

How Well Do Structured Controversy Debates—and a Dash of “Fake History”—Promote Historical Thinking?

David L. Ellis and Mark A. Vincent
Augustana College

THE MOMENT many students begin to get excited about history is when they realize that much historical knowledge is contested, enmeshed in a productive argument that never ends and therefore endlessly fascinates. Such an epiphany might occur at any time. But a frequent catalyst is classroom debate, a formalized contest that dramatizes and vitalizes the challenge of constructing historical knowledge. A central problem for many instructors, therefore, is to design debates that enhance the engagement of all students, both those involved directly in debate and those who listen actively. This article describes how, with the help of Mark Vincent, a psychologist with expertise in cognition and statistical analysis, historian David Ellis turned student-led debates from the weakest to one of the strongest links in his courses through introducing both a structured controversy format and elements of liminal and ludic learning. Data collected over several years demonstrate broad gains in historical and critical thinking in all measured categories. All student groups saw gains, and in comparison to a more traditional debate format, the structured controversy format moved closer to parity in gains among men and women.

The Problem

The central problem was that a series of student-led class debates built into introductory, intermediate, and advanced undergraduate history courses taught by Ellis worked well for good students—as all pedagogies do—but produced mediocre results at best for other students. In other words, the debates weren't working as they were designed.

Why the debates were not working properly was not clear. The debates seemed to reflect some of the best practices in the Scholarship of Teaching and Learning (SoTL) course design. They fit into scaffolded assignments, requiring students to synthesize previously covered material and integrate new material. The debates were also, to use the terminology of Grant Wiggins and Jay McTighe, “backward designed,” with specific assignments interrelated and ordered so that they “could be logically inferred from the results sought.”¹ Debate topics and questions were structured to complement the overall course goals (especially students' acquisition of skills of critical and historical thought, and of oral and written expression) and the specific goals for a formal, written essay (written for the class after the debate). Furthermore, the debates entailed “significant learning,” for the debates not only had the abstract benefit of helping students think about the meaning of the past and how we shape that meaning, but also the more concrete benefit of dress rehearsal for a historical essay they would write on the same topic using the same sources used in the debate.²

To be more specific, the debates fit into an architecture common to many of Ellis' courses, including introductory history courses designed for all undergraduates, as well as intermediate and advanced courses designed mainly for history majors and minors. In each course, there were three main learning goals for students. First, students were expected to master a historical narrative that “covered” significant historical developments. Second, students were expected to “uncover” the way historians “do” history, learning skills that pertain to how historians evaluate evidence from primary sources and how they construct arguments. Like our Augustana College colleague, Lendol Calder, Ellis made this goal of “uncovering” the dual nature of historical writing—that historians seek to understand a past that does not change, and yet, in interpreting it, historians

always change our perceptions of that past and of the evidence that underpins those perceptions—into the main objective of the course.³ Finally, students were expected to develop familiarity with some important historiographical controversies, struggling with the ways in which historians have contested selected elements of the past.

These three learning goals for students corresponded to the organization of each unit of the course. At the outset of a unit on a given topic, students were provided with one or more guiding questions for the week. For instance, the overarching question that linked all of the readings and discussion of material on the French Revolution and the Napoleonic Era was, “What was the relationship of Napoleon to the French Revolution?” Students then encountered three different kinds of reading that helped them to address that question. Students first read and discussed a traditional textbook narrative connected to the topic. The two-fold aim was to give students a factually rich, narrative framework to a past with which most were unfamiliar, and to point out problematic aspects of the narrative approach. Students then read ten to twenty excerpts from primary sources (accounts from historical eyewitnesses), and through interaction with the instructor and other students began to construct their own answers to the guiding question for the week. Finally, students read usefully problematic (and often conflicting) answers historians had given to the unit’s guiding question (or some selected aspect of it) and participated in a student-led class debate, either as members of one of two opposing debate teams (with each team consisting of two to four students), or as audience members who were tasked with acting as a hostile jury, peppering the debate teams with questions or critical comments. Finally, students wrote a formal essay in which they addressed a question related to yet more sharply defined than the guiding question for the week.⁴

Using a trial-and-error approach, Ellis had tried several variations of the debate format, experimenting with varying degrees of student agency. For example, in one iteration, students on one debate team were assigned to identify and defend a historian’s interpretation of the specific primary sources they had worked with earlier in the unit, while an opposing debate team was assigned to identify and defend the conflicting interpretation of a different historian who had worked with the same kind of evidence. The intention here was to provide both teams with solid, clear, and plausible historical

theses, and for students to learn the contested and contingent nature of historical knowledge through the juxtaposition of theses and the use of evidence, along with discovering the value of alternate interpretations of the evidence.

