

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

The Nebraska Educator: A Student-Led Journal Department of Teaching, Learning and Teacher Education

10-2020

The Current State of Assessing Historical Thinking: A Literature Analysis

Taylor S. Hamblin

University of Nebraska-Lincoln, taylor.hamblin@huskers.unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/nebeducator>



Part of the [Teacher Education and Professional Development Commons](#)

Hamblin, Taylor S., "The Current State of Assessing Historical Thinking: A Literature Analysis" (2020). *The Nebraska Educator: A Student-Led Journal*. 50.

<https://digitalcommons.unl.edu/nebeducator/50>

This Article is brought to you for free and open access by the Department of Teaching, Learning and Teacher Education at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in The Nebraska Educator: A Student-Led Journal by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

The Current State of Assessing Historical Thinking: A Literature Analysis

Taylor Hamblin

Department of Teaching, Learning, and Teacher Education

University of Nebraska-Lincoln

Abstract

In American schools since the mid 2000's, social studies departments and state departments of education have created goals and updated standards prioritizing critical thinking engagement. Promotion of critical thinking has created a wealth of scholarship on developing a specific type of critical thinking, or cognition, called historical thinking. Imperative to the promotion of teaching historical thinking is in how teachers can assess the inquiries that make it up. Unfortunately, standardized social studies assessments have failed to measure the acquisition of the new historical thinking standards. In order to improve the assessment practices of history teachers, I wish to do two things: (1) switch the focus from recall-memorization assessments to those that will focus on a diverse array of historical thinking inquiries; and (2) improve the assessments that we currently use to measure historical thinking skills. In order to accomplish these two objectives, I will examine empirical research studies that focused on how students obtain historical thinking skills. From the data, I ascertain how practicing teachers and researchers currently measure historical thinking skills, and from that, propose improvements.

Keywords: historical thinking, social studies education, history education, assessment, cognition

doi: 10.32873/unl.dc.ne008

In American schools since the mid 2000's, social studies departments and state departments of education have created goals and updated standards prioritizing critical thinking engagement. Promoting critical thinking is not unique to social studies, but it has created a wealth of scholarship on developing a specific type of critical thinking called historical thinking or historical reasoning. This type of thinking involves using inquiry to answer historical questions, like "how important was religion in Colonial America?", and evaluate social concepts, like religion, global citizenship, or cause & effect. In addition, many social studies educators want historical inquiry to focus on "real world" questions that students truly want to know (Selywn, 2014). Unfortunately, standardized social studies assessments have failed to measure the acquisition of the new standards and goals because teachers (or anyone who assesses) historical thinking skills are unlikely to be successful with current practices and assessment tools (Reich, 2009; Shemilt, 2018).

Difficulties arise from the cost of assessing higher order thinking, preparing preservice educators to competently teach social studies cognition, and changing traditional testing practices (Shemilt, 2018). Despite the struggle of assessing cognitive processes in social studies courses, there is tremendous value in creating these assessments. Improved cognitive processes in social studies courses could result in improved civic engagement, student motivation & agency, and better career prospects for students (McGrew, 2018, National Council for the Social Studies, 2013, Selwyn, 2014). An example of assessment that promotes civic engagement is when my 8th grade students created bills for a Mock Senate. During the Spring of 2019, my teenage students questioned, researched, debated, and voted on bills that they created: these bills went onto school admin for consideration. While students completed these bills, I assessed them to determine their

ability to source documents, corroborate evidence, and make inferences over arguments and reasoning (Hamblin, 2019).

Despite my success of assessing students' cognitive abilities while they completed a project, the previously mentioned issues hinder the assessment of historical thinking. To examine this issue, I searched for scholarly empirical studies that examined historical thinking assessments or examined how historical thinking could be measured. In addition, I focused on scholars who were research leaders in studying historical thinking, such as Sam Wineburg, Peter Seixas, John Lee, Bruce VanSledright, David Hicks, and Peter Doolittle. From these searches I found studies defining historical thinking, or reviewing the philosophy of it, but I found none that directly examined how assessments could be scaffolded for students. A lack of finding research examining scaffolding is troubling because any assessment of cognitive ability must follow some type of growth model, like Jerome Bruner's *spiraling curriculum* (Bruner, 1960).

According to the glossary of education reform, "scaffolding refers to a variety of instructional techniques used to move students progressively toward stronger understanding and, ultimately, greater independence in the learning process" (2015). Scaffolding is a means to ease students into a concept, curriculum, behavior, or idea and is a fundamental part of teaching and assessing historical thinking. As an example, let's say an elementary school teacher wants to teach their second graders how to find helpful sources to answer questions. This teacher might have a specific current event that they want to cover, but they will first break-up the lesson into interactive activities that teach the young students important concepts, like finding good sources. A teacher could ask students to write down the five most trustworthy people that could describe the student's identity and experiences. Students would enjoy writing down family members, teachers, and close friends. Finally, the teacher asks why the people listed are trustworthy

sources. This interactive activity then expands until students complete the standard, such as the C3 Framework's D1.5.K-2 "Determine the kinds of sources that will be helpful in answering compelling and supporting questions" (National Council for the Social Studies, 2013, p. 25).

