Epistemic beliefs about the value of integrating information across multiple documents in history

Jennifer Wiley\(^a,\ast\), Thomas D. Griffin\(^b\), Brent Steffens\(^b\), M. Anne Britt\(^b\)

\(^a\) University of Illinois at Chicago, USA
\(^b\) Northern Illinois University, USA

**ARTICLE INFO**

**Keywords:**
- Multiple-document comprehension
- Epistemic beliefs
- Epistemology
- Task model
- Reader goals

**ABSTRACT**

Individual differences in epistemic dispositions may affect learning from multiple-document inquiry tasks by prompting different students to have different task and activity models. Students with epistemic beliefs that are more appropriate for the required activities may view a multiple-document inquiry task as an exercise in corroboration, seeking coherence, and looking for evidence to support claims, whereas students with less-appropriate epistemic beliefs may see the goal as simply finding the “right” answer verbatim within the documents. This paper describes attempts to develop an assessment for this subset of epistemic beliefs about the value of engaging in integration of evidence when learning from multiple documents in history. Across three experiments, the measure was shown to be reliable and valid. It also uniquely predicted multiple-document comprehension in history above and beyond beliefs about the simplicity and certainty of knowledge, and accounted for differences in prior instruction and experience with document-based questions.

1. Introduction

A number of studies have demonstrated that tasks which prompt students to engage in constructive activities, such as generating an argument or explanation from multiple documents, or considering intertextual connections across a set of texts, can foster deeper comprehension than tasks that prompt students to simply describe, recall, or summarize what they have read (Britt & Sommer, 2004; Cerdán & Vidal-Abarca, 2008; Wiley & Voss, 1999). However, some research suggests that such tasks may be less effective when students are epistemically naïve. Griffin, Wiley, Britt, and Salas (2012) found that both individual differences in thinking dispositions and reading skill contributed to middle school students’ ability to learn from multiple documents when prompted to write an essay “explaining how and why recent patterns of global temperature are different from what has been observed in the past.” Students who showed a stronger disposition toward CLEAR thinking (Commitment to Logic, Evidence, and Reasoning) demonstrated better learning outcomes. These results are consistent with the theoretical suggestion that individual differences in epistemic dispositions may affect learning from multiple-document inquiry tasks by prompting different students to have different task models (Brüten, Britt, Strømsø, & Rouet, 2011; Britt & Rouet, 2012) and activity models (Wiley, Jaeger, & Griffin, 2018). As described below, the task model that students hold is what determines their approach to multiple-document comprehension. The activity model determines the kinds of behaviors that a student engages in to achieve their goals. Students with epistemic beliefs that are more appropriate for the activities required during a multiple-document inquiry task will have an activity model that includes the need to integrate information across sources and reason about the relevance of the information, whereas students with less activity-appropriate beliefs may see the goal of an inquiry task as finding the document that provides the “right” answer to the question. This paper describes attempts to develop an assessment for this subset of epistemic beliefs about the value of engaging in integration of information when constructing historical explanations in a multiple-document context.

1.1. The role of the task model in multiple-document comprehension

In multiple-document inquiry tasks, readers are given stand-alone documents (or excerpts) and task instructions (e.g., write an essay “explaining how and why recent patterns of global temperature are different from what has been observed in the past”). To create a representation of the content and to complete the task, the reader has to regulate their reading by deciding whether to read a document, what order to read the documents in, when to reread, and how to read. In a multiple-document context, the separate documents will generally not provide coherence cues that could help the reader to make inferences.
across documents and create an integrated representation of the content. Even more challenging, task instructions can be ambiguous and are rarely detailed enough to lead readers to know what the end goal should be and how to read to achieve their goals.

According to RESOLV model (Britt, Rouet, & Durik, 2018; Rouet, Britt, & Durik, 2017) which represents an updated version of the Documents Model Framework (Britt & Rouet, 2012; Rouet & Britt, 2011), task instructions are interpreted by the reader to create a task model which guides the reader's processing, decisions, and actions. The task model is a set of reader-generated goals including a representation of the end goal (e.g., “My explanation should include multiple causal factors”) and a set of activities (sometimes referred to as an activity model, Wiley et al., 2018) that can be used for achieving those goals (e.g., “I need to think about how everything in each document might be related to X”). These goals and actions are initially derived from cues in the context (e.g., who is giving the task, what resources are provided) and the knowledge structures activated from those cues (e.g., the structure of an explanation schema). The task model is highly dynamic, and the evolving set of hierarchical goals are updated and revised based on sub-goals being satisfied or the presentation of obstacles or impasses during reading. The reader's task model becomes especially important in multiple-document inquiry tasks when there is a mismatch between the original purposes for which each of the individual texts were written, and the assigned task that requires repurposing the information to answer a new question (see Britt et al., 2018). In that context, it is important for the sub-goals associated with the activity model to include establishing logical or causal connections and coherence across document boundaries.

Readers’ task models will be partially determined by their interpretation of cues in the context (e.g., instructions, hints, resources provided). The extent to which readers activate an appropriate set of goals and plans, and the extent to which they choose to engage in appropriate activities to achieve those goals and plans, will partially determine their success at multiple-document comprehension. In particular, the reader has to appreciate that a key set of activities required for multiple-document inquiry involve integrating information (Barzilai & Zohar, 2012; Cerdán & Vidal-Abarca, 2008; Bukavina & Daneman, 1996; Wiley et al., 2018; Wiley & Voss, 1999). Epistemic beliefs that might inform this particular activity model component of the broader task model construct provides the focus for the current work.

### 1.2. Epistemic beliefs and perceptions of tasks

Given the important role of the task model in determining how readers engage in multiple-document comprehension, researchers have been interested in the beliefs and perceptions that students hold in relation to their assigned tasks. One primary context in which skills of multiple-document inquiry are needed is in the context of answering document-based questions (DBQs) in history, which are intended to reflect some of the processes of historical inquiry and disciplinary practices of historians (Rouet, Britt, Mason, & Perfetti, 1996; Wiley & Voss, 1996; Wineburg, 1991). In an early attempt to assess students’ perceptions of what it means to “take the role of a historian” while engaging in multiple-document inquiry tasks, Voss and colleagues surveyed student perceptions (conceptions and misconceptions) about history and historical explanation. They found that students viewed historical accounts and history as being certain, preferred single-cause and simple explanations, and did not appreciate complexity, subjectivity, or interpretation (Voss, Wiley, & Carretero, 1995; Voss, Wiley, & Kennet, 1998). These perceptions may directly relate to the task models and activity models that students may adopt when they engage in answering multiple-document inquiry questions within the domain of history.

Measures of personal epistemology that assess various beliefs about the nature of knowledge and the process of knowing are also likely to shape the task and activity models a person forms when presented with learning tasks involving multiple-document inquiry. Bråten et al. (2011) discussed how science-related epistemological beliefs could be integrated with the Documents Model Framework to explain learning from multiple texts (Britt & Rouet, 2012; Rouet & Britt, 2011). According to Bråten et al. (2011), the less one holds a simplistic belief about knowledge, the more they will define the goal state of a multiple-source task as one of creating an integrated, coherent representation by means of corroboration. Also the extent to which one holds a belief in the certainty of knowledge, the more they will rely on a single text to find the answer. Thus, beliefs about knowledge and knowing will affect how a person interprets the goal state for a task, as well as the standards, criteria and appropriate sub-goals for achieving that goal state. The RESOLV model (Britt et al., 2018; Rouet, Britt, & Durik, 2017) also added a role for context in interpreting and valuing the goals so that the effort needed would or would not pass a benefit-cost analysis. This addition is important because integration can be effortful and challenging. Thus, according to RESOLV, when initially reading the task instructions, cues from the context can differentially impact which epistemic beliefs get activated. For example, the same task presented by a history teacher will be interpreted with a different epistemic lens than if the task were presented by a literature teacher or peer. Also, readers’ general beliefs about the type of task will affect their interpretation of the desired goal state and specific criteria for what constitutes a “good answer” to a particular inquiry question. Readers can interpret “explaning how and why” as an answer-seeking task or as an explanation-construction task, which will in turn determine the types of activities and strategies they employ (e.g., “search the text for the answer” versus “examine the text for information that might help explain X”). In other words, task models inform activity models, and both could be shaped by readers’ epistemic beliefs regarding the nature of knowledge and the process of knowing. The extent to which readers engage in intertextual integration will partly depend on readers’ beliefs about the nature of knowing (how knowledge is achieved). Readers who assume that knowledge is simple and certain are unlikely to consider multiple causes in their explanations. On the other hand, beliefs that integration is important when constructing explanations of historical events should make the reader more likely to consider information from multiple documents rather than just seek out the document with the best answer.

This application of epistemic beliefs toward the construction of the task and activity models that guide multiple-document inquiry corresponds to discussions of how epistemic beliefs relate to self-regulated learning more generally, as expressed in the COPES (conditions, operations, products, evaluation, standards) model (Winne & Hadwin, 1998). Conditions include the objective constraints of the task, as well as the learners’ own epistemic beliefs about the task goals and relevant strategies or activities needed to accomplish the goals. These in turn determine the specific operations that are executed, which then result in products (e.g., knowledge representations) which are evaluated against standards to determine whether the operations need to be modified. More recent adaptations of the COPES model have emphasized the critical role that epistemic beliefs play in creating the standards that drive metacognitive processes involved in the initial selection of, and adaptive revisions to, the operations and activities that learners enact. In particular, Bromme, Pieschl, and Stahl (2010, p. 15) argued that “sophisticated” epistemological beliefs allow for better or more adequate apprehension of the content and thereby also more appropriate standards for goal setting.

