But when asked whether this was also apparent from their high performances, they had to admit that this was not the case. The results actually fell short of their expectations.

Students studying framework knowledge necessarily have to be confronted with the same framework repeatedly. Did the students find that this became boring? Four students thought it was, but only to a certain extent:

Repeating is always useful of course, but at a certain moment we had to do a test paper on the eras 1 to 7, and then we were really like: now we are dealing with era 1 to 3 for so-and-so many times. Now we have enough of it.

Twelve students however, stressed that the advantages of repeating the framework outweighed the disadvantages, for example Lynn:

I certainly think that repeating the framework is an advantage. I think it would have been easier if we would have had the same framework also in junior secondary education. Because at that time, history to me looked more like one big mess, without any idea of the dimensions of it.

(For these students, the 10 eras had not yet been employed during their junior secondary education, due the moment in time when this curriculum

had been introduced). And Anne: "I like to learn history this way very much. The old division of periods we used to have in junior secondary was less surveyable than this one." According to Rutger, one easily forgot what was studied at school, but with the 10 eras, that was different: "Because one is repeating and repeating, I think it sticks to mind a little better. Not all the details any more, but one knows more or less what happened and in which order."

How useful was the framework, in the opinion of students? All of the 20 students questioned agreed that it was useful.

Anne:

I prefer to have global knowledge of the whole past instead of two topics that one rote learns and forgets within a few weeks. For instance, when you are studying at university, and you are confronted with something from prehistory or the Middle Ages, you also know something about that, and not only about recent wars.

Marije:

By this method I have acquired a clear image of history, and if for instance you have only studied the Cold War or the Netherlands East Indies, then you are really missing something. Sometimes, when I am watching the news, I think: hey, this is like something that happened in that time too. So you can explain things historically.

And Lynn:

My mother only studied Vietnam, and so she knows everything about that, but some time ago I asked her something about another topic and she really didn't know anything about it; really super stupid remarks she made . . . So I think it is all right to learn more general knowledge. What use is it to me if I can explain how Ho-Chi-what's his name did exactly this or that? With this history we learn just what we need for general insights, one understands how things have originated, and what relations in the world are like. I think that is very good for your later life.

Six students said the framework gave them opportunities to draw long lines through the centuries. Two voluntarily gave a substantiation of this statement by explaining in a short overview what was, according to them, the relationship between the empire of Napoleon and the outbreak of the First World War. Another student mentioned a survey of the phenomenon of "politics" through all the eras. "I like to have this kind of overview very much."

Was the framework fun to learn? Ten out of 20 students thought it was more fun than studying special topics for a longer period of time.

I think it is a bit of a shame to study a topic you don't like for two years, while there is so much more to history that you might find interesting. So I think it is a privilege to have more general knowledge instead of knowing very much about one thing.

Another student made the comparison with what other students had to learn at other schools:

That other program, I don't like that very much. I have seen the books of my old friends at other schools, but what we had to do was more fun to learn. In principle all of our class thought this was fun to learn, everyone was content with it. You can do much more with it, I think, and you are not studying one topic all the time. If you are finished with something, you go to something else.

The final question was whether students would advise the government to introduce the new curriculum or not: 17 out of 20 students reacted positively to this question. So it is perhaps illustrative to quote one of the exceptional students who had some doubts:

On the one hand, I think it is a good idea to have general knowledge of all history. On the other hand, I think it is quite a task for students to master all this. At least, in my case it was like that. I struggled a bit with this idea of general features that one has to apply to special cases. That is no fun way to learn for me. So just give me two topics that I can study thoroughly and ask me questions about those. But that is a purely personal point of view, because I learn better that way. If I look upon this question from the point of view of national interest, then I think this new curriculum should be introduced.

CONCLUSION

Is it possible for students to create a comprehensive framework which they can use to contextualize new and unknown cases of historical events, developments and phenomena? In general, the answer to this question appears to be positive, based on the results of this experiment. The level of performance of students taking the experimental examinations did not differ greatly from the level of performance of students taking the traditional examinations in the same years, at least at the higher level of secondary education. At the intermediate level, students seemed to have more difficulty in building up and applying a framework. The concrete examples students were confronted with in the experimental examinations were unknown to them and had to be related to specific features of 10 eras and explained in this way. On the whole, students at the higher level were able to do this well enough to perform at a level comparable to the level of the traditional examinations. Intermediate-level students scored a little lower, but still at an acceptable level.

