

Assessing Ways of Seeing the Past: Analysis of the Use of Historical Images and Student Performance in the NAEP U.S. History Assessment

Yonghee Suh and Leslie W. Grant

Old Dominion University and The College of William and Mary

ASSESSING STUDENTS' HISTORICAL UNDERSTANDING has been a long-standing challenge in history education. One of the widely used tools for accomplishing this task is the large-scale standardized test, the results of which are used as an indicator of student knowledge and skills in the social sciences/history. At the state level, eleven states currently administer social studies/history assessments to elementary students, fifteen states to middle school students, and twenty to high school students.¹ At the national level, the National Assessment of Educational Progress (NAEP), often referred to as the "nation's report card," is considered as the provider of a "common yardstick" by which to compare student knowledge and skills nationwide in social sciences/history along with many other content areas.² Despite their usefulness to educators and the public alike, large-scale assessments in the form of standardized testing at state and national levels have been criticized for addressing fragmented factual knowledge rather than students' historical and inquiry skills.³

In search of ways to address the problem of focusing on fact-based questions, it was Document-Based Questions (DBQs) that these large-scale assessments have incorporated. The National Assessment Governing Board, which developed the framework for the NAEP, created the DBQs because, in order "to reflect the richness of history and historical sources,

the assessment should use a variety of stimulus materials appropriate to U.S. history.⁴ Using historical knowledge, students are expected to interpret the meaning of primary sources (e.g., speeches, letters, diaries, quantitative materials, political cartoons, photos, and paintings) and relate the meaning of the sources to the historical context, given the students' historical understanding. Thus, ideally, DBQs require students to exhibit knowledge of history while demonstrating critical analysis skills.⁵ The question before us, then, is whether the NAEP uses DBQs in a way that calls for students to do more than just regurgitate historical facts.

To answer this question, this study explores how the NAEP uses primary sources—specifically visual images—to assess eighth and twelfth grade U.S. history students' historical knowledge and thinking skills. Though scores on the NAEP have traditionally served as a barometer for both researchers and the public to judge the quality of teaching and learning in U.S. history classrooms,⁶ very few studies have been conducted regarding the questions themselves.⁷ In our study, we neither idealize the NAEP assessment nor criticize it. We use it as a case for studying the use of historical images in assessing students' historical thinking skills on large-scale assessments, and for exploring the possibilities and challenges of using visual images to teach students how to think historically. In this regard, we sought to answer the following questions regarding the U.S. history section of the NAEP: How are visual images used as primary sources in the NAEP assessment? To what degree are visual images used to assess students' historical thinking skills?

Why Visual Images?

We focus specifically on the use of visual images in the NAEP for two reasons. First, the need for students to think critically about visual images is especially crucial in a contemporary society that is visually inundated.⁸ We are living in a society where the ways in which we interact and interpret visual images are important life skills. In particular, students in the United States run into and are influenced by information through visual images—perhaps more than verbal text—every day. They might purchase a product by watching a TV commercial where no words—only images—present the product. They might watch Presidential election campaigns on YouTube. They might play video games. They might view images of Pocahontas created by Disney more frequently than the historical portraits of Pocahontas. All of these images send out messages that are complex⁹ and, therefore, rather than practicing to blindly “absorb” the information, students need to practice critically analyzing the information presented by visual images. We, as history educators, also must appropriately and accurately assess the thinking skills that students need to analyze the visual images.

Second, visual images require our attention in instructing and assessing students' historical thinking skills because they are a type of historical evidence that is appropriate for certain historical topics and time periods. Peter Burke, for instance, eloquently summed up the need for using visual images as historical evidence.¹⁰ He suggested that visual images have been used as historical evidence by historians to explore time periods where "written documents are sparse or non-existent"¹¹ or to study "non-verbal experiences or knowledge of past cultures."¹² Especially in light of the relatively recent emphasis on the history of culture and mentalities, everyday life, and ordinary people such as women and social minority groups, he argued:

It would not have been possible for them [historians] to carry out research in these relatively new fields if they had limited themselves to traditional sources such as official documents, produced by administrations and preserved in their archives.¹³

However, engaging with and being critical of the stories and meanings represented in visual images is not an easy task. Scholars in history education argue that there are challenges in using visual images to teach history. Students often have a hard time in understanding that visual accuracy does not necessarily guarantee historical accuracy.¹⁴ Marcy Singer Gabella, for instance, studied an eleventh-grade U.S. history class where visual images such as photography, film, and painting were used as main teaching resources.¹⁵ She reported that students assumed that the past as represented in photographs was accurate without recognizing the perspective of the photographer behind the camera lens, whereas they assumed the past as represented in paintings was less accurate since they knew a painter created the painting. Another challenge posed by visual images is that interpreting visual images requires careful analytic skills. Researchers suggest that there are different types of communication tools that visual images use, such as color, tone, and mood—deciphering these clues is the key to successfully understanding the messages that the image delivers to the viewers.¹⁶ These communication tools are hard to teach in current history classrooms when history is being taught by heavily relying on textbooks and written materials.¹⁷