### 1.3. Effects of epistemic beliefs and student learning in multiple-document contexts

These theoretical frameworks that explicate how epistemic beliefs can impact learning from text have empirical support. Bromme et al. (2010) found that epistemic beliefs predicted the degree to which students matched the type of learning strategy they endorsed for various learning tasks that differed in complexity. Epistemic beliefs favoring the
oversimplification of complex and ill-structured knowledge can contribute to weak conceptual understanding, a failure to integrate new knowledge with prior knowledge, and poor comprehension (Coullson, Feltonvich, & Spiro, 1989; Schommer, 1990). Jacobson and Spiro (1995) extended these findings into a multiple-document context and found that only readers who rejected simplistic epistemic beliefs and endorsed integration of information benefitted from completing a multiple-document activity on the Impact of Technology on Culture presented in a hypertext environment. Similarly, Rukavina and Daneman (1996) found that readers with more “mature” epistemic beliefs about knowledge (who rejected that one should seek simple answers and avoid integration) performed better on test items that required integrating ideas across two separate texts representing competing theories about dinosaur extinction. Further, Bråten and his colleagues have shown that individual differences on their topic-specific epistemological belief questionnaire (TSEQ, Bråten & Strømso, 2009) related to simplicity and certainty have consequences for student learning from different multiple-document inquiry activities. For example, while argument or explanation-based prompts have led to better learning outcomes than prompts that encourage students to summarize the information provided in a set of documents (Hemmerich & Wiley, 2002; Le Bigot & Rouet, 2007; Naumann, Wechsung, & Krems, 2009; Wiley, 2001; Wiley et al., 2009; Wiley & Voss, 1996, 1999), this effect seems to interact with measures of topic-specific epistemology. Bråten and Strømso (2009) found that undergraduates who considered knowledge about climate change to be certain and unchanging experienced worse learning outcomes from an argument-writing task using multiple documents. Similarly, Gil, Bråten, Vidal-Abarca, and Strømsø (2010) found that undergraduates who believed knowledge about climate change to be simple and certain experienced worse learning outcomes from an opinion-writing prompt than a summarization prompt. Because epistemic beliefs related to the simplicity and certainty of knowledge have been found to have reliable effects on multiple-document comprehension in prior studies, they are included as part of the scales used in the current line of research. In addition, the novel goal for the present line of work was to validate a subscale of epistemic beliefs that focused on the value of integration of information across multiple documents as a key activity that readers need to engage in while attempting to learn from multiple-document inquiry tasks.

1.4. Goals of the current research

The main goal for the present line of work was to develop and validate a subscale of epistemic beliefs that targeted the value of integration of information across multiple documents and the key activities that readers need to engage in while attempting to learn from multiple documents in history. Since the epistemic belief scales were devised for their relevance to multiple-document inquiry and the activity models that are most appropriate for what is expected of students on such tasks, the term activity-appropriate beliefs will be used to refer to the direction of responses that are assumed to be optimal in this context. This label is in contrast to common terms in the literature such as “sophisticated beliefs” which ignore context-dependence and carry unnecessary additional assumptions.

In a first experiment, the reliability of the two subscales (Simplicity/ Certainty, and Integration) is established. To provide construct validity, all three experiments compare the epistemic beliefs of students who either took or are currently taking Advanced Placement (AP) History courses against students who have not. Advanced Placement refers to a specific college-level curriculum taken by a subset of United States high school students, developed by domain experts and college educators recruited by the College Board. The curriculum prepares students for a rigorous exam which includes a document-based question (DBQ) in which students are given 1 h to write an essay that answers a specific inquiry question using the information in a set of approximately 7 documents. The College Board provides teachers with a guidebook for using DBQs in the classroom to “help [students] build critical thinking skills for the rest of their lives” (Spoehr & Fraker, 1993, p. 3). The inquiry questions often focus on discussing causal factors (including the broader context and conditions) that brought about some event, and the documents consist mostly of primary sources that were not written to directly address the inquiry question, requiring students to reason about how the contents relate to the question and can be integrated to construct an explanation. These characteristics correspond closely to the multiple-document inquiry task used in Experiments 2 and 3 to provide an additional test of validity for the subscales, first in college students (Experiment 2), and then in middle school and high school students (Experiment 3). If the epistemic beliefs captured by these two subscales are useful for engaging in these types of learning activities, then one would expect that students who have engaged in this type of learning in their AP History courses should not only perform better on such tasks, but should also have epistemic beliefs that are more suited to these tasks. In turn, these beliefs should partially account for their better performance on multiple-document inquiry tasks. Thus, this series of studies used experience in AP History, and performance on a multiple-document inquiry task, to test for the construct validity of the two epistemic subscales.

2. Experiment 1

Several different surveys and instruments have been developed by a number of researchers to tap a range of epistemic beliefs about the nature of knowledge and knowing. Some measures explore beliefs about the nature of knowledge and the nature of knowing at a domain-general level (Mason & Boscolo, 2004; Rukavina & Daneman, 1996; Schommer, 1990; Spiro, Feltonvich, & Coulson, 1996; Weinstock & Cronin, 2003). Others have used more domain-specific measures with items that focus on a domain such as science, math, genetics, medicine (Bråten, Ferguson, Strømsø, & Anmarkrud, 2013; Conley, Pintrich, Vekiri, & Harrison, 2004; Elder, 2002; Greene, Azevedo, & Torney-Purta, 2008; Hofer, 2004; Kienhues, Stadler, & Bromme, 2011; Mason, Boldrin, & Ariasi, 2010) or in the domain of history (Buehl, Alexander, & Murphy, 2002; Maggioni, Vansledright, & Alexander, 2009; Stoel, van Drie, & Van Boxtel, 2017). Relationships have also been found between learning and topic-specific epistemic beliefs on a variety of science topics including climate change (Bråten & Strømsø, 2009; Bråten, Strømsø, & Samuelstuen, 2008; Gil et al., 2010; Muis et al., 2015; Trevor, Muis, Pekrun, Sinatra, & Muijselaar, 2017), desalination of seawater in Israel (Barzilai & Eshet-Alkalai, 2015), health risks of chocolate consumption and negative impacts of Red Sea fish farming (Barzilai & Zohar, 2012), and nutrition and health (Barzilai & Ka’adan, 2017).

There is also variety in the kinds of measures that have been used to assess epistemological beliefs. Developmental theorists such as Kuhn and Weinstock (2002) have used prompts to elicit responses which are classified along a stage-like progression from absolutist to multiplist to evaluativist. One measure using a similar method (Beliefs about Learning and Teaching History Questionnaire; Maggioni et al., 2009) was originally developed to address how history is taught and learned (rather than the nature of historical knowledge). This measure includes a set of statements that are endorsed on a scale of agreement corresponding to three progressive stages: copier (i.e., realistic and absolutist level), borrower (i.e., multiplist level), and criterialist (i.e., evaluativist level). Stoel, van Drie, & Van Boxtel, 2017 adapted this measure to assess whether changes can be seen in students’ subjectivist and criterialist beliefs following explicit instruction about epistemological ideas. Others have used open-ended methods such as semi-structured interviews (e.g., King & Kitchener, 2004; Muis et al., 2015), think-aloud protocols (e.g., Barzilai & Zohar, 2012; Bråten, Ferguson, Strømsø, & Anmarkrud, 2014; Ferguson, Bråten, & Strømsø, 2012; Mason, Ariasi, & Boldrin, 2011), and retrospective interviews (Mason et al., 2010; Pluta, Chinn, & Duncan, 2011) to classify student epistemic beliefs.
In contrast to these open-ended and stage-based approaches are measures in which students rate their endorsement with statements that capture different dimensions of epistemic beliefs (e.g., Hofer & Pintrich, 1997; Schommer, 1990). The measures developed for the present research adopted this approach, following prior work on perceptions of history and historical explanation (Voss et al., 1995, 1998), as well as studies using epistemology scales that have demonstrated relations among simplicity and certainty beliefs and learning from multiple texts in science (Bråten et al., 2008; Bråten & Strømsø, 2009). In the measures developed for present line of work, past items were selected and adapted to focus on evaluating beliefs about the actions, behaviors, or activities that one might engage in when attempting to develop an accurate explanation from multiple documents in history. In other words, the items developed for the new Integration subscale were designed to tap the reader’s activity model. What was emphasized in developing items for this subscale was the value of reasoning about how various pieces of information are related, making connections, and integrating ideas across documents to construct an explanation. These activities are highly relevant for constructing complex multi-causal explanations of historical events from multiple documents, and reflect activities historians engage in during their inquiries (Wineburg, 1991) as well as activities that students are encouraged to engage in to support historical understanding (Lee & Ashby, 2000; Lee, Dickinson, & Ashby, 2001). Items from prior Simplicity and Certainty subscales were also adapted to be more about the simplicity and certainty of explanations and causes, rather than about facts, claims, or knowledge. Further, instead of proposing a specific topic as a point of reference for the responses, the survey administered in these studies used a more general, domain-level context of learning from multiple documents in history. The hypotheses tested in this study were:

H1. Simplicity/Certainty items and Integration items would form two distinct and reliable subscales.

H2. Subscale scores were expected to differ between students who had taken AP History and those who had not, with AP History students having more activity-appropriate epistemic beliefs.

2.1. Method

2.1.1. Participants

A sample of 553 first-year college students at a large public university in the United States completed the scale for course credit in Introduction to Psychology. The sample self-reported as 62% female and 38% male. Ages ranged from 17 to 26 with 85% between the ages of 18 and 20. Students identified as 38% Asian/Indian, 10% Black/African American, 28% Hispanic/Latinx, and 32% White/Caucasian. (Students could select as many descriptors as they wished.) The scale was administered as part of mass testing at the beginning of the semester. A small sample of participants (n = 27) also completed the same survey again later in the semester to provide test-retest reliability.

