



EVALUATION OF WORLD HISTORY PROJECT

Emi Iwatani, PhD, Angela Hardy, MA, Barbara Means, PhD,
Shelton Daal, MPP, and Xin Wei, PhD

December 1, 2023

A Digital Promise Report

Evaluation of World History Project



Suggested Citation

Iwatani, E., Hardy, A., Means, M., Daal, S., & Wei, X. (2023). Evaluation of World History Project. Digital Promise. <https://doi.org/10.51388/20.500.12265/199>

Acknowledgments

This research was supported by funding from Gates Ventures. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funder. We thank the OER Project team, especially Dr. Rachel Phillips, Mike Weisensel, and Angelina Meadows, for their collaboration in conducting this research, Dr. Abby Reisman for valuable comments on the draft report, as well as Bob Regan, Gabrielle Lue, our assignment scoring collaborators, and all the teachers and students who shared their views and experiences with us.

Contact Information

eiwatani@digitalpromise.org and ahardy@digitalpromise.org

Digital Promise
1001 Connecticut Avenue NW, Suite 935,
Washington, DC 20036

<https://digitalpromise.org/>

Cover image: Detail of North Africa from the 'Catalan Atlas', attributed to the mapmaker Cresques Abraham. © Royal Geographical Society via Getty Images.

Author Biographies



Dr. Emi Iwatani is a Senior Learning Sciences Researcher at Digital Promise who leads educational design research and evaluations in secondary school settings. Her recent projects partner with educators to deepen discussions in high school world history, design heritage-resonant computing pathways, and reimagine assessment. Dr. Iwatani was formerly a science and writing teacher at Boston Arts Academy, and Education Director at Voices Against Violence Pittsburgh. She holds a Ph.D. in Education Research Methodology and an M.A. in History and Philosophy of Science, both from the University of Pittsburgh, an M.A. in Philosophy from Boston University, an M.Ed. in Science Education from Wake Forest University, and a Sc.B. in Biology from Brown University.



Angela Hardy is an Education Researcher in Digital Promise's Center for Inclusive Innovation. In her current role, she co-designs and evaluates instructional coaching programs and leads co-design of curricular resources to support deeper discussions in world history classrooms. Before joining Digital Promise, for eight years, Angela taught World History, U.S. History, World Cultures and English in schools with high populations of students from low income backgrounds. She holds a B.A. in history and French from the University of North Texas, a 4-8 generalist certification, 8-12 history and social studies certifications, and an M.A. in International Education Policy Analysis from Stanford University.



Dr. Barbara Means, Principal Learning Sciences Researcher at Digital Promise, studies the implementation and impacts of innovative education approaches supported by digital technology. Currently, she is working with postsecondary instructors to incorporate equity-minded pedagogy and digital learning technology into high-enrollment gateway courses. Dr. Means has authored or edited seven books on learning and technology, and has served on many study panels related to education research for the National Academies of Sciences, Engineering and Medicine, including the panels that produced *How People Learn I* and *How People Learn II*. Dr. Means earned her undergraduate degree in Psychology from Stanford University and her Ph.D. in Educational Psychology from the University of California, Berkeley.



Shelton Daal is a Senior Education Research Associate with the Learning Sciences Research team at Digital Promise. His current research is centered in co-design, working with educators to develop innovative approaches to world history curriculum design and to reimagine evaluation through the use of alumni survey data. His former work includes serving as a research associate for the Judge Alexander Williams, Jr. Center for Education, Justice, and Ethics, where he conducted a needs assessment on the digital divide in the Prince George's County K-12 school system. Shelton has a Master of Public Policy with a specialization in social and education policy and a Bachelor of Arts in Sociology from the University of Maryland, College Park.



Dr. Xin Wei is a Senior Quantitative Researcher at Digital Promise's Learning Sciences Research team specializing in experimental design, statistical analysis, and machine learning techniques for evaluating and enhancing instruction, interventions, assessments, and policies. Her research spans various topics aimed at improving outcomes for students from underserved backgrounds. She holds a Ph.D. in Educational Psychology with a concentration in statistics and measurement as well as an M.S. in Statistics, both from Stanford University.

Table of Contents

Executive Summary.....	iii
Introduction.....	1
Background.....	1
Evaluation questions and population.....	4
Methods.....	5
Findings.....	11
A. Whether World History Project helps teachers provide more learning opportunities for historical thinking skills.....	11
B. Whether students who use World History Project learn more historical thinking skills relative to business-as-usual curricular activities.....	18
C. Explanations for the patterns observed in learning opportunities and outcomes.....	25
D. Whether students who learn from World History Project find world history more relevant, engaging, and/or perspective changing.....	30
E. Usability of World History Project.....	39
Discussion & Implications.....	42
Summary of findings and discussion.....	42
Implications for World History Project curriculum developers.....	43
Implications for teachers and school districts considering adoption.....	45
Implications for future research.....	46
Concluding remarks.....	47
References.....	48
Appendix A: Methodological details.....	52
Sample & data collection details.....	52
Qualitative data analysis details.....	55
Appendix B: Data tables.....	56
Rubric scoring reliability.....	56
Historical thinking skills rubric score descriptive statistics.....	56
Hierarchical linear model result details for rubric score comparisons.....	61

Executive Summary

Purpose and background

Our study examines the potential for *World History Project*, a free online high school world history curriculum, to produce learning benefits, particularly around historical thinking. This year-long curriculum, created by OER Project in partnership with educators and historians, includes units with readings, videos, and historical thinking skill activities. It launched in November 2019, and the site is currently accessed by thousands of teachers across the United States.

Over the 2022-23 school year, we examined the curriculum's impact in 9th or 10th grade on-level or honors world history classes, in public schools across the United States. Our evaluation questions were:

1. Does *World History Project* help teachers provide more opportunities for students to learn historical thinking skills, relative to business-as-usual curricular materials?
2. Do students using *World History Project* learn more historical thinking skills relative to those learning from business-as-usual curricular materials?
3. Why might have *World History Project* impacted teachers and students in some skills and not others?
4. Do students using *World History Project* find world history more relevant, engaging, and/or perspective changing?
5. How useable is *World History Project*, especially for teachers new to teaching world history?

Methods

Our overall study approach was to compare characteristics and outcomes of two groups of curriculum implementations: classrooms that used materials from *World History Project* for at least 60% of their classroom instructional time (e.g., “*World History Project* adopters,” “students learning from *World History Project*”), and classrooms that did not rely on any resources from *World History Project* (e.g., “comparison curricula users,” “students learning from comparison curricula”). The comparison group consisted of teachers who were drawing not just from traditional textbooks, but also from a variety of web resources (e.g., Newsela, Teachers Pay Teachers, Stanford History Education Project, Khan Academy, Smithsonian Learning Lab).

We compared these two groups over the 2022-23 school year. To answer the first two research questions, we collected 268 curricular activity samples and 1,509 samples of student work from 25 teachers, and compared how these scored on 4-point “historical thinking skills” rubrics that we had developed and for which we had collected some validity evidence (Hardy & Iwatani, 2021; Hardy et al., 2021/2023; Iwatani et al., 2021). These rubrics indicated the relative emphasis that each artifact placed on: historical argumentation, historical causation, historical comparison, historical contextualization, continuity and change over time in history, and sourcing. Each artifact was triple-scored by trained scorers who were blinded to the study

condition. A rubric score of “0” indicated that the activity does not explicitly call for students to employ the skill, or that the student did not explicitly demonstrate use of that skill; while a score of “3” indicated that the activity calls for students to provide an extended explanation, or that the student explicitly demonstrated the skill to a large extent. We collected both “everyday activities” (activities for learning purposes completed in about one class period, e.g., notes, worksheets) and “summative activities” (extended activities to demonstrate learning, e.g., tests, projects, essays), and conducted separate analyses for each.

To answer the remaining research questions we conducted and analyzed teacher interviews (N=25), teacher surveys (N=10), student pre- and post-surveys (N=293), and two student focus groups. We also conducted deep dive examinations into the curricular activities and student work that were submitted.

Significance

Our study is significant in its relevance to research and practice.

Perhaps most importantly, *World History Project* plays a pivotal role in addressing a significant gap within the open educational resource curriculum market dedicated to world history. In essence, *World History Project* currently stands alone in providing comprehensive, skills-based, and freely accessible curricula for high school world history that is aligned to multiple course-specific content standards. OER Project’s unique commitment to this mission, their capacity to make continual and large-scale revisions, and substantial resources already dedicated to this undertaking, make a thorough examination of their work important.

Additionally, world history teaching and learning is understudied, with strikingly few empirical studies centered on secondary world history classrooms (Girard & Harris, 2018). Few studies have explored the impact of specific history curricula or programs on opportunities for historical thinking (Epstein & Salinas, 2018). Our study helps fill this knowledge gap, in part by applying rubrics for multiple dimensions of historical thinking to analyze curricular assignments and student work for its authentic intellectual demand. Related, our study is unique and important given its relatively large scale, comprehensive mixed- and multi-methods approach, high data quality, detailed findings and potential for impact.

We are optimistic that the knowledge generated by this report can be translated into improved tools and experiences for world history teachers and students, since as a provider of open-source, online material, OER Project is positioned to continuously refine their content and to improve their curriculum and professional learning with annual updates and overhauls.

Findings

Section A. Whether World History Project helps teachers provide more learning opportunities for historical thinking skills

We found statistically significant impact estimates on learning opportunities for historical thinking skills, favoring the use of *World History Project* in: summative activities overall ($M_{WHP} = 1.77$, $M_{Comp} = 1.29$, $\beta = .56$, $SE = .24$, $p = .018$, Hedges’ $g = .66$), *change and continuity over time in history* summative activities ($M_{WHP} = .43$, $M_{Comp} = .13$, $\beta = .31$, $SE = .14$, $p = .026$, Hedges’ $g =$

.70) and *change and continuity over time in history* everyday activities ($M_{WHP} = .24$, $M_{Comp} = .07$, $\beta = .18$, $SE = .06$, $p = .002$, Hedges' $g = .41$).

We also found seemingly notable, but not statistically significant, impact estimates on learning opportunities for *historical argumentation* summative activities and *historical comparison* summative activities. For these, the use of *World History Project* was associated with a .52 point and .23 point increase in rubric scores, and in differences of .39 and .30 standard deviations, respectively. No notable impacts were detected for the remaining historical thinking skills.

The data suggests that

- Teachers who adopted *World History Project* assigned summative activities with a higher emphasis on historical thinking skills, relative to teachers using business-as-usual curricula. They provided roughly comparable learning opportunities through the everyday activities.
- Teachers using *World History Project* provided more learning opportunities in *continuity and change over time in history* (through both summative and everyday activities), and possibly in *historical argumentation* and *historical comparison* (through summative activities), relative to teachers using business-as-usual curricula. They provided roughly comparable learning opportunities for the other historical thinking skills.

Section B. Whether students who use *World History Project* learn more historical thinking skills relative to those exposed to business-as-usual curricular activities

We found statistically significant impact estimates on student outcomes for historical thinking skills, favoring the use of *World History Project* in: *continuity over time in history* summative student work ($M_{WHP} = .21$, $M_{Comp} = .06$, $\beta = .15$, $SE = 0.06$, $p = .014$, Hedges' $g = .60$) and everyday student work ($M_{WHP} = .11$, $M_{Comp} = .04$, $\beta = .07$, $SE = 0.03$, $p = .032$, Hedges' $g = .36$).

We also found seemingly notable, but not statistically significant, impact estimates on student outcomes for *historical argumentation* summative activities, and summative activities overall. For these, the use of *World History Project* was associated with a .26 point and .32 point increase in rubric scores, and in differences of .39 and .36 standard deviations, respectively. No notable impacts were detected for the remaining historical thinking skills.

The data suggests that

- Students who learned using *World History Project*, relative to those who learned from business-as-usual curricula, demonstrated more understanding of *continuity and change over time in history*, in both everyday and summative assignments.
- Students learning from *World History Project* demonstrated comparable competency in the other five historical thinking skills when compared to students learning from comparison curricula.

Section C. Explanations for the patterns observed in learning opportunities and outcomes

Analysis of a subsample of lessons showed that *World History Project* activities that scored relatively high on continuity and change over time in history were unique, relative to the comparison activities, in placing intentional emphases on building this skill, providing an array of materials and scaffolds that appear to guide students to try out a process that historians might utilize for such analysis. For example, *World History Project* provides unit overview articles and mini-article series that are specifically designed for students to pick-up on important changes and continuities happening across large units of time.

The analysis also revealed some factors that were likely getting in the way of students' learning of *continuity and change over time*, helping to explain why the student outcomes were not as high as the opportunities provided. For example, two recurring conceptions intended to build skills on *continuity and change over time* appeared under-defined, especially for a beginner learner (and teacher), and in some cases may reinforce some misconceptions. We also observed potential issues with engagement and accessibility, elaborated in sections D and E.

Further series of analyses explored possible reasons teachers using *World History Project* provided comparable (but not necessarily more) opportunities for students to learn many of the historical thinking skills, relative to comparison teachers. Likely reasons included teachers needing more experience implementing the curriculum in order to become comfortable enacting it, and some difficulties related to the content and format/organization of *World History Project* that limited robust use by teachers and students (sections D and E provide related and further information).

Section D. Whether students who use *World History Project* find world history more relevant, engaging, and/or perspective changing

Student surveys and focus groups showed a somewhat negative trend for students who learned from *World History Project*. For example, they were more likely in the spring to describe their class as “boring,” and less likely to report that they want to know what lies behind the story when they study a conflict in history. Our data suggests that

- *World History Project* does not engage students more than business-as-usual curricula, at least when a teacher implements it for the first time.
- *World History Project* students' negative perceptions of their class, relative to students receiving business-as-usual curricula, appear to be related to the greater demands placed upon them for reading, writing, and original analysis.

This highlights a tension that often exists in mandated learning, between rigor and enjoyment. Ideally, schoolwork is both rigorous and enjoyable, but especially to developing adolescents, it is difficult to provide the right balance; the right amount of enjoyment to ease the strain of the rigor, and the right amount of rigor to ensure healthy development and growth.

Section E. Usability of World History Project, especially for teachers new to teaching world history

All interviewed teachers reported that they found *World History Project* to be overwhelming in their initial year of adoption, due to the abundance of available resources. Many also reported that they often felt the need to make substantial modifications to *World History Project* resources to suit their students' comprehension levels, and that the process of making these modifications can be time-consuming. A review of a purposive sample of 26 activities confirmed that teachers were making substantial modifications to the materials before presenting them to students, which fell under three main categories: (1) creation of assignment scaffolds, possibly to improve accessibility; (2) creation of slide decks to motivate students and explain assignment instructions; (3) transformation of PDFs to Word formats that teachers and students can more easily edit.

In spite of common challenges, some teachers expressed appreciation and reported positive outcomes from using the *World History Project*, including the opportunity to enhance student engagement and to present a more comprehensive view of global history.

Implications

Implications for Curriculum Designers

We recommend that curriculum developers keep many of the curriculum's core aspects intact, including its comprehensive global world history focus, intentional design to support historical thinking skills, and development approach of working closely and iteratively with educators and historians. In particular, we recommend prioritizing collaboration with teachers who represent a broad range of contexts and student populations when implementing the changes suggested below. In addition, we recommend they

- **Continue to consider and improve usability for teachers:** Consider ways to help teachers avoid feelings of overload during the initial phases of curriculum, such as providing a “quick start” overview of key resources to help new teachers avoid feeling overwhelmed with resource curation. Emphasize that not all resources need to be used, encouraging teachers to choose what aligns best with their goals. Make curriculum resources easily editable to support teachers' ability to adapt materials to better fit their context and students' needs.
- **Enhance engagement and accessibility for students:** Consider ways to support teachers in making their instruction more engaging and relevant. Possible avenues for this include highlighting activities that other world history teachers have found resonated the most with their students, adjusting material difficulty levels to address a broader range of student needs, and incorporating additional scaffolds in lesson plans.
- **Consider how historical thinking skill activities align with content-related goals:** Recognize that educators may allocate different amounts of time for skill-building based on perceived goals related to standards (e.g., a pressure to cover content). Consider how existing historical thinking skill activities align with content-related goals, and explore ways to make those connections stronger.

Implications for Practitioners

Educators seeking a world history curriculum that takes a truly global approach and embeds scaffolds for historical thinking skills may want to consider adopting the *World History Project*, while noting that as of the 2022-23 school year the curriculum had areas that could be improved.

We recommend that **social studies administrators** who adopt the curriculum take the following approaches:

- Clearly communicate why the curriculum was chosen and its benefits, as well as expectations around which specific resources and activities teachers are expected to adopt.
- Provide support and give teachers time for curriculum exploration before implementation.
- Consider appointing a curriculum head or experienced teacher for organizing and curating resources into a sample district course plan.
- Offer guidance on emphasizing specific historical thinking skills, considering vertical alignment across grade levels and meaningful connections to required content.
- Provide opportunities for professional development opportunities and communities of practice.

We recommend that **teachers** who adopt the curriculum take the following approaches:

- Understand that a holistic understanding of the curriculum structure takes time.
- Review example course plans and map out a general plan for the year.
- Align curriculum use with state standards and goals for students.
- Connect with the OER Project online community to gain insights from experienced teachers.

Implications for Researchers

This study has several implications for future research in history education. Teachers, even those who are experienced and motivated, struggle with teaching historical thinking skills, suggesting the need for more research on how to improve curricula and support teachers and students effectively. For example, how do teachers modify curriculum around historical thinking, and what are theoretical and practical implications? We believe these types of questions can be satisfactorily addressed only by research that positions practitioners and students at the center of the inquiry.

Additionally, finding the right balance between academic rigor and enjoyable, engaging teaching methods is essential. Investigating what motivates and engages high school students in their world history courses is vital for designing effective curricula.

The study also highlights the unique opportunities world history education offers for students to develop historical thinking skills (e.g., by comparing different contexts and enduring themes). More research is needed on how to design historical thinking activities for world history in particular, and how to do so in ways that students perceive as relevant and interesting.

Conclusion

This evaluation of the *World History Project* highlights both strengths and areas that could be improved. While the curriculum has shown promise in supporting teachers to give more emphasis to some historical thinking skills, it faces challenges in terms of usability and accessibility to teachers and students. The findings emphasize the importance of providing teachers with time, support, and resources to make the curriculum effective. Curriculum developers can help by prioritizing usability, refining the balance between rigor and enjoyment, and offering more support for effective adaptation by teachers.

Given that *World History Project* is the most comprehensive, skills-based, and cost-effective (free) source of learning materials for the discipline to date that is aligned to multiple course-specific state content standards, it has the potential for a significant impact. This study stresses the importance of centering the experiences and perspectives of teachers and students that the curriculum aims to support most in future efforts to refine the curriculum. We expect that our feedback will act as a driving force for ongoing efforts to refine *World History Project*, with the ultimate goal of empowering teachers to skillfully teach world history and fostering students' understanding and appreciation of the subject.