Did the new program influence the students' concepts of comprehensive narratives of the past? As far as could be established with the data available, the experimental group students were largely able to describe the use of a framework to which specific phenomena could be related. They conceived

of the framework, however, still largely as “knowledge,” important knowledge to display in an examination. They did not think of knowledge of the past as a narrative for which there might be alternatives. For example, the students who were disappointed not to be questioned about “important events” like the French Revolution and the World Wars, expressed the opinion that these events were important “as such.” They would probably have difficulty with a perspective that would make these events seem less important. What does seem to have been accomplished is the insight by students that history can be useful to understand things that one has not been confronted with before. The inclination to forget the subject matter studied was less with the experimental-group students than with the comparison-group students.

In addition, some teachers continued to see history as knowledge about the past and continued to define a level of performance as the ability to reproduce knowledge and do something with it, albeit in this case largely with framework knowledge. The criticism by some teachers inside the project, and certainly of some commentators from outside, was that the general framework did not provide enough substantive knowledge to study. Mastery of more substantive knowledge would, in their view, imply a higher level of historical learning. This was also the implication of a critical remark like “these examinations were about all but knowledge” and the remark of one teacher that more substantive knowledge in the examination program would have taken away his main criticism. So the project did not succeed completely in transferring the concept of frame of reference knowledge to all the teachers involved; some stuck to the idea that reproduction of large quantities of complicated looking substantive knowledge represented a high level of historical learning. Some teachers within the project, however, took a different stance: they said that more could have been done by comparing cases from two or more eras, for example.

As to the more practical questions related to the project, teachers seemed to be able to cope well with a novel and not too clearly defined situation. Of course we have to take into account that teachers participated voluntarily in the project and were predominantly positively inclined toward the new curriculum. The positive impression that arose from the teacher interviews probably has to be put into perspective. But at least, a different style of teaching history had been shown to be feasible.

The matter of substantive knowledge continued to dominate the discussions after the completion of the project. It was decided to adopt the new curriculum, but the syllabus has been extended by the addition of more specific substantive knowledge pertaining to three (intermediate level) or four (higher level) “historical contexts”⁴ about which factual questions can be asked. These historical contexts are meant to be permanent additions to the framework of eras, and not annually changing topics. In this hybrid form it became acceptable enough to prescribe the new curriculum nationally. The first examination in the new style will take place in 2015.

APPENDIX

TABLE 5.5 Results of the Questionnaire for Students

		2006 Intermediate	2010 Intermediate	2010 Higher
1. I found this examination easy to make.	exp.	2.43	2.38	2.67**
	comp.	2.48	2.51*	2.43
2. The questions of this examination were like I expected them to be.	exp.	3.02	3.01	3.38
	comp.	3.09	3.17*	3.28
3. The most important aspects of the subject matter I studied were tested in this examination.	exp.	3.09	3.03	3.27
	comp.	3.29**	3.11	3.26
4. The examination questions were clear.	exp.	2.97	3.06	3.34
	comp.	3.30**	3.06	3.22
5. The questions in this examination were very similar to each other.	exp.	2.71*	2.76*	2.93**
	comp.	2.57	2.60	2.59
6. The source material in the examination was clear.	exp.	3.79	3.58**	3.62
	comp.	3.72	3.36	3.44
7. I have been able to show well in this examination how much historical knowledge I have.	exp.	3.03	2.83	3.21
	comp.	3.05	3.01*	3.02
8. I have been able to show well in this examination how much historical understanding I have.	exp.	3.33	3.23	3.61*
	comp.	3.25	3.21	3.36
9. I have been able to show well in this examination how well I can apply historical skills.	exp.	3.20	3.15	3.43
	comp.	3.18	3.17	3.28
10. I consider the knowledge tested in this examination important.	exp.	3.21	3.00	3.44
	comp.	3.11	3.19*	3.30
11. I consider the historical skills which were tested in this examination important.	exp.	3.19	3.14	3.40
	comp.	3.13	3.22	3.32
12. The examination was well adapted to the subject matter we dealt with in history lessons.	exp.	3.34	3.30	3.64
	comp.	3.59**	3.49*	3.77
13. I consider the quantity of subject matter we had to study for this examination too little.	exp.	2.03	2.08	2.30
	comp.	1.96	2.24*	2.49*
14. The history lessons during the last two years were generally interesting.	exp.	3.39	3.47**	3.58
	comp.	3.53*	3.16	3.71