The difficulties of this approach soon became evident. Most students had surprisingly great difficulty in discerning the actual thesis of “their” assigned historian, and many students seemed detached from the process because they had not understood or were not sympathetic to the thesis they were tasked to defend. In another variation, students on one debate team were asked fairly open-ended questions (such as “To what extent was race an important factor in causing New Imperialism?”) and invited to show why the evidence supported their own nuanced, sophisticated thesis, while students on the opposing debate team were tasked with the same charge, as long as their thesis was significantly different from the other team’s. The intention here was largely the same, but was predicated on the notion that allowing students more flexibility in determining their own theses would help them take ownership of the debate. The two main challenges with this approach were that some students developed problematic and simplistic theses (giving, for instance, answers to the above question that were tantamount to saying racist thinking was *the* cause of New Imperialism, or it had *no* causal effect at all on New Imperialism), and some teams developed theses that were indeed different from each other in ways that the instructor—yet all too few students listening to the debate—found interesting. The result was as indicated above. While some of those debates worked well to help students synthesize and critically re-evaluate what they had learned about the topic, some debates were flops. Even those debates that worked well in the end required the instructor to flog the discussion along by posing provocative questions that a more engaged student audience could have produced on its own.

Towards a Solution

Holding other parts of the course and the unit organization relatively steady, Ellis continued to tinker with the debates, tweaking them by introducing two changes suggested to him by Holly Swyers (Lake Forest College): “structured controversy” for debate team members and a more formal system of note-taking for

audience members. Structured controversy has a history of success in a variety of disciplines, ranging from history to biology.⁵ In the discipline of history, Elizabeth Green Musselman has asserted that using this technique led students “to see themselves as co-producers of a historical knowledge that is never final,” and that “debates emphasize the open-ended quality of historical scholarship and the importance of discussion with colleagues in formulating more sophisticated understandings of history.”⁶ Ellis’ work builds on Musselman’s claims, offering statistical evidence of the effectiveness of structured controversy in producing gains in student learning. However, Ellis’ work adds a ludic twist. One common approach to a structured controversy is to provide a specific debate resolution (or, in Musselman’s case, specific prompts from scholars with contending views) and then to invite students to test the claims through a close reading of primary sources, performing analytical moves that typify historical thinking. Such attempts are predicated on the power of the *good* example in learning: the exemplary practice of skills of critical thought and argumentation can lead debaters and their audiences to consolidate learning gains.

We don’t dispute that predicate, but Ellis’ approach is based on harnessing the power of the *bad* example in learning. He began by crafting a specific (and deliberately problematic) debate resolution that a pro team affirmed and a con team contested. For instance, in a unit of a course broadly designed to have students analyze continuities and discontinuities in the development of communism in Russia (and, later, the USSR) by considering such factors as the intentions of communist leaders, the intended and unintended effects of their policies, their freedom (and lack thereof) to shape revolutionary outcomes, etc., the debate resolution was, “Resolved: Stalin broke with and betrayed Lenin’s revolution.” To take another example, in a course unit dedicated to the examination of the complex intermixture of various economic, political, and social causes (and effects) in the cloth trade in late medieval and early modern Europe, with wide variations notably not only over time, but also across different states and cultures, the debate resolution was provocatively given as, “Resolved: In legislation and disputes about the production of and trade in cloth, economic factors clearly trumped any other concerns, such as politics and social order.” In another case, a course unit whose aim was to have students wrestle with the rich interplay of the

causes of New Imperialism, the debate resolution was problematically worded as, “Resolved: Racial considerations outweigh any other single factor in explaining what caused ‘New Imperialism.’”