Introduction

Student acquisition of historical thinking skills is considered to be important by educators, social studies education scholars, and nationally-recognized social studies standards (Keirn, 2018; Lévesque & Clark, 2018), but detecting and measuring thinking skills is not easy (Seixas & Ercikan, 2015a; Shemilt, 2018). Our study examines the potential for *World History Project*, an online, year-long world history curriculum, developed by a group of educators and expert historians, to produce learning benefits, particularly around historical thinking. This curriculum, created by OER Project in partnership with educators and historians,¹ includes units with readings, videos, and historical thinking skill activities. It launched in November 2019, and the site is currently accessed by thousands of teachers across the United States.²

Background

Over the past few decades, the paradigm of teaching history has shifted. Instead of viewing history merely as a narrative filled with facts that students need to memorize and recite, scholars and educators now commonly emphasize the application of historical knowledge along with skills involved in “doing” history through engaging in disciplinary practices such as interpreting historical texts (Keirn, 2018; Lévesque & Clark, 2018). *World History Project*, an online high school world history curriculum published by OER Project in late 2019, was created to foster the development of such **historical thinking skills**, or **skills that students employ when they analyze and apply historical content knowledge to interpret the past** (e.g., make historical arguments, identify patterns over time, analyze cause-and-effect).

In 2018, we were invited by curriculum developer OER Project, to evaluate the impact of *World History Project*. At the time of our initial engagement, the curriculum was still under development and not publicly available. To learn about the curriculum, to discern whether (and when) it might be ready to be evaluated, and to conceptualize an evaluation design, we interviewed three curriculum developers and 11 world history teachers, including eight teachers who advised the curriculum design. We also examined literature and artifacts, including the curricular materials under development, pilot teacher comments and feedback, social studies frameworks, and articles on academic conceptions and learning progressions of historical thinking skills.

Through these activities, we came to understand that *World History Project* has a number of promising and distinctive features including these:

¹ For example, leading social studies education scholar Bob Bain was actively involved in designing the curriculum’s organization and driving narrative, and leading world historians (e.g., Trevor Getz) developed curricular content (e.g., articles, videos). Dozens of teachers have been consulted to pilot materials and provide lesson ideas. In addition, the OER Project advisory board consists of both scholars and school-level practitioners with various expertise (<https://www.oerproject.com/Advisory-Board>).

² For the 2022-23 school year, 4,706 teachers logged in at least 20 times, with another 4,888 logging in 6-19 times, and another 19,000 logging in 1-5 times.

- comprehensive set of standards-aligned resources that aim to be truly global³ in both authorship and content, through inclusion of sources and stories from around the world,
- spiraling supports for students' historical thinking skill development, including structures and activities to get students beyond simply memorizing a broad range of discrete facts,
- extensive pedagogical guidance for teachers,
- and a cohesive curriculum that frees teachers of the requirement to search for supplementary material, while still providing options for customization to meet their curricular needs and preferences.

In addition, *World History Project* is intended to be “free, for all, forever,” making available supports to schools and teachers who might otherwise not be able to afford up-to-date, high-quality curricular resources.

Our initial engagement and investigation suggested that the curriculum possessed both evaluability and merits for evaluation. We determined it was evaluable because the materials were unique and defined well enough such that teachers can adopt it with some reasonable fidelity as it stands, and if they adopt *World History Project*, teachers likely would teach somewhat differently than when using conventional textbooks and other commonly used curricular resources.

Furthermore, we considered *World History Project* evaluation-worthy because its curriculum emphases appeared to align closely with widely recognized U.S. educational frameworks and standards (e.g., C3 Framework for Social Studies, 2013; the Common Core State Standards for English Language Arts & Literacy in History/Social Studies, 2010), and conceptualizations of historical thinking skills endorsed by North American scholarship in history education (e.g., National Research Council, 2005; Seixas & Ercikan, 2015b; Reisman, 2015; Korber & Meyer-Hamme, 2015). Our confidence in its evaluation worthiness was further bolstered by our interviews with world history teachers, including those who had piloted the curriculum and possessed in-depth knowledge of its intricacies. These teachers expressed genuine appreciation for *World History Project's* objectives and were optimistic about its potential impact.

When considering which outcomes to measure, we thought it would be valuable to focus on the curriculum's impact on the teaching and learning of historical thinking skills and student engagement. Evaluating the impact on the teaching and learning of historical thinking skills is crucial because these skills hold high value within the educational community, and it is *World History Project's* express aim to enhance their teaching and learning.

Additionally, we recognized the value in examining the curriculum's impact on student engagement. Engagement is widely acknowledged as a crucial precursor to effective learning (National Academies, 2018). We were also curious about how students would respond to *World*

³ OER Project's efforts to date have included making a conscious effort to feature international scholars and people of color as experts in videos (the most visible parts of the course), and focus attention on individuals in world history in ways that authentically featured perspectives from below and from-the-margins that students could use to support, extend, or challenge master narratives. Their board members have broad representation including scholars from Latin America and South and West Africa, and efforts continue to include writers from diverse backgrounds.

History Project's aim to be truly global, in terms of who is represented in history and who is telling or authoring the historical narratives.

Previous research (e.g., Reisman et al., 2016) has found that even when history teachers are provided with educative curricula, their subject matter knowledge and pedagogical content knowledge influences how and what they teach. Given the strong and unique emphasis that *World History Project* places on historical thinking skills and supporting understanding of global perspectives, we wanted to investigate how usable the curriculum would feel to teachers, especially those early in their careers or new to teaching world history content.

In terms of how to measure the curriculum's impact on teaching and learning of historical thinking skills, given the difficulty of administering meaningful standardized achievement tests in this subject area, we decided to use an approach that involves sampling and analyzing curricular assignments and student work for its authentic intellectual demand, using subject-specific rubrics (see Joyce, Gitomer & Iaconangelo, 2018 for a recent review of this methodology). The evaluation approach entails collecting classroom artifacts to understand whether students are given tasks that are considered educationally meaningful and important. Studies in English language arts, mathematics, and science have shown that such rubric scoring can be done reliably and that the scores given to the intellectual demands of assignments tend to be associated with student achievement scores (e.g., Mitchell et al., 2005; Newmann et al., 2001; Wenzel et al., 2002). This approach does not seem to have been used in history curriculum evaluation or at least has not been a prominent approach (see Shemilt, 2018; Epstein & Salinas, 2018 for reviews on research methodologies in history education), and existing studies of history curriculum that have included some analysis of assignments and student work have generally focused on only one type of historical thinking (e.g., sourcing: Britt & Aglinskas, 2002; historical argumentation: De La Paz, et al., 2014).

We created rubrics for six historical thinking skills: historical argumentation, causation, comparison, contextualization, continuity and change over time, and sourcing (Hardy et al., 2021/2023). There were two rubrics for each thinking skill. One set of rubrics was for evaluating teacher-assigned activities (e.g., an essay prompt), and the other was for evaluating the student work produced during those activities (e.g., a written essay). We decided on the six historical thinking skills to measure based on an extensive literature review (e.g., Korber & Meyer-Hamme, 2015; Brookhart, 2015; van Boxtel & van Drie, 2018), examination of leading social studies standards, and some initial validation work. These rubrics enabled trained scorers to review teacher activities and student work to assign a rating between 0 and 3 (called "progressions" or "levels") for each historical thinking skill, depending on how advanced the activity or student work was along that skill dimension. Details of the development process and initial validity evidence collected on these rubrics can be seen in Hardy & Iwatani (2021), and Iwatani et al. (2021).

After two years of development, *World History Project* curriculum launched in November 2019, and 19,000 unique teachers visited the site between August and October 2020, likely largely in response to the COVID-19 pandemic that sharply increased the demand for free online curricular resources.

Due to the COVID-19 pandemic, the evaluation timing was pushed back two years to begin in fall 2022. The catastrophic disruption in teaching and learning caused by the pandemic influenced our evaluation approach in some ways but not others. We still decided it was important to focus on the outcomes as initially conceptualized, but we took much more seriously the participation burden on teachers and districts and decided to recruit at the teacher level rather than the district level. That is, we solicited willing teachers first, then asked their districts for permission, rather than recruiting districts and asking them to help recruit their teachers, potentially (even if unintentionally) pressuring teachers to participate in the study because it is a district initiative. This aligned with OER Project's strategic decision at that point, to focus on recruiting individual teachers rather than districts. We also revisited our initial expectations for physical site visits and substituted virtual interviews, focus groups, and a closer examination of the submitted classroom artifacts, which seemed more practically feasible in the context of a pandemic.

Evaluation questions and population

As noted previously, the objective of this evaluation was to understand how *World History Project* impacts learning opportunities and outcomes for six different historical thinking skills that are valued in history education. We also aimed to understand impacts on student engagement and educator perceptions of the curriculum's usability.

Our evaluation questions were as follows:

1. Does *World History Project* help teachers provide more opportunities for students to learn historical thinking skills relative to business-as-usual curricular materials?
2. Do students using *World History Project* learn more historical thinking skills relative to those learning from business-as-usual curricular materials?
3. Why might have *World History Project* impacted teachers and students in some skills and not others?
4. Do students using *World History Project* find world history more relevant, engaging, and/or perspective changing?
5. How usable is *World History Project*, especially for teachers new to teaching world history?

We wanted to answer these questions about teachers and students in ninth or 10th grade on-level or honors world history classes (*not* Advanced Placement or International Baccalaureate world history), in public schools across the United States, where the teacher has some motivation or interest in adopting new skills-focused curricula. Furthermore, we wanted the comparison to be between users of *World History Project* and teachers who were teaching world history in "typical" ways. As we recruited and talked to different teachers, we realized that this latter group consisted of teachers who were drawing not just from traditional textbooks but also from a variety of web resources (e.g., Newsela, Teachers Pay Teachers, Stanford History Education Project, Khan Academy, C3 Teachers, Smithsonian Learning Lab).

Methods

To answer our questions, over the 2022-23 school year, we studied teachers and students using *World History Project*, as well as a comparison (or “business-as-usual”) group, using surveys, interviews, focus groups, and artifact analysis, including analysis of student performance on world history daily classwork and summative activities.

Our overall study approach was to compare characteristics and outcomes of two groups of curriculum implementations: classrooms that used materials from *World History Project* for at least 60 percent of their classroom instructional time (e.g., “*World History Project* adopters,” “students learning from *World History Project*”), and classrooms that did not rely on any resources from *World History Project* (e.g., “comparison curricula users,” “students learning from comparison curricula”). We sampled and collected information from teachers, students, and classrooms after gaining requisite institutional review board and district approvals and participant consent/assent for research.

Starting in spring 2022, we recruited teachers who would be teaching on-level/honors ninth/10th grade world history for the following school year and were willing to be randomized into using *World History Project* or conducting business as usual. After realizing toward the end of spring that we may not reach our desired sample size, we expanded our recruitment to include teachers who would continue their existing teaching practice without asking them to submit to random assignment (i.e., teachers who were already using *World History Project* or comparison curricula, and would continue to do so for the 2022-23 school year).

The initial recruitment process yielded 16 teachers who were willing to be randomly assigned to *World History Project* or business as usual, and the second stage recruited 18 teachers who were not randomly assigned. All of these teachers were requested to submit examples of lessons (which we refer to as “activities” in this study) and student work at the end of each semester, and were invited to an end-of-semester interview and to help us organize a focus group for their students. In addition, the teachers who were randomly assigned were requested to conduct pre- and post- student surveys and to complete a teacher survey.

Notably, three teachers in the random assignment group notified us early in fall 2023 that they would be withdrawing from the study. While all of them cited personal reasons for their withdrawal (e.g., family illness), their exit interviews (N=2) suggested it would be instructive to examine barriers and promoters of curriculum use. Therefore, we added our fifth research question about usability and interviewed six additional teachers specifically about that.

Our data sources are summarized in Exhibit 1, and our sample sizes are described in Exhibits 2 through 4. Twenty-five teachers ultimately provided classroom artifacts and participated in various other aspects of the study, and six additional teachers provided perspectives on usability (Exhibits 2 and 3). Our quantitative analysis approaches are described in conjunction with our findings at the beginning of each section. Additional methodological details, including our qualitative analysis approaches and data tables, are provided in Appendices A and B.

Exhibit 1

Data sources

Source	Protocol / data description
Teacher survey	Online survey about classroom setting, world history instruction (including historical thinking skills instruction), student engagement, use and perception of curricular materials/course, and teacher background.
Student survey	Online survey about student-centered teaching of history, historical thinking skills, engagement in history class and in learning history, course satisfaction and feedback, student background.
Student focus group	Semi-structured protocol on internalization of historical narrative/frame, student motivation and engagement, and implementation of historical thinking skills instruction.
Teacher interview	Semi-structured protocol on background and participation, student-centered teaching of world history, impact of the curriculum on students' historical thinking skills, and engagement.
Examples of classroom activities	Classroom activity artifacts submitted by teachers in two batches (fall and spring), broken into two categories: "everyday activities" (activities for learning purposes completed in about one class period, e.g., notes, worksheets) and "summative activities" (extended activities to demonstrate learning, e.g., tests, projects, essays). Teachers were requested to share 4 "everyday" and 2 "summative" activities each semester, and asked to submit what was typical for their class. All submissions included a cover sheet that described the assignment, class, main sources of information students were expected to draw from to complete the assignment, independent/group work, grading criteria, and proportion of students who exceeded and met teachers' expectations.
Examples of student Work	Teachers were instructed to submit six examples of student work to correspond with their respective activities submitted. If there was variation in the quality, they were instructed to submit two samples each of the following: (1) A work or work that exceeds expectations (2) B work or work that meets expectations (3) C work or work that does not meet expectations

Exhibit 2

Sample sizes and recruitment sources for activity and student work collection

Curriculum group		N teacher (attrition)	N activities	N student work samples
<i>World History Project</i>	Randomly assigned to <i>World History Project</i> as main curricular resource	4 (4)	42	241
	Had already been using <i>World History Project</i> as main curricular resource	7 (2)	77	445
Comparison	Randomly assigned to continue business as usual	6 (2)	83	374
	Had not been using <i>World History Project</i> as main curricular resource	8 (1)	66	449
Total		25 (9)	268	1,509

Note. Teachers were recruited through Digital Promise and OER Project outreach and contacts. Attrition refers to teachers who were recruited for the study but did not submit any activities or student work.

Exhibit 3

Sample sizes of teacher interviews

Curriculum group		N teacher interviews	N teacher surveys
<i>World History Project</i>	Randomly assigned to <i>World History Project</i> as main curricular resource	2	4
	Had already been using <i>World History Project</i> as main curricular resource	4	-
Comparison	Randomly assigned to continue business as usual	4	6
	Had not been using <i>World History Project</i> as main curricular resource	7	-
Usability Interviews	Teachers using <i>World History Project</i> , given questions specifically about the usability of the curriculum	6	-
Exit interviews	Teachers who decided to drop out of the study (in fall 2022) were asked about their experiences with the curriculum and reasons for dropping out	2	-
Total		25	10

Exhibit 4

Sample sizes of student perspective data collection

Curriculum group	N student survey fall 2022	N student survey spring 2023	N student surveys fall 2022 & spring 2023	N student focus groups
<i>World History Project</i>	199 students from 9 classrooms from 5 teachers	88 students from 6 classrooms from 3 teachers	Approximately 110 students (131 fall and 88 spring) from 6 classrooms from 3 teachers	1 focus group (2 students)
Comparison	202 students from 12 classrooms from 8 teachers	163 students from 8 classrooms from 5 teachers	Approximately 162 students (161 fall and 163 spring) from 8 classrooms from 5 teachers	1 focus group (5 students)
Total	401 students from 21 classrooms from 13 teachers	251 students from 14 classrooms from 8 teachers	Approximately 293 students from 14 classrooms from 8 teachers	2 focus groups (7 students)

Note. The student survey counts in the fourth column overlap with the counts in the second and third columns of this table.

Limitations

This study has several limitations. First, it is important to remember that our findings apply to teachers who have some initial interest in adopting a curriculum that emphasizes historical thinking skills. In other words, we only studied teachers who *wanted* to use *World History Project* (so they had either adopted it on their own accord, or were willing to participate in a study that might randomly assign them to adopt it), so we cannot say how teachers and students would respond if teachers who had no initial interest were requested to adopt it.

Second, the examination of historical thinking skills was limited to six specific skill competencies and thus did not include other potential outcomes such as historical empathy and content knowledge. These unexplored elements may also play significant roles in shaping students' historical understanding but were not measured in this research.

Third, our study has a limited sample size, especially for student focus groups and student surveys (when considered at the teacher level), but also when it comes to the number of assignments and student work samples as representing what happens in a classroom within a year. This impacts the power of our statistical tests and the generalizability of some of our findings, but we do believe that the multiple ways in which we triangulated our findings helped reduce sample size-related threats to validity.

Furthermore, the study was conducted during a time marked by the ongoing COVID-19 pandemic. The exact nature and extent of the effects of COVID-19 on both teachers and

students remain uncertain, particularly when it comes to student engagement, and hence the COVID context may impact the generalizability of the study.

Last but not least, it is important to recognize that an educative curriculum alone likely has limitations in fostering historical thinking skills if not matched with a strong foundation in pedagogical knowledge. The study does not delve deeply into the pedagogical strategies employed by teachers, which may have significant implications for the effectiveness of the curriculum.

Significance of this study

Limitations notwithstanding, this study has some unique features that we believe make it an important contribution to the field of social studies education and education practice.

Perhaps most importantly, *World History Project* plays a pivotal role in addressing a significant gap within the open educational resource curriculum market dedicated to world history. In essence, *World History Project* currently stands alone in providing comprehensive, skills-based, and freely accessible curricula for high school world history that is aligned to multiple course-specific state content standards. OER Project's unique commitment to this mission, their capacity to make continual and large-scale revisions, and substantial resources already dedicated to this undertaking, make a thorough examination of their work important.

Additionally, our study aims to describe and help improve world history education in public schools. Although world history is a required course for a majority of U.S. students (Keirn, 2018), world history teaching and learning is understudied, with strikingly few empirical studies centered on secondary world history classrooms (Girard & Harris, 2018). Relatedly, while there is a substantial body of literature exploring historical thinking skills, fewer studies have explored the impact of specific history curricula or programs on opportunities for historical thinking (Epstein & Salinas, 2018). This report is the first to investigate this topic in world history in a systematic way by applying rubrics for multiple dimensions of historical thinking to analyze curricular assignments and student work for its authentic intellectual demand.