(continued)

TABLE 5.5 Results of the Questionnaire for Students (continued)

		2006 Intermediate	2010 Intermediate	2010 Higher
15. Compared to junior secondary education, I have learned many new things during the last two years.	exp.	2.18**	3.70	3.80
	comp.	1.98	3.69	3.99
16. I think history is a difficult subject.	exp.	2.83	3.03	2.70
	comp.	3.05*	3.13	3.11**
17. Next year I'll have forgotten most of the history I have studied the last two years.	exp.	2.71	2.91	2.56
	comp.	3.04**	3.07	2.89**
By Means of the History Lessons of the Last Two Years:				
18. I have acquired a good survey from Antiquity to now.	exp.	3.73**	3.78**	4.03**
	comp.	2.97	3.22	3.24
19. I understand better how and why societies can change.	exp.	3.55	3.62**	3.77
	comp.	3.45	3.37	3.74
20. I can better contextualize events in the past in the right context.	exp.	3.54**	3.66	3.91
	comp.	3.31	3.55	3.76
21. I understand the world of today better.	exp.	2.92	3.07	3.22
	comp.	3.06*	3.03	3.27
22. I have learned to think more critically.	exp.	2.88	3.01	3.12
	comp.	3.13**	3.12	3.24
23. I have learned things that I can apply to other subjects.	exp.	3.07	3.14	3.64
	comp.	3.10	3.15	3.60
24. I have learned things that can be useful for further study or for a profession.	exp.	2.58	2.61	3.03
	comp.	2.49	2.49	2.96
25. I have learned things that actually everybody should know.	exp.	3.32	3.42**	3.82
	comp.	3.34	3.14	3.88
26. I have started to consider history more interesting.	exp.	3.19	3.31	3.55
	comp.	3.39**	3.25	3.71
27. I have started to consider history a more important subject.	exp.	3.06	3.12	3.38
	comp.	3.17	2.96	3.46

* $p < 0.05$, ** $p < 0.005$

Note: Mean results of Likert-scales 1–5: (strongly) disagree–(strongly) agree.

Background shading in cells where differences between the experiment group and the comparison group are statistically significant.

NOTES

1. Some of the experiments included inquiries into the way students dealt with evidence, causation and continuity, and change. Others were assessments of their control of outline historical knowledge at the beginning of the project and at the end of each academic year.
2. The questionnaire was composed in 2006 by Carla van Boxtel and Jannet van Drie (University of Amsterdam). The elaboration and interpretation of the data was completed by the present author.
3. The normal practice in the Netherlands is that national examinations are first graded by the students' own teachers, using a scoring instruction provided by CITO. The work is then sent to another school, unknown to the first one, where teachers repeat the grading of the work. In cases where gradings differ, the teachers of the two schools have to reach an agreement by negotiation. In the examination project described here, the same procedure was applied.
4. An example of these three/four "contexts" is "The Cold War 1945–1991" about which a description of 536 words was added to the syllabus, containing the names of Stalin, Truman, Khrushchev, Kennedy, Brezhnev, and Gorbachev and concepts like spheres of influence, containment policy, arms race, détente, peaceful co-existence, glasnost, and perestroika. As advisable "concrete examples" the addition mentions the Potsdam Conference, the Marshall Plan, the Berlin Blockade, Senator McCarthy's speech about communists in the United States, the Hungarian Revolt, the building of the Berlin Wall, the Cuban Missile Crisis, the Prague Spring, the signing of SALT-I and the fall of the Berlin Wall. The Cold War in Asia was added as an extra context only for the higher level, which indicates that the idea that a larger quantity of reproductive knowledge is an indicator of a higher level of performance still persists.