Ellis’ twist on structured controversy was to offer each team the option to try to win the audience over by adopting an approach closer to sophistry than history—and challenging the audience to catch the sophists in the act. Debate team members were empowered and explicitly invited to manipulate the evidence, pull quotes from primary sources out of context, and generally—as the familiar saying goes—torture the evidence to confess to their thesis. It was the job of the audience to identify and critique the bad historical thinking the debaters were invited to exercise. This use of structured controversy debates thus took what Sam Wineburg rightly identified as a problem, namely, that “we contort the past to fit the predetermined meanings we have already assigned to it,” and, in unmasking bad historical thinking, made visible how seductively natural bad historical thinking is. This, therefore, laid the groundwork for students to grasp that good historical thinking is “neither a natural process nor something that springs automatically from psychological development.”⁷

In this approach, structured controversy debates became somewhat liminal experiences, in the sense that no one quite knew at any given moment how the class debate would develop, exactly which pieces of evidence would be presented, how the evidence would be interpreted and contested, whether debaters would deliberately twist the evidence, what questions and follow-ups the audience would generate, and how the debate would be resolved. Both audience and debate teams second-guessed both the well-founded and the deliberately not well-founded claims that were advanced, having added reason to be skeptical of others’ use of evidence. The debates were also ludic in the sense that debaters were directed to play assigned roles for or against the resolution (rather than to represent their own, actual views), and to use, if they so chose, bad historical moves to fool the opposing team and the audience—who were in turn challenged to call out those bad historical moves. Students had to display their intellectual sophistication by artfully practicing sophistry, cleverly detecting it in others, or both.

How do the debates fit with other elements of Ellis’ courses? As with his previous approach, he used structured controversy debate as a scaffolded assignment. In a typical course unit in an introductory

course, the final product demonstrating evidence of student learning would be a scholarly essay making a persuasive, evidence-based argument that explains change (or continuity) over time to a skeptical audience. To reach that goal, Ellis usually starts a unit with a reading from a textbook on the unit topic (such as the French Revolution or the rise of National Socialism), with the aim of introducing students to the broad contours of the topic and some of the important names, terms, dates, and concepts that will figure in later assignments. Although many facts about the past are thus addressed, the main aim at that stage in the course is not to achieve comprehensive coverage per se, but to provide enough background for students to join a scholarly conversation about a problematic aspect of the broader topic. A broadly formulated guiding question (concerning, for example, the problem of authority and legitimacy in the different phases of the French Revolution, the causes of the failure of the Weimar Republic, or the relationship of feminism to anti-colonialist and anti-imperialist movements in the 1960s) helped students focus their textbook reading.

In a second phase of the unit, usually lasting two or three class periods, Ellis has students discuss and analyze both a few excerpts of scholarly sources and numerous primary sources connected to the debate topic. He asks students to look back to the guiding question and ahead to the debate resolution, and, taking at least one quote from a scholarly reading as a prompt for the eventual essay, to begin writing paragraphs for a rough draft based on a close and critical reading of primary sources to provide evidence for their response to the scholarly source. This approach to writing is indebted to Gerald Graff and Cathy Birkenstein's *They Say, I Say*, a book that all first-year students work with extensively at Augustana, and which, therefore, provides a common vocabulary across disciplines and throughout students' college careers to articulate expectations in different disciplines.⁸

Since students' essays were to be assessed using a grading rubric adapted from Thomas Andrew and Flannery Burke's five C's of historical thinking,⁹ Ellis stressed in class that the student-generated paragraphs not only required students to find telling quotes in the written primary sources to back up their views, but also that the quality of the evidence should be evaluated critically by assessing the context of the author (especially the degree of reliability); context of the historical moment (especially how representative the

primary sources are); complexity (especially by explaining why each student's view is better supported by evidence or less problematic than other plausible readings of the evidence); multiple causality; and contingency (especially by analyzing how much agency historical actors possessed and how much they were constrained by factors they had little or no control over).

Having completed a substantial portion of a rough draft that connects to the debate resolution, and having a good sense of the primary sources each debate team will likely draw upon, students then enter the next phase of the unit—a class period devoted exclusively to the debate experience. Following the debate, students then work on revising their rough drafts so that the essay they produce is connected to the debate topic, but is more carefully defined in its scope and claims. After Ellis began using the structured controversy debates, he noticed a marked improvement not only in student engagement throughout the course, but also higher quality and greater attention to nuance in students' essays, with particular gains in the ability of students to contextualize primary sources critically, to anticipate potential criticisms, and, in responding directly to them, to make their overall thesis more complex and more compelling.