Based on my analysis of research that examined assessment of historical thinking, there are five questions that are important to answer: (1) How do we currently measure historical thinking? (2) How can we take valid data from assessments over inquiry and historical thinking? (3) What scaffolding exists when testing for historical thinking and how can we improve it? (4) How does inquiry and assessment motivate students? (5) Why is assessing historical thinking and inquiry-based learning difficult? Through this literary analysis, I will answer these five questions and will use one to two empirical research article(s) to serve as an exemplar for answering each question.

Before answering the questions, it is foremost to provide context of the current state of historical thinking in academic research. Over the last twenty years, researchers have developed assessments for historical thinking. Assessing historical thinking is part of a broader movement of inquiry assessments. Inquiry assessment is not isolated to just history and the social studies disciplines; Stephanie Corliss (2011) examined how her science projects, which required students to solve basic to complex problems using a rational system, improved student content knowledge and their ability to apply concepts over multiple disciplines. If multiple disciplines are moving towards inquiry assessment, then there must be a broader force pushing social studies educators to redefine their purpose in education.

Fortunately, there are institutions and individuals researching historical thinking skills and producing teaching resources based on their findings. The founders and leaders of these groups have published several theoretical and practical studies that help define historical thinking

and the measurement of it. For example, Peter Seixas, a leader in studying cognitive abilities in history, wrote “A Model of Historical Thinking” (Seixas, 2012) which defined and explained historical consciousness, historical thinking, and the teaching of history in Canadian and United States schools. Seixas and others have defined historical thinking and created teaching resources, but the definitions are heterogeneous, and sometimes contradictory. Furthermore, the analysis they provide in their studies does not consider how historical inquiry assessments need to be scaffolded. An absence of analysis into inquiry assessment and scaffolding challenges the likelihood that teachers will use historical thinking as a curriculum and pedagogical tool.

The move to promoting more inquiry-based assessments, like testing for historical thinking, became clear because of the creation of the C3 Framework (2013). The C3 Framework is a set of standards that unites the disciplines of History, Civics, Geography, and Economics. Assessment designers created the C3 Framework to help improve students’ abilities to “recognize social problems; ask good questions and develop robust investigations into them; considering possible solutions and consequences; separate evidence-based claims from parochial opinions; and communicate and act upon what they learn” (National Council for the Social Studies, 2013, p. 6).

The objectives within C3 provide teachers with a structure to vertically align social studies disciplines across multiple grades. Missing from the framework is assessment advice, measuring devices, and scaffolding. The authors of C3 admitted this gap in their framework, and believed it would be “smart, thoughtful, and imaginative teachers” who need to find ways to make the framework adaptable (NCSS, 2013, p. 15). It is wise for education policy and standard makers to collaborate with teachers to make decisions on implementing curriculum in the

classroom. However, for this to work to be successful, school districts must give teachers time during the workday to be lucrative in producing implementation strategies.

Unfortunately, the C3 Framework lacks practical advice for teachers who may feel overburdened with other educational duties. John Lee was one of the writers for C3 and has developed inquiry assessments to make up for the gap. In an unpublished chapter titled “Assessing Inquiry,” Lee created a lesson plan, which focused on inquiry assessment, compelling questions, supporting questions, formative questions, featured sources, summative performance tasks, and taking informed action (Lee, manuscript submitted for publication). I include Lee’s unpublished chapter because it demonstrates that creating inquiry assessment is an ongoing issue. Lee’s lesson plan is a step in the right direction as it gives practicing teachers plenty of practical advice. Lacking is scaffolding advice.

The writers of the C3 Framework, and later Lee in his unpublished article, create lesson plans for general education students who are supposedly at the same academic level. This lack of differentiation makes it difficult to implement these lessons because of the realities of teaching. In order to be successful in my own practice, I felt that I needed to differentiate within the classroom environment through student choice, alternative primary sources, second- and third attempts on summative assessments, and choice boards. Going into the future, researchers need to do more to advise teachers and give resources on scaffolding historical thinking assessments.

Question 1: How do we currently measure historical thinking skills?

First, there is still confusion in what historical thinking skills are. In order to measure historical thinking skills, there must be categorizations of the different types of cognitive processes which comprise it. A clear breakdown is especially important, because as Stephane Levesque and Penny Clark pointed out in their handbook chapter about historical thinking

definitions; "...if the ability to think historically should go beyond the mere mastery of factual knowledge about the past ("know that"), it is still unclear as to what the alleged connections between "history" and "thinking" actually means in conceptual and practical terms ("know how") (2018, p. 119). Fortunately, researchers have been clarifying the connection between history and thinking by classifying historical thinking skills. For example, Peter Doolittle and David Hicks breakdown these skills into six separate inquiry categories (Virginia Tech), which are summarizing, contextualizing, inferring, monitoring, corroborating, and interpretation. Doolittle and Hicks are not the only researchers to promote a set of skills, Wineburg also did, and included skills that are different, like sourcing, which focuses analysis on a document or artifact's creator and the circumstances of its creation (Wineburg, 2001).

Although we have researchers describing the "know how" of historical thinking, and corresponding skills, there does need to be a greater effort in separating those skills so teachers may test for them. If researchers do not further bracket these skills, then it will be as Denis Shemilt points out, "it may not be possible to make secure assessments of students' historical consciousness," which impacts one's ability to historically think (2018, p. 453). For instance, Sam Wineburg and Sara McGrew, the author of the upcoming exemplary article in this paper, created activities that test specific skills (McGrew et al., 2018). One such assessment measures a student's ability to source a painting (Stanford History Education Group). The assessment only requires students to successfully identify that the painter created the painting at a different date than the event that the painting depicts (in this case the First Thanksgiving). Sourcing, and other historical thinking skills are more complicated than simply identifying a date; therefore, there needs to be further categorization of each skill (Appendix A).