2.1.2. Materials and procedure

The epistemic beliefs scale was created by modifying and adapting several items from the “Perceptions of History” scale used by Voss et al. (1998), and combining them with items adapted from the climate change TSEBQ developed by Bråten and Strømsø (2009, original items 5, 11, 19, 20, 23, 30, 31, 34, 48). Several items that were selected overlapped across the two sources, such as whether explanations are stable or change over time. Six items were intended to capture whether students held naïve, activity-inappropriate beliefs that historical explanations are simple and certain (Simplicity/Certainty subscale). Another six items were intended to capture whether students held activity-appropriate beliefs about the value of engaging in integration activities, relying on multiple sources, and relying on evidence when constructing historical explanations (Integration subscale). The items are shown in Table 1.

Table 1.

Instead of the topic-specific “climate change” prompt used by Bråten and Strømsø (2009) to measure topic-specific epistemological beliefs in science, in these studies students were instructed to answer each item in relation to “knowledge about history and how one comes to understand the causes of historical events.” The items were also written to be more about historical explanations, rather than about scientific claims, facts or information, and less about “trust.” Several of these items were also similar to items used by Ferguson and Bråten (2013) who had adapted several of the Bråten and Strømsø (2009) items to target beliefs about corroboration in science. Again, the focus of the Ferguson and Bråten items was on when one should trust scientific claims or facts, whereas the focus in the Integration items created for this scale was on the activities one should engage in when constructing historical explanations. In contrast to other prior work (Bråten & Strømsø, 2011, MTSI, Multiple-Text Strategy Inventory), students were asked to think about what actions “should” be taken, rather than retrospectively reporting which strategies (such as comparing across multiple sources) they might have actually used.

In this study, all responses were made using a 1 to 6 scale with 1 meaning “Strongly Disagree” and 6 meaning “Strongly Agree.” Due to the desire to test the scale with younger populations, the original 10-point scale used in Bråten and Strømsø (2009) was condensed to limit the number of categories that needed to be considered (Conley, Elder, Hofer, & Pintrich, 2004; Elder, 2002; Schommer-Aikins, Mau, Brookhart, & Hutter, 2000). Importantly, the absence of a middle response alternative was reserved to avoid satisfying behaviors (Kosnik, 1999). Initial versions of the items were piloted on middle school, high school and college students, and further revised based on student responses and teacher feedback about confusing phrases or difficult vocabulary. After completing the epistemic belief items, participants were asked to complete a background survey in which they indicated their ACT composite scores and whether they had taken AP History.

2.2. Results

2.2.1. Principal Component Analysis for subscales

Following Bråten and Strømsø (2009), responses for all items were submitted to a Principal Component Analysis (PCA) using oblique rotation (because the two components were expected to be correlated). The PCA resulted in only 2 components with Eigenvalues above 2. When a two-component solution was forced, the solution explained 46.12% of the total sample variation. The item loadings are shown in Table 1. Both components included 6 items which loaded highly (> 0.48) on only that component, and low on the other component (< 0.18). The items loading on the two components corresponded to the two intended subscales for Simplicity/Certainty (Eigenvalue 3.17) and Integration (Eigenvalue 2.36).

The internal reliability estimate (Cronbach’s alpha) was 0.79 for items on the Simplicity/Certainty subscale and 0.70 for items on the Integration subscale. For the 27 students who completed all survey items at both time-points, the test-retest reliability was 0.71 for the Simplicity/Certainty subscale and 0.60 for the Integration subscale.

2.2.2. AP status and epistemic belief scores

In addition, the construct validity of the subscales was demonstrated by testing for differences between students who had taken AP History in high school and those who had not. As shown in Fig. 1, a 2 (AP History, non-AP History) X 2 (Simplicity/Certainty, Integration) mixed ANOVA revealed there was a significant interaction, F(1, 551) = 16.28, p < .001, ηp^2 = 0.03, between AP status and subscale.1 Students who

---

1 Although the main effect for AP status is not interpretable due to the scoring being related in opposite directions to activity-appropriate beliefs for the two
took AP History were more likely to reject the idea that explanations in history are simple and certain, and to endorse the need to integrate information, than students who had not taken AP History. Follow-up tests confirmed that differences were significant on both subscales (Simplicity/Certainty: $F(1, 551) = 11.50, p = .001, \eta^2_p = 0.02$; Integration: $F(1, 551) = 6.23, p = .01, \eta^2_p = 0.01$).

Of course, this result could have been merely due to a self-selection bias in which students with generally higher academic skills or motivation are more likely to take AP courses. Students who had taken AP History did have higher ACT composite scores ($M = 25.14, SD = 3.67$) than students who had not taken AP History ($M = 23.55, SD = 3.26$), $F(1, 551) = 1395.51, p < .001$. Students who had been trained in multiple-document inquiry in history and who had previously taken an AP History course reporting more activity-appropriate epistemic beliefs on both subscales compared to those who had not taken AP History, even after accounting for general ability level. Thus, both scales were sensitive to epistemic differences that one would expect between students who had been trained in multiple-document inquiry in history and those who had not.

### Table 1

Epistemic belief items and loadings from principal component analysis.

<table>
<thead>
<tr>
<th>Item</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The best explanations in history are those that stick just to the one major cause that most directly lead to the event.</td>
<td>.738</td>
<td>.074</td>
</tr>
<tr>
<td>2. Most historical events are due to a single cause.</td>
<td>.601</td>
<td>-.040</td>
</tr>
<tr>
<td>3. You can be certain that historical explanations are true.</td>
<td>.764</td>
<td>.101</td>
</tr>
<tr>
<td>4. Good explanations in history are always indisputable.</td>
<td>.751</td>
<td>.090</td>
</tr>
<tr>
<td>5. There is only one good historical explanation that can be written from a set of facts.</td>
<td>.658</td>
<td>-.117</td>
</tr>
<tr>
<td>6. Historical explanations should not change in light of new information.</td>
<td>.658</td>
<td>-.119</td>
</tr>
<tr>
<td>7. To understand the causes of historical events, you need to connect evidence using reasoning.</td>
<td>-.016</td>
<td>.744</td>
</tr>
<tr>
<td>8. History is best understood by acting like a detective and connecting all the known facts.</td>
<td>.948</td>
<td>.632</td>
</tr>
<tr>
<td>9. When you read about history, you should have the most trust in explanations based on evidence.</td>
<td>.173</td>
<td>.483</td>
</tr>
<tr>
<td>10. When you read about history, you have to check what other sources say.</td>
<td>-.134</td>
<td>.666</td>
</tr>
<tr>
<td>11. To find out whether what you read is accurate, you should compare multiple sources.</td>
<td>-.173</td>
<td>.628</td>
</tr>
<tr>
<td>12. When you read about something new in history, you have to think about other things you have learned about the topic.</td>
<td>-.001</td>
<td>.659</td>
</tr>
</tbody>
</table>

Note: Correlation between components: $r = −0.09$.

![Fig. 1. Average scores on Simplicity/Certainty and Integration subscales in Experiment 1. Error bars represent standard errors.](image)

3. Experiment 2

The results of Experiment 1 suggested that the 12-item survey was promising as a measure of students’ epistemic beliefs concerning the nature of historical knowledge as simple and certain, as well as the kinds of integration activities that are required to develop knowledge in history. The new goal for Experiment 2 was to test whether the two subscales would be seen to relate to performance on a DBQ on a historical topic. In this case, the students were given a multiple-document inquiry unit about the Scopes Trial and asked to write about the factors that brought about the Scopes Trial. The responses that were written based on this task were scored to provide measures of performance. The hypotheses tested in this study were:

**H1.** Epistemic beliefs were expected to differ between students who had taken AP History and those who had not, with AP History students having more activity-appropriate epistemic beliefs.

**H2.** Epistemic beliefs were expected to uniquely predict performance on a multiple-document inquiry task.

3.1. Method

3.1.1. Participants

A sample of 151 first-year college students at a large public university in the United States completed the study for course credit in Introduction to Psychology. Demographic information was not collected on this sample, but it was from the same population used for Experiment 1. A subsample of 56 indicated they had taken AP History in high school, while 95 did not.

3.1.2. Materials and procedure

Students were given a multiple-document inquiry unit about the Scopes Trial in which an American high school teacher was tried for illegally teaching human evolution in a state school. The instruction that students were given for this task was to “Write an essay about the factors that brought about the Scopes Trial and what contributed to it..."
occurring when (1925) and where (Dayton, TN) it did.” Once they were done with the task, all materials were collected and students completed a final booklet which included the 12 epistemic belief items used in Experiment 1. Participants also provided demographic information including whether they had taken AP History in high school. The session was scheduled for 1 h and students were allowed to read and write their essays at their own pace as long as they spent a minimum of a half-hour on the inquiry task and were done within the hour, which approximates the 1-h time limit on the DBQ for AP exams.

The inquiry task and documents used in this study were adapted from the Scopes Trial unit in Historical Thinking Matters (Rosenzweig & Wineburg, 2005), recognized with the 2008 James Harvey Robinson Prize for an Outstanding Teaching Aid by the American Historical Association. The secondary sources are written by professional historians with no indicators that they vary in reliability. The primary sources do vary in perspective, but these reflect the motivations and potential biases of the people who caused the historical events to unfold as they did. Thus, the bias present in any source is not an indication that the content is “unreliable” and should be discounted, but rather that it reflects the differences in perspectives that may be important to include when explaining the historical event. The unit was created to reflect what is typical of AP exam DBQs (Spoehr & Fraker, 1993), where a limited number of documents have already been prescreened and selected for the students to consider, with all documents designed to provide some relevant information or context that students should integrate into their essay. Therefore, the task did not include all activities of professional historians or students engaging in less constrained internet research tasks, such as deciding that a source is so unreliable in its factual claims that the contents of the document should be discounted.