Third, in the context of social studies education research, our study is exceptionally large in scale. It includes a strong quantitative and quasi-experimental component, which allows for some of the findings to be generalizable. It also includes some detailed qualitative analysis that offers nuanced explanations for the quantitative findings. Most empirical studies in world history and social studies education are qualitative, with relatively small sample sizes (e.g., investigating one or two classrooms) (Epstein & Salinas, 2018).

We are particularly proud of the quality, variety, and scope of the data that was collected for this study. Our largest efforts went to collecting and making meaning out of the 268 activities and associated 1,509 samples of student work from 25 world history teachers. Not only did we triple-score each activity and student work sample on multiple rubrics with generally very consistent scoring across raters on most dimensions (see Appendix B, Table B1), we also collected and considered information about each lesson, teacher, and classroom. The quantitative analysis for this study strand involved merging 10 different datasets that collectively

held about 55,000 data cells, most of which were manually entered by researchers, teachers, or scorers. To this quantitative strand, we added nuance and validity by qualitatively analyzing a subset of activity and student work samples and by triangulating with teachers interviews, student focus groups, and teacher and student surveys.

Finally, we are optimistic that the knowledge generated by this report can be translated into improved tools and experiences for world history teachers and students, since as a provider of open-source, online material, OER Project is positioned to continuously refine their content and to improve their curriculum and professional learning with annual updates and overhauls.

Findings

Our findings are organized into sections addressing five topics: (A) Whether *World History Project* helps teachers provide more opportunities for students to learn historical thinking skills, relative to business-as-usual curricular activities; (B) Whether students who use *World History Project* learn more historical thinking skills relative to those exposed to business-as-usual curricular activities; (C) Explanations for the patterns observed in learning opportunities and outcomes; (D) Whether students who use *World History Project* find world history more relevant, engaging, and/or perspective changing; (E) Usability of *World History Project* for teachers new to teaching world history.

A. Whether *World History Project* helps teachers provide more **learning opportunities** for historical thinking skills

To examine whether *World History Project* helps teachers provide more opportunities to learn historical thinking skills relative to business-as-usual curricular activities, we examined 119 curricular activity samples collected from *World History Project* teachers (i.e., teachers using *World History Project* as their main curriculum) and 149 curricular activity samples collected from comparison teachers (i.e., teachers not at all using *World History Project*), in terms of their relative emphasis on six different historical thinking skills.

Trained scorers who were blinded to the teacher's condition scored each activity using the historical thinking skills rubrics mentioned in the introduction. Most activities were scored by at least three scorers, and the average of their scores was taken as the final score for each historical thinking skill for each activity. In general, a rubric score of "0" indicates that the activity does not explicitly call for students to employ the skill; a score of "1" indicates that the activity explicitly calls for students to employ the skill very briefly (e.g., by selecting, listing, or briefly describing); a score of "2" indicates the activity explicitly calls for students to briefly describe and explain; while a score of "3" indicates that the activity calls for students to provide an extended explanation. An example rubric is provided as Exhibit A1. See Hardy et al. (2021/2023) for the full set.

Exhibit A1

Example rubric for examining historical thinking skills in high school world history activities

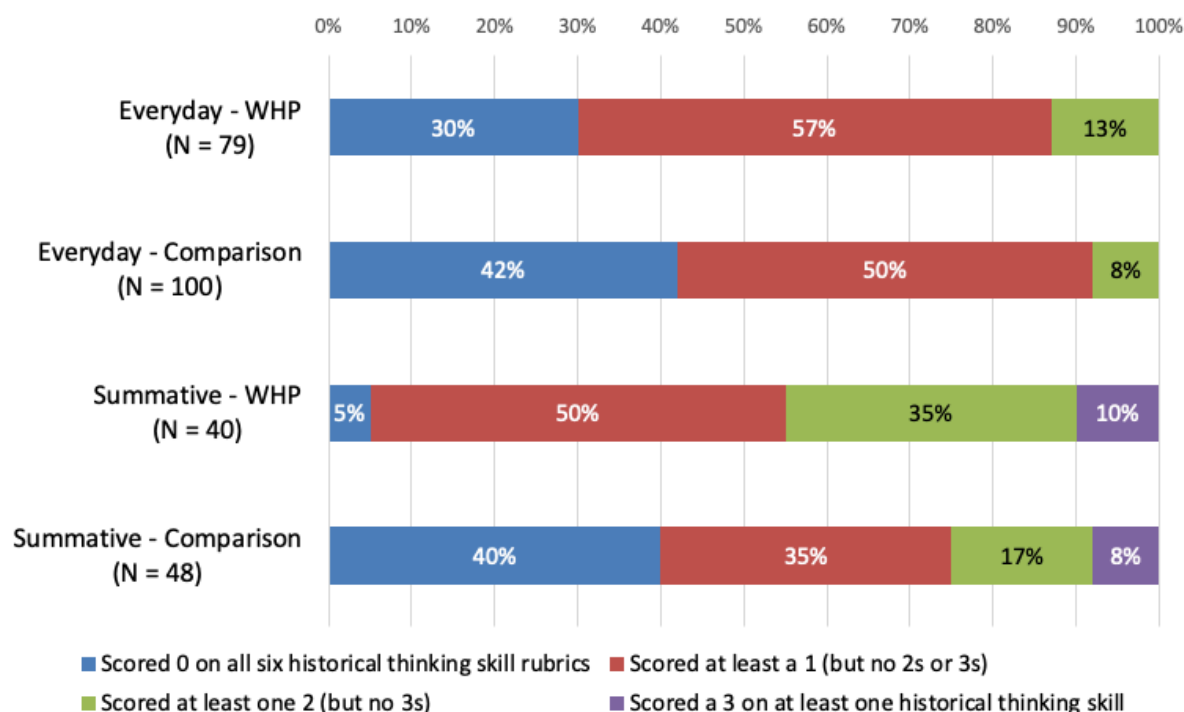
A3. HISTORICAL COMPARISON: Activity explicitly calls for students to describe and explain similarities and differences between historical developments and processes, regions, eras, or other focal areas.

0	1	2	3	Level 3 Examples
Activity does not explicitly call for students to describe similarities and differences between historical developments, processes, regions, eras, or other focal areas.	The activity explicitly calls for students to select, list, or describe similarities and/or differences between historical developments, processes, regions, eras, or other focal areas.	The activity explicitly calls for students to describe similarities and/or differences between historical developments, processes, regions, eras, or other focal areas and prompts students to provide brief explanations of why the similarities and/or differences existed.	The activity explicitly calls for students to describe similarities and/or differences between historical developments, processes, regions, eras, or other focal areas and prompts students to provide extended explanations of why the similarities and/or differences existed.	For example, the activity may prompt students to extend their analysis by <ul style="list-style-type: none"> evaluating the relative historical significance of particular similarities or differences and/or exploring the connection between similarities and differences within and across different categories (e.g., political, religious, geographic).

Differences across groups overall. It would not be reasonable or appropriate to expect every class activity to require the exercise of every historical thinking skill at the highest level. Accordingly, we first asked whether assignments were requiring *any* historical thinking at level 1, 2 or 3. Exhibit A2 shows the highest thinking skill requirement score in any skill category for both “everyday” activities and “summative” activities. The blue bars represent the proportion of activities that scored all 0s (i.e., did not call for any of the six historical thinking skills we examined); red bars represent activities that scored a 1 on at least one historical thinking skill rubric (emergence of that historical thinking skill learning opportunity) but not anything higher; green bars represent the proportion of activities that scored a 2 on at least one historical thinking skill rubric (presence of that historical thinking skill learning opportunity) but not anything higher; and purple bars show the proportion of activities that scored a 3 on at least one historical thinking rubric (presence of at least one rigorous historical thinking skill learning opportunity).

Exhibit A2

Proportion of activities, by activity type and curriculum condition, that scored all 0s, at least one 1 (but no 2s or 3s), at least one 2 (but no 3s), and at least one 3, across the six historical thinking skills



Note: We considered that the lesson scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

There is a stark difference in how much more the summative activities (e.g., essays, tests, projects), assigned by teachers who use *World History Project* as their main curriculum, emphasized historical thinking skills, relative to summative activities assigned by teachers who did not use *World History Project*. **Of the summative assignments submitted by *World History Project* teachers, almost all (95%) emphasized at least one of the historical thinking skills we examined, compared to less than two-thirds (60%) of the summative assignments submitted by comparison teachers.** We estimated the statistical likelihood of this difference by comparing the highest score that an assignment scored on any of the historical thinking skills rubrics, across *World History Project* and comparison activities, controlling for score clustering within teachers, years of teaching experience, percent of students receiving free/reduced price lunch at the school, and whether the teacher was randomly assigned to their teaching condition. The difference was statistically significant ($p = .018$), where the use of *World History Project* was associated on average with a .56 point increase in rubric scores, and the difference was approximately two-thirds of a standard deviation (Hedges' $g = .66$), which is considered a moderate effect size in education research.

A weaker but similar pattern seemed evident in our sample for “everyday” activities (e.g., worksheets), where 70% of everyday classwork submitted by *World History Project* teachers

emphasized at least one historical thinking skill, compared to 58% of activities submitted by comparison teachers. A small effect size was estimated (Hedges' $g = .26$) although the result was not statistically significant. See Appendix B tables B14 and B15 for additional details.

Differences across groups for each historical thinking skill. Next, we examined data for individual historical thinking skills to ascertain which historical thinking skills are supported more consistently in *World History Project*. Exhibit A3 shows the average rubric score on each historical thinking skill for the “everyday activities” (e.g., worksheets), while Exhibit A4 shows scores on each historical thinking skill for the “summative activities” (e.g., essays, tests, projects). Again, the comparison was conducted using a two-level hierarchical linear regression model that controlled for clustering of scores within teacher submitting the assignment, years of teaching experience, percent free/reduced price lunch at the school, and whether the teacher was randomly assigned to their teaching condition. The effect sizes (Hedges' g) for rubric score comparisons are displayed in Exhibits A5 and A6. Additional data tables are provided in Appendix B.

We found that on average, **both everyday and summative activities from *World History Project* classrooms scored higher on the continuity and change over time learning opportunity rubric** than did activities from the comparison classrooms. For everyday activities the use of *World History Project* was associated, on average, with a .18 point higher rubric score, which is a difference of .41 standard deviations and generally regarded as a small to moderate effect size in education research ($M_{WHP} = .24$, $M_{Comp} = .07$, $\beta = .18$, $SE = .06$, $p = .002$, Hedges' $g = .41$). For summative activities, the use of *World History Project* was associated with the a .31 point higher rubric score, which is a difference of .70 standard deviations, and considered a moderate to large effect size in education research ($M_{WHP} = .43$, $M_{Comp} = .13$, $\beta = .31$, $SE = .14$, $p = .026$, Hedges' $g = .70$).

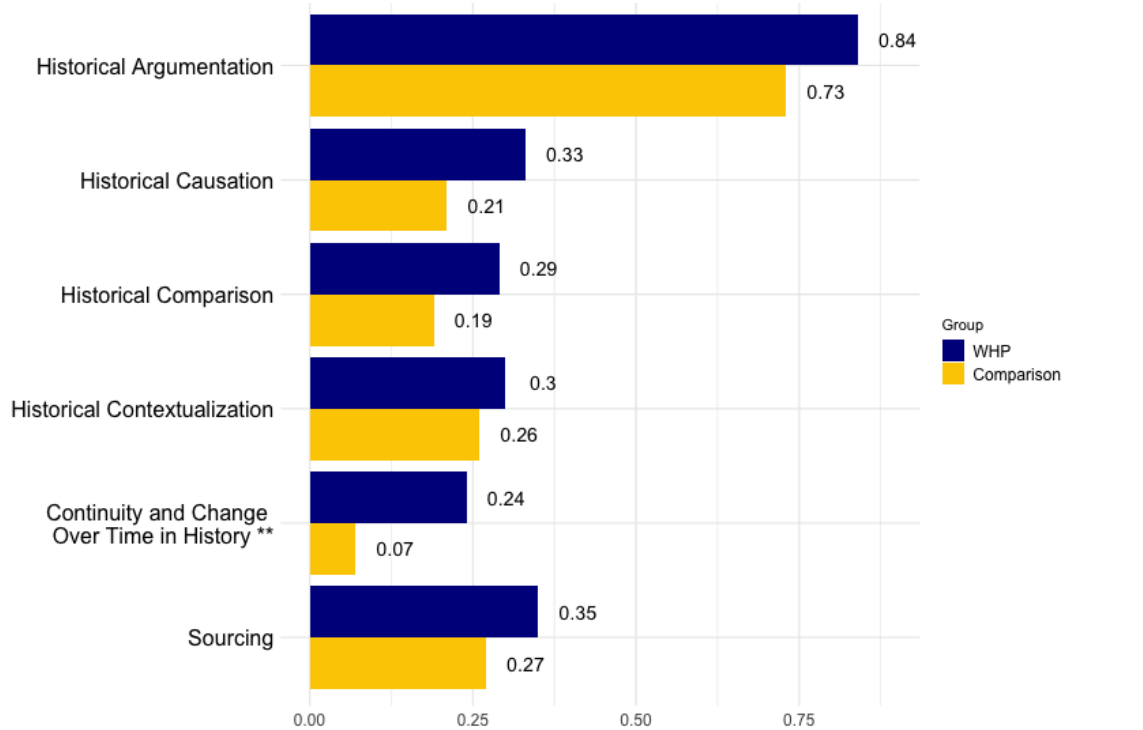
While there were **no other statistically significant differences on individual historical thinking skill rubric scores across the two groups**, Exhibits A3 and A4 show that the mean rubric scores were generally higher for *World History Project* classroom activities. Among these, impact estimates on learning opportunities for *historical argumentation* summative activities and *historical comparison* summative activities were seemingly notable, with the use of *World History Project* was associated with a .52 point and .23 point increase in rubric scores, and in differences of .39 and .30 standard deviations, respectively.

Summary of findings. To summarize, data suggests that

- Teachers who adopted *World History Project* assigned summative activities with a higher emphasis on historical thinking skills, relative to teachers using business-as-usual curricula. They provided roughly comparable learning opportunities through the everyday activities.
- Teachers using *World History Project* provided more learning opportunities in *continuity and change over time in history* (through both summative and everyday activities), and possibly in *historical argumentation* and *historical comparison* (through summative activities), relative to teachers using business-as-usual curricula. They provided roughly comparable learning opportunities for the other historical thinking skills.

Exhibit A3

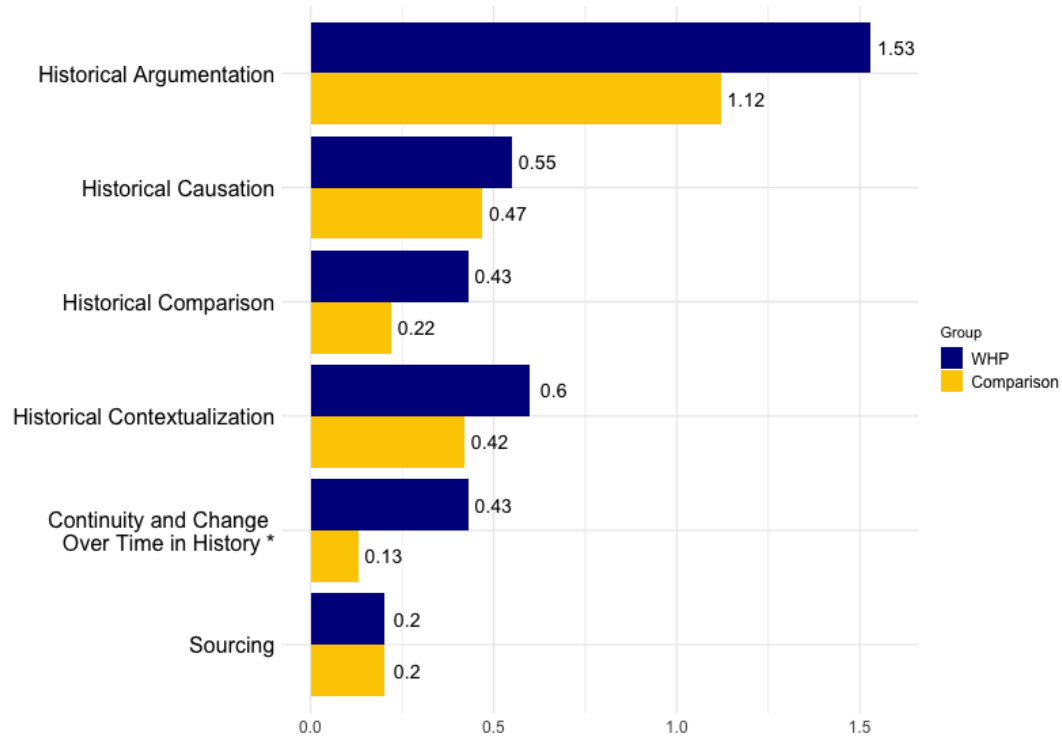
Mean rubric scores for everyday curricular activities in world history



Note. Compares 79 activities from *World History Project* teachers and 100 activities from comparison teachers, controlling for clustering within teacher, years of teaching experience, percent free/reduced price lunch at the school, and whether the teacher was randomly assigned to their teaching condition. p = .298, .294, .342, .617, .002 and .250, from top to bottom.

Exhibit A4

Mean rubric scores for summative curricular activities in world history



Note. Compares 40 summative activities by *World History Project* teachers and 48 by comparison teachers, controlling for the same covariates as Exhibit A3. $p = .066, .682, .097, .298, .026$, and $.968$, from top to bottom.