While it is challenging to teach and assess students' historical thinking through visual images, it is worth noting how visual images often have been used as a research tool to investigate students' historical understanding, as an alternative to quizzing them on their specific knowledge of dates and people. In studying students' chronological thinking skills and their judgment of historical significance, Keith Barton used a set of captioned pictures and asked students to order them chronologically and to explain their choices.¹⁸ Sam Wineburg and his colleagues, on the other hand,

used a set of photos to inquire about what different generations—parents and students—remember and know about the Vietnam War, one of the most pivotal historical events in the parents’ generation.¹⁹ Using photos as prompts, they asked both the parents and the students to think aloud about what the photos reminded them of about the war, as a way to explore the two generations’ understanding of the war. Overall, their findings suggest that visual images serve as a vehicle to assess students’ historical understanding and thinking in a more comprehensive manner than tests that measure mere listings of names and dates.

To summarize, visual sources play a unique role in testing the historical thinking skills of students because images are by nature comprehensive and able to simultaneously show many different cultural and historical details, which must be inferred or carefully observed. Since correctly and fully evaluating a historical image necessitates historical thinking skills, image questions in the NAEP should reflect this. Thus, it will be critical to know how the NAEP uses visual images in assessing students’ historical knowledge and skills.

Methods

This study analyzed the use of visual images as historical evidence on the NAEP U.S. history assessment. More specifically, we identified what students are expected to know and be able to do when asked to answer history questions with visual images. In doing so, we wanted to ascertain the historical thinking skills required when responding to questions that use visual images. To meet this goal, we asked the following research questions:

- How many questions on the NAEP assessment use visual images?
- What kinds of visual images (e.g., prints, paintings, or photos) are used?
- What types of questions were asked using visual images?
- To what degree are questions with visual images more or less difficult than questions without visual images?
- What cognitive level is required to respond to questions with visual images?

Data Source and Analysis

The researchers accessed released questions from the 2001, 2006, and 2010 NAEP U.S. history assessment for grades eight and twelve. The released questions yielded a database with a total of 246 U.S. history questions. Forty-nine questions with visual images were selected for the study. To sample the questions for the analysis, we defined a primary source

as “an artifact or document that was created during the time period under study.”²⁰ Given this definition, we retrieved questions from our database that used visual images such as paintings, photographs, and political cartoons as primary sources.²¹ According to this definition, we did not include charts, graphs, and maps that display data as part of our analysis.

We employed a mixed methods strategy using both quantitative and qualitative approaches.²² First, we coded each question according to the coding categories provided by the NAEP—the question type, the historical theme, the cognitive level, and the difficulty level.²³ Once we had organized each question into these categories, we continued the study with quantitative data analysis, which included descriptive statistics in terms of frequency counts.²⁴ Chi-square tests of independence were conducted to answer research questions related to whether items differed by the characteristics stated for the use of visual images. By examining the observed versus the expected frequencies, we were able to determine the ways in which items differed based on the use of visual images.

Next, qualitative data analysis was used to explore further the cognitive levels required to respond to questions that included visual images. The NAEP separates the cognitive levels into two categories: 1) historical knowledge and perspectives, and 2) historical interpretation and analysis.²⁵ However, we wanted a more in-depth analysis than the NAEP cognitive categories offered, and wanted to identify the type of historical thinking that is required when responding to the questions. To do this, we created what we term the Visual Image Coding Taxonomy (VICT) and conducted document analysis of the questions with visual images.²⁶ We developed the coding taxonomy by drawing on previous research, including Sam Wineburg’s work on three heuristics of sourcing, contextualizing, and corroborating,²⁷ Peter Seixas’ work on historical significance of visual images,²⁸ and Stephen Mucher’s and Walter Werner’s work on the limitations of visual images to comprehensively portray a given historical time period.²⁹ The lead author coded all the questions by sorting each NAEP visual image question into its respective VICT coding category. The second author verified the reliability of the first coder by separately coding half of the questions. An initial inter-rater agreement of 95% was the result, and differences were discussed to achieve 100% agreement. **Figure 1** provides the Visual Image Coding Taxonomy that we constructed for the analysis.