The document set included 7 short text excerpts: the ACLU press release about the case; an excerpt of a trial speech by the defense; a New York Times article; an editorial from the Chicago Defender (Moran, 2002); an excerpt from Anti-intellectualism in American Life (Hofstadter, 1963); an excerpt from The Klan’s Fight for Americanism (Evans, 1926); an excerpt from The Most Sinister Movement in the United States (Straton, 1925); 2 non-text documents (a graph showing rural/urban population changes and a political cartoon); and one background excerpt from the chapter on the 1920s (The Roaring Twenties: American Life Changes and a Political Cartoon); and one background excerpt from the book The Most Sinister Movement in the United States (Straton, 1925).

Fig. 2 illustrates the causal factors available in the documents that responses were coded for, and some of the connections that could be drawn among them. Many factors were indirect, and mediated or moderated by other factors. For example, technological, economic, and other societal changes in the early 1900s heightened fears of change and threats to traditional values, especially among protestant fundamentalists in rural and southern regions. This occurred around the same time that laws began to require high school education and as the theory of evolution became part of the curriculum. This combination of events not only led to laws like the Butler Act forbidding the teaching of evolution in Tennessee, but also to attacks on immigrants and intellectuals that prompted a group of lawyers to form the ACLU. The ACLU was founded just a few years before it brought a legal challenge against the Butler Act. Meanwhile, industrialization and urbanization led to economic hardship in rural farming areas like Dayton, TN, which prompted local businessmen to encourage a teacher, John Scopes, to answer the ACLU ad in order to generate a media circus and tourism revenue.

This causal model reflects an idealized possible account that might be generated by a professional historian using these documents. It was not expected that such a model be generated by any one student, especially within the allotted time. The selection of a unit that provided students with the opportunity to notice and integrate numerous complex causal relations was intentional in order to avoid both ceiling effects and floor effects.

Students were presented with all documents on separate pieces of paper in a pocket folder. The folder also contained the inquiry prompt and writing paper. Students had access to all documents while writing their responses to the inquiry prompt.

3.1.3. Coding

One way that the quality of each written response was assessed was by coding for the coverage of the causal model of the event (CAUSES, the number of causal concepts mentioned). Two independent raters scored approximately half of the responses (n = 78) with good inter-rater reliability, ICC (2, 1) = 0.80. One rater scored the remaining responses. Fig. 2 served as the model against which responses were scored.

An additional measure derived from this initial coding counted only mentions of distal causal concepts including the broader context of societal changes and background conditions that enabled the focal event of the Scopes Trial (CONTEXT, i.e., the Scopes Trial was caused by a cultural conflict between religion and science, and economic changes that were happening at the time). This measure did not include the more immediate, direct, or proximal causes of the Scopes Trial “happening when and where it did” (i.e., Scopes was on trial because he broke the law, or because he answered the ACLU ad), which had high semantic overlap with the essay prompt itself. The non-underline concepts in Fig. 2 were the concepts that were coded as distal causes that provided broader context.

Both the overall number of causal concepts and the number of contextual factors in the response provided measures of implicit integration across documents. The documents and inquiry task were specifically designed so students would need to take concepts presented in the documents for varied rhetorical purposes and repurpose them to address the given inquiry prompt. In other words, mentioning multiple causal concepts taken from various documents in one’s response to the given prompt is an act of integration. This is particularly true for the CONTEXT measure, because the contextual factors came from documents that had no explicit connection to the essay prompt, since they made no mention of Scopes, the law he broke, the trial, or Dayton, TN. Because contextual factors are a subset of total causal concepts, the two measures are likely to be very highly correlated. Nevertheless, both are analyzed as each provides important information. The presence of distal factors in the responses reflects the student attempting contextualization of the event. At the same time, the total number of causal concepts shows the extent to which any effects are independent of which causes are designated as distal versus proximal. Neither of these measures required that students used explicit rhetorical markers of integration in their responses.

To complement these implicit measures of integration, another code was used to score for collective expressions. These were explicit rhetorical markers including phrases signaling the collection of ideas that were being brought together (“several reasons”, “many causes”, “both factors”) or seriation (“first”, “second”, “third”), COLLECT, ICC (2, 1) = 0.81. The use of such devices may be more dependent upon students’ composition and writing skills, which may or may not directly relate to the beliefs about the value of integrating information when constructing an explanation as reflected in the epistemic belief scales.

The responses written by students during this brief task were quite short (generally between 200 and 300 words) and did not represent polished written products. However, to provide overlap with prior research that has coded for the quality of students’ written reports generated as part of more extended instructional units with multiple sources in history, the responses were also scored for several other features. Inter-rater reliability on these new codes was established by having two independent raters score a subset of 60 essays. The
remainder of the responses (for both remaining experiments) were then scored for each of these codes by one of the raters. The new codes assessed whether or not the responses included: evidence of sourcing, SOURCE, ICC (2, 1) = 0.85 (citations or references to particular documents as providing information), quotations, QUOTE, ICC (2, 1) = 0.98 (phrases quoted from documents using quotation marks), or the presence of a basic argument structure by including a main claim or thesis statement about why the Scopes Trial happened when and where it did, CLAIM, ICC (2, 1) = 0.85. A final code was based on whether the response included explicit statements of corroboration or conflict, CORROB, ICC (2, 1) = 0.88 (“several documents stated”, “many sources agreed”, “others disagreed”). This measure was intended as a complement for the other measures of integration described above, particularly the contextualization code (CONTEXT). While both entail processing more information, corroboration is more about seeking out redundancies to establish reliability and trust, whereas contextualization is about seeking out new information and relevant factors to add to one’s causal understanding.

3.2. Results

The goals for analyses in Experiment 2 were, first, to test for the replication of the results found in Experiment 1, that epistemic beliefs as measured by these subscales would differ across AP and non-AP History students. The second goal for analyses was to test if scores on the subscales would be predictive of performance on a multiple-document inquiry task, and if so whether they contributed unique variance beyond AP status.

3.2.1. AP status and epistemic belief scores

As shown in Fig. 3, a 2 (AP History, non-AP History) X 2 (Simplicity/Certainty, Integration) mixed ANOVA largely replicated the results of Experiment 1. Importantly, there was a significant interaction between AP status and subscale, F(1, 149) = 9.62, p < .01, \( \eta_p^2 = 0.06 \). Students who had taken AP History had higher Integration scores but lower Simplicity/Certainty scores than non-AP students. Follow-up tests showed significant differences on the Integration subscale, F(1, 149) = 16.27, p < .001, \( \eta_p^2 = 0.10 \), but not the Simplicity/Certainty subscale, F < 1.

3.2.2. AP status and essay quality

As shown in Table 2, a one-way between-participants ANOVA revealed that students who had taken AP History included more causal concepts in their written responses than students who did not take AP History (\( \eta_p^2 = 0.07 \)). AP History students also included more of the broader contextual factors in their written responses (\( \eta_p^2 = 0.05 \)). None of the other response coding measures showed significant differences between the AP and non-AP History students.

Relations among the various coding measures are shown in Table 3. Responses that included more causal concepts and more contextual factors were more likely to include explicit statements of corroboration. At the same time, the correlation between corroboration and contextualization measures was only modest with most of the variance being unshared, supporting the distinction between these behaviors. Students who integrated more causal concepts into their responses also tended to use more collective expressions (although this correlation was not significant, p = .06). The presence of a causal claim, sourcing, and quoting were not related to the overall number of causal concepts or

---

**Fig. 2.** Causal model for the Scopes Trial unit. Arrows represent only a subset of potential connections that could be drawn between concepts. Underlined concepts are proximal causes. Non-underlined concepts represent background factors.

**Fig. 3.** Average scores on Simplicity/Certainty and Integration subscales in Experiment 2. Error bars represent standard errors.

Students who had taken AP History had higher Integration scores but lower Simplicity/Certainty scores than non-AP students. Follow-up tests showed significant differences on the Integration subscale, F(1, 149) = 16.27, p < .001, \( \eta_p^2 = 0.10 \), but not the Simplicity/Certainty subscale, F < 1.

3.2.2. AP status and essay quality

As shown in Table 2, a one-way between-participants ANOVA revealed that students who had taken AP History included more causal concepts in their written responses than students who did not take AP History (\( \eta_p^2 = 0.07 \)). AP History students also included more of the broader contextual factors in their written responses (\( \eta_p^2 = 0.05 \)). None of the other response coding measures showed significant differences between the AP and non-AP History students.

Relations among the various coding measures are shown in Table 3. Responses that included more causal concepts and more contextual factors were more likely to include explicit statements of corroboration. At the same time, the correlation between corroboration and contextualization measures was only modest with most of the variance being unshared, supporting the distinction between these behaviors. Students who integrated more causal concepts into their responses also tended to use more collective expressions (although this correlation was not significant, p = .06). The presence of a causal claim, sourcing, and quoting were not related to the overall number of causal concepts or
contextual factors included in the response, however markers of corroboration were often associated with sourcing. Sourcing and quoting were related to each other. Rhetorical collective expressions were often used in responses with a main causal claim or thesis statement, which led to relations among these writing quality codes.