Exhibit A5

Effect sizes (Hedges' g) for everyday curricular activities

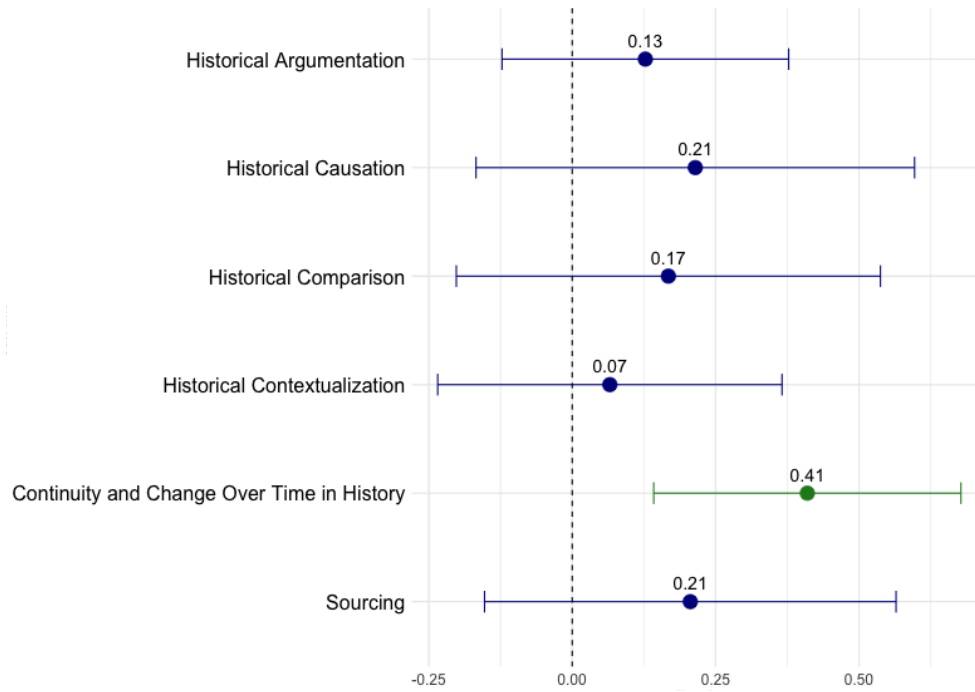
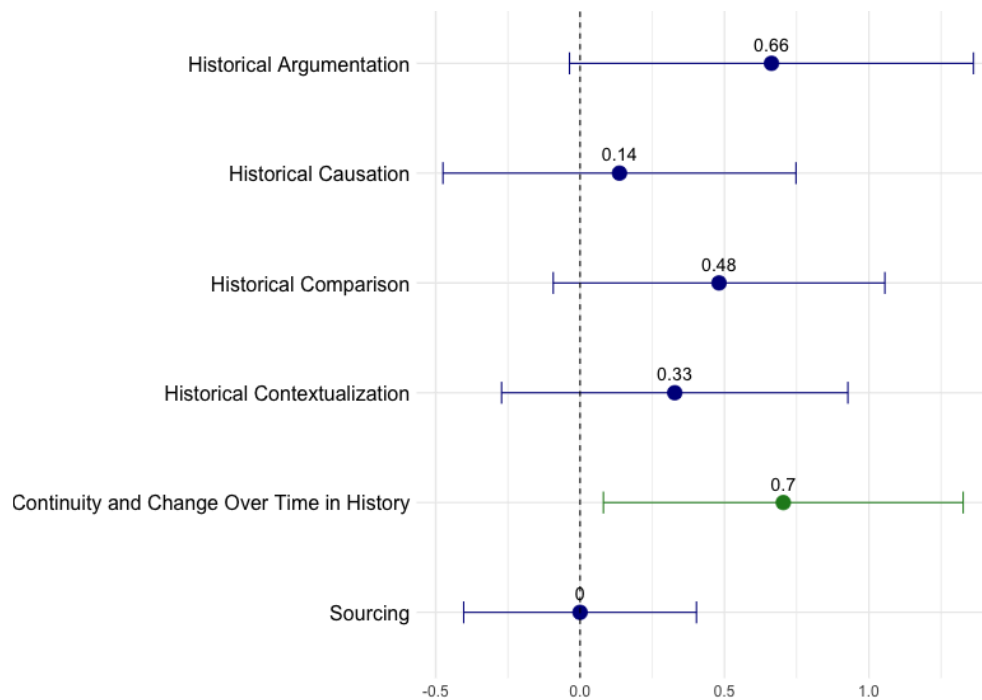


Exhibit A6

Effect sizes (Hedges' g) for summative curricular activities



B. Whether students who use *World History Project* learn more historical thinking skills relative to business-as-usual curricular activities

To examine whether students who use *World History Project* learn more historical thinking skills relative to those experiencing business-as-usual curricular activities, we analyzed rubric scores of the student work samples for each activity collected. The same scorers who scored the activities scored the student work samples, using the same process as the scoring of the activities (Section A). Most student work samples were scored by three scorers. We combined these sample scores with information provided by the teacher about how representative the student work was to estimate an *average student work score* for the class.⁴

Thus, a rubric score of “0” suggests that the average student in the class did not employ the skill; a score of “1” indicates that the average student may have very briefly employed the skill (e.g., by correctly selecting something from a list that presumes application of the historical thinking skill); a score of “2” indicates the average student provided a brief explanation that suggests application of the historical thinking skill (e.g., by providing a correct one-sentence analysis); while a score of “3” indicates that students used the skill to a large extent by providing an extended explanation. Students generally score no higher than what the assignment provides the opportunity for (e.g., if the assignment rubric is a “2,” the corresponding student rubric scores are “2” or lower). An example rubric is provided as Exhibit B1. See Hardy et al. (2021/2023) for the full set. Incorrect responses were not considered as valid demonstrations of the skill.

Exhibit B1

Example rubric for examining historical thinking skills in high school world history student work

S3. HISTORICAL COMPARISON: Students described and explained similarities and differences between historical developments, processes, regions, eras, or other focal areas. The student response and any evidence/reasoning provided is generally accurate and relevant.

0	1	2	3	Level 3 Examples
The student does not describe similarities and differences of developments, processes, regions, eras, or other focal areas.	The student selects, lists, or describes similarities and/or differences of developments, processes, regions, eras, or other focal areas.	The student describes similarities and differences between the foci of comparison, and provides a brief analysis of reasons for these similarities and/or differences.	The student describes similarities and differences between the foci of comparison, and provides an extended analysis of the reasons for these similarities and/or differences.	For example, the student may <ul style="list-style-type: none">• extend their analysis by evaluating the relative historical significance of particular similarities or differences and/or• exploring the connection between similarities and differences within and across different categories (e.g., political, religious, geographic).

⁴ Generally, teachers submitted two examples of student work that exceeded expectations (or scored an “A”), two examples of student work that met expectations (or scored a “B”), and two examples of student work that did not meet expectations (or scored a “C”). Teachers also reported on what percentage of students in their class exceeded, met, or did not meet expectations on the assignment. We weighed the student scores based on these percentages—for example, if a teacher reported that 20% scored As, 20% scored Bs, and 60% scored Cs, we weighted the rubric scores of the “C” work samples to count three times as much as the scores of the A or B work samples.

Differences across groups overall. Exhibit B2 shows the distribution of highest thinking skill scores that the average student work received for each of the “everyday” and “summative” activities submitted by *World History Project* adopters and comparison teachers. The blue bars represent the proportion of activities where the average student in the class scored all 0s on the student work rubric (i.e., did not appear to use any of the six examined historical thinking skills in the assignment); red bars represent the proportion of activities where the average student work scored a 1 on at least one historical thinking skill rubric (i.e., used at least one historical thinking skill to a small extent) but not anything higher; green bars represent the proportion of activities where the average student work scored a 2 on at least one historical thinking skill rubric (i.e., used at least one historical thinking skill to a moderate extent) but not anything higher; and purple bars show the proportion of activities where the average student work scored a 3 on at least one historical thinking skill rubric (i.e., used at least one historical thinking skill to a substantial extent).

Regardless of whether students were learning from *World History Project* or comparison curricula, the vast majority of activities scored no more than a “1” on the student work rubric. The extent to which students demonstrated historical thinking at the highest levels (“2” or “3”) was very low for everyday activities (2–3%, with none scoring a “3” for either group), and not too much higher for the summative activities (8% for students in *World History Project* classrooms and 13% for students in comparison classrooms). A much larger proportion of student work scored at least and at most a “1” (40–80% depending on the group and assignment type).

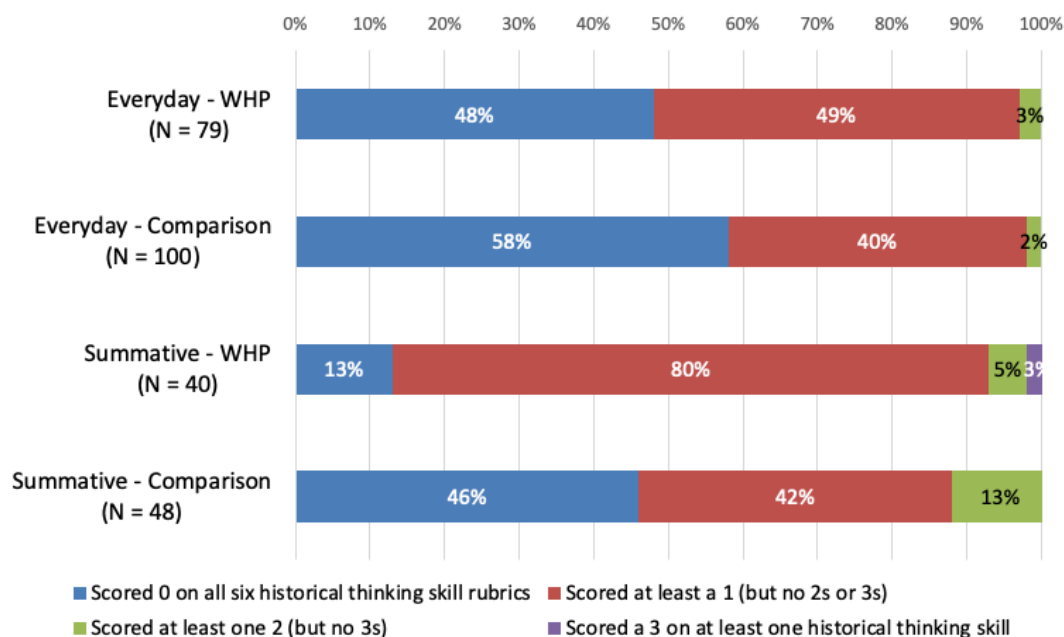
Statistically, there was no difference across curriculum groups in terms of the overall patterns in student outcomes in historical thinking skills, although the impact estimates were positive in favor of *World History Project* (for everyday activities, $\beta = .13$, $p = .168$, Hedges’ $g = .29$; for summative activities $\beta = .32$, $p = .056$, Hedges’ $g = .36$; see Appendix B tables B16 and B17 for additional details).⁵

Furthermore, when comparing Exhibit A2 (learning opportunities) with Exhibit B2 (student outcomes), we find that there is a very large difference among score distributions for the summative activities submitted by *World History Project* adopters. While 5%, 50%, 35%, and 10% of summative activities scored 0, 1, 2 and 3, respectively on the learning opportunities rubric, the student outcome scores distributions were 13%, 80%, 5% and 3%, respectively. This suggests that even when *World History Project* is providing more opportunities for students to practice historical thinking skills at a high level, the students are not necessarily rising to the occasion to meet those opportunities. We investigate and discuss possible reasons for this further in the next section (Section C) by looking at the assignments and student work in more detail.

⁵ As in the previous section, we estimated the statistical likelihood of this trend by comparing the highest score that an assignment scored on any of the historical thinking skills student work rubric, across *World History Project* and comparison activities, controlling for score clustering within teachers, years of teaching experience, percent free/reduced price lunch at the school, and whether the teacher was randomly assigned to their teaching condition.

Exhibit B2

Proportion of average student work rubric scores, by activity type and curriculum condition, that scored all 0s, at least one 1 (but no 2s or 3s), at least one 2 (but no 3s), and at least one 3, across the six historical thinking skills



Note: We considered that the lesson scored a 0, 1, 2, and 3 on a rubric if the weighted average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Differences across groups for each historical thinking skill. How do the student work scores differ across curricular conditions, if at all, for each historical thinking skill? Exhibit B3 shows the mean rubric score on each historical thinking skill for the everyday activities, while Exhibit B4 shows scores on each historical thinking skill for the summative activities. The effect sizes (Hedges' g) for rubric score comparisons are displayed in Exhibits B5 and B6. Additional data tables are provided in Appendix B.

On average, **both everyday and summative student work from *World History Project* classrooms scored higher on the *continuity and change over time* learning opportunity rubric** than did activities from the comparison classrooms. For everyday activities the use of *World History Project* was associated, on average, with a .07 point higher rubric score, which is a difference of .36 standard deviations and generally regarded as a small effect size in education research ($M_{\text{WHP}} = .11$, $M_{\text{Comp}} = .04$, $\beta = .07$, $SE = .03$, $p = .032$, Hedges' $g = .36$). For summative activities, the use of *World History Project* was associated with the a .15 point higher rubric score, which is a difference of .60 standard deviations, and considered a moderate effect

size in education research ($M_{WHP} = .21$, $M_{Comp} = .06$, $\beta = .15$, $SE = .14$, $p = .014$, Hedges' $g = .60$).

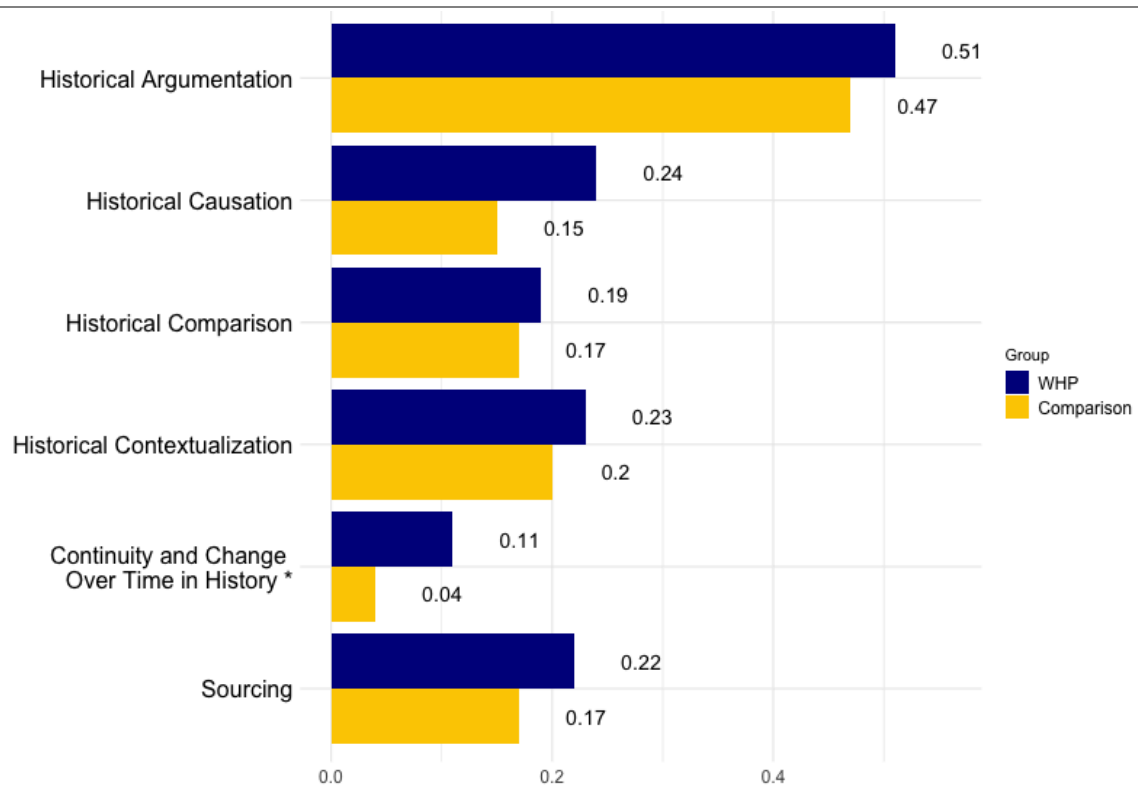
There were **no other statistically significant differences on individual historical thinking skill rubric scores across the two groups.**

Summary of findings

- Overall, students who learned from *World History Project* demonstrated comparable historical thinking skills, relative to those who learned from business-as-usual curricula. The estimated impacts were higher for *World History Project* by about a third of a standard deviation, but not statistically significant.
- Students who learned using *World History Project*, relative to those who learned from business-as-usual curricula, demonstrated more understanding of *continuity and change over time in history*, in both everyday and summative assignments.
- Students learning from *World History Project* demonstrated comparable competency in the other five historical thinking skills when compared to students learning from comparison curricula.

Exhibit B3

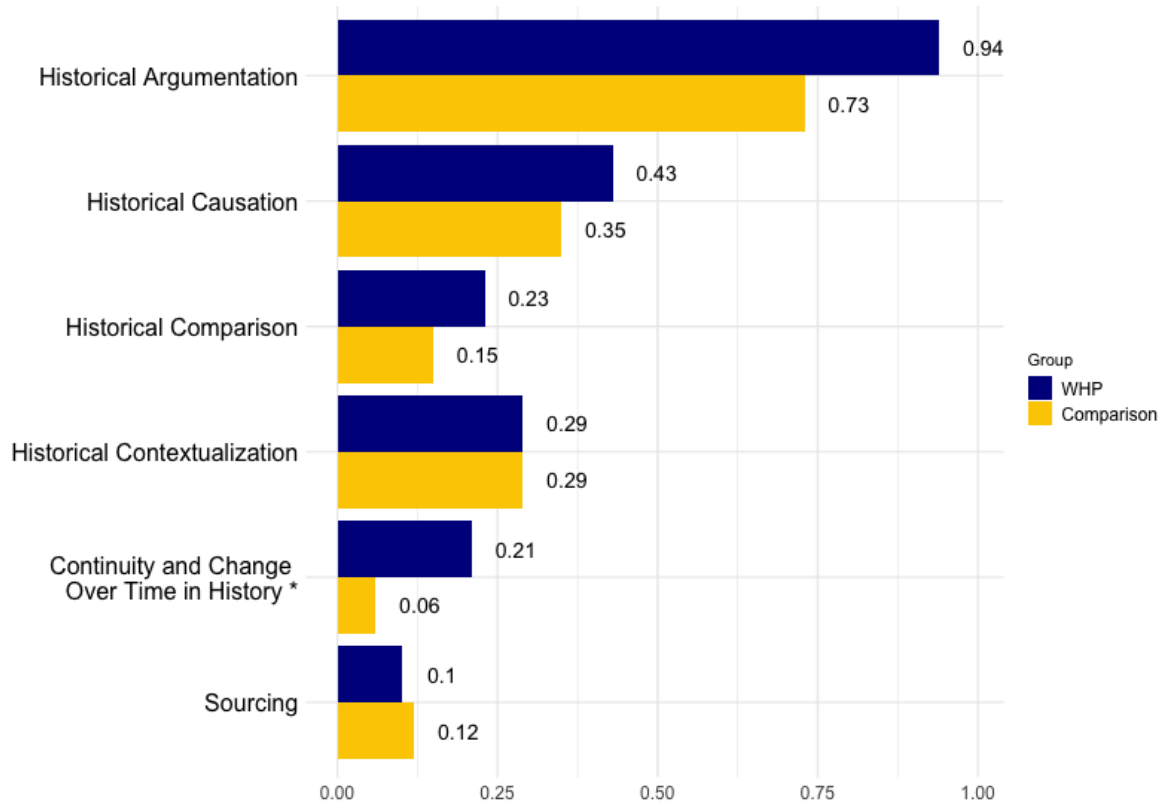
Mean rubric scores for everyday student work samples in world history



Note. Compares 79 activities by *World History Project* teachers and 100 activities from comparison teachers, controlling for clustering within teacher, years of teaching experience, percent free/reduced price lunch at the school, and whether the teacher was randomly assigned to their teaching condition. p = .549, .271, .764, .697, .032, .358, from top to bottom.