Findings

Out of the 246 questions, 49 used visual images and 197 did not. Five different types of images were used as primary sources, including 20

VICT Coding Categories	Questions Drawn from the Literature	Examples of NAEP Questions
Observing	What historical incident does this image depict?	“The poster refers to...” “What event was depicted in the photograph above?”
Sourcing	Who created this image and why? What is the message the author tried to communicate with this image?	“What was the main message of the poster?” “The poster above was printed during the second World War. Why was it printed?”
Contextualizing	What inference can you draw from the image about the historical incident and time period it portrays?	“The cartoon above was drawn after the end of the 1937 sit-down strike at the General Motors automobile plant in Flint, Michigan. What conclusion about the outcome of the strike can be drawn from the cartoon?”
Corroborating	How do different visual images portray the same historical incident or time period in similar or different manners and why?	“The top picture is a painting that shows settlers crossing the Great Plains. Below the painting is a photograph of the same topic. What important differences do you notice between the photograph and the painting?”
Historical Significance	Is this source important to understand the historical incident that it portrays? Why or why not?	[No example was found.]
Limitations	What do I still not know? What doesn't this image tell me about the historical incident or time period that it portrays? Do other sources help me understand this one better? If so, what other sources should I look at?	“The remains of this Sinaguan cliff house tell us something about the way ancient people lived in what is now the southwestern part of the United States. Which of the activities below would be the best way to learn how the Sinaguan people lived in ancient times?”

Figure 1: The Visual Image Coding Taxonomy (VICT), with coding categories developed for the NAEP question analysis.³⁰

NAEP Coding Categories	Visual Image		No Visual Image		χ^2
	<i>n</i>	%	<i>n</i>	%	
Question Type					6.226*
Multiple Choice	29	11.8	138	56.1	
Short constructed Response	17	6.9	38	15.4	
Extended constructed Response	3	1.2	21	8.5	
Historical Theme					19.131*
Changing Continuity and American Democracy	6	2.4	64	26.0	
Changing Role of America in the World	18	7.3	31	12.6	
Economic and Technological Changes	15	6.1	38	15.4	
Gathering of People, Cultures, and Ideas	10	4.1	64	26.0	
Cognitive Level					16.536*
Historical Knowledge and Perspective	6	2.4	86	35.0	
Historical Analysis and Interpretation	43	17.5	111	45.1	
Difficulty Level					8.660*
Easy	23	9.3	54	22.0	
Medium	16	6.5	66	26.8	
Hard	10	4.1	77	31.3	

* $p < .05$

Figure 2: Frequencies and Chi-squares of questions with and without visual images, with coding categories provided by NAEP.

posters, 15 photos, 9 cartoons, 4 paintings, and 1 case of multiple visual images for a set of questions. There were 29 multiple-choice questions, 17 short constructed response questions, and 2 extended constructed response questions that referred to visual images. **Figure 2** provides the specific analyses.

Analysis of the Use of Visual Images through the NAEP Coding Categories

Our quantitative analysis using the NAEP framework suggests that NAEP questions that use historical images tend to be short constructed response questions. We also found that NAEP questions with historical images tend to be asked about two themes, including 1) Changing Role of America in the World and 2) Economic and Technical Changes. Furthermore, they measure historical analysis and interpretation and are easier for students as defined by student performance on the questions.

Question Type

NAEP uses three types of questions on tests: multiple-choice, short constructed response, and extended constructed response. We used Fisher's Exact Test since one of the expected count cells contained less than 5.³¹ The results indicated that the question type is related to whether visual images were used. The question type differed by the use of visual images, $\chi^2(2, n=246) = 6.226, p = .033$. Multiple-choice questions with visual images occurred less frequently (of the 246 questions, 11.8% were multiple-choice questions with images compared to 56.1% without). Short constructed response questions with visual images occurred less frequently (6.9% with images compared to 15.4% without). Extended constructed response questions with visual images also occurred less frequently (1.2% with images compared to 8.5% without).

Historical Theme

The NAEP includes four historical themes used as a framework in creating questions: 1) Changing Continuity and American Democracy; 2) Changing Role of America in the World; 3) Economic and Technological Changes; and 4) Gathering of People, Cultures, and Ideas. The theme of the items slightly differed by the use of visual images, $\chi^2(3, n=246) = 19.131, p = .000$. Items with visual images that assessed the first theme occurred less frequently compared to the other themes, and items without visual images occurred more frequently overall.

Cognitive Level

The cognitive level of items differed by the use of visual images, $\chi^2(1, n=246) = 16.536, p = .000$. Items with visual images in the historical analysis and interpretation cognitive level occurred more frequently than those of the historical knowledge and perspective level (17.5% compared to 2.4%, respectively).

Difficulty Level

The difficulty of items differed by the use of visual images. Easy items are those for which 60% or more students responded correctly, medium for which 40% to 59% of students responded correctly, and hard for which less than 40% of students responded correctly. The analysis indicated that difficulty level was related to whether visual images were used. The difficulty of items differed by the use of visual images, $\chi^2(2, n=246) = 8.660, p = .013$. Easy items with visual images occurred more frequently and easy items without visual images occurred less frequently. Hard items with visual images occurred less frequently and items without visual images occurred more frequently overall.