### 3.2.3. Relations between epistemic belief scores and essay quality

The relations between each of the coding measures and the scores on the two subscales are also shown in Table 3. Three measures of performance on the inquiry task related to scores on the epistemic belief subscales: number of causal concepts included in the responses (CAUSES), number of background factors included in the responses (CONTEXT), and the presence of statements of corroboration (CORROB). The other 4 coding measures did not relate to epistemic beliefs as measured by these subscales in this sample.

To test for unique relations between these epistemic beliefs and performance in the multiple-document inquiry task, both subscales were entered into a simultaneous regression as predictors for three performance measures that were related to beliefs in Table 3. For the number of causal concepts that students integrated into their responses, the regression model including both subscales provided a significant fit for the data, $R^2 = 0.16$, $F(2, 148) = 13.76$, $p < .001$. As shown in Fig. 4, both Simplicity/Certainty ($\beta = -0.27$, $t = -3.59$, $p < .001$) and Integration ($\beta = 0.28$, $t = 3.65$, $p < .001$) subscales contributed significant unique variance, but with opposite relationships to quality.

In addition, when AP status was added to the model, there was a significant change in $R^2 = 0.03$, $F(2, 148) = 5.71$, $p = .02$. Even though AP status was a significant predictor in this analysis ($\beta = -0.19$, $t = 2.39$, $p = .02$), both the Simplicity/Certainty ($\beta = -0.27$, $t = -3.56$, $p < .001$) and Integration ($\beta = 0.22$, $t = 2.77$, $p < .01$) subscales still contributed significant unique variance.

Similar results were seen when predicting the number of contextual factors that were included in essays. The model with just the subscales was significant, $R^2 = 0.11$, $F(2, 148) = 8.67$, $p < .001$. Adding AP status led to significant $R^2$ change = $0.02$, $F(3, 147) = 3.99$, $p = .05$. AP status was a significant unique predictor in this analysis ($\beta = -0.16$, $t = 2.00$, $p = .05$), as well as both the Simplicity/Certainty ($\beta = -0.23$, $t = -2.96$, $p < .01$) and the Integration ($\beta = 0.16$, $t = 2.01$, $p = .05$) subscales.

When a binary logistic regression was performed predicting the presence of statements of corroboration, the Integration subscale emerged as the only significant unique predictor (Integration subscale, Wald = 6.15, $p = .01$; Simplicity/Certainty subscale, Wald = 2.41, $p = .12$). AP History status was not a significant predictor (Wald = 0.12, $p = .73$).

### 3.3. Summary

These results replicate the finding that scores on the two subscales are sensitive to whether students have taken AP History. College stu-

---

**Table 2**

Average Quality Scores (with Standard Error) or Proportions for AP and non-AP History Students in Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>AP History</th>
<th>Non-AP History</th>
<th>Statistic</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAUSES</td>
<td>7.52 (.34)</td>
<td>6.06 (.26)</td>
<td>$F(1, 149) = 11.53$</td>
<td>.001</td>
</tr>
<tr>
<td>CONTEXT</td>
<td>5.39 (.36)</td>
<td>4.24 (.23)</td>
<td>$F(1, 149) = 7.88$</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Corroboration</td>
<td>0.66</td>
<td>0.56</td>
<td>$\chi^2 = 1.85$</td>
<td>.21</td>
</tr>
<tr>
<td>Sourcing</td>
<td>0.48</td>
<td>0.39</td>
<td>$\chi^2 = 1.24$</td>
<td>.27</td>
</tr>
<tr>
<td>Quotations</td>
<td>0.52</td>
<td>0.41</td>
<td>$\chi^2 = 1.64$</td>
<td>.20</td>
</tr>
<tr>
<td>Claim</td>
<td>0.50</td>
<td>0.54</td>
<td>$\chi^2 = 0.19$</td>
<td>.66</td>
</tr>
<tr>
<td>Collective</td>
<td>0.38</td>
<td>0.44</td>
<td>$\chi^2 = 0.65$</td>
<td>.42</td>
</tr>
</tbody>
</table>

**Table 3**

Correlations among quality and epistemic belief scores in Experiment 2.

<table>
<thead>
<tr>
<th>CAUSES</th>
<th>CONTEXT</th>
<th>CORROB</th>
<th>SOURCE</th>
<th>QUOTE</th>
<th>CLAIM</th>
<th>COLLECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplicity/Certainty</td>
<td>-.29**</td>
<td>-.24**</td>
<td>-.13</td>
<td>.04</td>
<td>-.05</td>
<td>.01</td>
</tr>
<tr>
<td>Integration</td>
<td>.29**</td>
<td>.23**</td>
<td>.23**</td>
<td>.04</td>
<td>-.06</td>
<td>.05</td>
</tr>
<tr>
<td>CAUSES</td>
<td>-.87**</td>
<td>-.</td>
<td>-.</td>
<td>.08</td>
<td>.05</td>
<td>.16*</td>
</tr>
<tr>
<td>CONTEXT</td>
<td>-.23**</td>
<td>.24**</td>
<td>-.</td>
<td>.15</td>
<td>.12</td>
<td>.09</td>
</tr>
<tr>
<td>CORROB</td>
<td>.02</td>
<td>.04</td>
<td>.13</td>
<td>.09</td>
<td>.04</td>
<td>.35**</td>
</tr>
<tr>
<td>SOURCE</td>
<td>-.03</td>
<td>-.01</td>
<td>-.07</td>
<td>.30*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUOTE</td>
<td>-.02</td>
<td>.04</td>
<td>.13</td>
<td>.09</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>CLAIM</td>
<td>.15</td>
<td>.12</td>
<td>.09</td>
<td>.01</td>
<td>-.06</td>
<td>.35**</td>
</tr>
</tbody>
</table>

Note: **p < .01, *p < .05, p = .06.
beliefs. Presenting a main thesis statement and expressions of collection may be more indicative of writing skill or past instruction in composition, while sourcing and quoting behaviors are characteristics of historical writing that may be more related to justification beliefs which were not assessed. It is possible that these features of student writing may sometimes correlate with epistemic beliefs about simplicity, certainty, and need for integration, but they did not directly relate to these subscales in this sample.

4. Experiment 3

The results of Experiment 2 showed that both the Simplicity/Certainty items and the Integration items predicted several measures of the quality of responses written by undergraduate students as part of a multiple-document inquiry task in history. This provides some evidence for the predictive validity of these scales. The new goal for Experiment 3 was to examine how high school students, including students who were currently enrolled in AP History courses, would respond to the same survey items, and to test whether the two subscales could be seen to uniquely relate to performance on a document-based inquiry question for these students as well. Further, based on prior work suggesting that epistemic beliefs about knowledge may be less developed in younger students (Conley et al., 2004; Elder, 2002; Schommer-Aikins et al., 2000) and work showing developmental progressions in beliefs about history and historical explanation (Lee & Ashby, 2000; Lee et al., 2001), it was important to test whether the subscales would predict performance on a multiple-document inquiry task in history among middle school students.

The hypotheses tested in this study were:

**H1.** Epistemic beliefs were expected to differ between middle school and high school students, with high school students having more activity-appropriate epistemic beliefs.

**H2.** Epistemic beliefs were expected to differ between high school students who were currently enrolled in AP History and those who were not, with AP History students having more activity-appropriate epistemic beliefs.

**H3.** Epistemic beliefs were expected to uniquely predict performance on a multiple-document inquiry task.

4.1. Method

4.1.1. Participants

A sample of 325 middle school students and 130 high school students (82 taking non-AP US History and 48 in AP US History) from public schools in Massachusetts and Illinois participated in the study as part of their history classes. Middle school students were aged between 11 and 14, and high school students were between 15 and 17. The sample self-reported as 45% male and 55% female, and identified as 9% Asian/Indian, 10% Black/African-American, 47% Hispanic/Latinx, and 37% White/Caucasian. (Students could select as many descriptors as they wished.)

4.1.2. Materials and procedure

Students completed the same historical inquiry unit on the Scopes Trial as in Experiment 2. The materials were again presented to students in a folder, however this study was completed over two 50 min periods, with approximately 75 min being allotted for the students to read the documents and write their responses. Although students were encouraged to read through the document set on Day 1 and to write their responses on Day 2, they could begin writing on Day 1 if they chose. Students had the documents available as they wrote their responses. Fifteen minutes before the end of the period on the second day, students were asked to finish their responses and the folders were collected. Students then completed a final booklet which included the same epistemic belief survey items as used in Experiments 1 and 2. Participants also provided demographic information including two questions about their prior experience with learning from multiple documents. The first question asked for a yes or no response to “Have you ever done this type of task to learn about history before?” Responses to the second question “How different was this type of task from the way you usually learn about history?” were on a 1 to 6 scale with 6 meaning very different. The number of causal concepts included in the responses were scored using the same criteria as in Experiment 2 by two independent raters. Two independent raters scored approximately half of the essays with good inter-rater reliability (Middle school, n = 183, ICC (2, 1) = 0.86; High school, n = 63, ICC (2, 1) = 0.81). One rater scored the remaining responses.

4.2. Results

The goals for analyses in Experiment 3 were, first, to test for the replication of the results found in Experiments 1 and 2, that epistemic beliefs as measured by these subscales would differ across student levels. The second goal for analyses was to test for the replication of the relations that were seen between epistemic beliefs and inquiry task performance in Experiment 2. Specifically, the analyses tested if scores on the subscales were predictive for middle school and high school students on the same measures of performance that were significant for
4.2.1. Student level and epistemic belief scores

As shown in Fig. 5, a 3 (AP History high school students, Non-AP History high school students, Middle school students) X 2 (Simplicity/Certainty, Integration) mixed ANOVA replicated the results of the prior experiments. There was a significant interaction between student level and subscale, with a significant difference in average scores on Simplicity/Certainty and Integration subscales between high school students and non-high school students. Post-hoc tests revealed that AP students had lower Simplicity/Certainty scores than non-AP students, who in turn had lower scores than middle school students. For Integration scores, AP students had higher scores than non-AP students, who in turn had higher scores than middle school students.