Exhibit B4

Mean rubric scores for summative student work samples in world history



Note. Compares 40 summative activities by *World History Project* teachers and 48 by comparison teachers, controlling for the same covariates as Exhibit B3. $p = .187, .535, .234, .865, .014$, and $.653$, from top to bottom.

Exhibit B5

Effect sizes (Hedges' g) for everyday student work samples

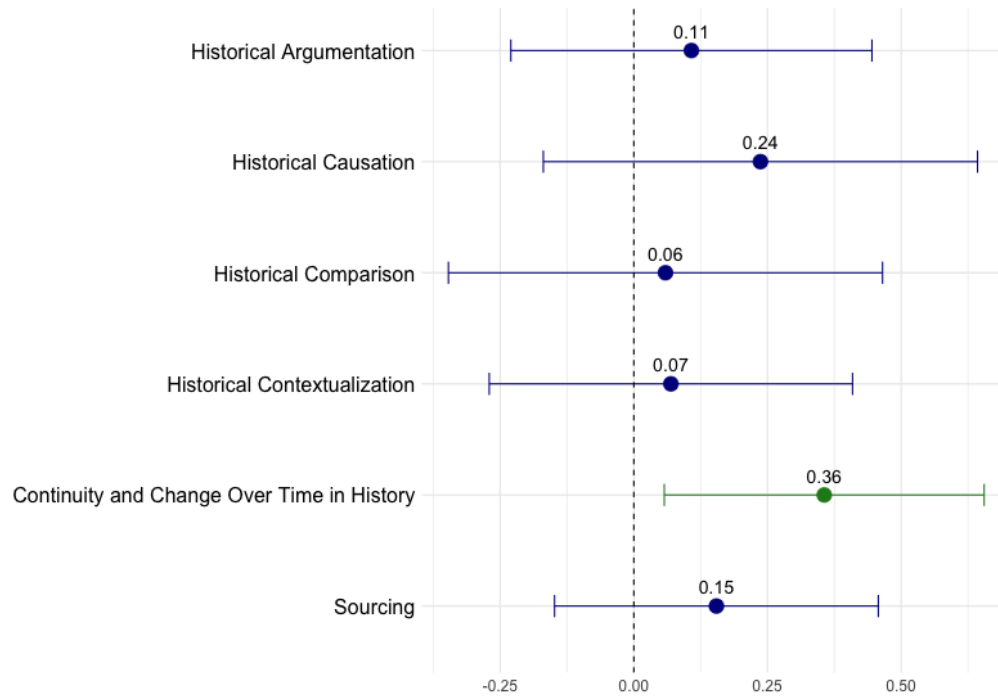
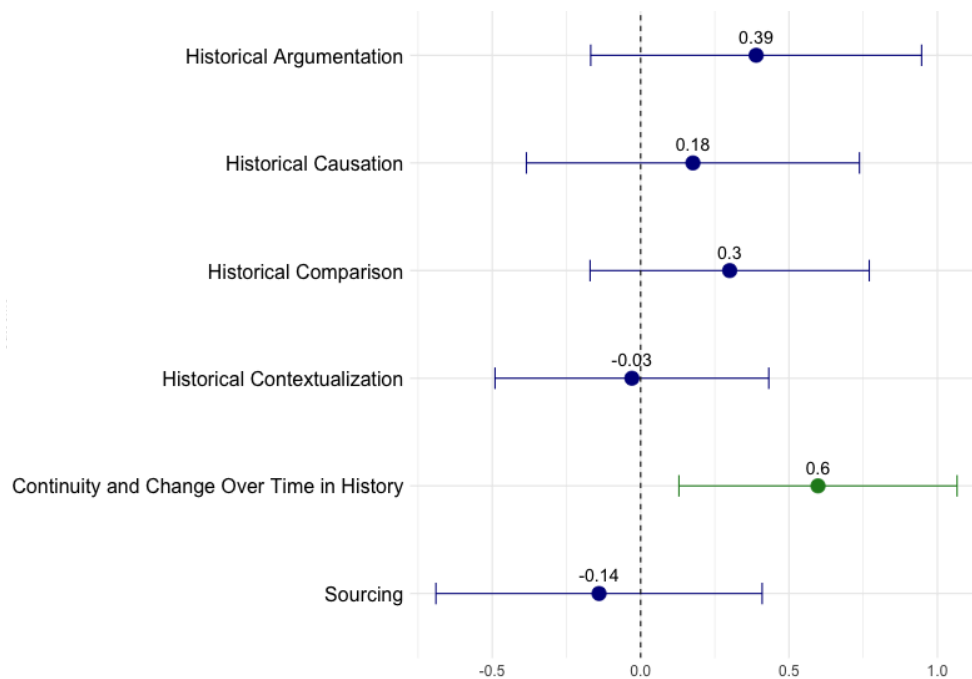


Exhibit B6

Effect sizes (Hedges' g) for summative student work samples



C. Explanations for the patterns observed in learning opportunities and outcomes

To understand the quantitative trends in rubric scores, we examined the characteristics of submitted activities and student work and surveyed other available evidence. We explored two questions:

- Why were *continuity and change over time* learning opportunities scores and student outcome scores higher in classrooms that adopted *World History Project*?
- Why were rubric score differences in learning opportunities and student outcomes not observed in the other historical thinking skills, despite the explicit emphasis *World History Project* places on supporting these skills?

Possible reasons for score differences across curricular groups in continuity and change over time in history. To explain the patterns about *continuity and change over time*, we closely examined all activities that scored at least a “1” on this thinking skill. Our initial review focused on identifying the source and type of activity, in part to check whether the teachers in the *World History Project* group were in fact using materials from *World History Project* to teach *continuity and change*, and using the activities as intended. We found that of the 25 activities submitted by nine *World History Project* adopters that scored a 1 or higher on CCOT (17 had scored a “1”, 7 had scored a “2,” and 1 had scored a “3”), all of them were authored by *World History Project*. These activities came from various parts of the curriculum and were generally one of the following: a skills progression activity (e.g., sorting factors into continuity or change), reading comprehension activities on articles that focused on *continuity and change over time*, a document-based or long essay question that focused on *continuity and change across time*.

In contrast, the vast majority of the activities submitted by comparison curricula users appeared teacher-created (eight total, submitted by six teachers, where seven activities had scored a “1” and one had scored a “2”). Two required students to search for information to respond to at least one open-ended question regarding *change over time*, one was a long essay question, one was a brief document-based question, one was a timeline activity, and the others were opportunities where students responded to short-answer questions where at least one question was on the topic of continuity and/or change. Notably, none appeared to be *intentionally* designed to provide students with an opportunity to understand continuity and change over time (more on this below). Overall, this initial review suggested that ***World History Project* may help teachers provide more learning opportunities in continuity and change over time in history, because it is currently the only world history curriculum in the market that provides multiple activities that emphasize this skill.**

To further understand how and why the students learned or did not learn about *continuity and change over time* from these activities, we conducted a deeper examination of 10 sets of activities and student work. For each curricular group, we examined two activities that appeared successful and two that appeared unsuccessful, in terms of student outcomes on *continuity and change over time*, given the learning opportunity. We also examined two additional activities from *World History Project* with a skill building focus.

For each lesson, we examined all materials provided—the lesson overview, materials related to the activity, and examples of student work. We reviewed classroom characteristics (e.g., grade level), teacher characteristics (e.g., years of teaching), lesson objectives, lesson flow, student tasks and guidance provided related to understanding of *continuity and change*, and evidence of student insights related to *continuity and change*. We also examined some measures related to assignment length and reading level, as well as noting ways that teachers modified the activity, and potential sources of confusion or misconceptions.

This examination revealed clear differences between *World History Project* and comparison materials in how *continuity and change over time* was introduced. **The comparison materials made no explicit mention about continuities and changes over time in history, nor provided materials or scaffolds that support students to identify these.** For example, in one activity, students scored high on the student rubric for *continuity and change* because they correctly identified changes in navigational technologies during the “age of exploration,” in response to several reading comprehension questions about an article on that topic. In another activity, students were asked to conduct independent research on achievements of an ancient civilization and how that evolved over time. No further guidance was provided than these questions, at least according to the materials and lesson overview submitted. Students who “exceeded expectations,” according to their teacher, listed innovations and potential impacts of that innovation, without mentioning changes or continuities.

In stark contrast, *World History Project* activities that scored relatively high on *continuity and change over time in history* had intentional emphases and an array of materials and scaffolds that appear to guide students to try out a process that historians might utilize for such analysis. These were most evident in the *change and continuity* skill building activities. One activity involved asking students to analyze a list of historically significant factors that might have changed or stayed the same during the period where empires were forming (e.g., “people were farmers,” “some communities were treated differently/unequally”). They asked students to consider whether this was a continuity or change. Then they asked students to identify what type of factor this is (namely, does this factor relate to how goods were produced and distributed, about the nature of communities, or about networks of exchange). Finally, students were asked to evaluate the extent to which they believed the continuity or change was positive or negative.

We closely examined three activities that followed this general pattern, which differed in the period of time and location, and whether students identified continuities and changes themselves or were provided a list to sort. One activity, designed to be the introductory activity to continuity and change, asked students to focus on farming in Iowa across four time periods. Another asked students to identify changes and continuities within a unit they had just studied. A third asked students to identify changes and continuities across two units—one that they just studied and another that they are about to study.

All three activities asked students to consider what changed and what remained the same and asked students to identify whether these factors related to (i) distribution and production of goods, (ii) how ideas and innovations tended to be interconnected and exchanged, and (iii) the norms and structures of communities. Not only did these *World History Project* activities provide

specific guidance on how to analyze *continuity and change*, they provided articles that appeared to support students to be successful in this analysis and be able to see the general point, such as unit overviews that focus on key changes and continuities, and targeted information about farming in Iowa across time.

This analysis also suggested some factors that were likely getting in the way of students' learning of *continuity and change over time* in *World History Project* classrooms, helping to explain why the student outcomes were not as high as the opportunities provided.

First, two central conceptions intended to build skills on *continuity and change over time* appeared under-defined, especially for a beginner learner (and teacher), and in some cases may reinforce some misconceptions.

The first construct is what *World History Project* calls “frames,” namely the themes of “networks, communities, and production and distribution.” We observed that teachers and students were referring to these terms without defining them (e.g., teachers were asking “How did networks of exchange connect societies, and how were communities changed by these connections?”) and potentially leading students to believe that these were self-evident terms and make ill-defined claims (e.g., “Networks of exchange and societies were connected in ways by use of trading cities, ports, and the weather”). While we agree it is very beneficial, and likely necessary, to provide students and teachers with organizing principles through which to analyze continuities and changes, it might be more conducive to teaching and learning if these were more explicitly stated each time they are mentioned (e.g., rather than asking students to consider “networks,” ask them to consider “how ideas and innovations tended to be interconnected and exchanged”).

The second conception that was under-defined and potentially misleading is about asking students to identify changes and continuities as “positive or negative.” Presumably whether a change or continuity in history is good or bad depends on the perspective and problem-definition, and part of the value of learning about history is that we learn this nuance. Thus, instead of asking students to provide an absolute normative evaluation, we suggest asking students to consider and articulate nuance. For example, have them reflect on *who* might have considered something to be positive/negative and why or what we might learn from the ways of the past even if people decided not to continue with those ways.

This deeper dive analysis also suggested that teachers were modifying the activities to reduce the difficulty level and increase accessibility. For example, we observed that one teacher converted introductory information about *continuity and change*, written at an 11th grade level, to a series of slides written at an eighth grade level. This involved the teacher highlighting *World History Project's* articulation of the definition and purpose of *change and continuity over time*, that we suspected students may have skipped over if they were simply provided the handout. We also observed that another teacher converted a four-page PDF into two pages (which reduces photocopying by half). We discuss the idea of modifications again in Section E.

Finally, this analysis made us wonder about student engagement. While the activities from the comparison curriculum were not particularly inspiring in terms of how students were going to learn how to identify important continuities and changes over time in history, there were two activities (by the same teacher) that included a few videos clips that inspired some awe and

wonder, from the perspective of the researchers, about history that we did not feel when reviewing materials from *World History Project*. Two such video clips were from *The History Channel* and combined visually compelling reenactments and curated narrative that explained the significance of factors (e.g., writing, agriculture) that contributed to historical achievements or changes. The other video clips were individual posts on YouTube and had no narration but visually transported the viewer to different locations across time and the world. These videos had a different quality to videos we observed within *World History Project* (which appear to focus on conveying a lot of information verbally). We discuss student engagement further in Section D.

Possible reasons for score differences not being observed across curricular groups for most other historical thinking skills (i.e., argumentation, causation, contextualization, comparison, and sourcing)

For *argumentation, causation, contextualization, comparison, and sourcing*, the quantitative analysis did not show clear advantages for *World History Project*. From the perspective of the curriculum designer, this may come as a disappointment, particularly given that they are each spiraled throughout the course. We tried to identify why *World History Project* did not score higher by thinking of major reasons for this and identifying their likelihood given the available evidence. We considered implementation factors, reasons associated with the curriculum, factors external to the curriculum, and instrumentation.

Exhibit C1 shows a summary of our analysis. We did not think that implementation fidelity or quality of the rubrics were major factors that impacted the results. We found it most likely that teachers need more time and experience with *World History Project* to internalize it and implement it well. This need was a common refrain among teachers who were interviewed (see Section E), and aligns with an understanding among scholars that quality enactment of new curriculum is a complex and multifaceted process (e.g., Ball et al., 1996; Dietiker et al., 2018) that can take teachers more a single year (e.g., Obara et al., 2010).

In addition, difficulties in content and format of the curriculum likely contributed to the intended impacts not being realized as strongly as expected. Teachers often modified *World History Project* materials for a variety of reasons, including reducing the difficulty level and making the content more accessible to students. Some of this was discussed in the previous section (on *continuity and change over time*), and more will be discussed in the curriculum usability section (Section E) below.

For some specific historical thinking skills, we found it somewhat likely that external factors, such as availability of other skill-focused curricular materials, or the extent to which a skill lends itself to teaching of required content, may impact the adoption of *World History Project*. For some skills like causation, there are likely other curricula that many teachers/districts have been using that work well for them, so *World History Project* may not necessarily provide a relative advantage. While for some other skills like contextualization, even if *World History Project* were to provide activities, they might not be readily adopted if these skills were considered less directly relevant to content standards for which many teachers feel accountable.

Exhibit C1

Examination of why rubric score differences across curricular type were not observed across many historical thinking skills

Evaluation Question: Why were rubric scores comparable across curricular types, despite the explicit emphases *World History Project* places in supporting these skills?

Possible reason	Likelihood of reason	Rationale for likelihood rating
There are <u>external pressures that make the teaching of historical thinking skills difficult</u> for all teachers, regardless of curriculum.	Likely, especially for contextualization and sourcing	Overall, teachers were not implementing at a high level (Appendix B). State standards generally have a very heavy content focus, so the amount of time that teachers feel they can dedicate to teaching any particular skill is limited. In addition, historical thinking skills are difficult to teach regardless of curricula (National Research Council, 2005), and for some historical thinking skills, such as contextualization, there is scant research evidence on how to teach these effectively (Huijgen et al. 2019; Van Boxtel et al., 2018). Furthermore, argumentation, causation, and comparison appear to have much more direct connections to curricular content standards, with contextualization and sourcing having fewer direct ties.
<u>Study participants did not adequately implement</u> the curriculum.	Not so likely	Teacher interviews and surveys suggested that the vast majority of study teachers took implementation seriously and tried to follow the study request (e.g., used <i>World History Project</i> for at least 60% of their instruction), and felt successful in that regard. Review of the submitted activities support that teachers were implementing the curriculum, generally speaking (but see the three rows below for some difficulties and complications related to implementation).
<u>Teachers needed more time and experience with the curriculum</u> to learn and internalize it.	Very likely	Teachers who were new to the curriculum reported this to be the case (see Section E). The conceptual organization of the curricula requires teachers to understand and teach in a way that students can remember and build on a wide range of skills over time, which generally takes practice.
<u>Content difficulty</u> might get in the way of teaching and learning using <i>World History Project</i> .	Likely	<i>World History Project</i> expects a lot of independent reading, on extended and complex topics that students (and potentially many teachers) have little/no background on, which could be challenging, especially for English Learners. Teachers have been creating scaffolds and adaptations to reduce the difficulty for students. See Section E for more.
<u>Format difficulty</u> might get in the way of teaching and learning using <i>World History Project</i> .	Likely	Teacher interviews and review of activities indicate the teachers have struggled to navigate the site and select what is relevant for their students and in several cases were reformatting PDFs into slides or Word documents. See Section E for more.

<u>Other established curricula</u> are available to teach these historical thinking skills equally well.	Somewhat likely for some skills	<i>The DBQ Project</i> and Stanford History Education Group's <i>Reading Like a Historian</i> were often mentioned in teacher surveys and interviews. Activities that were submitted from these scored high on argumentation, causation, and, in some cases, sourcing.
<u>Our research rubrics</u> were too crude or otherwise insufficient to detect meaningful differences.	Not so likely	Rubrics were consistently applied for the most part (Appendix B, Table B1). During follow-up examinations of activities and student work, nearly always, the rubric scores matched our expectations in terms of the learning opportunities, and rubrics scores have so far not felt too crude or harsh.

Summary of findings. To summarize, the data suggest the following:

- Adoption of *World History Project* helps teachers provide more learning opportunities in *continuity and change over time in history* because it is currently the only world history curriculum in the market that provides multiple activities that intentionally emphasize this skill.
- A unique affordance of *World History Project* is that it supports - or is positioned to support - students to see the big picture and understand important themes related to world history, which are difficult to glean when relying on traditional textbooks or encyclopedia articles that chronicle specific events.
- Adopters of *World History Project* provided comparable, but not necessarily more, opportunities for students to learn many of the historical thinking skills, likely because:
 - Teachers needed more experience implementing *World History Project* in order to become comfortable.
 - Some difficulties related to the content and format/organization of *World History Project* limit its usability to teachers and students (see also sections D and E).

D. Whether students who learn from *World History Project* find world history more relevant, engaging, and/or perspective changing

To examine whether students using *World History Project* find world history more relevant and/or engaging, we conducted pre- and post-surveys⁶ with students in the classes randomized to curriculum treatment. It's important to keep in mind both that all teachers in this random group were open to using the *World History Project* curriculum, and that those assigned to use it were implementing it for the first time.

⁶ Teachers administered pre-surveys during the first month or two of instruction, between mid-September and the end of October, and administered post-surveys during the last month of instruction (mid-December for the one teacher in our random sample who taught world history as a double-blocked semester course, and May for the remaining teachers who taught the course over the course of a full school year).