Although there are descriptions of inquiry types and skills for historical thinking most history assessments do not adequately measure them, because they primarily utilize multiple choice questions to test factual recall. The questions on these standardized tests, at their best, only measure aspects of factual recall, but often even fail at doing that (Reich, 2009). The popularity of multiple choice is not surprising since they are simple to use, and as Denis Shemilt points out in his article *Assessment of Learning in History Education* (2018), teachers believed multiple choice tests “improved the reliability” of assessments (p. 449). There is potential for multiple choice tests to measure historical thinking, Bruce VanSeldright argued convincingly that they could in his book *The Challenge of Rethinking History Education* (2011), but the issues in creating a reliable multiple-choice assessment are the same as creating any assessment that examines historical thinking. Creating, administering, and grading appropriate exams is expensive, as well as being time consuming. Current practicing teachers are unaware or unprepared to create and administer such assessments. In addition, researchers like Gabriel Reich suggest that recall tests cannot accurately measure historical knowledge (Reich, 2009). When primarily using multiple choice assessments, teachers are like a baker measuring ingredients using inches and feet, they are using the wrong measurement to assess historical thinking.

A popular alternative to factual recall tests is document-based questions (DBQs), which prompt students to analyze several primary documents, form a thesis, and defend it. In social studies courses, the most widespread use of DBQs is in the U.S. History Advanced Placement exam. If multiple choice recall tests are like measuring how much flour goes into the bowl in inches, then DBQs are using a jackhammer to mix the ingredients. DBQs are substantial questions, and require over an hour to complete; therefore, students use several types of inquiry and skills to form their answers. Due to the extent of these questions, it is unclear as to what

particular historical thinking skill is being measured. Analysis of questions that test historical thinking and what they actually measure is necessary because assessments have a tremendous influence over classroom curriculum and pedagogy. Since assessments are necessary, some researchers examined how they can be created in order to properly measure historical thinking.

Exemplar Article #1

Sarah McGrew and her team of Stanford researchers created short assessment tasks that measured students' ability to search for, evaluate, and verify online information. McGrew et al. work within the Stanford History Education Group (SHEG), which has developed online assessment tools that measure students' historical thinking ability. McGrew created assessment tasks for her research study "Can Students Evaluate Online Sources? Learning from Assessments of Civic Online Reasoning" (2018), which were similar to history lesson plans that SHEG authors created.¹ Although the study focuses on civic reasoning, many of the inquiries are similar to historical thinking because they use many of the same inquiry skills found within the historical thinking skills, like sourcing a document or artifact.

McGrew created fifteen assessment tasks, which focused on three constructs; "Who is behind the information? What is the evidence? What do other sources say?" These questions are similar to Wineburg's (2001) historical thinking skills, which are respectively, sourcing, critical thinking, and corroboration. 405 middle school students, 348 high school students, and 141 college students from twelve different states comprised the participant pool, and researchers collected 2,616 responses from this group. Once given the analysis sheets and online sources, students struggled to successfully evaluate online claims, sources and evidence. McGrew believed curriculum materials must be better in order to support students "civic online reasoning

¹ These assessment tasks can be seen at <https://sheg.stanford.edu/history-assessments>.

competencies” (McGrew, 2018, p. 165-166). In the case of the McGrew study, “better” seemed to mean “just need to exist.”

The McGrew study is helpful in understanding how to measure historical thinking skills because it provides a measuring framework for assessing online civic skills. McGrew measured civic skills in a similar way to how SHEG researchers did historical thinking skills. For example, McGrew used the exact same categories, such as “sourcing” that SHEG used in their Beyond the Bubble history assessments. It is beneficial to use the same types of skills in both civics and history because it means educators can use common assessments across social studies disciplines. Having common assessment measurements benefits curriculum like the C3 Framework because it demonstrates how it is possible to vertically align assessment across social studies disciplines.

Along with the benefit of vertical alignment, assessments that measure skills are necessary because students misinterpret information. McGrew identified commonly held misinterpretations of students, such as always trusting “news” sources, even when they are clearly biased (2018, 193). These types of misinterpretations exist across social studies disciplines. If teachers want to ensure their assessments are going to improve cognitive processes, then they must collaborate in how they teach and assess skills. This collaboration is imperative as students only have an average of one year of civics and will take history sporadically throughout their secondary education. SHEG and McGrew are developing reasoning and thinking skills, which could be the uniting force to vertically align social studies departments. Researchers must continue to improve and model pedagogical methods like scaffolding to encourage historical thinking as a standard practice.

Exemplar Article #2

Gabriel A. Reich has a high school history teacher background, where he grew increasingly frustrated with standardized testing over historical knowledge. Reich is also a historian, who focuses on how Americans, especially the young, learn about the Civil War and how myths form a significant part of our historical consciousness. Due to his background and frustrations, it is not a surprise that Reich examined how high school students choose answers on a set of multiple-choice questions. Reich wanted to know if students used historical reasoning (Reich's term) when they selected A, B, C, or D. The researcher focused on a class of urban 10th grade students who had to take a high-stakes exam at the end of the year in order to earn a social studies credit that they needed to pass high school. Reich used the questions from New York State's Global History and Geography Regents Exam, which is a required test in order to complete high school.