Summary of Findings from the Analysis by the NAEP Categories

Our analysis through the NAEP categories suggests that the use of visual images is not evenly distributed among question types, cognitive levels, or difficulty levels. Most of all, our findings suggest that, using the two cognitive levels from NAEP, questions that used visual images were more likely to assess historical analysis and interpretation rather than historical knowledge and perspective.

Furthermore, questions using visual images and asking historical analysis and interpretation questions were not necessarily more difficult than those asking historical knowledge and perspective questions. Considering the fact that we often equate questions that require higher levels of cognition with greater difficulty because they tend to measure more complex outcomes,³² these findings are quite surprising. We speculate that the scaffolding provided by the visual makes complex ideas more accessible to students. This finding also aligns with others who found that visual images assist struggling readers.³³ We also speculated that, given the historical themes such as the Changing Role of America in the World or Economic and Technological Changes involve contemporary historical events such as World War II and scientific inventions, historical images might be frequently used in the questions on these themes because of the easier access to visual images as primary sources from the time period.

Though the findings from the analysis through the NAEP coding categories are promising for using visual images to assess students' historical thinking skills, analysis through the visual image coding taxonomy, however, suggests that there are some limitations in the ways questions that used visual images measured the knowledge and skills necessary to analyze those images.

VICT Coding Categories	<i>n</i>	%
Observation	16	32.7
Sourcing	12	24.5
Contextualizing	19	38.8
Corroborating	2	4.1
Limitations	1	0.2
Historical Significance	0	0

Figure 3: Frequencies and percentages of questions with visual images, ranked in descending order, with coding categories provided by VICT.

A Close Look at the Questions with the Visual Image Coding Taxonomy

As shown in **Figure 3**, our findings indicate that questions with historical images were asked most frequently to contextualize (19 questions) and observe what happened in the picture (16 questions). Questions were also asked to source the images by identifying the intentions of the authors (12 questions) and corroborate events/perspectives portrayed in the image (2 questions). One question asked students to consider the limitations of the visual images as a primary source, while none of the questions asked students to judge the historical significance of the visual images. Figure 3 shows patterns that we identified across the questions with visual images.

Observation

Sixteen questions asked students to “observe” the images and associate these images with historical events and concepts. One eighth-grade question, for instance, presented a photograph showing two African American women chatting in front of a theater that displays a sign reading “Colored Entrance.” The question asked students, “What social policy is reflected in the photograph?” Another question used for both eighth and twelfth grades displayed a poster of a man raising a big fist over the words, “Avenge December 7.” The question asked students to identify the historical event referenced in the poster. Two issues can be identified with both the photograph of the women and the poster of the man. First, neither of these questions provides images with enough clues that students can utilize to associate what they see in the images with historical events or concepts. Neither of these questions offers the accurate source of the images. The only information that the questions provide are the names of the archives, such as *The New York Times* for the photograph and the

National Archives for the World War II poster. When and by whom it was created, which would serve as clues for students to figure out what the images represent, are not found in either of these questions—or, in fact, for any of questions in this category. Second, it is quite questionable whether using visual images for these questions is the best way to measure students' historical knowledge and understanding. Both these questions and the majority of the visual image questions on the NAEP can be answered even without having students look at the images. For example, another question presents a photograph of Cesar Chavez marching with farmers, one of whom is holding a flag of the United Farm Workers of America. The question asks, "Why did Cesar Chavez, pictured above, ask people to stop buying grapes in the 1960s?" This question can be answered without the photograph if the students know who Cesar Chavez is and what he did.

Sourcing

We also found 12 questions that asked students to "source" the images. The pattern of these questions was almost identical. Ten of the 12 questions read: "Why was it [the visual image] created? What was its message?" Similar to the observation questions, half of the sourcing questions (3 out of 6) for the eighth-grade U.S. history NAEP did not present the precise origin of the image. As an example, one political cartoon showed an underwater shark labeled "discrimination" approaching a small ship with a flag saying "immigration." The question under the image reads: "What point is the cartoonist trying to make?" Except for the information under the cartoon, "Joel Pette, *Lexington Herald-Leader*," which indicates the name of the cartoonist and the newspaper in which it appeared, no clues were offered about what year the cartoon was created or what specific event the cartoonist made a point about. Discrimination against immigrants has a long history in the United States and the rest of the world. Depending on the time period, a variety of push-and-pull factors in domestic and international contexts both encouraged and discouraged immigration policies. Without any relevant clues or any other historical evidence, students would be unable to demonstrate their historical understanding of the issue; in this case, what specific immigration issues of what time period are being portrayed by this cartoonist, and what messages he wanted to communicate with the viewers.