4.2.2. Student level and essay quality

As shown in Table 4, a one-way between-participants ANOVA revealed that the number of causal concepts included in written responses varied significantly by student level, $F(2, 452) = 3.37, p = .04, \eta^2 = 0.02$. Follow-up tests showed that AP History high school students included more causal concepts than the other two groups of students, who did not differ from each other.

The number of broader contextual factors included in the written responses also varied by student level, $F(2, 452) = 4.39, p = .01, \eta^2 = 0.02$. AP History students also included more of the broader contextual factors in their written responses than the other two groups of students, who did not differ from each other.

Student level also affected the likelihood of making a statement about corroboration ($\chi^2 (2) = 56.38, p < .001$), sourcing ($\chi^2 (2) = 37.68, p < .001$), quoting ($\chi^2 (2) = 14.64, p = .001$), stating a causal claim ($\chi^2 (2) = 104.75, p < .001$), and including collective expressions ($\chi^2 (2) = 112.25, p < .001$). For each coding measure, follow-up chi-squares were performed and group differences are indicated in Table 4.

Relations among the coding measures are shown in Table 5. Because of the larger sample size, and perhaps because the high school and middle school essays were written in the context of history classes with the discipline-related expectancies they imbue, significant positive relations were seen among all of the coding measures in this sample. However, what is most important for the goal of this study was testing whether the significant relations that were seen in the undergraduate sample in Experiment 2 (CAUSES, CONTEXT, CORROB) were replicated in Experiment 3.

4.2.3. Relations between epistemic belief scores and essay quality

As shown in Table 5, the three performance measures that significantly related to subscale scores in Experiment 2 were also related to subscale scores in Experiment 3. As a test for replication of the results found in Experiment 2, regression analyses simultaneously entered both subscales as predictors for each of these three measures, adding student level (middle school vs. high school) in a second step. Both subscales were unique predictors of number of causal concepts ($R^2 = .03, F(2, 452) = 6.60, p = .001$; Simplicity/Certainty $\beta = −0.11, t = −2.40, p = .02$; Integration $\beta = 0.14, t = 2.96, p < .01$, with no significant $R^2$ change for adding student level, $0.003, F(1, 451) = 1.30, p = .25$). Both subscales uniquely predicted number of contextual factors ($R^2 = .04, F(2, 452) = 10.31, p < .001$; Simplicity/Certainty $\beta = −0.14, t = −3.05, p < .01$; Integration $\beta = 0.17, t = 3.67, p < .001$, with no significant $R^2$ change for adding student level, $0.000, F(1, 451) = 0.00, p = .97$). Both subscales uniquely predicted the likelihood of including statements of corroboration (Simplicity/Certainty Wald = 13.79, $p < .001$; Integration Wald = 25.26, $p < .001$; Student level Wald 9.77, $p < .01$). These results replicate that both subscales uniquely predicted performance on these measures, over and above any effects of student level.

As can be seen in Table 6, these relations were stronger for high school students than for middle school students. Additional analyses tested whether each subscale predicted the measures of essay quality at each student level. For each measure of essay quality, two separate regressions were performed (one for each subscale), in which epistemic beliefs and a dummy-coded term for grade level (middle school versus high school) were entered, with the interaction term added in a second step. Scatterplots corresponding to the regressions for number of causal concepts included in the response are shown in Fig. 6. For Simplicity/Certainty beliefs, the initial regression model did not provide a significant fit for the data at $R^2 = .01, F(2, 452) = 2.20, p = .11$. Adding the interaction term resulted in a significant model, $R^2 = .03, F(3, 451) = 4.19, p < .01$. Student level was a significant predictor, $\beta = 0.43, t = 2.63, p = .01$. Simplicity/Certainty beliefs were not a significant predictor in the middle school referent group, $\beta = −0.00, t = −0.03, p = .98$. The interaction term was significant, $\beta = −0.43, t = −2.85, p < .01$, indicating that Simplicity/Certainty beliefs were a significant negative predictor among high school students.

For Integration beliefs, the initial model provided a significant fit, $R^2 = .02, F(2, 452) = 3.69, p = .03$. Adding the interaction term resulted in significant $R^2$ change = 0.01, $F(1, 451) = 6.30, p = .01$. Student level was a significant predictor, $\beta = −0.70, t = −2.44, p = .02$. Integration beliefs were not a significant predictor for the middle school referent group, $\beta = .07, t = 1.23, p = .22$. The interaction term was significant ($\beta = 0.73, t = 2.51, p = .01$), indicating that Integration beliefs were a positive predictor of the number of

![Fig. 5. Average scores on Simplicity/Certainty and Integration subscales by student level in Experiment 3. Error bars represent standard errors.](image)

concepts included in responses of high school students.

A parallel set of regressions were done predicting the number of contextual factors that were included in the essays. For Simplicity/Certainty beliefs, the initial regression model provided a significant fit for the data at $R^2 = 0.02$, $F(2, 452) = 4.01$, $p = .02$. Adding the interaction term resulted in significant $R^2$ change = 0.02, $F(1, 451) = 8.23$, $p < .01$. Student level was a significant predictor, $\beta = 0.52$, $t = 3.14$, $p < .01$. Simplicity/Certainty beliefs were not a significant predictor in the middle school referent group, $\beta = 0.07$, $t = 0.10$, $p = .92$. However, the interaction term was significant, $\beta = 0.45$, $t = -2.97$, $p < .01$, indicating that Simplicity/Certainty beliefs were a significant negative predictor among high school students. For Integration beliefs, the initial model including just beliefs and student level provided a significant fit for the data, $R^2 = 0.03$, $F(2, 452) = 6.65$, $p = .001$. The model including the interaction term was not a better fit, $R^2$ change = 0.01, $F(1, 451) = 3.39$, $p = .07$. Student level was not a significant unique predictor ($\beta = 0.07$, $t = 1.47$, $p = .14$). Integration beliefs were a significant predictor of the number of contextual factors for both high school and middle school students, $\beta = 0.14$, $t = 2.96$, $p < .01$.

A final set of analyses were done using binary logistic regressions predicting the presence of statements of corroboration. For the Simplicity/Certainty subscale, both student level (Wald = 22.39, $p < .01$) and beliefs (Wald = 9.40, $p < .01$) were significant unique predictors of corroboration statements. The interaction term was not significant when included in the model (Wald = 0.50, $p = .48$). For the Integration subscale, both student level (Wald = 35.42, $p < .001$) and beliefs (Wald = 21.61, $p < .001$) were significant unique predictors of corroboration statements. The interaction term was not significant when included in the model (Wald = 0.12, $p = .73$).

These results largely replicate the main findings of Experiment 2. Both subscales were unique predictors of several qualities of the written responses (CAUSES, CONTEXT, CORROB) when the sample was considered as a whole, and even when student level was included in the models. However, when the high school and middle school students were considered separately, the subscales were stronger predictors of performance for the high school students and were not predictive for middle school students in some cases.

Although there are indications that scores on the epistemic belief items were predicting some qualities of the written responses for middle school students (amount of contextualization and corroboration), the

| Table 5: Correlations among quality scores and epistemic belief scores in Experiment 3. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | CAUSES          | CONTEXT         | CORROB          | SOURCE          | QUOTE           | CLAIM           | COLLECT         |
| Simplicity/Certainty | -.10*          | -.12**          | -.28**          | -.16**          | -.11*           | -.35**          | -.34**          |
| Integration     | .13**           | .16**           | .27**           | .16**           | .13*            | .22**           | .19**           |
| CAUSES          |                | -.12**          |                | -.28**          |                |                |                |
| CONTEXT         | .87**           |                |                |                |                |                |                |
| CORROB          | .15**           | .27**           |                |                |                |                |                |
| SOURCE          | .13**           | .17**           | .29**           |                |                |                |                |
| QUOTE           | .11*            | .12*            | .18**           | .70**           |                |                |                |
| CLAIM           | .18**           | .25**           | .60**           | .30**           | .22**           |                |                |
| COLLECT         | .19**           | .24**           | .39**           | .25**           | .19**           | .55**           |                |

Note: **$p < .01$, *$p < .05$. |

| Table 6: Correlations among Epistemic Beliefs and Quality Scores for High School and Middle School students in Experiment 3. |
|-----------------|-----------------|-----------------|-----------------|
|                | CAUSES          | CONTEXT         | CORROB          |
| High School Simplicity/Certainty | -.27**          | -.27**          | -.23*           |
| Integration     | .27**           | .22*            | .22*            |
| Middle School Simplicity/Certainty | -.00            | .01             | -.10*           |
| Integration     | .07             | .11*            | .23**           |

Note: **$p < .01$, *$p < .05$, ∧$p = .06$. |

Fig. 6. Number of causal concepts as predicted by epistemic belief scores for middle school (solid lines and circles) and high school students (dotted lines and triangles) in Experiment 3.
weaker relationships in general could be due to these younger students having not yet formed epistemic beliefs on which to base their responses. Instead, they may have been responding idiosyncratically. This is consistent with middle school students’ similar middle-of-the-scale scores for activity-appropriate beliefs on the Integration scale and activity-inappropriate beliefs on the Simplicity/Certainty scale.