Based on students' answers, and interviews with students afterward, Reich determined that students were using test-wise thinking skills to select correct answers. For example, in many of the interviews, the researchers found that students would eliminate answers because they used a similar response on a different question, or the student knew a certain name did not fit the era they were studying. In these situations of answering test questions, historical thinking and factual recall played little part in how students choose correct answers. Reich did not see students using skills like sourcing, corroboration, continuity, or contextualization when they answered their multiple-choice questions.

Reich's research suggests that multiple choice tests do not accurately measure students' knowledge of history. It would be useful to use Reich's methods of examining multiple choice questions to examine the tests that VanSledright (2011) created. VanSledright created multiple

choice answers that he weighted, some were more correct than others, and students analyzed all of the answers to choose the most accurate one. Also, Reich developed useful categorizations for the skills that students used to select the correct answer. The skills were “test-wiseness,” “literacy,” and “domains-history content.” If the question is “how do we measure historical thinking skills,” then it is beneficial to know what may interfere with measuring them and what does not qualify as a skill. If students use skills outside of historical thinking to answer questions, then researchers and teachers must identify and manage those skills in a way where they do not interfere with assessment.

Reich’s study serves as a warning sign before designing standardized multiple-choice exams. The research suggests that multiple choice tests do not assess historical thinking, but they also do not accurately measure a student’s ability to recall content information. The ineffectiveness of standardized multiple-choice assessments is not a new revelation, but it is important to state because governments and teachers subjugate millions of students to these exams each year. This is especially disturbing in social studies courses because many states do not require standardized assessments of historical thinking or knowledge of history, but many social studies teachers still rely on exams that solely contain simple recall of factual information. Despite not having standardized tests, the testing culture has adversely affected social studies teachers enough where they have created their own standardized testing regime.

Question 2: How can we take valid data from assessments over inquiry and historical thinking?

In the last section analyzing the first question, I mentioned how DBQ assessments cannot accurately measure specific cognitive processes, like the historical thinking skill “sourcing.” Additionally, Reich demonstrated how students use unintended thinking processes to answer multiple choice questions. Exacerbating the issue of creating tests, confounding factors can

compromise the data pulled from tests. For example, Adam Wallace examined motivation and belief in oneself when examining National History Day projects (Wallace, 1987), in which Wallace suggested that students who believed in their historical thinking ability did better on assessments. Additionally, Lee and Ashby focused on students' interpretation of tone, theme, and timescale as they grew older; therefore, a student's age might compromise their ability to think historically. Due to the complexity of assessing historical thinking, assessment creators must be deliberate in identifying what they are measuring and the possible issues with their evaluations.

Exemplar Article #1

Sam Wineburg has been working on being deliberate since the late 1990s; he has routinely published work on the topic over the last 30 years (Wineburg, 1991; 1997, 2001; 2009; Smith et al., 2019). Wineburg received his doctorate in Psychological Studies in Education, which is in part why he is thoroughly invested in investigating how students think about and learn history. Examples of Wineburg's investment are his numerous publications, one of which is his book *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past* (2001). In the book, Wineburg examines many reasons why students think about the past differently, such as the power of one's gender on historical thinking, which may affect how students measure when completing assessments that test historical thinking.

Wineburg wrote a chapter called "Picturing the Past" (2001) where he focused on the question: "how do boys and girls picture the past?" The researcher asked students to draw pictures of different historical figures, like Pilgrims, Western Settlers, and Hippies, in order to see how they pictured these people. Wineburg quantified the images based on gender, number of people, and types of actions the historical figures were committing. Additionally, Wineburg and his assistants conducted interviews to let students explain their reasoning behind the drawings.

Male students drew predominantly male characters, isolated or alone, and were more likely to be engaged in violence. Girls were more likely to draw female characters than boys, but their female to male ratio was 50/50. Girls also drew more groups of people such as families. Most concerning was the girls' propensity to fill their “historical world” with more men than would be realistic, like drawing a community square with only men. The researchers wondered, do girls do this when they are reading textbook accounts of historical events. If girls and boys have different outlooks of gender in history, how might this affect their historical thinking? Would it be fair to examine boys and girls using the same prompts and rubrics?

If young girls and boys are “seeing” a different historical world, then they will likely interpret primary sources and historical arguments differently. If this is the case, then gender may benefit or hinder students’ mastery over concepts. For example, a female student may be less likely to disassociate violence or discrimination from other actions that historical people committed. This is problematic because girls may conclude historical figures and events are invalid sources to use in arguments because they do not meet present moral standards, like American Founding Fathers owning slaves, Free Blacks settling on Native American lands, blaming Adolf Hitler for Germany’s anti-Semitism. Girls are not the only ones at risk, students from certain religious groups could believe that historical figures such as Thomas Jefferson or Martin Luther King Jr., who despite their many accomplishments, most likely committed adultery.

When designing assessments, creators must be careful to isolate cognitive processes from each other. Test creators can bracket historical thinking skills by breaking down current historical thinking inquiry types, like sourcing, into several different sub-skills (Appendix A). One such sub-skill is “identified the category of the source.” Assessments must explicitly

measure this sourcing subskill, along with the six others, in order to get a more accurate understanding of how well students are doing with the more general skill of sourcing. Furthermore, there needs to be an understanding of how students stumble into theoretical pitfalls, and how mastery may look differently for various groups of students. Due to these reasons, researchers should design and research scaffolds in order to help students reach mastery over skills and concepts. For example, a simple pedagogical scaffold would be to model analysis of primary documents.² A curriculum scaffold would be to have multiple types of the same reading, which would differ in reading level and/or theme.