Contextualizing

We found that there were 19 questions that asked students to "contextualize" the images. We found that questions in this category—along with those that asked students to "corroborate," which we will discuss later—best assess students' historical thinking skills, although these questions should be crafted more carefully. For instance, presenting

a painting, one question reads: “The picture above shows farming on a Georgia cotton plantation before the Civil War. Using your knowledge of history and evidence from the picture, explain two important differences between farming on large plantations and farming on small farms in the South before the Civil War.” Another question with a painting of the Western frontier reads: “Paintings like this one helped to make people want to go West. How was life on the frontier different from what the painting shows?” The second question for this image further asks students to look carefully at the painting of a Western town and describe three specific things they see in the painting that could have made people want to become settlers in the West. Both questions have strength in terms of explicitly suggesting specific strategies that students can use to answer the question (“using your knowledge of history and evidence from the picture”) and requiring students to gather specific evidence from the painting to support their response (describing three specific things they see in the painting that could have made people want to become settlers in the West). However, we noticed that, similar to other questions, there appears to be one big hole across these contextualizing questions in assessing students’ historical thinking. Neither of these questions offered source information for these images, such as when they were created and by whom, and, thus, none asked students about the validity and reliability of the visual images as primary sources—that is, whether the images are trustworthy, and what makes the viewer trust or distrust the past as represented in the images by examining the perspectives of the creator. Instead, all of these questions assume that the creators of the images tell the “truth” about the past.

Corroborating

Two questions were coded as “corroborating.” One of these questions presented two images—a painting that shows a family of settlers crossing the Great Plains and a photograph of the same topic, although the two images represent the settlers in a dramatically different manner. The facial expressions of the pioneer family in the painting look peaceful and calm, and they look affluent, having a number of cows following their wagon. On the other hand, the pioneer family in the photograph looks dull and tired in a barren field without any cows in sight. The question asked students to identify the key differences they noticed between the photograph and the painting. The scoring guide listed that the credited response would notice that:

- different emotions are portrayed (e.g., the people look glamorous in the painting, while they look drab in the photo; or the people look happy in the painting, while they look sad in the photo)
- more poverty, sadness, and hunger is shown in the photo

As this scoring rubric suggests, this question requires students to demonstrate their thinking skills first by observing each image, then by comparing and contrasting (“corroborating”) the two images of the pioneer family, and finally by pinpointing the differences in representing the settlers. This question, however, has the same problems as discussed earlier in other questions. Neither of these images was presented with accurate source information about when and where it was created. The scoring guide also assumes that the image portrayed in the photograph is more accurate than the painting, which might be true in this case. However, we cannot presume that photographs always offer more accurate information than paintings since the photographer’s perspective is embedded in their photograph just as a painter’s perspective translates into their painting.

Historical Significance and Limitation

Finally, one question was coded to ask students to identify the limitations of an image as a primary source in the context of the time period. Presenting a photograph of a Sinaguan cliff house, the multiple-choice question reads: “The remains of this Sinaguan cliff house tell us something about the way ancient people lived in what is now the southwestern part of the United States. Which of the activities below would be the best way to learn how the Sinaguan people lived in ancient times?” The possible answers list:

- a. study letters and diaries left in the cliff houses
- b. talk to people living near the cliff houses
- c. study tools, bones, and pottery left in the cliff houses
- d. camp out in the cliff houses for a couple of days

This question aimed to assess two things: whether students know Sinaguan people had a written system and, if they did not, what historical evidence along with cliff houses students need to look for in order to investigate Sinaguan people’s lives. We found that this question is appropriate. Sinaguan people are pre-Columbian Native Americans who occupied central Arizona and left no written record about themselves. Thus, a visual image such as a photo that represents cliff dwellings would be an appropriate source for students to inquire about their lives. In addition, by asking students to identify “the best way to learn how the Sinaguan people lived in the ancient times,” this question requires students to consider various artifacts as historical evidence in case we do not have written records.

We, however, noticed that this question, too, has weaknesses. Similar to other questions, it failed to offer historically accurate clues that students would need in order to explore the photograph, such as source

information of the image and historical context useful to students who perhaps have never learned about the Sinaguan people. Some students in the United States—those in Virginia, for instance—have never been taught about Sinaguan people in their curriculum. Instead, they learn “how the contributions of ancient China and Egypt have influenced the present world in terms of architecture, inventions, calendar, and written language.”³⁴ How far ago do we consider “ancient”? Ancient times as referred to in the NAEP question compared to what students in Virginia learn are obviously not the same time period. Even if a majority of these students responded to the question correctly, it is doubtful how accurately this question measured how much these students know and understand about the time period during which Sinaguan people lived, and how capable students were of using relevant clues to identify what other sources to use when *ahistorical* clues arise.