REFERENCES

- CG (1998). *Commissie geschiedenisonderwijs: het verleden in de toekomst* [Commission on history education: The past in the future]. The Hague, the Netherlands: Ministry of Education and Sciences.
- CHMV (2001). *Commissie historische en maatschappelijke vorming: Verleden, heden en toekomst* [Commission on historical and social education: past, present, and future.] Enschede, the Netherlands: SLO.
- CITO (2011a). *Examen havo 2011, geschiedenis. Vragenboekje* [Examination, intermediate level 2011, history. Booklet with questions]. Retrieved February 21, 2013, from http://www.cito.nl/static/ce/ex2011_havovwo/bestanden/HA-1021-a-11-1-o.pdf. *Examen havo 2011, geschiedenis. Bronnenboekje* [Examination, intermediate level 2011, history. Booklet with sources]. Retrieved February 21, 2013, from http://www.cito.nl/static/ce/ex2011_havovwo/bestanden/HA-1021-a-11-1-b.pdf
- CITO (2011b). *Examen vwo 2011, geschiedenis. Vragenboekje* [Examination, higher level 2011, history. Booklet with questions]. Retrieved February 22, 2013, from

- http://www.cito.nl/static/ce/ex2011_havovwo/bestanden/VW-1021-a-11-1-o.pdf. *Examen vwo 2011, geschiedenis. Bronnenboekje* [Examination, higher level 2011, history. Booklet with sources]. Retrieved February 22, 2013, from http://www.cito.nl/static/ce/ex2011_havovwo/bestanden/VW-1021-a-11-1-b.pdf
- Dunn, R. E. (2000). Constructing world history in the classroom. In P.N. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching and learning history*, pp. 121–140. New York, NY: New York University Press.
- Dunn, R. E. (2009). The two world histories. In L. Symcox, & A. Wilschut (Eds.), *National history standards: The problem of the canon and the future of teaching history*, pp. 55–69. Charlotte, NC: Information Age.
- Granatstein, J. L. (1998). *Who killed Canadian history?* Toronto, Canada: HarperCollins.
- Howson, J., & Shemilt, D. (2011). Frameworks of knowledge: Dilemmas and debates. In I. Davis (Ed.), *Debates in history teaching*, pp. 73–83. London, England: Routledge.
- Lee, P. (2005). Historical literacy: Theory and research. *International Journal of History Learning, Teaching, and Research*, 5 (1).
- Lee, P., & Howson, J. (2009). Two out of five did not know that Henry VIII had six wives: History education, historical literacy and historical consciousness. In L. Symcox, & A. Wilschut, (Eds.), *National history standards: The problem of the canon and the future of teaching history*, pp. 117–140. Charlotte, NC: Information Age.
- Phillips, R. (1998). *History teaching, nationhood and the state*. London, England: Cassell.
- Shemilt, D. (2009). Drinking an ocean and pissing a cupful: How adolescents make sense of history. In L. Symcox, & A. Wilschut, (Eds.), *National history standards: The problem of the canon and the future of teaching history*, pp. 141–210. Charlotte, NC: Information Age.
- Symcox, L. (2002), *Whose history? The struggle for national standards in American classrooms*. New York, NY: Teacher's College Press.
- Wilschut, A. (2007). *Handreiking schoolexamen geschiedenis havo/vwo* [Manual for the school examination history, intermediate and higher level]. Enschede, the Netherlands: SLO. Retrieved February 22, 2013, from http://www.slo.nl/downloads/archief/Handreiking_geschiedenis_DEFINITIEF.pdf/download.
- Wilschut, A. (2012). *Images of time: The role of an historical consciousness of time in learning history*. Charlotte, NC: Information Age.
- Wilschut, A., Boom, S., Van Boxtel, C., & Van Dric, J. (2006). *Geschiedenis examineren. Rapport pilotproject CHMV-examen geschiedenis havo 2006* [Examining History. Report of the pilot project CHMV-examination history, intermediate level 2006]. Amsterdam, the Netherlands: Hogeschool van Amsterdam.
- Wilschut, A., Boom, S., & Van der Kaap, A. (2010). *Geschiedenis examineren op niveau. Rapport van het pilotproject CHMV-examen geschiedenis havo en vwo 2010* [Examining history at higher levels. Report of the pilot project CHMV-examination history, intermediate and higher levels 2010]. Amsterdam, the Netherlands: Hogeschool van Amsterdam.

SECTION II