Question 3: What scaffolding exists when testing for historical thinking and how can we improve it?

Scaffolds must be incorporated when researching the mastery of historical thinking skills. Scaffolds are the curriculum designs or pedagogical methods used to help students reach mastery over a certain skill or set of information and producing scaffolds is a time-consuming challenge. Creating scaffolds does not stop at teachers adapting curriculum for differences in reading levels or learning disabilities. There are additional socioeconomic, political, and natural circumstances affecting students that teachers must address by scaffolding curriculum. A United States teacher may need to adapt curriculum for students who do not speak English, are from cultural groups which represent “the enemy” in dominant, conservative, American narratives, like American Indians and Muslims, and students who simply cannot afford to ride the bus in the winter (Attewell, 2011, Renn, 2013). There are additional circumstances than the ones listed and teachers knowing about these confounding factors does not help when they have little time to build appropriate scaffolds. The education system is unlikely to change soon enough to meet the

² An example of this modeling can be watched at <https://www.youtube.com/watch?v=ib8R6T4qsJM&t=54s>

needs of diverse learners; social studies educators must individually implement curriculum and pedagogy that will help scaffold assessment measuring efforts.

After reviewing fifteen research studies over creating inquiry assessments, I found that some researchers address scaffolding in some way. For instance, in Monika Waldis's (2015) research study, German students mastered thinking skills with greater ease when they analyzed history that they were familiar with, such as "The Nazi Boycott of Jewish Businesses," compared to "Trade Relations with Japan." Waldis found students are more comfortable using historical thinking skills with familiar topics, which can support their willingness to take theoretical leaps necessary for historical thinking. Waldis concluded that teachers should use familiar topics as a scaffold to help students learn new historical thinking skills. Although Waldis and a few others mention scaffolding assessments of historical thinking directly or indirectly, researchers have not considered scaffolding enough to help practitioners create learning models for differentiated classrooms.

The lack of consideration towards practical application of scaffolding historical thinking assessments underlies a deeper issue. Teachers are struggling to apply historical thinking to their curriculum because their students are at different cognitive levels and possess different identities and backgrounds. For example, teachers may have a difficult time applying Waldis's research because there was no suggestion as to how assessments could be familiar to all students. Waldis did not describe how her finding of *familiarity promoting superior cognitive development* could apply to helping students progress through a historical thinking model. The two following exemplar articles also demonstrate how researchers did not consider the differentiated needs of students nor ideas on scaffolding while researching historical thinking.

Exemplar Article #1

Examination of the article “Fostering Analysis in Historical Inquiry Through Multimedia Embedded Scaffolding” (2008) suggests researchers do not consider how familiar students are with the historical topic. In their article, David Hicks and Peter Doolittle developed a strategy for historical thinking called SCIM-C. The strategy stands for “summarizing,” “contextualizing,” “inferring,” “monitoring,” “corroborating,” and “interpretation.” These historical thinking skills are similar to how Sam Wineburg, Sarah McGrew, and Monika Waldis define their skills. Students and teachers can use the SCIM-C strategy as a rational system to answer historical questions and analyze documents and artifacts. First publishing their ideas for the strategy in 2004, Hicks and Doolittle join other scholars in arguing for a greater evidence-based approach in history education. In their study, Hicks and Doolittle report their findings and answer the question; does the SCIM Historical Inquiry Tutorial foster the development of historical source analysis?

Seventy-seven college undergraduates compromised the study; they were enrolled in a general studies health education course. Researchers chose participating students from the health course because they would have little knowledge of historical procedures. The study introduced the SCIM strategy (the researchers removed the C for this study) to the students over three instructional periods, and researchers assessed student’s knowledge using a *single* open-ended question. Based on the teaching of the SCIM strategy, many students applied their newfound skills as part of a cognitively sophisticated process of analyzing sources.

Despite the success of numerous students, Hicks and Doolittle found that students applied historical thinking skills unevenly. This unevenness could be due to students not receiving differentiated assessments. Using Monika Waldis’s theory on familiarity, students could have

underperformed because they lacked knowledge of the historical time period used for the assessment questions. There is a chance that other confounding factors could have skewed data, like the factors mentioned under the second question of this paper. Whether it was lack of familiarity with historical topics, or something else, researchers and teachers should consider scaffolds to help students equitably reach mastery. If this is the case, then Hicks and Doolittle have data that does not truly show mastery of isolated thinking skills, but more a relationship between skills and knowledge of historical content. If K-12 teachers used the same strategy as Hicks and Doolittle to assess younger students, then they would need to differentiate the content in the assessments. Just one of these differentiated scaffolds would be allowing students to choose content that the teacher will use to assess them.

Exemplar Article #2

Along with familiarity with a topic, the identity and background of someone can affect their ability to master historical thinking skills. Instead of seeing knowledge and perspective as affecting the ability to historically think, Peter Lee and Rosalyn Ashby attributed age as a more prominent factor in their study “Progression in Historical Understanding Among Student Ages 7-14” (2000). As one of their central tasks, Lee and Ashby examine how students change their perceptions of history as they age. The philosophy of this research falls in line with Jean Piaget’s ideas of students learning through a cognition model as they grow older, with strict limits on what a student can do at a certain age.