Using Visual Images in Assessing Students’ Historical Thinking

In this study, we were most interested in the cognitive levels required for questions that used visual images. Based on the two cognitive levels from the NAEP, our analysis found that questions that used visual images were more likely to assess historical analysis and interpretation rather than historical knowledge and perspective, and in a way that students found relatively easier to answer than questions using written sources. Using the visual images taxonomy, however, we found that even if the NAEP intends to use visual images as primary sources and measure students’ historical analysis and interpretation skills, in many cases, it either ends up measuring only basic knowledge or fails to create a context where students can adequately demonstrate their historical thinking skills. The “Avenge” poster, for instance, does not ask students to demonstrate their thinking skills. Rather, it asks students to tell what date the attack on Pearl Harbor took place. These questions tend to ask only basic knowledge about the historical event or time period without offering appropriate source information. This means that even if the NAEP highlights the importance of introducing to students “the process of historical inquiry,”³⁵ the type of questions asked on the test are still limited in their ability to measure students’ historical thinking skills. This issue requires us to come back to the notion of the problems of assessing students’ factual knowledge rather than historical thinking skills, and supports others’ assertions that large-scale history assessments are limited in assessing the thinking required of students.³⁶

Findings of the study also suggest a few pedagogical suggestions in using visual images to assess students’ historical thinking skills. First,

when visual images are used either for instruction or assessment, their source information must be provided. This includes the information about when the visual image was created, the identity of the author, as well as, if necessary, background information about the historical incident that the image portrays. This source information should be provided because it offers students a context for analyzing the image. Depending on the time when the image was created (e.g., before or after the historical event it portrays), and in which context the image was examined, the image can be classified as a primary or secondary source and, thus, the reliability of the image as a historical source can be determined.³⁷ Guarding this kind of key information, as we saw in the NAEP questions, won't allow students to arrive at their answer by means of analyzing either contextual information or the evidence from the visual image, even if they get the correct answers.³⁸

Second, the selection of visual images and the crafting of questions to ask around those visual images must be done systematically. In other words, the visual image questions must be asked in a sequence that allows students to demonstrate historical thinking rather than merely recall factual information. The "Avenge" poster question could have been further developed to ask students to show more sophisticated thinking skills, such as by asking about the intentions of the author ("sourcing"), how different groups of people in the U.S. would feel and respond to the poster ("contextualize"), and what background information or different sources students will need to evaluate the reliability of this poster ("corroborate"). To create a meaningful assessment with visual images, it will also be important for the teachers and assessors to be clear about the goals of the assessment—that is, which specific historical knowledge and thinking skills they would like to assess. We also argue that there are certain historical topics for which visual images are better representative of, and therefore they can serve as appropriate assessments of students' thinking skills. As noted earlier, visual images can serve as a more comprehensive source of information about the culture and everyday lives of ordinary people or the history of people in civilizations who are otherwise invisible in written records or artifacts. When selecting images for assessments, it will be helpful to predetermine the topics for which visual images would be better suited.

Third, most of the suggestions made here can be applicable to both visual images and written sources, although highlighted analytic skills are not taught as explicitly with visual images as with the written sources. Visual images must also be treated as significant sources of evidence. Just as they are with written sources, questions must be asked about the reliability and historical significance of visual images in understanding the historical

incident or time period. The pattern that we most frequently noticed across the visual image questions is that 24.5% of sourcing questions asked students about the intentions of the creators of the images without offering any information about the authors. Besides these questions, all the questions assume that the visual images mirror the historical incident or time period represented in those images. The questions about the plantation in Georgia, for instance, assume that what is represented in the painting will accurately portray to students what farming on large Southern plantations looked like before the Civil War. However, it is not historically rigorous to ask students identify the differences between farming on a large plantation versus a small farm given the painting, without asking first to examine any information about when the painting was created and by whom, and judge the reliability of the source.

Lastly, researchers note that when teachers use visual images in their instruction or assessment, selecting visual images for the topic is the first challenge.³⁹ Even if the visual image questions in the NAEP reveal limitations, the large database of sample questions that the NAEP shares in public⁴⁰ with some modification discussed earlier would be useful for the teachers and administrators who are interested in using visual images in their assessments. This database of sample questions is easy to navigate by grade, type of questions, difficulty, and testing year. Students' responses for each item are also compared across the state and selected school districts; and analysis of students' responses is shared on the site. In addition, several online resources serve as good places to learn about how to select and use visual images in historical instruction and assessment (see Appendix).

Conclusion

In analyzing the use of visual images as historical evidence in the NAEP U.S. History assessment, findings from this study add to the growing body of literature on how large-scale assessments measure students' historical thinking skills. Previous studies reported that large-scale assessments tend to not assess students' higher order thinking skills,⁴¹ and, especially when using multiple-choice items, it is even harder for these assessments to accurately measure students' historical knowledge and thinking skills.⁴² Echoing with these previous findings, this study indicates that the NAEP questions, despite intentions, tend to either end up assessing factual information rather than historical thinking skills or do not offer sufficient clues for students to demonstrate their historical thinking skills. However, our findings also indicate that the range of historical knowledge and thinking skills that are associated with interpreting historical images is complicated and comprehensive, which we observed in a few examples of

the NAEP questions. When used systematically and rigorously,⁴³ historical images serve as effective tools that allow us to assess students' historical understanding even in large-scale assessments.