Differences across groups in student interest, enjoyment, and relevance

The survey asked students to indicate the degree to which statements about interest and enjoyment in world history described them (*not at all like me, a little bit like me, somewhat like me, quite a bit like me, or exactly like me*). We found that compared to students using business-as-usual curricula, a smaller percentage of *World History Project* students agreed in the spring semester that world history is one of their favorite subjects to study, that the class was enjoyable, and that learning about world history is important for their future (see Exhibit D1). However, **the only statistically significant change⁷ from fall to spring was in the percentage of students who felt that learning about world history topics would be important for their future, with a 13% decrease in the proportion of *World History Project* students who strongly identified with the statement, compared to a 6% increase in strong agreement with this sentiment among students in the comparison group.**

⁷ We estimated the statistical likelihood of the difference we see in the change across time across student groups by conducting two-level hierarchical linear modeling of the spring survey responses, controlling for clustering at the classroom level, and the mean pre-survey results at the classroom level. We assumed the Likert response options were equidistant across levels.

Exhibit D1

Comparing World History Project and business-as-usual students' interest and enjoyment in world history class in the fall and spring

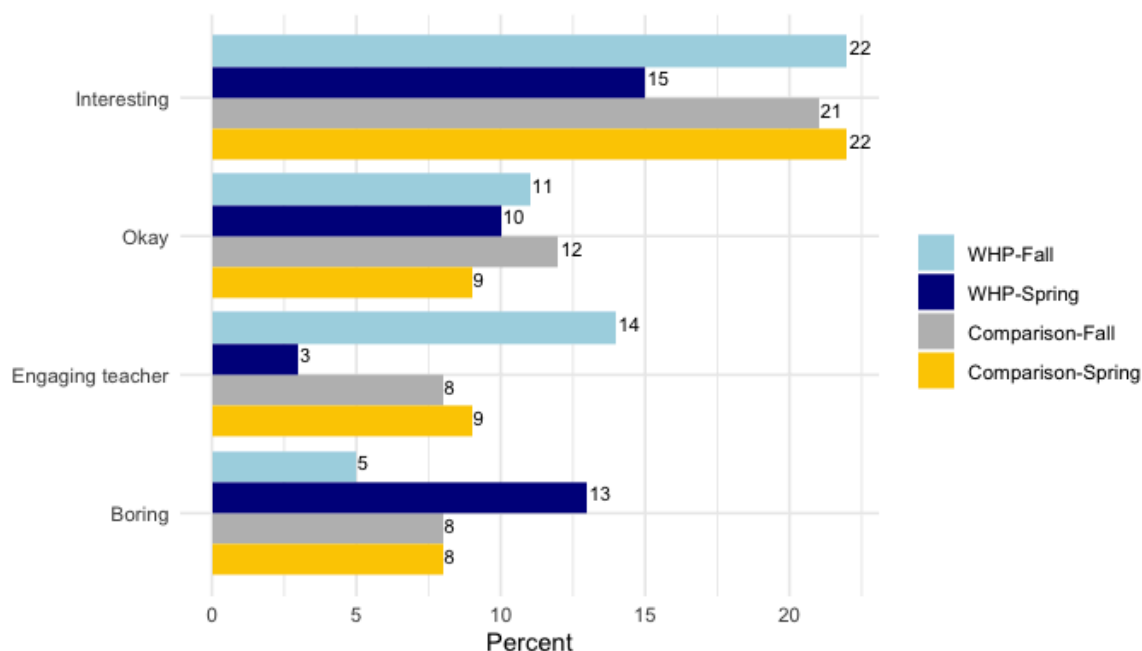
Statement	Students who strongly identified with the statement (selected <i>quite a bit like me</i> or <i>exactly like me</i>)				p-value
	World History Project (fall)	World History Project (spring)	Comparison (fall)	Comparison (Spring)	
World history is one of my favorite subjects to study.	24%	19%	27%	31%	.180
I enjoy doing schoolwork about world history.	18%	17%	22%	21%	.650
I enjoy discussing world history topics with others.	25%	20%	30%	30%	.503
I think that world history helps me understand what is happening in the world around me.	40%	27%	31%	37%	.164
I think that learning about world history topics will be important for my future.	37%	24%	36%	42%	.033*

Note. Includes responses from 292 students who took the pre-survey and 251 students who took the post-survey from 14 classrooms that were randomly assigned to curricular condition. Statistically significant differences in the post-test scores between the *World History Project* and comparison students highlighted in gray, *p < 0.05. Survey items were adapted from U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2018 U.S. History Assessment.

We also asked students to select up to three words/phrases that best described their world history class, from several choices (e.g., too easy, too hard, didn't learn much, okay, learned a lot, relevant, fun, uninteresting). Among *World History Project* students, 63% of the words selected in the fall were positive, compared to 60% for the comparison group students. In the spring, the percentage of positive words selected by *World History Project* students decreased to 45%, while the percentage of positive words selected by the comparison group increased to 65%. As seen in Exhibit D2, **at the end of the year, more *World History Project* students described their class as “boring” and fewer described it as “interesting.”**

Exhibit D2.

Percent of top four words that World History Project and comparison students selected to describe their class



Note. Compares responses from 292 students who took the pre-survey and 251 students who took the post-survey from 14 classrooms that were randomly assigned to curricular condition. Students could select up to three words.

Open-ended survey items related to interest and relevance told a similar story. In open-ended responses, a slightly greater proportion of *World History Project* than comparison students said the most recent unit they studied in world history was not related to their everyday life (32% of *World History Project* responses vs. 25% of comparison responses), that nothing in their world history class was interesting (11% of *World History Project* responses vs. 6% of comparison responses), and that historical information from other parts of the world is not relevant to how we live today (18% of *World History Project* responses vs. 3% of comparison responses).

Open-ended responses from *World History Project* and comparison students who did find world history interesting and relevant were similar. For example, when asked to share the most interesting thing they learned from their world history class during the past year, about two-thirds of students in each group named a topic related to war or genocide.

As another illustration, about the same percentage of *World History Project* (32%) and comparison (31%) students said that history is relevant because learning about the past helps us understand change over time and the development of our contemporary circumstances (e.g., a *World History Project* student responded, “Some of the information is relevant to how we live

today because it has shaped how we live and do things,” while a comparison student wrote, “I think it’s relevant so we know where things come from and where we come from”). Similarly, although more comparison (41%) than *World History Project* (25%) students wrote that knowledge about the past can help us tackle contemporary issues or make fewer mistakes today, the level of complexity and substance of open-ended responses from the two groups were comparable. For instance, one comparison student shared, “It is relevant because without information of how something bad happened, we could repeat the same mistakes,” while one *World History Project* student wrote, “Historical information is relevant to how we live today because we can see the thing people have done wrong in the past so we cannot relive horrible things from the past.”

Differences across groups in historical perspective taking (i.e., the tendency to consider multiple viewpoints before arriving at a conclusion about historical events)

We also asked surveyed students to indicate the degree to which statements about perspective-taking in history described them (not at all like me, a little bit like me, somewhat like me, quite a bit like me, or exactly like me). While the two groups of students responded similarly to baseline questions around perspective taking, at the end of the year, fewer students who had learned using *World History Project* agreed that the items were “exactly like me” or “quite a bit like me” (see Exhibit D3). However, the only statistically significant changes⁸ from fall to spring were in the percentage of students who strongly identified with the statements “I form opinions about historical events only after I have information from more than one source” (30% of *World History Project* students in the spring compared to 51% of comparison students), and “I want to know what lies behind the story when I study a conflict in history” (43% of *World History Project* students in the spring compared to 54% of comparison students).

⁸ See previous note for details on statistical methods.

Exhibit D3

Comparing World History Project and business-as-usual students changes in perspective taking from the fall to spring

Statement	Students who strongly identified with the statement (selected <i>quite a bit like me</i> , or <i>exactly like me</i>)				p-value
	World History Project (fall)	World History Project (spring)	Comparison (fall)	Comparison (spring)	
I need to know the history leading up to an event to truly understand it.	62%	33%	65%	55%	.081
I try to understand others better by imagining how things look from their perspective.	58%	36%	60%	51%	.119
I try to look at everybody's side of a disagreement before I make a decision.	61%	41%	58%	60%	.060
I think that there is more than one side to every question, and I try to look at all of them.	63%	40%	54%	60%	.099
I form opinions about historical events only after I have information from more than one source.	44%	30%	52%	51%	.005*
I want to know what lies behind the story when I study a conflict in history.	56%	43%	60%	54%	.034*

Note. Includes responses from 292 students who took the pre-survey and 251 students who took the post-survey from 14 classrooms that were randomly assigned to curricular condition. Statistically significant differences in the post-test scores between the *World History Project* and Comparison students highlighted in gray, *p < 0.05. Survey items were adapted from U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2018 U.S. History Assessment.

Teacher interviews and student focus groups suggest that the emphasis on reading, writing, and historical skill building in *World History Project* is related to these patterns around student engagement and perspective taking.

To help contextualize these quantitative trends, we analyzed 17 teacher interview transcripts, including interviews with two teachers who were randomly assigned to teach using *World*

History Project, four who were randomly assigned to business-as-usual, four who used *World History Project* without having been randomly assigned, and seven who used other curricular materials without having been randomly assigned.

Compared to the curricula they used previously, the teachers randomly assigned to use *World History Project* during the study year reported that their students were exposed to a broader range of historical perspectives and sources and that students particularly enjoyed the inclusion of assets such as graphic biographies that centered stories that are often left untold. Both teachers also reported that the cohesion of materials provided more consistent opportunities and support for incorporating historical thinking skills, including those related to perspective taking. At the same time, they shared that their students found the amount of reading and writing in *World History Project* tedious and struggled at times with the difficulty level of activities.

Similar themes emerged in interviews with teachers who used *World History Project* but were not randomly assigned. For example, one teacher expressed how much her students had grown by using scaffolded historical thinking activities but at the same time noted how intensely they disliked being asked to demonstrate their mastery through written expression:

“I have been able to take things that would be just an activity and help my kids learn how to express themselves better [through writing] even though they hate, I mean, they hate it. Every assessment, I get asked ‘Can’t we just do a multiple choice [test]?’ They want [multiple choice options] so bad. I’m like, ‘No, it’s good to be written. You’re gonna show me what you know. Not regurgitate for me what you know.’”

We heard similar pushback against reading, writing, and original analysis in focus groups conducted with students from two classrooms, one where the teacher was assigned to use *World History Project* and the other where the teacher was assigned to business-as-usual. See the case study vignette below for more details.

Comparing perspectives of students in two classrooms: A case study

To hear more directly from students about what they perceive as interesting, relevant, and meaningful in their world history class, we conducted virtual focus groups with students from two classrooms: one where the teacher was using *World History Project* for the first time and another where the teacher reported teaching using a wide variety of curricular resources (e.g., an outdated textbook, primary sources from *The DBQ Project*, activities from *World History*

For Us All, and teacher-created materials). Both educators had over 15 years of teaching experience and worked in settings where about 50 percent of their students received free or reduced-price lunch. Two students shared their experiences about the *World History Project* class, while five students shared experiences about the comparison class. While these two classrooms cannot represent the sample as a whole, the cases illustrate the tension world history teachers can face when balancing a desire to make classroom activities appealing and interesting for students while also academically rigorous and provide a lens for interpreting the trends that emerged in student surveys.

A typical day in the comparison classroom

The teacher starts class with a 10–15-minute presentation, pausing frequently for students to pose and respond to questions. Next, students use their textbooks or a copy of the teacher’s slide deck to complete a worksheet or a summary chart related to the lecture, where they record information such as definitions of keywords. Students find the lectures “interesting” and feel the teacher is “very creative with the way she sets things up” by including illustrations and colorful backgrounds in her slide decks and incorporating humor in her delivery. In each unit, the teacher also provides opportunities for collaborative group work, such as a project where students were asked to design a theme park with rides and attractions related to the Protestant Reformation. Students recalled few opportunities for engaging in specific historical thinking skills, although after several probing questions they did remember completing Venn diagrams on more than one occasion. For example, they recalled comparing knights and samurai: “Training is somewhat similar. Their armor is different.”

A typical day in the World History Project classroom

Rather than lecturing, the teacher uses a variety of sources to present various viewpoints and evidence to students and asks them to grapple with the content, providing support as needed. Students shared that they were asked to read, annotate, and interpret passages about four times per week and engage independently or collaboratively to develop their own interpretations using historical evidence. As one student put it, “There was no PowerPoint or a slide or anything, we figured [it] out while we were reading it. And then, if we couldn’t figure it out, she would’ve helped us or something. But she mostly lets us figure it out on our own and use the passages to use it. But then, if we need help, she’s there to advise us. She doesn’t give us the answer right away, but she gives us clues to figure out the answer.” Students shared a desire for additional scaffolding in historical thinking activities, particularly document-based questions, with one student pointing out, “They seem too intimidating. So I mean, there’s no effort put in if it seems too hard for someone to do.”

Comparing student interest in their world history class

We asked students in each group to rate their world history class on a scale of 1 to 5, with 5

representing “very, very interesting” and 1 representing “not interesting at all.” Both of the *World History Project* students gave their class a 3, while all comparison students gave their class a 4 or 5. Although students learning from *World History Project* were satisfied with the range and depth of content they were exposed to, they assigned a lower rating because they wished they were assigned less reading and that their teacher incorporated a wider range of activities. One student explained, “Some of the information was useful and I learned some stuff that I didn’t know (...) it was some good reading, but it was just too much for a student, I think, in my opinion.” Her classmate elaborated that not only were students expected to engage in “a lot of reading,” but they were also asked to draw conclusions about “the actual impact of things like the revolution and imperialism and industrialization and transoceanic connections” without their teacher explicitly telling them “how all of it fits together.”

On the other hand, students in the comparison classroom found their world history class highly interesting because they appreciated the teacher’s presentation style and learning compelling facts. For example, one student was excited to learn that, “Vikings didn’t actually wear horned helmets. They just wore rounded, it’s like a regular helmet,” while another found it “really cool” to learn that in the Renaissance, “canvases were usually made of vellum, which is a form of animal skin.” Additionally, students in the comparison class also found it “helpful” and “less time-consuming” that their teacher pointed them to sections of their textbook where they could find answers when filling out worksheets and provided them the “main point,” and they enjoyed having opportunities to express their creativity in projects through color and design.

Comparing relevance of world history to students

Both students in the *World History Project* classroom focus group agreed that the content they learned in their class was meaningful for their everyday lives. Students cited examples of how Russia’s invasion of Ukraine echoed previous historical conflicts and themes of imperialism and how learning about the history of slavery provided context for understanding contemporary systems of oppression. One student explained, “In our curriculum, there has been a big center around slavery and different forms of it and different types of it and who all did it, because basically every country did it. And I think that is very important and that is very prevalent [sic] to our everyday life.” On the contrary, five of the six students in the comparison classroom shared that although they found their class enjoyable and interesting, they could not think of any particular way in which the content they learned was relevant for their life.

These cases highlight the tension between attending to student enjoyment and interest while challenging students to think historically in ways that they might push back against. The teacher using *World History Project* for the first time incorporated a focus on reading, writing, and argumentation that students felt was too intense, and interest declined as a result. At the same time, the ways in which students in the classroom using *World History Project* engaged with historical content seems to have provided more opportunities for making connections

between events of the past and their own lives and for engaging in meaningful historical interpretation. Alternately, the comparison teacher who used a variety of materials to supplement an out-of-date textbook was able to maintain student interest through her charismatic delivery during direct teaching, and by piquing students' curiosity about the past by incorporating historical facts that she knew students would find engaging. However, students in her class seem to have had fewer opportunities to interrogate multiple perspectives, build interpretations, make connections, and see the past as more than just a series of facts.

Summary of findings. To summarize, the data suggests the following:

- *World History Project* does not engage students more than business-as-usual curricula, at least when a teacher implements it for the first time.
- *World History Project* students' negative perceptions of their class, relative to students receiving business-as-usual curricula, appears to be related to the greater demands placed upon them for reading, writing, and original analysis.

E. Usability of *World History Project*

To explore the challenges and benefits experienced by teachers new to *World History Project*, we analyzed 13 teacher interviews. Five interviews were with teachers who were part of the larger study of learning opportunities and student work, which included questions about their overall experience with *World History Project*. Six interviews were with an additional set of teachers recruited to better understand usability. Among them, five were relatively new to (within five years of) teaching world history, and all were using *World History Project* as their main curriculum. Their interview focused more specifically on the curriculum's usability. Two final interviews were with teachers who were randomly assigned to teach using *World History Project* but elected to drop out of the study after trying it for a few weeks in the fall.

We examined the interview transcripts to identify themes on challenges and benefits associated with the use of *World History Project*, as well as how and why teachers made modifications to *World History Project* materials. We also examined teacher surveys to triangulate some of these findings.

Usability challenges of *World History Project*

All 11 interviewees who had used *World History Project* for the entire school year described feeling overwhelmed during their first year of using it. The sheer volume of resources *World History Project* offers was frequently cited as the cause of these sentiments. Some teachers specifically brought up the large number of articles per unit and expressed frustration around how to determine what materials they should use in their instruction. Most teachers noted that over time, they became more familiar with the platform and that this reduced their initial sense of

overload. However, teachers' learning curves varied, and it appears to take at least a year for teachers to feel comfortable navigating and selecting among *World History Project* resources.

Similar sentiments came up for the two teachers who were randomly assigned to *World History Project* and elected to drop out of the study after trying it for a few weeks in the fall. They appreciated the value of the wealth of articles, activities, and resources in *World History Project* but wished for a more digestible presentation. They said that the time that it took them to sit down, absorb the necessary professional development materials, and then adapt the curriculum for their particular students posed too much of a hurdle for them to continue using the curriculum.

How and why teachers made modifications to World History Project materials

Among the 11 interviewed teachers who were using *World History Project* as their main curriculum, eight acknowledged the need or desire to make modifications to *World History Project* materials. Reasons for this were consistently tied to improving student understanding. Teachers highlighted that the difficulty of some resources posed challenges for students, requiring adjustments to be made. For instance, several teachers mentioned adapting complex activities into smaller, more manageable segments to enhance student engagement and learning. Notably, a few teachers reported that certain asset types, such as graphic biographies and document-based questions, had to be excluded entirely from their teaching plans due to their high level of complexity.

Teacher responses showcased a wide spectrum of teacher-initiated modifications, ranging from copying and pasting materials to adjust formatting, to translating worksheets to meet language needs, to crafting custom materials aligned with students' abilities. Teacher surveys echoed the challenges of modifying materials. One teacher shared, "I created/provided graphic organizers to help guide their reading or guided reading questions but at times, it seemed like it was a lot, and students kind of checked out." The curricular activity samples submitted by teachers using *World History Project* tell a similar story with many teachers creating additional slides and documents to provide further explanation and guidance for their students.