In the main investigation, Lee and Ashby collected responses from 320 children between the ages of seven and fourteen. They also interviewed 1/3 of the students in order to determine the reasoning behind their interpretations of history. Students responded to questions by examining secondary source accounts of Romans in Briton, but each story differed in theme,

tone, and timescale. As students got older, they described differences in the stories based on their dates and abstract concepts. Lee and Ashby took this observation and theorized that students' progress in their formation of history as they age. Lee and Ashby use several practical codes when measuring students' historical thinking. Some of these codes were "selection," "legitimate viewpoint," "intentional distortion," "mistakes," and "opinion unexplained" (Lee & Ashby, p. 58).

The disadvantage with Lee and Ashby's findings is they do not consider how confounding factors like familiarity with a topic, student perspectives, or the amount of knowledge a student possesses may affect their ability to historically think. Simply using age as a factor in how students develop prevents educators and researchers from developing scaffolds to assist students in mastering historical thinking skills. Since Lee and Ashby believe age is a factor in how well students can understand abstract concepts, they perhaps did not see a reason to formulate any steps or methods that would help students progress through a cognitive learning model. They also did not apply existing models, like Jerome Bruner's *spiralizing curriculum*, which has been effective in helping students learn deeper concepts (Bruner, 1960). Researchers benefit from a model like Bruner's because if a researcher or teacher is able to measure progression of historical thinking, then scaffolds can be likely built between each step to help with the advancement of cognition.

Question 4: How does inquiry assessment motivate students?

A reason teachers scaffold a lesson or activity is because it motivates students to accomplish tasks which lead to mastery over skills, concepts, and information. In essence, if students believe they can climb the mountain, even if it is difficult, they are more likely to start down the trail. Like scaffolding, inquiry assessments also motivate students because it involves

them learning knowledge through the active creation of it, not just simply receiving it. Inquiry assessments can be brief prompts that ask students something like; “Should Abraham Lincoln have signed the Emancipation Proclamation?” Questions like this require students to go beyond simple recall of information, they must think about why historical actors committed their actions.

Inquiry assessments can also be large-scale, such as the National History Day projects that thousands of students complete each year. In these projects, students form questions, read sources, and develop a historical argument. Throughout these extended projects, students engage in deeper cognitive processes in order to answer inquiry questions. Inquiring is often more strenuous than receiving a lecture or copying notes, yet inquiry activities that promote the advancement of cognitive processes tend to motivate students more. This is especially true during extended inquiries like the kind National History Day provides.³ Through inquiry, students create knowledge by relating new information to their own perspectives, beliefs, and ideas. It is from forming knowledge through the combination of self, new information, and cognitive processes that motivates students. National History Day, and other large-scale projects, provide opportunities for students to create their own inquiry assessments, thus adding another layer of agency and motivation.

Students being comfortable with topics is key when assessing their historical thinking ability. Students shaping inquiry assessments can provide motivation in an environment where students are uncomfortable or uncertain about the skills they are trying to obtain. For example, at the beginning of a project, students likely will not understand the importance of historical thinking skill like “contextualization”, or what it truly means to detect bias in a secondary

³ National History Day is an organization that sponsors a competition between project-based learning (PBL) style history projects. The organization also promotes a specific type of curriculum for teachers to use.

source. Students need perseverance to understand abstract topics, which teachers can promote by allowing students to shape curriculum and the assessments over their inquiry.

Exemplar Article #1

David Wallace's qualitative analysis of National History Day is an apt research study that demonstrates the power of motivation. Wallace was a professor of history at Cleveland State University in Ohio. In the mid-1980s, National History Day was operated in Cleveland, and Wallace was one of the earliest history professors who helped the program. In 1987, Wallace completed a qualitative research project and wrote the article "The Past as Experience: A Qualitative Assessment of National History Day." The purpose of this research was to describe and evaluate National History Day as an education program, and to describe its implications for teaching history.

Wallace sent a questionnaire to 1,500 students who were state winners in the National History Day program. The questionnaire focused on the perception of the students on how they viewed their cognitive ability because of National History Day. Not only did student participation in National History Day result in increased excitement and engagement with the history curriculum, students believed the program fostered new skills useful for historical research. These skills were evident in the explanations of the students, especially when they described how their theories and evidence involved forces of culture, politics, and economics. The most significant limitation to Wallace's study is the survey was sent only to state winners, who most likely possessed social advantages compared to their less victorious peers. The lack of a more balanced participatory group may skew data since Wallace did not include students who did not move on past the local and state contests.

Wallace's study connects learning historical skills, historical projects, and student motivation to one another. The importance of students obtaining historical skills does not lie within students receiving some score on a test, although there are studies that suggest National History Day, and other extended history projects, do support better standardized test scores (Monaco et al., 2009; Parker et. al, 2013; Sloan & Rockman, 2010). The more important benefit is students creating products that demonstrate their ability to inquire through historical thinking. Finally, completion of the projects motivated students because they were proud of the skills they had learned, and they wanted to demonstrate them outside of their classroom.

Question 5: Why Is It Difficult to Assess Historical Thinking and Inquiry Based Learning?