Notes

1. The Editorial Projects in Education (EPE) Research Center, "Quality Counts, 2012: The Global Challenge: Education in a Competitive World," *Education Week*, 12 January 2012, <<http://www.edweek.org/ew/toc/2012/01/12/>>.
2. National Center for Education Statistics, *National Assessment of Educational Progress* (Washington D.C.: National Center for Education Statistics, 2011), <<http://nces.ed.gov/nationsreportcard/>>.
3. S. G. Grant, Jill M. Gradwell, and Sandra K. Cimbricz, "A Question of Authenticity: The Document-Based Question as an Assessment of Students' Knowledge of History," *Journal of Curriculum and Supervision* 19, no. 4 (Summer 2004): 309-337; Avner Segall, "Teachers' Perceptions of the Impact of State-Mandated Standardized Testing: The Michigan Educational Assessment Program (MEAP) as a Case Study of Consequences," *Theory and Research in Social Education* 31, no. 3 (Summer 2003): 287-325; Stephanie van Hover, David Hicks, and William Irwin, "Beginning Teachers' Thinking Historically? Negotiating the Context of Virginia's High-Stakes Tests," *International Journal of Social Education* 22, no.1 (Spring/Summer 2007): 85-115; Gabriel A. Reich, "Testing Historical Knowledge: Standards, Multiple-Choice Questions and Student Reasoning," *Theory and Research in Social Education* 37, no. 3 (Summer 2009): 352-360.
4. National Assessment Governing Board, *U.S. History Framework for the 2010 National Assessment of Educational Progress* (Washington, D.C.: National Assessment Governing Board, 2010), 45, <<http://www.nagb.org/content/nagb/assets/documents/publications/frameworks/historyframework.pdf>>.
5. Eric Rothchild, "The Impact of the Document-Based Question on the Teaching of United States History," *The History Teacher* 33, no. 4 (August 2000): 495-500.
6. Richard J. Paxton, "A Deafening Silence: History Textbooks and the Students who Read Them," *Review of Educational Research* 69 no. 3 (Fall 1999): 315-339; Diane Ravitch and Chester E. Finn, Jr., *What do Our 17-Year-Olds Know? A Report on the First National Assessment of History and Literature* (New York: Harper & Row, 1987); Bruce VanSledright, "Confronting History's Interpretive Paradox While Teaching Fifth Graders to Investigate the Past," *American Educational Research Journal* 39, no. 4 (Winter 2002): 1089-1115.
7. Rebecca Zwick and Kadriye Ercikan, "Analysis of Differential Item Functioning in the NAEP History Assessment," *Journal of Educational Measurement* 26, no. 1 (Spring 1989): 55-66.
8. John Berger, *Ways of Seeing: Based on the BBC Television Series* (London, United Kingdom: Penguin Books, 1990); Douglas Kellner, "Critical Perspectives on Visual Imagery in Media and Cyberculture," *Journal of Visual Literacy* 22, no. 1 (April 2002): 81-90; Walter Werner, "Reading Visual Texts," *Theory and Research in Social Education* 30, no. 3 (Summer 2002): 410-428.
9. Berger; Kellner.