This use of structured controversy debates fits not only with the other assignments in the course and serves as a penultimate stage in helping students craft their own essays, it also serves the primary goal in Ellis' courses: learning to think historically. Admittedly, any attempt to characterize historical thinking is fraught: its elements and even the language to describe them are contested by historians. Ellis' own conception of what it means to teach students how to think historically, as one might gather from his adaptation of Andrews and Burke, fits with the work of Lendol Calder and Tracy Steffes, who have usefully elaborated “not a comprehensive list of markers defining expertise in history,” but rather “concepts and competencies” that are integral to “*essential* outcomes.”¹⁰ Our truncated distillation of the five essential concepts that students, according to these authors, must grasp, are (1) that history “is an interpretive account” that has “conclusions presented in the form of a narrative or an analytic argument”; (2) that there exists a “complex relationship between past and present,” such that historians are engaged in “[m]ediating between the pastness of the past and the presentness of the past”; (3) that it is important to “understand the nature, potential, and limits”

of primary and secondary sources; (4) that “historical accounts are multiple and layered, avoiding monocausal explanations and reductionist thinking”; and (5) “that [h]istorians generally regard something as significant if (a) it affects change or continuity with meaningful consequences, for many people, over a long period of time or if (b) it is revealing, leading us to understand other subjects in history and contemporary life in new ways, or was important at some stage in history within the collective memory of a group or groups.”¹¹ Therefore, Calder and Steffes argued, students should learn as core competencies to “evaluate historical accounts,” “interpret primary sources,” “apply chronological reasoning,” “contextualize,” and “construct acceptable historical accounts”—although “a definitive list of learning outcomes for history is, of course, a chimera.”¹²

Structured controversy debates also fit with the work of Joel M. Sipress and David J. Voelker, who have contended that “[a]n important step toward a new signature pedagogy for the introductory history course would therefore be to render visible both to ourselves and our students, the centrality of argument to the practice of history.”¹³ This claim echoes their assertion elsewhere that the introductory survey course should be transformed by shifting away from a coverage model toward an argument-based approach whose benefits are at least two-fold: students not only get better at doing history (in part, because argumentation trumps fact-acquisition as a motivational factor), but also become better citizens in the process. They have argued that the “process of debate and argument provides a model of an engaged civic community into which introductory history students can be invited at a novice level. An argument-based history course,” they add, “particularly when housed within a broader curriculum that emphasizes engaged learning, may encourage students to incorporate historical modes of thinking into their daily lives.”¹⁴ We concur, and would add that this variety of structured controversy debates, in which heightened skepticism results from the known possibility that debaters might use bad historical methods, or “fake history,” is especially relevant in the era of “deepfakes,” potentially “fake news,” and potentially fake assertions of “fake news.”

Ellis’ use of structured controversy is also consistent with the “pedagogies of positionality” advocated by Frances A. Maher and Mary Kay Thompson Tetrault. Writing in 1994, they found that “positional understandings are discouraged or blocked in the

academic environment” in part because “[i]n most institutions the academic disciplines, as traditionally formulated, still hold sway as modes of structuring and transmitting knowledge as well as repositories of the accumulated wisdom of trained scholars.”¹⁵ Although this version of structured controversy debates features a competitive element, the formalized nature of the debate teams’ contestation and the playfully couched permission to exhibit bad historical thinking to fool the audience lowers the emotional and psychological threshold of competition: the debaters, as well as the audience members, are ostensibly merely playing assigned roles rather than making their actual selves vulnerable. Additionally, this pedagogical approach presents learning as an intellectual game, albeit with serious learning outcomes, in which all participants begin on a more level playing field since the radical skepticism displayed toward all speakers (on the debate teams as well as in the audience) may work somewhat against any perceptions of privilege that might be attached to any given participant.

Student comments on anonymous course evaluation forms supported instructor impressions about the effective and enjoyable impact of the debates on student learning. In one representative course section, when asked to identify the most and least helpful parts of the course, not all students commented, but 17 out of 27 students (63%) volunteered that debates were one of the most helpful elements (while none identified the debates as the least helpful). Individual students made comments such as: “Debates made you immerse yourself in the material and made it personally relevant,” “Debates increased interest in class,” “Debates were a fun way to learn the material,” “I learned a lot about...how to study better and write better papers and speak in front of the class,” and “The most helpful were the team debates, class discussions, and the primary source book. All three allowed us to examine history on a first-hand account and to judge for ourselves what we wanted to believe.”