Although I do not have much space left, I believe it's important to briefly write about this fifth question. It is laborious to measure students' cognitive abilities in history. When considering assessing historical thinking and inquiry, teachers face a lack of instructional time to assess, pressure to satisfy standardized high-stakes tests, and cultural issues with focusing on inquiry over recall/memorization. Monika Waldis's research study elaborates on another problem: the coding mechanisms that assessors need to measure historical thinking are convoluted.

Exemplar Article #1

The use of how narrative changed in quality and structure based on a student's ability to think historically interested Waldis. In order to assess students' historical thinking, Waldis asked German students to produce a narrative based on analysis of primary sources. The study used two topics, "Trade Relations with Japan" and "The Nazi Boycott of Jewish Businesses," which were seldom and often taught respectively in Germany. The teacher gave students exam booklets with primary sources inside of them and gave students as much time as they needed to complete

the analysis questions. The participant sample included 193 high school students from nine classrooms in three different towns. Waldis and the team of researchers developed quality features of narrative assessments, which included concepts such as “value judgments” and “quality of making historical references” (Waldis, 2015, p. 122).

Student answers were structurally heterogeneous; for example, some answers were one sentence long while the longest was thirty-three sentences. The raters of the narrative answers distributed low numbers in the category of *normative cogency* because students did not support their values with evidence and reasoning from the primary sources. This finding is disturbing as it suggests history courses are not educating students to provide evidence to support their own opinions. If teachers are not teaching the concept of using evidence, it could be because the task is more difficult than researchers realized. Waldis used highly detailed coding mechanisms, similar to other research articles but much more complex, and categories when measuring student answers, which resulted in discovering that responses lacked certain qualities. If these complex coding mechanisms are the only way to accurately measure students' acquisition of thinking skills, then teachers will not have the training or background to successfully lead students through historical thinking exercises.

Conclusion

Considering educators did not treat historical thinking seriously until the mid- to late-1990s, there has been an impressive amount of research in how we measure and define historical thinking, on what data researchers can lift from research experiments, and in how inquiry motivates students to succeed. There were even pleasant surprises after I analyzed the studies over historical thinking. Before I examined all of the research, I believed there was an overabundance of focus on high school and undergraduate students in their abilities to

historically think. My position was due to the lesson plans available at historical thinking websites like Beyond the Bubble, SCIM-C, The Library of Congress, and The National Archives, which nearly solely cater to older students. In this literature analysis, I found a variety of research studies that used participatory students in elementary school, middle school, high school, and college undergraduate courses.

Research on historical thinking is incomplete because it is difficult to implement the recommended practices into most of K-16 education. Although researchers have created categories of historical thinking, they have not scaffolded historical thinking skills, nor found ways to simplify them for younger students, or students not at grade level. For instance, different groups of researchers have described how sourcing is an important skill for students to have (McGrew, 2018, Wineburg 2001, Hicks 2008). The assessments I reviewed did not explain sourcing beyond a short definition, along with some complex examples of how teachers can assess sourcing. Teachers will need scaffolded strategies in order for students to work with these skills, such as modeling or simplified versions of sourcing assessments (Shemilt, 2018).

Currently, only a few studies addressed scaffolding historical thinking and how assessment can be tailored to meet student needs. Researchers examining students from multiple age groups gives me hope that researchers will soon conduct studies that examine the relationship between scaffolding practices and the degree to which students can master historical thinking. There are two potential reasons why such studies may not be forthcoming.

First, there simply has not been a lot of time for the inquiry models and beliefs to seep into the education system. For example, the unifying standards document, the C3 Framework, was published in 2013, only seven years ago. For all their goals, the authors of the C3 Framework mainly wanted to create a document that could help social studies departments

vertically align their classrooms based on inquiry. Second, the coding mechanisms, assessment rubrics, and overall process of measuring historical thinking is arduous. Teaching students how to historically think is difficult. Without context, students cannot base their development of questions or their conclusions on social realities of history, economics, geography, or whatever teachers use as curriculum.

Final Questions and Future Research

When considering questions for future research, I wonder what methods of scaffolding historical thinking have researchers tested and what were the results? Based on those results, what additional scaffolds do teachers need? My principal assumption is scaffolding should start with general education and then branch out to other areas. For example, a teacher assessing a specific historical thinking skill, like sourcing, could use alternative resources depending on students' reading levels. One way of doing this is alternating the reading Lexile level of certain primary source documents, which will enable students with a lower reading level to access the key information of the document. These students would then be less distracted by the words that are no longer prevalent in modern language and can better show their analysis abilities in identifying the author and detecting their bias. Finally, there seems to be many confounding factors, like familiarity with a topic, that can skew data of mastering historical thinking skills. Researchers must identify and control these confounding factors in order to get more accurate results.

Appendix A

Sourcing can be classified as a skill that focuses on the analysis of a document or artifact's creator and the circumstances of its creation. There are sub-skills to this type of inquiry which are listed below.

Skill #1: Identified the category of the source.

Skill #2: Identified the date and creator of the source.

Skill #3: Identified if the source is primary or secondary.

Skill #4: Described the audience of the source.

Skill #5: Described the purpose of the source.

Skill #6: Described characteristics, bias, or perspectives of the source's creator.

Skill #7: Described the trustworthiness of the source.