10. Peter Burke, *Eyewitnessing: The Uses of Images as Historical Evidence* (Ithaca, NY: Cornell University Press, 2008).
11. *Ibid.*, 10.
12. *Ibid.*, 13.
13. *Ibid.*, 9.
14. Sam Wineburg, Daisy Martin, and Chauncey Monte-Sano, *Reading Like a Historian: Teaching Literacy in Middle and High School History Classrooms* (New York: Teachers College Press, 2012).
15. Marcy Singer Gabella, "Beyond the Looking Glass: Bringing Students into the Conversation of Historical Inquiry," *Theory and Research in Social Education* 22, no. 3 (Summer 1994): 340-363; Marcy Singer Gabella, "The Art(s) of Historical Sense," *Journal of Curriculum Studies* 27 no. 2 (March-April 1996): 139-163; Marcy Singer Gabella, "Formal Fascinations and Nagging Excerpts: The Challenge of the Arts to Curriculum and Inquiry," *Curriculum Inquiry* 28, no. 1 (Spring 1998): 27-56.
16. Dipti Desai, Jessica Hamlin, and Rachel Mattson, *History as Art. Art as History: Contemporary Art and Social Studies Education* (New York: Routledge, 2009).
17. James W. Loewen, *Teaching What Really Happened: How to Avoid the Tyranny of Textbooks and Get Students Excited About Doing History* (New York: Teachers College Press, 2010).
18. Keith C. Barton, "'Oh, That's a Tricky Piece!': Children, Mediated Action, and the Tools of Historical Time," *Elementary School Journal* 103, no. 2 (November 2002), 161.
19. Sam Wineburg, Susan Mosborg, Dan Porat, and Ariel Duncan, "Common Belief and the Cultural Curriculum: An Intergenerational Study of Historical Consciousness," *American Educational Research Journal* 44 no. 1 (2007): 40-76.
20. Helge S. Kragh, *An Introduction to the Historiography of Science* (Cambridge, United Kingdom: Cambridge University Press, 1989).
21. Keith C. Barton, "A Picture's Worth: Analyzing Historical Photographs in the Elementary Grades," *Social Education* 65, no. 5 (September 2001): 278-283; Gabella, "Beyond the Looking Glass," "The Art(s) of Historical Sense," "Formal Fascinations"; Desai, Hamlin, and Mattson, *History as Art*.
22. John W. Creswell, *Research Design: Quantitative, Qualitative and Mixed Methods Approach* (London, United Kingdom: SAGE Publications, 2009).
23. National Assessment Governing Board, *U.S. History Framework for the 2010 National Assessment of Educational Progress*.
24. Joyce P. Gall, M. D. Gall, & Walter R. Borg, *Applying Educational Research: How to Read, Do, and Use Research to Solve Problems of Practice*, sixth ed. (Boston, MA: Allyn and Bacon, 2010).
25. National Assessment Governing Board, *U.S. History Curriculum for the 2010 National Assessment of Educational Progress*, 31-34.
26. Grant, Gradwell, and Cimbricz.
27. Sam Wineburg, *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past* (Philadelphia, PA: Temple University Press, 2011).
28. Peter Seixas, "Students' Understanding of Historical Significance," *Theory and Research in Social Education* 23, no. 3 (Summer 1994): 281-304.
29. Stephen Mucher, "Building a Culture of Evidence through the Professional Development," *The History Teacher* 40, no. 2 (February 2007): 265-273; Walter Werner, "Reading Authorship into Texts," *Theory and Research in Social Education* 28, no. 2 (Spring 2000): 193-219.
30. Mucher; Seixas; Werner, "Reading Authorship into Texts."
31. Gall, Gall, and Borg, *Applying Educational Research*.

32. C. Keith Waugh and Norman E. Gronlund, *Assessment of Student Achievement*, tenth ed. (Boston, MA: Pearson, 2013).
33. Sam Wineburg and Daisy Martin, "Tampering with History: Adapting Primary Sources for Struggling Readers," *Social Education* 73, no. 5 (September 2009): 212-216; Terrie L. Epstein, "Sometimes a Shining Moment: High School Students' Representations of History through the Arts," *Social Education* 58, no. 3 (March 1994): 136-141.
34. Board of Education Commonwealth of Virginia, *History and Social Sciences Virginia Standards of Learning for Virginia Public Schools* (Richmond, VA: Board of Education Commonwealth of Virginia, 2008), 4.
35. National Assessment Governing Board, *U.S. History Framework for the 2010 National Assessment of Educational Progress*, vi.
36. Grant, Gradwell, and Cimbricz; Segall; Van Hover, Hicks, and Irwin; Reich.
37. Keith C. Barton, "Primary Sources in History—Breaking through the Myths," *Phi Delta Kappan* 86, no. 10 (June 2005): 745.
38. Peter Seixas, "Student Teachers Thinking Historically," *Theory and Research in Social Education* 26, no. 3 (Summer 1998): 310-341.
39. Ibid.
40. National Center for Education Statistics, "NAEP Questions Tool," <<http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=history>>.
41. Grant, Gradwell, and Cimbricz.
42. Reich.
43. Veronica Boix Mansilla, "Assessing Expert Interdisciplinary Work at the Frontier: An Empirical Exploration," *Research Evaluation* 15, no. 1 (April 2006): 17-29.

Appendix

Online Resources for Visual Images in Historical Instruction and Assessment

National Archive: *Teaching with Documents*

<<http://www.archives.gov/education/lessons/>>

Lesson plans and activities for teaching with primary documents.

Smithsonian Museum: *Smithsonian Source*

<<http://www.smithsoniansource.org>>

Resources for teaching American history through primary documents.

Library of Congress: *The Teaching with Primary Sources Program*

<<http://www.loc.gov/teachers/tps/>>

Free sets of primary sources and supplementary materials to teach American history.

C-SPAN: *C-SPAN Classroom*

<<http://www.c-spanclassroom.org/>>

Free primary source materials for social studies teachers.

Stanford University: *Beyond the Bubble*

<<http://beyondthebubble.stanford.edu/>>

Easy-to-use assessments using the Library of Congress digital archive.