To better understand the nature of these teacher modifications, we examined 22 activities in which students scored above 1 in the historical thinking skill of argumentation. Among these, 16 out of 22 activities were modified. We also assessed an additional four activities that had been flagged by two researchers as having noteworthy modifications. We observed that these modifications fell into several main categories:

Teacher-created scaffolds: In 10 instances, teachers created scaffolds often in the form of graphic organizers or organizing tools designed to assist students in tackling specific aspects of the lesson, like breaking down document-based questions and organizing notes. We hypothesize that teachers have created these scaffolds because students may need additional support to successfully complete assignments, possibly due to their complexity. However, this doesn't necessarily imply that *World History Project* should universally provide more graphic organizers, as these tools are often tailored to the unique needs of a teacher's context. Nonetheless, it could indicate the need for more accessible resources.

Use of slide decks: In nine cases, teachers developed slide decks to further explain assignments. Converting activities into PowerPoint presentations served to visually simplify instructions, making them more comprehensible. An associated recommendation for curriculum developers would be to consider creating starter slides for teachers. This would enable educators to readily customize and display relevant information.

Transformation into editable Word documents: In three instances, activities were turned into editable Word documents. This approach aimed to make it easier for students to fill in worksheets without having to print PDFs. It was also used to streamline complex worksheets and assignments from the *World History Project* platform.

Other modifications included teachers pulling together resources from across the platform to add to their lesson or changing the structure or prompt of a lesson.

Benefits experienced by teachers through their use of World History Project

Several positive reactions to *World History Project* did emerge from these teacher interviews. One teacher highlighted the transformation in classroom dynamics, stating that the curriculum facilitated a shift from traditional teacher-centered instruction to a more student-centered environment. This shift was attributed to the diverse range of activities incorporated into the curriculum, including video content, reading materials, and graphic biographies. Another teacher noted an increase in student engagement with the activities provided in the curriculum (note: this report by an individual teacher conflicts with the majority of student responses described in Section D).

Additionally, several teachers highlighted the curriculum's wide range of materials that seek to highlight various socioeconomic, racial, and gender backgrounds and decenter European perspectives and narratives. *World History Project* was also sometimes described as a flexible tool that allowed teachers to take ownership of their teaching, aligning with their specific goals, whether focused on reading, writing, historical thinking skills, or other pedagogical aims.

Summary of findings

- Teachers, particularly those who are new to teaching world history, often find *World History Project* overwhelming in their initial year of adoption, due to the abundance of available resources.
- Teachers often feel a need to make substantial modifications to *World History Project* resources to suit their students' comprehension levels, and the process of making these modifications can be time-consuming.
- In spite of common challenges, some teachers reported positive outcomes from using the *World History Project*, including the opportunity to enhance student engagement and to present a more comprehensive view of global history.

Discussion & Implications

Summary of findings and discussion

Traditional history textbooks focus on historical knowledge (e.g., periods, themes, regions, people, and events), with little opportunity for students to analyze and apply that content using historical thinking skills (e.g., making historical arguments, identifying patterns over time, establishing relationships of cause-and-effect). While a number of open-access resources exist for supplementing history instruction to include individual historical thinking skills, such as sourcing (e.g., Library of Congress, National Archives, Stanford History Education Group's *Reading Like a Historian*), *World History Project* is one of the first freely available world history curriculum to embed scaffolded historical thinking progressions for a range of skills within a comprehensive course. This evaluation has explored whether and to what extent using *World History Project* helps teachers provide more opportunities for historical thinking and helps students learn more historical thinking skills and find world history more relevant, engaging, and perspective changing. In addition, we explored educator perceptions of the curriculum's usability.

Our evaluation surfaced several key findings regarding the impact of *World History Project* in its early years of adoption by teachers who have some baseline interest in using a curriculum that emphasizes historical thinking skills. We found that this curriculum appears to support teachers to provide more learning opportunities about *continuity and change over time in history* and to help students to better learn this complex skill. *World History Project* has unique affordances toward this end, likely because it was intentionally designed to support these skills, in collaboration with historians and world history education experts.

At the same time, perhaps somewhat disappointingly for a curriculum that aimed to support the teaching and learning of a broad set of historical thinking skills, we observed *World History Project* as being comparable with comparison curricula and not having a strong relative advantage on five other historical thinking skills. There are many plausible reasons for this. Studies on curriculum adoption (the majority of which have focused on mathematics) have found that it generally takes teachers more than a single year to feel comfortable with a new curriculum, even if the curriculum is not particularly rigorous and even if the curriculum is paired with quality professional learning (Obara et al., 2010). A variety of sources also suggested that this initial version of *World History Project* was likely pitched at too high a level for many high school teachers to use without modification in their on-level ninth and 10th grade classes. These factors likely contributed to our not detecting differential impacts.

Our evaluation also suggested that while *World History Project* supports the development of students' attention to *continuity and change over time*, students tend to find the lessons uninteresting and can end up liking world history less by the end of the second semester, at least when the curriculum is taught by a teacher for the very first time. Here again, we see the importance of providing teachers with adequate time to become comfortable with implementation and to adapt the materials for their local context and the needs, strengths, and identities of themselves and their students.

Our findings about student engagement highlight a tension that often exists in mandated learning, between rigor and enjoyment. Ideally, schoolwork is both rigorous and enjoyable, but especially in a developing adolescent brain, it is difficult to provide the right balance; the right amount of enjoyment to ease the strain of the rigor, and the right amount of rigor to ensure healthy development and growth. When teachers emphasize reading, writing, and skills practice without providing students a clear understanding of the purpose of those practices, adjusting or modifying the materials, or diversifying the types of activities they plan from day to day, students are likely to find the course stressful (as we heard in our focus group). In the next round of its development, *World History Project* may want to use student and teacher feedback and expertise to strike a better balance between rigor and enjoyment.

World History Project materials are designed to be educative and support teachers' learning and understanding of how to help students progress in their historical thinking. In interviews, teachers expressed a desire for additional time for reflection and planning as they interacted with these educative features for the first time. Additionally, teachers shared that materials often required adjustments to match students' comprehension levels, which can be time-consuming for teachers. This, compounded with *World History Project's* wealth of resources, highlights how new teachers may feel overwhelmed adopting the curriculum. Despite these challenges, some educators reported positive outcomes, including increased student engagement and the ability to offer a more comprehensive global history curriculum. Research tells us that teacher proficiency and comfort with a new curriculum increases gradually over time (Obara et al., 2010), and our findings provide guidance for further research and development in ways to support teachers new to the curriculum who may feel overwhelmed by the quantity of ancillary materials and instructions.

Implications for *World History Project* curriculum developers

The findings presented have significant implications for developers of *World History Project* and perhaps more generally for curriculum developers in the field of history education. One theme that emerged is the **importance of prioritizing usability for teachers, especially those who are new to teaching or to teaching world history**. To address this in *World History Project*, curriculum developers may want to consider providing a “quick start” overview of key resources, which can help teachers avoid feelings of overload during the initial phases of curriculum implementation. Given the vast amount of resources the curriculum offers and how overwhelmed this causes teachers to feel, it seems important to emphasize OER Project's guidance that teachers are not expected to use every resource provided and instead are recommended to pick and choose what aligns best with their teaching goals and students' needs. That said, teachers new to *World History Project* are likely to struggle with the curation of resources for a wide range of reasons. In response, we recommend OER Project provide concrete guidance to teachers who may not want to or have the bandwidth to curate resources on their own. This might be accomplished, for example, by making sample course plans more accessible and readily available, as they can serve as invaluable templates for educators. It may also help teachers to design curriculum resources to be more easily editable (e.g., Word

documents rather than PDF), to ease the burden on teachers trying to adapt materials to suit their specific contexts.

Consider ways to support teachers in making their instruction more engaging and relevant for students. *World History Project* is a rigorous curriculum that challenges teachers and students alike. Teachers, especially those new to using the curriculum, need additional support in understanding how to contextualize, adapt, and deliver *World History Project* materials in ways that feel most engaging and relevant for their student population. For example, teachers may struggle to decide how much reading to assign, and students may push back against receiving one reading assignment after another. In curating resources for teachers new to the curriculum, we recommend highlighting activities that teachers across various contexts have found most engaging for their students.

Additionally, curriculum developers may want to consider ways to improve accessibility for students. This includes adjusting the difficulty level of materials to meet the diverse needs of students. Similarly, language translations across the platform for graphic biographies, primary source selections, and articles across all Lexile levels, could be beneficial to support diverse student populations. It would be beneficial to incorporate scaffolds in lesson plans to support students in developing their historical thinking skills progressively. Additionally, based on the modifications made by current *World History Project* teachers, developers may want to consider creating starter slide decks for and editable Word documents of their materials. To implement these changes, we recommend collaboration with teachers from a wide variety of different contexts, with the goal of creating resources that are adaptable and effective in diverse classroom settings.

Another implication pertains to the emphasis on historical thinking skills within the curriculum. Curriculum developers may want to consider the possibility that educators want to spend relatively more or less time on skill-building based on how they perceive the activity to contribute to content-related goals (e.g., historical causes might be more emphasized in state content standards than contextualization or sourcing, so teachers may be more inclined to teach the former activities). If the goal is to promote higher order learning and skills, it may be beneficial to take these differential teacher motivations as a given and design skill building activities partly in service of teachers' perceived compliance goals. For example, a *continuity and change* activity that helps students successfully compare and understand the gist of two eras could be valued by teachers because it reinforces what was already conveyed about the older era, previews what is about to be conveyed about the more contemporary era, and seems to allow the student to do better in state exams that will likely ask about major facts and trends.

Finally, we recommend that *World History Project* maintain its core aspirations and features, including its focus on providing a comprehensive global world history curriculum, focus on historical thinking skills, and its development approach of working closely and iteratively with educators and historians. *World History Project's* global approach can support students to gain a broader understanding of historical events and perspectives from various cultures and regions worldwide. Emphasizing critical historical skills such as historical argumentation, causation, comparison, and contextualization equips students with the necessary tools to think critically and analytically about historical events and their significance.

By soliciting feedback from those directly involved in the teaching and learning process, *World History Project* can continue to evolve and adapt to the changing needs of educators and students.

Implications for teachers and school districts considering adoption

Potential curriculum adopters, particularly those seeking cohesive materials and resources that take a global (as opposed to a more regional or “Western Civilization”) approach, and incorporate scaffolds for disciplinary understanding of historical thinking skills, may want to consider *World History Project*. However, potential adopters should be aware that the curriculum, as it stood in the 2022–23 school year, had areas of needed improvement.

Implications for social studies administrators. Our usability findings suggest that teachers, especially those with less content matter knowledge and pedagogical content knowledge, will need support in exploring the *World History Project* resources and making decisions about which materials to use and how to adapt them for their setting. We recommend that administrators provide adequate time for teachers to explore the curriculum before beginning to implement it and consider having a curriculum head or highly skilled and experienced world history teacher curate OER Project resources (e.g., historical thinking skill progression placemats, sample course plans) that provide an organizational framework for teachers new to the curriculum. Depending on their background, teachers may not have received prior professional learning related to fostering historical thinking. For this reason, we recommend that administrators be explicit about why they selected the curriculum and what they see as its benefits. To avoid overwhelming teachers, we recommend that administrators be explicit about how many of the resources and activities they expect teachers to incorporate into their instruction and consider developing a sample district course plan with recommended assets.

In particular, we suggest that district social studies administrators who adopt the curriculum for their school or district provide clear guidance on the extent to which they expect teachers to emphasize particular historical thinking skills and activities. In making these decisions, we advise administrators to highlight connections to required content and consider vertical alignment across grade levels. For example, in a district where history teachers have never been expected to explicitly teach historical thinking skills, an administrator could consider what scaffolds and supports (both within and outside of OER Project) might be most useful for teachers and students. Conversely, in a district where world history teachers have been expected to center their instruction on historical thinking and where students have engaged in activities targeting specific historical thinking skills (e.g., sourcing historical documents) throughout their elementary and middle school years, scaffolds for those particular skills might be removed, modified, or used at the teacher’s discretion.

Administrators could also carve out time during district-provided professional development for teachers to engage in professional learning opportunities offered by OER Project (e.g., webinars, conferences) and for teachers to meet in communities of practice to discuss lesson modifications and adaptations and analyze samples of student work.

Implications for teachers. Teachers adopting *World History Project* for the first time should understand that gaining a holistic understanding of the structure of the course (e.g., frames, historical thinking skills, location of resources) will not happen overnight. Our usability findings suggest that some of this understanding will develop naturally over time as teachers learn by using the resources in their instruction. That said, we recommend that before beginning implementation teachers review the example course plan(s) suggested by other teachers in their state and map out a general plan for the year. We strongly recommend that teachers consider the goals they have for their students (e.g., mastering state standards, developing competency in particular disciplinary skills, making connections to cultural histories and lived experiences), as a lens for deciding which *World History Project* assets to use and which to omit. We further recommend that teachers, especially those who may be the only teacher in their school or district using *World History Project*, turn to the online community discussion board to connect with a community of practice and gain insights from teachers who are more experienced with the curriculum.

Implications for future research

Our study has several implications for future research. First, there are technical and practical questions that are important to pursue regarding historical thinking skills education for world history and in history education more generally. We found that teachers find it difficult to teach historical thinking skills, even those who are experienced and are motivated to teach such skills, and motivated to use a curriculum that is designed specifically to help teachers excel in this regard. While more time with the curriculum is likely to help, our findings suggest that there are ways that the curriculum itself can be stronger, and for that, more research evidence is needed on the ways that world history teachers implement historical thinking scaffolds and tools and the ways in which students use those supports, including the challenges that they encounter.

Thus the research we believe is very much needed is practical and tactical, tied directly to serving student and teacher needs at this point of history, in the post-COVID United States and more broadly. How can we better support world history teachers today and their high school aged students? To direct attention to our current context, if a curriculum developer is convinced about the importance of teaching historical thinking skills and has resources and capacity to dedicate toward its development, what specifically should be in that curriculum, and what professional learning opportunities can best support enactment of the curriculum and be delivered at scale? For example, is it important to articulate enduring “frames”? Why (or why not) and how (or how not)? And how and to what extent should skills like *contextualization* and *continuity and change* be taught? Do they need to be taught separately or at the same time as other skills and content? What is both feasible, viable, and motivating in ninth and 10th grade non-IB, non-AP classrooms? How do teachers adopt and modify curriculum, and what are theoretical and practical implications (see Fogo, Reisman, & Breakstone, 2019 on how such research is needed and scant)? We believe that such research questions and ideas occur at the intersection of theory and practice, and can be satisfactorily addressed only by research that positions practitioners, classrooms and students at the center of the inquiry.

Related, our research also highlights the need to explore the balance between rigor and enjoyment in world history education and to investigate what high school students feel to be motivating, relevant, and meaningful to their course of study. In our experience working with educators and being former teachers ourselves, student fulfillment and joy is one of the primary motivations for teachers to engage in their careers and for taking risks to expand their pedagogical repertoire. Investigating how to strike a good balance between academic rigor and enjoyable, engaging teaching methods is essential for designing effective and motivating history curricula.

Finally, our study encourages a deeper examination of the importance of studying world history in high school. It seems important to explore the unique ways in which world history provides opportunities for students to learn and appreciate the value of historical thinking skills (see Bain 2011, Girard & Harris, 2018, Harris 2021, and Shreiner & Zwart 2020, for thought-leadership on this topic). Unlike in U.S. History, world history survey classes must naturally introduce multiple histories from diverse regions and time periods, which likely necessitates teachers to introduce some organizing principles through which students can engage and make meaning of the vast amount of historical evidence and narratives. It provides an opportunity for students to think about enduring themes across different contexts and analyze what is important, which aligns with the practices of historical thinking. This opens up avenues for research into the comparative approaches of teaching world history versus U.S. history and the reasons behind students' perceptions of relevance and interest in these subjects.

Concluding remarks

This evaluation of *World History Project* has shed light on both its strengths and areas in need of improvement. While the curriculum has shown promise in supporting teachers to give more emphasis to some historical thinking skills, it faces challenges in terms of usability and accessibility to teachers and students. The findings emphasize the importance of providing teachers with the time, support, and resources necessary to become comfortable with the curriculum and to tailor it to their students' needs. The curriculum developers can play a supportive role in addressing many of the observed challenges by prioritizing accessibility, refining the balance between rigor and enjoyment, and offering additional support for teachers to contextualize and adapt the materials effectively.

It is important to continue to direct research and development efforts toward improving the quality and usefulness of *World History Project*. As the most comprehensive, extensive, and affordable (free) learning material in this subject area to date, and given how difficult it is to effectively teach a survey course on the history of the world, the curriculum's potential reach and impact is large. The study also underscores the need to continue the design and research, centering the experiences and perspectives of teachers and students that the curriculum aims to support most. As we reflect on this study, we hold great optimism for the continued evolution of the curriculum, given the expertise, capacity, and strong desire of its designers to support teachers. We anticipate that our feedback will serve as a catalyst for ongoing efforts to refine the curriculum, ultimately empowering teachers to effectively instruct world history and enabling students to understand and appreciate the value of learning world history.