References

- Attewell, Paul and David Lavin (2011). "The Other 75%: College Education Beyond the Elite." In E. Lageman's and H. Lewis's (Eds.) *What is College For? The Public Purpose of Higher Education*. New York: Teachers College Press.
- Bruner, Jerome (1960). *The Process of Education*. Cambridge: Harvard University Press.
- Corliss, Stephanie B., and Marcia C. Linn (2011). Assessing Learning from Inquiry Science Instruction. In D. Robinson and G. Schraw (Eds). *Assessment of Higher Order Thinking Skills*, 219-243.
- Hamblin, Taylor S. [History Forge]. (2019). *Mock Senate Explanation* [Video] YouTube. Retrieved from <https://www.youtube.com/watch?v=7I-ADV LZz4s&t=776s>
- Hicks, David, and Peter E. Doolittle (2008). Fostering Analysis in Historical Inquiry Through Multimedia Embedded Scaffolding. *Theory and Research in Social Education*, 36(3), 206-232.
- Lee, John (unpublished chapter). *Assessing Inquiry*. In the author's possession.
- Lee, P., & Ashby, R. (2000). Progression in Historical Understanding Among Student Ages 7-14. In P.N. Stearns, P. Seixas, and S. Wineburg (Eds.), *Knowing Teaching & Learning History: National and International Perspectives*, 45-94. New York: New York University Press.
- Levesque, Stephane and Penney Clark (2018). Historical Thinking: Definitions and Educational Applications. In S. Metzger's and L. Harris's (Eds.) *The Wiley Handbook of History Teaching and Learning*, 119-148. Hoboken, NJ: John Wiley & Sons.
- McGrew, Sarah, Joel Breakstone, Teresa Ortega, Mark Smith, and Sam Wineburg. (2018). Can Students Evaluate Online Sources? Learning from Assessments of Civic Online Reasoning. *Theory & Research in Social Education*, 46(2), 165-193.
- Monaco, Giuseppe, Bo Lu, & Megan Wood (2009). Impact of the National History Day in Ohio Program on Students' Performances: Pilot Evaluation Project. *Journal of Museum Education*, 34(1), 79-96.
- National Council for the Social Studies (NCSS). *The College, Career, and Civic Life (C3) Framework for Social Studies State Standards: Guidance for Enhancing the Rigor of K-12 Civics, Economics, Geography, and History* (2013). Silver Spring, Maryland: NCSS.

- Parker, Walter C., Jane Lo, Angeline Yeo, Sheila Valencia, Diem Ngyuyen, Robert Abbott, Susan Nolen, John Bransford, Nancy Vye (2013). Beyond Breadth-Speed Test: Toward Deeper Knowing and Engagement in an Advanced Placement Course. *American Educational Research Journal*, 50(6), 1424-1459.
- Reich, G. A. (2009) Testing historical knowledge: Standards, multiple-choice questions and student reasoning. *Theory & Research in Social Education*, 37(3), 325-360.
- Renn, Kristen and Robert Reason. "Characteristics of College Students in The United States". In K. Renn's and R. Reason's *College Students in the United States: Characteristics, Experiences, and Outcomes*, 3-27. San Francisco, CA: Jossey-Bass.
- Shemilt, Denis (2018). Assessment of Learning in History Education: Past, Present, and Possible Futures. In S. Metzger's and L. Harris's (Eds.) *The Wiley Handbook of History Teaching and Learning*, 449-472. Hoboken, NJ: John Wiley & Sons.
- Selwyn, Doug (2014). *Why Inquiry?* In E. Ross's *The Social Studies Curriculum*, 267-288. New York: State University Press.
- Sloan, Kay & Saul Rockman (2010). National History Day Works: Findings from the National History Program Evaluation. Rockman et al.
- Smith, M., Breakstone, J., & Wineburg, S. (2019). History Assessments of Thinking: A Validity Study. *Cognition and Instruction* 37(1), 118-144.
- Stanford History Education Group (n.d.). The First Thanksgiving. Retrieved from <https://sheg.stanford.edu/history-assessments/first-thanksgiving>
- The Glossary of Education Reform (2015). Scaffolding. Retrieved from <https://www.edglossary.org/scaffolding/>
- VanSledright, Bruce (2011). *The Challenge of Rethinking History Education: On Practices, Theories, and Policy*. New York: Routledge.
- VanSledright, Bruce (2014). *Assessing Historical Thinking & Understanding: Innovative Designs for New Standards*. New York: Routledge.
- Virginia Tech. *SCIM-C: Historical Inquiry*. Retrieved from <http://www.historicalinquiry.com/>.
- Waldis, Monika, et al. (2015). Material-Based and Open-Ended Writing Tasks for Assessing Narrative Among Students. In K. Ercikan and P. Seixas (Eds.), *New Directions in*

- Assessing Historical Thinking* (pp. 117-131). New York: Routledge.
- Wallace, David Adams (1987). The Past as Experience: A Qualitative Assessment of National History Day, *The History Teacher*, 20(2), 179-242.
- Wineburg, Sam (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 93-87.
- Wineburg, Sam (1997). Beyond “breadth and depth”: Subject matter knowledge and assessment. *Theory into Practice*, 36(4), 255-261.
- Wineburg, Sam (2001). Picturing the Past. In *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past* (pp. 113-136). Philadelphia: Temple University Press.
- Wineburg, Sam (2009). Tampering with history: Adapting primary sources for struggling readers. *Social Education*, 73(5), 212-216.