The lack of well-formed beliefs can lead to a response bias, such as acquiescence, where responders either tend to agree or disagree with all items regardless of the content (Knowles & Nathan, 1997; Krosnick, 1999). Acquiescence in responding would result in a positive correlation between subscales that should otherwise have either no correlation or show a modest negative correlation (due to the scales having opposing relationships to performing historical explanation tasks). Consistent with this hypothesis, the subscales showed a moderate positive correlation among middle school students (r = 0.41, p < .001), but a negative correlation among high school students (r = −0.28, p = .001). Thus, it seems likely that many middle school students may have lacked pre-formed epistemic beliefs on which to base their responses to the survey items, giving way to a stronger influence of an acquiescence response bias.

Middle school students were also less likely to report having ever done this type of task before (47.8%) compared to non-AP high school students (67.1%) and AP high school students (81.3%). χ²(2, 452) = 24.87, p < .001, and found the task to be more different than the tasks they usually did in history class (F(2, 447) = 20.10, p < .001, η² = 0.08). Follow-up tests revealed middle school students reported that the Scopes Trial unit was more different from their usual activities (M = 3.83, SD = 1.56) than did non-AP History high school students (M = 3.24, SD = 1.29), who in turn found the task more different than did AP History high school students (M = 2.44, SD = 1.49).

### 4.2.4. AP status and epistemic belief scores

A final set of analyses tested for a replication of the differences observed between AP and non-AP History students in Experiment 2. To examine whether AP status and each of the subscales accounted for unique variance in essay scores, a new set of regression models were run only for high school students, parallel to the analyses used in Experiment 2 with undergraduate students. When AP status (AP History or non-AP History) and both subscales were entered into a simultaneous regression predicting number of causal concepts integrated into responses, the model provided a significant fit for the data, R² = 0.11, F(2, 127) = 7.95, p < .001. Both subscales were significant unique predictors (Simplicity/Certainty: β = −0.21, t = −2.42, p = .02; Integration: β = 0.21, t = 2.35, p = .02). However, there was no additional unique variance explained by AP History status (R² change = 0.01, F(1, 126) = 0.68, p = .41).

Similar results were seen for AP Status and the prediction of contextual factors and corroboration. Adding AP Status in a second step of a regression did not lead to a significant change in the variance explained over an initial model including both epistemic belief subscales for contextual factors (R² change = 0.00, F(1, 126) = 0.21, p = .65). For corroboration statements, AP status did not account for significant variance when added to the binary logistic regression (Wald = 0.64, p = .42).

Thus, even though AP high school students included more causal concepts and more contextual features in their written responses, and were more likely to include a statement of corroboration than non-AP high school students, these differences were well accounted for by differences in epistemic beliefs.

### 4.3. Summary

The results of Experiment 3 were largely consistent with the findings from Experiments 1 and 2. The effect of being currently enrolled in AP History on epistemic beliefs was slightly larger in this younger (and presumably less homogeneous) sample than was the effect of having taken an AP History course in the two studies with college students, although the effect remained quite small in magnitude. Nevertheless, the differences due to AP status were significant, and high school students enrolled in AP History held more activity-appropriate epistemic beliefs and wrote higher quality responses than other high school students and middle school students. Moreover, the epistemic beliefs surveyed by the two subscales were independent predictors of performance for high school students. Further analysis of high school students’ performance revealed that the differences between AP History students and non-AP History students were not significant when epistemic beliefs were added as a predictor. This result means that once epistemic beliefs were taken into account, AP status no longer predicted performance on the multiple-document inquiry task on the Scopes Trial (i.e., non-AP students with activity-appropriate beliefs could do just as well as AP students). Therefore, differences in epistemic beliefs provide a plausible account of the performance differences between AP students and non-AP students.

Although this study did not have a measure of general academic skills that could be used as a control, Experiment 1 showed that such a measure (ACT composite) was only modestly related to Simplicity/Certainty and Integration scores among college students, and controlling for ACT did not account for the differences between AP and non-AP students in epistemic beliefs. Similarly, if the effects of AP status were due to differences in general academic skills in Experiment 3, controlling for epistemic beliefs would not have eliminated differences in performance between AP and non-AP students. In contrast, the opposite result was found. This suggests that the differences in performance are better explained by differences in epistemic beliefs. In addition, Experiment 3 showed that middle school students were less likely to report holding activity-appropriate epistemic beliefs, and variance in these measures did not predict performance among those students as well as it did for older students, possibly because younger students are responding with an acquiescence bias due to lacking explicit beliefs to guide their responses to the scale items.

### 5. General discussion

The aim of the current set of studies was to validate a new measure reflecting two dimensions of epistemic beliefs that are highly relevant to comprehension from multiple documents in history: the nature of historical explanations as simple and certain, as well as the value of engaging in integration as one attempts to develop explanations in history. Results of these studies provided evidence of internal reliability and test-retest reliability of the scales. In addition, the current work demonstrated the construct, convergent, and predictive validity of these scales by showing differences between AP and non-AP history students across all three experiments, and by showing that the scales uniquely predict several measures of performance on a multiple-document inquiry task about the Scopes Trial across Experiments 2 and 3 even when student level was taken into account.

The Scopes Trial unit was selected because it seemed representative of the kinds of DBQs that are used on the AP History exam, as well as of multiple-document inquiry tasks that may be used as part of history courses. The unit consisted of multiple documents of different types including primary and secondary sources. Importantly, the answer to the inquiry prompt was not contained in a single text, and many texts did not even mention the trial. Instead, each document discussed events, motives or conditions that could be either directly or indirectly related to the event unfolding in the time, place, and manner that it did. As such, completing this inquiry task required students to repurpose, integrate, and justify the inclusion of various pieces of information. Although there is no reason to suspect that the Scopes Trial unit would be the only multiple-document inquiry task that might show positive relations with these scales, it is a limitation that only a single learning activity was actually investigated in these studies. An important next step for research is to overcome this limitation by testing for the
relation of these epistemic belief items with learning from multiple-document inquiry tasks on different topics.

Another limitation is that scores on the epistemic belief items did not reliably predict performance among the middle school students. Additional analyses indicated that for this younger population, the scales were more likely capturing variance in response biases than meaningful variance in actual epistemic beliefs. One possible reason for this would be because the middle school students in this study had not yet developed beliefs about the nature of knowledge and explanation in history (Lee & Ashby, 2000; Lee et al., 2001). This interpretation is consistent with the observation that these students were also less likely to have reported any experience in learning from multiple documents. Furthermore, this lack of consistency in middle school student responses also aligns with research using scenario-based measures of epistemic beliefs (Barzilai & Zohar, 2012). More work is needed with students in this age range to understand if they lack epistemological theories, or if there might be some other explanation (i.e., the wording of these items was too difficult) that led to poor measurement of epistemic beliefs for younger students.

One concern about the use of self-reported belief measures is the extent to which they require that students have explicitly formed epistemic ideas. Some researchers have argued for a construct of epistemic cognition that includes not only such beliefs but also any of the processes whereby one is seeking to verify and justify their knowledge, or attempting to actually enact the activities described in the belief measures (Greene et al., 2008). It is possible that more direct measures of processing (or epistemic cognition-in-action, VanSledright & Maggioni, 2016) would capture more reliable variance in younger students’ implicit epistemologies that have not yet been formulated into declarative beliefs. The current studies did not include a direct measure of online epistemic cognition. However, the successful prediction of the inquiry product rests upon the assumption that the measured epistemic beliefs influenced epistemic cognition and enacted activities in a manner that then impacted the responses that students produced while performing the inquiry task. The nature of the multiple-document inquiry task, where the causal factors are presented in documents that do not directly address the inquiry question, was designed to require these type of activities (e.g., integration). Thus, the findings lend general support to the assumption that, at least among older students, these measured epistemic beliefs impact epistemic cognition in ways that impact learning within this context.

These epistemic scales and the inquiry tasks they were designed to correspond to reflect an analytic approach to constructing causal explanations for historical events. This is a core aspect of historical scholarship that is directly tied to the DBQs used on advanced exams in the United States that emphasize constructing cause/effect historical explanations (National Governors Association Center for Best Practices, & Council of Chief State School Officers, 2010). However, it is important to note that such explanation is not the only form of historical activity. The present scales may have less relevance for other tasks such as considering historical significance, engaging in historical empathy, or hermeneutic interpretation of historical philosophical or religious texts (VanSledright & Maggioni, 2016).

5.1. “Integration” and “justification” in multiple-document tasks for science and history

Prior work on epistemic beliefs and multiple-document inquiry has typically paired Simplicity/Certainty subscales with Justification subscales (Bråten & Strømsø, 2009; Ferguson et al., 2012; Hofer, 2000; Mason et al., 2010; Muis et al., 2015; Schommer, 1990). That work has mostly focused upon the task of learning about a science topic, typically in a context analogous to internet searches with secondary sources that vary starkly in expertise and directly contradict each other about factual claims. A central feature of such tasks is deciding which information to trust and include in one’s understanding of the phenomena versus which information can be disregarded as unscientific. For example, Bråten and Stremmo (2009) had students read several documents by authors of varying scientific credentials that presented directly conflicting arguments about the topic of climate change, such that the students’ primary task was to determine which conflicting explanation they should accept or reject. In fact, students were not given a specific inquiry question to think about, and instead were simply told to learn “about climate change.” In such a task, “justification” beliefs related to source authority and reliability would be central, and there is less need to integrate pieces of information across documents.