References

- Bain, R. B. (2011). Considering the “hidden” challenges of teaching and learning world history. *Historically Speaking*, 12(3), 20–22. <https://doi.org/10.1353/hsp.2011.0034>
- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is — or might be — the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, 25(9), 6–14. <https://doi.org/10.2307/1177151>
- Britt, M. A., & Aglinskas, C. (2002). Improving students’ ability to identify and use source information. *Cognition and Instruction*, 20(4), 485–522. https://doi.org/10.1207/S1532690XCI2004_2
- Brookhart, S. (2015). Commentary: Assessment of historical thinking in practice. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 1–13). Routledge/Taylor & Francis Group.
- De La Paz, S., Felton, M. K., Monte-Sano, C., Croninger, R., Jackson, C., Deogracias, J. S., & Hoffman, B. P. (2014). Developing historical reading and writing with adolescent readers: Effects on student learning. *Theory & Research in Social Education*, 42(2), 228–274. <https://doi.org/10.1080/00933104.2014.908754>
- Dietiker, L., Males, L. M., Amador, J. M., & Earnest, D. (2018). Research commentary: Curricular noticing: A framework to describe teachers’ interactions with curriculum materials. *Journal for Research in Mathematics Education*, 49(5), 521–532.
- Epstein, T. & Salinas, C.S. (2018). Research methodologies in history education. In Metzger, S.A. & Harris, L.M. (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 61–92). John Wiley & Sons, Inc.
- Fogo, B., Reisman, A. & Breakstone, J. (2019). Teacher adaptation of document-based history curricula: results of the Reading Like a Historian curriculum-use survey, *Journal of Curriculum Studies*, 51(1), 62–83. <https://doi.org/10.1080/00220272.2018.1550586>
- Girard, B. & Harris, L.M. (2018). Global and world history education. In Metzger, S.A. & Harris, L.M. (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 253–279). New York, NY: John Wiley & Sons, Inc.
- Hardy, A. & Iwatani, E. (2021). Rubrics for examining historical thinking skills in high school world history activities and student work: Construct validity evidence from the literature. Digital Promise. <https://doi.org/10.51388/20.500.12265/113>

- Hardy, A., Iwatani, E., Means, B., & Seylar, J. (2023) Rubrics for examining historical thinking skills in high school world history activities and student work (2nd ed). Digital Promise. (Original work published 2021) <https://doi.org/10.51388/20.500.12265/111>
- Harris, L. M. (2012). Conceptual devices in the work of world historians. *Cognition and Instruction*, 30(4), 312–358. <https://doi.org/10.1080/07370008.2012.716884>
- Huijgen, T., Holthuis, P., van Boxtel, C., & de Grift, W. (2019). Promoting historical contextualisation in classrooms: An observational study. *Educational Studies*, 45(4), 456-479. <https://doi.org/10.1080/03055698.2018.1509771>
- Iwatani, E., Means, B., Seylar, J., & Hardy, A. (2021). Rubrics on historical thinking skills for assignments and student work: Initial validity evidence. Digital Promise. <https://doi.org/10.51388/20.500.12265/112>
- Joyce, J., Gitomer, D.H., & Iaconangelo, C.J. (2018). Classroom assignments as measures of teaching quality. *Learning and Instruction*, 54, 48-61. <https://doi.org/10.1016/j.learninstruc.2017.08.001>
- Keirn, T. (2018). History curriculum, standards, and assessment policies and politics: U.S. experiences. In Metzger, S.A. & Harris, L.M. (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 119–148). John Wiley & Sons, Inc.
- Korber, A. & Meyer-Hamme, J. (2015). Historical thinking, competencies, and their measurement: Challenges and approaches. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 1–13). Routledge/Taylor & Francis Group.
- Lévesque, S., & Clark, P. (2018). Historical thinking: Definitions and educational applications. In S. A. Metzger & L. McArthur Harris (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 119–148). John Wiley & Sons, Inc.
- Mitchell, K., Shkolnik, J., Song, M., Uekawa, K., Murphy, R., Garet, M., & Means, B. (2005). Rigor, relevance, and results: The quality of teacher assignments and student work in new and conventional high schools. *Washington, DC: American Institutes for Research and SRI International*. <https://www.sri.com/wp-content/uploads/2022/05/Rigor-Relevance-And-Results-The-Quality-Of-Teacher-Assignments-And-Student-Work-In-New-And-Conventional-High-Schools.pdf>
- National Council for the Social Studies (NCSS). (2013). *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*. <https://www.socialstudies.org/system/files/2022/c3-framework-for-social-studies-rev0617.2.pdf>

- National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*.
<http://www.corestandards.org/ELA-Literacy/>
- National Research Council. (2005). *How students learn: History in the classroom*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/11100>
- Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). Authentic intellectual work and standardized tests: Conflict or coexistence? Improving Chicago's schools.
<https://consortium.uchicago.edu/sites/default/files/2023-06/Authentic%20Intellectual%20Work-Jan2001-Consortium.pdf>
- Obara, S., & Sloan, M. (2010). Classroom experiences with new curriculum materials during the implementation of performance standards in mathematics: A case study of teachers coping with change. *International Journal of Science and Mathematics Education*, 8, 349-372. <https://doi.org/10.1007/s10763-009-9176-9>
- Reisman, A. (2015). The difficulty of assessing disciplinary historical reading. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 1–13). Routledge/Taylor & Francis Group.
- Reisman, A., & Fogo, B. (2016). Contributions of educative document-based curricular materials to quality of historical instruction. *Teaching and Teacher Education*, 59, 191–202.
<https://doi.org/10.1016/j.tate.2016.05.018>
- Seixas, P., & Ercikan, K. (2015a). Introduction. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 1–13). Routledge/Taylor & Francis Group.
- Seixas, P., Gibson, L., Ercikan, K. (2015b). A design process for assessing historical thinking. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 1–13). Routledge/Taylor & Francis Group.
- Shemilt, D. (2018). Assessment of learning in history education: Past, present, and possible futures. In Metzger, S.A. & Harris, L.M. (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 119–148). John Wiley & Sons, Inc.
- Shreiner, T.L., & Zwart, D.E. (2020) It's just different: Identifying features of disciplinary literacy unique to world history. *The History Teacher*, 53(3), 441-469.
<https://www.jstor.org/stable/27058609>
- Van Boxtel, C., & van Drie, J. (2018). Historical reasoning: Conceptualizations and educational applications. In S. A. Metzger & L. McArthur Harris (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 119–148). John Wiley & Sons, Inc.

Wenzel, S., Nagaoka, J., Morris, L., Billings, S., & Fendt, C. (2002). Documentation of the 1996-2002 Chicago Annenberg research project strand on authentic intellectual demand exhibited in assignments and student work. A technical process manual. *Consortium on Chicago School Research*.
<https://consortium.uchicago.edu/sites/default/files/2018-10/p67.pdf>

Appendix A: Methodological details

Sample & data collection details

Table A1

Characteristics of teachers who participated in the study

Characteristics	World History Project (N=14)	Comparison (N=19)
Randomly assigned to condition	36%	42%
Years of teaching experience	$M = 12$ ($SD = 4.9$)	$M = 13$ ($SD = 9.7$)
Years of experience teaching world history	$M = 7.5$ ($SD = 6.3$)	$M = 7.6$ ($SD = 7.0$)
Female	79%	79%
Coaches a sport	14%	5%
Teaches world history as a year-long course	79%	79%
Self-rating on competence in teaching historical thinking skills on a 1–5 scale	$M = 3.3$ ($SD = .91$)	$M = 3.3$ ($SD = .75$)
% Free/reduced-price lunch (school level)	$M = .46$ ($SD = .29$)	$M = .50$ ($SD = .24$)
% Black students (school level)	$M = .15$ ($SD = .16$)	$M = .17$ ($SD = .23$)
% Hispanic students (school level)	$M = .32$ ($SD = .23$)	$M = .20$ ($SD = .23$)
% White students (school level)	$M = .47$ ($SD = .30$)	$M = .48$ ($SD = .30$)
% Other racial demographic (school level)	$M = .06$ ($SD = .05$)	$M = .19$ ($SD = .28$)

Note. None of the differences between the two groups were statistically significant.

Table A2*Student survey participant characteristics*

Student characteristics	World History Project (67-69 students reporting)	Comparison (143-145 students reporting)
9th grade	56%	23%
10th grade	44%	77%
Female	46%	45%
Male	41%	38%
American Indian or Alaska Native	<5%	5%
Asian	<5%	8%
Black, African, or African American	19%	8%
Hispanic or Latino	41%	64%
Middle Eastern	<5%	<5%
Native Hawaiian or other Pacific Islander	<5%	<5%
White	24%	50%
A race/ethnicity not listed above	6%	6%
English is the only language spoken at home	45%	62%
English and another language are spoken at home	49%	37%
non-English language is the only language spoken at home	6%	1%

Note. Data from spring 2023 student survey, responses from students from classrooms that took the pre-survey and post-survey. Race/ethnicity categories add to over 100% since many students identified as belonging to more than one group.

Table A3*Everyday activity characteristics*

Everyday activity characteristics	World History Project (79 activities)	Comparison (100 activities)
% 9th grade activities	46%	32%
% 10th grade activities	41%	64%
% Independent work	51%	60%
% Group work	11%	11%
% Independent and group work	38%	27%
% Not graded or graded for completion	61%	46%
% Graded based on criteria established by the teacher	38%	50%
% Students who exceeded expectations (according to the teacher)	36%	50%

Note. Data from teacher lesson overview

Table A4*Summative activity characteristics*

Summative activity characteristics	World History Project (40 activities)	Comparison (48 activities)
% 9th grade activities	45%	31%
% 10th grade activities	38%	60%
% Independent work	80%	71%
% Group work	1%	6%
% Independent and group work	13%	19%
% Not graded or graded for completion	3%	10%
% Graded based on criteria established by the teacher	95%	85%
% Student who exceeded expectations (according to the teacher)	30%	40%

Note. Data from teacher lesson overview

Qualitative data analysis details

Open-ended survey item analysis

After familiarizing ourselves with the data, we conducted thematic analyses of open-ended responses for each individual question. An initial coding pass helped identify patterns within the data and define themes that emerged. We then organized each response by thematic category, quantifying the number of responses which fell under each of the categories. During that process, themes were refined in multiple passes. Most student responses were only assigned to one category. However, responses to the question which asked students to share an example of how history repeats itself were categorized both by the type of concept the student named, as well as the type of analysis/substantive explanation the student provided. In those cases, extended student responses that named a particular concept (e.g., war, genocide) were assigned to two categories.

Student focus group analysis

Two researchers from the team conducted each virtual student focus group. While one researcher led the discussion and asked probing questions, the other researcher took detailed notes and asked clarifying questions. After the focus group concluded, the two researchers met briefly to discuss emerging themes from the session. Next, researchers read and re-read focus groups transcripts, selecting excerpts of text that were consistent with the categories of meaningfulness, relevance, and impact, and recording memos on emerging themes within each category. Multiple coding passes were conducted to filter and focus features of the data relevant to the research questions. Finally, we created tables with a description of the properties of each theme including detailed examples and illustrative quotes. To avoid drowning out individual student voices, we also noted the percentage of students who were in consensus with the theme and percentage of students with a dissenting view.

Teacher interview analysis

During interviews, researchers took detailed notes. Following the interviews, they reviewed their transcripts and added to their notes, categorizing the data by the categories of sample and context, implementation fidelity, opportunity for learning historical thinking skills, evidence of student learning of historical thinking skills, student engagement and relevant, general sentiment about their curricular, and (where applicable) usability for teachers new to teaching world history. Next, the research team met in a series of analytical meetings to discuss and record summary memos on trends within each category for the randomized and non-randomized *World History Project* groups and for the randomized and non-randomized comparison groups.

Appendix B: Data tables

Rubric scoring reliability

Table B1

Rubric scoring reliability on historical thinking skills rubric scores (Cronbach's alpha)

Outcome	Reliability of activities rubric scores	Reliability of student work rubric scores
Historical argumentation	.849	.942
Historical causation	.835	.960
Historical comparison	.908	.953
Historical contextualization	.565	.927
Continuity and change over time in history	.860	.928
Sourcing	.679	.930

Historical thinking skills rubric score descriptive statistics

Table B2

Historical argumentation: Learning opportunity (activity rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	32%	56%	13%	0%
	Comparison (N = 100)	45%	41%	14%	0%
Summative	World History Project (N = 40)	15%	33%	40%	13%
	Comparison (N = 48)	31%	38%	17%	15%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B3

Historical causation: Learning opportunity (activity rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	71%	25%	4%	0%
	Comparison (N = 100)	83%	16%	1%	0%
Summative	World History Project (N = 40)	58%	38%	5%	0%
	Comparison (N = 48)	65%	29%	4%	2%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B4

Historical comparison: Learning opportunity (activity rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	86%	13%	1%	0%
	Comparison (N = 100)	90%	9%	1%	0%
Summative	World History Project (N = 40)	78%	18%	5%	0%
	Comparison (N = 48)	92%	6%	2%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B5

Historical contextualization: Learning opportunity (activity rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	78%	19%	3%	0%
	Comparison (N = 100)	77%	23%	0%	0%
Summative	World History Project (N = 40)	50%	45%	3%	3%
	Comparison (N = 48)	73%	21%	6%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B6

Continuity & change over time in history: Learning opportunity (activity rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	82%	13%	5%	0%
	Comparison (N = 100)	96%	4%	0%	0%
Summative	World History Project (N = 40)	73%	18%	8%	3%
	Comparison (N = 48)	92%	6%	2%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B7

Historical sourcing: Learning opportunity (activity rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	75%	19%	6%	0%
	Comparison (N = 100)	81%	16%	3%	0%
Summative	World History Project (N = 40)	85%	13%	3%	0%
	Comparison (N = 48)	83%	17%	0%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B8

Historical argumentation: Student outcome (student work rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	67%	29%	4%	0%
	Comparison (N = 100)	72%	26%	3%	0%
Summative	World History Project (N = 40)	35%	58%	5%	3%
	Comparison (N = 48)	52%	33%	15%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B9

Historical causation: Student outcome (student work rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	82%	16%	1%	0%
	Comparison (N = 100)	91%	9%	0%	0%
Summative	World History Project (N = 40)	63%	38%	0%	0%
	Comparison (N = 48)	77%	21%	2%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B10

Historical comparison: Student outcome (student work rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	87%	13%	0%	0%
	Comparison (N = 100)	90%	10%	0%	0%
Summative	World History Project (N = 40)	88%	13%	0%	0%
	Comparison (N = 48)	92%	8%	0%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B11

Historical contextualization: Student outcome (student work rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	90%	10%	0%	0%
	Comparison (N = 100)	90%	10%	0%	0%
Summative	World History Project (N = 40)	78%	23%	0%	0%
	Comparison (N = 48)	85%	15%	0%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B12

Continuity & change over time in history: Student outcome (student work rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	95%	5%	0%	0%
	Comparison (N = 100)	98%	2%	0%	0%
Summative	World History Project (N = 40)	85%	15%	0%	0%
	Comparison (N = 48)	98%	2%	0%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Table B13

Historical sourcing: Student outcome (student work rubric) score distribution by activity type and curriculum condition

Type	Curriculum condition	Score = 0	Score = 1	Score = 2	Score = 3
Everyday	World History Project (N = 79)	86%	14%	0%	0%
	Comparison (N = 100)	92%	8%	0%	0%
Summative	World History Project (N = 40)	95%	5%	0%	0%
	Comparison (N = 48)	98%	2%	0%	0%

Note: We considered that the activity scored a 0, 1, 2, and 3 on a rubric if the average score across raters was between 0–0.66, 0.67–1.66, 1.67–2.66, and 2.67–3, respectively.

Hierarchical linear model result details for rubric score comparisons

Table B14

Learning opportunity (activity rubric scores) for everyday activities, by group

Rubric scores for everyday activities	World History Project (N = 79)		Comparison (N=100)		HLM results			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	β	<i>SE</i>	<i>p</i>	<i>g</i>
Historical argumentation	.84	.79	.73	.78	.10	.10	.298	.13
Historical causation	.33	.54	.21	.49	.11	.10	.294	.21
Historical comparison	.29	.55	.19	.41	.08	.09	.342	.17
Historical contextualization	.30	.50	.26	.42	.03	.07	.617	.07
Change and continuity over time in history	.24	.58	.07	.28	.18	.06	.002	.41
Sourcing	.35	.49	.27	.39	.09	.08	.250	.21
<i>Highest rubric score</i>	<i>1.18</i>	<i>.79</i>	<i>.97</i>	<i>.91</i>	<i>.22</i>	<i>.11</i>	<i>.055</i>	<i>.26</i>

Note. The 2-level HLM results compare the mean differences using the comparison activity as the reference group, and controlling for score clustering within teachers, years of teaching experience, percent free/reduced price lunch at the school, and whether the teacher was randomly assigned to their teaching condition.

Table B15*Learning opportunity (activity rubric scores) for summative activities, by group*

Rubric scores for summative activities	World History Project (N = 40)		Comparison (N=48)		HLM results			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	β	<i>SE</i>	<i>p</i>	<i>g</i>
Historical argumentation	1.53	.79	1.12	.78	.52	.28	.066	.66
Historical causation	.55	.54	.47	.49	.07	.16	.682	.14
Historical comparison	.43	.55	.22	.41	.23	.14	.097	.48
Historical contextualization	.60	.50	.42	.42	.15	.14	.298	.33
Change and continuity over time in history	.43	.58	.13	.28	.31	.14	.026	.70
Sourcing	.20	.49	.20	.39	.00	.09	.968	.00
<i>Highest rubric score</i>	<i>1.77</i>	<i>.78</i>	<i>1.29</i>	<i>.91</i>	<i>.56</i>	<i>.24</i>	<i>.018</i>	<i>.66</i>

Note. The 2-level HLM results compare the mean differences using the comparison activity as the reference group, and controlling for score clustering within teachers, years of teaching experience, percent free/reduced price lunch at the school, and whether the teacher was randomly assigned to their teaching condition.

Table B16*Student outcomes (student work rubric scores) for everyday activities, by group*

Student work rubric scores for everyday activities	World History Project (N = 79)		Comparison (N=100)		HLM results			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	β	<i>SE</i>	<i>p</i>	<i>g</i>
Historical argumentation	.51	.43	.47	.49	.05	.08	.549	.11
Historical causation	.24	.38	.15	.30	.08	.07	.271	.24
Historical comparison	.19	.37	.17	.31	.02	.07	.764	.06
Historical contextualization	.23	.31	.20	.27	.02	.05	.697	.07
Change and continuity over time in history	.11	.25	.04	.14	.07	.03	.032	.36
Sourcing	.22	.34	.17	.31	.05	.05	.358	.15
<i>Highest rubric score</i>	.79	.40	.68	.48	.13	.09	.168	.29

Note. The 2-level HLM results compare the mean differences using the comparison activity as the reference group, and controlling for score clustering within teachers, years of teaching experience, percent free/reduced price lunch at the school, and whether the teacher was randomly assigned to their teaching condition.

Table B17

Student outcomes (student work rubric scores) for summative activities, by group

Student work rubric scores for summative activities	World History Project (N = 40)		Comparison (N=48)		HLM results			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	β	<i>SE</i>	<i>p</i>	<i>g</i>
Historical argumentation	.94	.64	.73	.69	.26	.19	.187	.39
Historical causation	.43	.42	.35	.48	.08	.13	.535	.18
Historical comparison	.23	.38	.15	.29	.10	.08	.234	.30
Historical contextualization	.29	.34	.29	.34	-.01	.08	.865	-.03
Change and continuity over time in history	.21	.31	.06	.19	.15	.06	.014	.60
Sourcing	.10	.23	.12	.20	-.03	.06	.653	-.14
<i>Highest rubric score</i>	<i>1.12</i>	<i>.49</i>	<i>.92</i>	<i>.82</i>	<i>.32</i>	<i>.17</i>	<i>.056</i>	<i>.36</i>

Note. The 2-level HLM results compare the mean differences using the comparison activity as the reference group, and controlling for score clustering within teachers, years of teaching experience, percent free/reduced price lunch at the school, and whether the teacher was randomly assigned to their teaching condition.