This is quite different from the kind of multiple-document inquiry tasks typically included on AP History exams and used in the present studies for which beliefs about integration would be more relevant. In these tasks, documents often have little direct conceptual overlap because they are written for various purposes. Most documents are primary sources written by and for the people involved with the event to be explained. Some documents precede the event, so they cannot make any mention of it, but rather provide vital contextual information about the human motives and conditions that are key causal factors that brought the event about. In fact, as with the current materials, some form of conflict in perspective between authors of multiple documents is often key for the historical event to be explained, rather than a signal that one or more of the documents is untrustworthy and thus its contents should be disregarded. For example, the first DBQ in a teachers’ guide produced by The College Board includes only primary documents that make no mention of the question that students are required to answer why the New England and Chesapeake settlements developed into two distinct societies by 1700 (Spoehr & Fraker, 1993). Although the included notes for teachers point out that one of the documents “contrasts sharply” (p. 14) with another, there is no actual contradiction of fact between those or any documents. Rather, the contrast is in the different attitudes and biased perspectives conveyed by the authors about their fellow settlers, and thus this contrast between sources is part of the explanation required by the DBQ rather than a conflict that signals one of the documents lacks credible information. For these reasons, the emphasis in these studies was on a subscale targeted at epistemic beliefs about the value of engaging in integration rather than justification.

In addition, the authors of primary documents in DBQs are rarely expert authorities on history itself, and these primary documents are not typically meant to be viewed as less reliable sources of relevant information than secondary sources that are written by historians. Rather primary and secondary sources provide different types of information (as opposed to more or less reliable information). Instead of being a signal about whether the document contents are useful, the source is essentially part of the conceptual content that students must figure out how to integrate into their response to the inquiry prompt. Therefore, while the activities of professional scientists and professional historians may have more in common, the typical multiple-document activities used in science and history instruction may tend to differ in ways that would moderate the role that different dimensions of epistemic beliefs play when students are constructing an explanation. Nevertheless, integration beliefs could also play a role in developing explanations from multiple-document science units, especially where some of the documents do not directly address a given inquiry prompt, but present logically related pieces of information. With slight modifications (e.g., replacing “historical events” with “scientific phenomena”) these current scales could be applied to the science domain.

5.2. Domain-specific beliefs instead of more general or more topic-specific beliefs

A recent meta-analysis by Greene, Cartiff, and Duke (2018) revealed
that while topic-specific scales and domain-general scales were predictive of academic achievement, neither was as strong of a predictor as domain-specific scales. In addition, the somewhat distinct nature of inquiry and forms of evidence-based argumentation in history versus science (especially the natural sciences) favors the use of domain-specific over general epistemology measures. In contrast, despite their predictive utility, topic-specific measures have questionable construct validity as measures of “epistemology” in the sense of theories of knowledge and knowing being distinct from actual knowledge. Suppose a person is aware of political factions that widely publish false pseudo-science about climate change and therefore that person endorses the importance of evaluating scientific credentials of sources, specifically on that topic more so than most others. Is that endorsement a measure of the person’s theories about knowledge and knowing (i.e., epistemology) or a measure of their prior knowledge of socio-political forces that shape discourse on that topic? Bromme, Kienhues, and Stahl (2008) pointed out a similar conceptual problem with topic-specific epistemology. Bråten and Stremos (2009) demonstrated that climate-change-specific epistemology is not entirely (although still partially) overlapping with having accurate prior knowledge of the current scientific theories of climate change. However, that does not demonstrate that these beliefs are “epistemology” rather than a reflection of prior knowledge about the political landscape that surrounds that particular topic, which would also be only partially overlapping with having a complete and accurate understanding of the current scientific theories.

In addition, there seems to be limited practical and pedagogical utility for topic-specific measures. A major goal of this area of research is to identify malleable factors that can be modified by instruction to improve student learning. One cannot expect that teachers will revisit and revise their instruction about the various dimensions of epistemological beliefs every time a new topic is introduced. Further, they could only do so in an effective way if researchers first identified which beliefs on every specific topic are the most relevant and in need of instruction. Based on these observations, domain-specific but topic-general measures seem to offer more promise for use as a pedagogical guide, as well as for formative and summative assessments of students’ epistemic beliefs.

5.3. Implications for individual differences in task models and activity models

These results are consistent with the theoretical suggestion that individual differences in epistemic dispositions may affect learning from multiple-document inquiry tasks by prompting different students to have different task models and activity models (Bråten et al., 2011; Britt & Rouet, 2012; Wiley et al., 2018). In addition to the subscales differing in the types of beliefs they reflect and their relationship to task-appropriate activities, they may also impact different stages of the learning process. One possibility is that Simplicity/Certainty beliefs impact learning via shaping learners’ goal state in their Task Models by setting the general parameters for what their resulting explanation of the phenomena should look like (e.g., should it include a single or multiple potential causes). In contrast, integration beliefs may more directly shape the Activity Model by indicating the types of behaviors needed to construct a good explanation (e.g., find and integrate relevant pieces of information from each document rather than just selecting the “best” document that contains “the right answer”). Although students’ task models and activity models are a plausible mediators of the observed effects of epistemic beliefs on learning outcomes, a direct test of this theory would require measuring these constructs. This would require measures reflecting the task and activity models that student actually held, which could be derived from online processing traces such as think-aloud protocols, cognitive interviews, or self-report responses about what students perceived the activity to be about and how they interpreted the scales. However, such measures were not collected in these studies. Even so, the results of the regression analyses showing the importance of these epistemic beliefs for performance suggest new directions for future interventions that directly seek to support or strengthen the learners’ task models or activity models as part of instruction.

5.4. Implications from relations between AP history experience and epistemic beliefs

5.4.1. Differences between AP history and non-AP history students as evidence for construct validity

Across all three studies, students who had taken or were taking AP History generally had more activity-appropriate scores on the epistemic belief items, supporting the construct validity of the measure. AP History courses in particular would be expected to expose students to historical practices of integrating information from multiple types of documents, increase their appreciation for the complex and uncertain nature of historical events, and thus prompt students to disavow beliefs favoring simplicity or certainty when engaging in historical explanation. Due to their experiences in AP History classes specifically with answering DBQs, AP History students would be expected to have both more activity-appropriate epistemic beliefs about historical explanations, and should also be better equipped to engage in document-based inquiry.

5.4.2. Differences between AP history and non-AP history students as showing promise for intervention

Another important implication of the relation that was seen between taking AP History and scores on these subscales is that these results are consistent with the malleability of the epistemic beliefs that promote better understanding from multiple-document inquiry. Although the difference between AP and non-AP students could have been due to self-selection tied to general academic skills, controlling for ACT would be expected to account for such differences. The fact that epistemic differences in Experiment 1 were independent of ACT gives some support to the proposition that experience with the multiple-document inquiry tasks that are common in AP History courses is what led these students to develop more activity-appropriate epistemic beliefs. Also, the fact that in Experiments 2 and 3 both belief scales uniquely predicted essay scores and accounted for differences between AP and non-AP students even when AP status was included in regression models, supports that the more appropriate beliefs of AP students are contributing to their better task performance. These results are consistent with other studies that have begun to show an influence of engaging in multiple-document inquiry on epistemic development (Bråten, Stremos, & Ferguson, 2016; Ferguson & Bråten, 2013; Kienhuis et al., 2011).

5.5. Concluding remarks

Multiple-document inquiry tasks entail critical reasoning and argumentation skills that reflect the kind of thinking done by historians themselves. These skills are critical for higher education in both history as well as in the STEM disciplines, and are important for academic success more generally. The fact that these dimensions of epistemic beliefs uniquely predict performance on multiple-document inquiry tasks and account for the benefits of having taken AP History courses suggests that they are important aspects of students’ approach to learning. Furthermore, the independence of the results from general skills or abilities suggests that these epistemic beliefs may be teachable and might be developed via experience with the kinds of multiple-document inquiry tasks that are typical in AP History courses. For example, the scales could be administered at the start of a semester and then again following an extended multiple-document inquiry task like the Scopes Trial unit. If the Scopes Trial unit were used during classroom instruction, teachers might consider simplifying the task by eliminating some of the documents and thus parts of the causal model.
Or, even better, they could spend more than 1 h on the task, and use discussions to help students to see the various connections, to understand the complexity of factors that contribute to historical events, and to appreciate the numerous ways that one can repurpose and interpret information from various documents into a coherent explanation. Students might also benefit from a discussion about how the items on the epistemic belief scales relate to the kinds of activities that are required when generating explanations as part of a multiple-document inquiry task. The goal would be to develop students’ epistemic beliefs and their appreciation for the need to integrate and repurpose information in various documents in order to construct an explanation, rather than to simply search for a correct answer intact within a single source. Integrating multiple-document inquiry tasks into instruction outside of AP History courses would help all students to benefit from such experiences, not just those who are given the opportunity or encouragement to enroll in AP coursework. Finally, while a number of studies have demonstrated that tasks which prompt students to engage in constructive activities, such as generating an argument or explanation from multiple documents, can foster deeper comprehension than tasks that prompt students to simply describe, recall, or summarize what they have read (Britt & Sommer, 2004; Cedérn & Vidal-Abarca, 2008; Wiley et al., 2018; Wiley & Voss, 1999), this research suggests that some epistemically naive students may require additional instruction to help them to adopt a more appropriate task model or activity model in order to benefit from these activities. These scales may be useful for identifying when that additional support may be needed.

Funding information

This research was supported in part by Grant R305F100007 from the Institute of Education Sciences, U.S. Department of Education. The authors thank Dylan Blaum, Sarah Davis, Tricia Guerrero, Morgan Hager, Jodi Hoard, Allison Jaeger, Tegan Michl, Carlos Salas, Andrew Taylor, Bianca Trigo, and James F. Voss for their contributions to this research as well as the other members of Project READi. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

Declaration of competing interest

None (other than grant funding).

References


comprehension. Journal of Educational Psychology, 82, 498-504. https://doi.org/10.1037/0222-6631.82.3.495.