Structured Controversy Debate in the Classroom

This style of structured controversy debate involved both a rigidly prescribed structure and opportunities for liminal experiences for debaters and audience members. The debates consisted of a setup followed by three distinct debate phases and a post-mortem. In

the setup prior to the debate, the two debate teams, each usually consisting of three or four students, took their seats, which had been arranged at the head of the room in rows facing each other and perpendicular to the audience. This use of the physical space was designed to heighten the sense of opposition between the two sides. Ellis reminded the entire class of the structure of the debate assignment (previously described for them in the syllabus, as well as discussed in the previous class) and distributed a sheet to the audience to help them track and evaluate each debate team's claims. Audience members were reminded that the instructor would feel free to take up these flowcharts of the debate, and treat them as a quiz for the day. In practice, he rarely took up the debate forms, instead instructing students at the end of the debate to take them home as an aid in writing a formal, traditional historical essay on a topic related to the debate, but defined by the students in a more careful and scholarly way. Debate team members were not required to complete the debate flowcharts, although many debaters did so on the grounds that the flowcharts helped them track and challenge their opponents' arguments in a comprehensive, systematic way. A formal countdown timer, complete with a bell that jangled annoyingly when the assigned time elapsed, was displayed on the classroom projection screen. Prior to class, Ellis enjoined debate team members to meet to prepare their cases carefully, so that each debate team member would use the time equitably and productively in each of the three phases of the debate.

Ellis also had instructed debate team members that they had a high degree of freedom to shape their presentations. Debate team members could choose, in consultation with each other, exactly how to proceed, including whether they wished only to speak or also to augment their spoken presentations by distributing handouts to the class, using PowerPoint, etc. Students also had total freedom in choosing—again in consultation with their team members—their own line of argumentation as well as their choice of evidence from primary and secondary sources. While most students chose to rely on evidence from primary and secondary sources previously assigned in the unit and discussed in group discussion in previous class meetings, students were also encouraged to bring in new evidence that the other debate team was unlikely to have seen, thereby adding to the understanding of the topic and wrong-footing their opponents. Ellis also had previously advised debate team members to try to take

ownership of the key terms in the problematic debate resolutions, since those who get to define the key terms of a debate have already gone a long way toward winning it, but also mainly to base their arguments on the use of evidence from primary sources—in part, so that the ensuing debates would not devolve into an unresolvable contest of definitions. Immediately prior to beginning first phase, Ellis reminded all students that the debate team members were merely playing a role and were being forced to defend or contest a specific resolution, adding that much of the learning would take place in the active responses from the audience. He also took the opportunity to remind all students of some of the basic rules of civil debate, including the fact that it is always fair to question the logic and use of evidence of others, but never fair to attack their individual dignity. Prior to the debate's start, students were reminded that since every student in the class would eventually be on a debate team, the audience should pay careful attention to what was worthy of emulation in their own debates that would follow later in the course.

In the first phase of the debate, pro team members had five minutes to present their carefully prepared case in favor of the resolution. Then, con team members had five minutes to respond immediately and extemporaneously to the argument they had just heard (but might, with good preparation, have anticipated) from the pro team. During these presentations, the audience made careful notes about what each team said, recorded each asserted point and the use of supporting evidence (especially quotes from properly sourced primary sources) by the pro team, evaluated how and how well the con team had responded, and determined which team had won each asserted point. The debate flowchart form used by the audience was directly drawn from Holly Swyers' work. The form featured the formal debate resolution printed at the top and was divided into columns to facilitate the audience in its task to track and evaluate the exchanges between the debate teams. After the exchange of debaters' views, a one-minute pause then followed, in which each audience member wrote a critical question or comment for the pro team, and the debate teams prepared for the second phase of the debate.

In the second phase of the debate, the debate teams swapped roles: the con team had five minutes to make its carefully prepared case, and the pro team had five minutes to respond immediately and extemporaneously to the argument they had just heard from

the con team. The audience, during the second phase, flipped their flowchart over to the second page and continued to track and evaluate the arguments and use of evidence by each debate team. After the exchange of views by the debate teams, the audience again took one minute to write a question or comment, this time for the con team, while the debate teams prepared for the third phase of the debate.

In this third phase of the debate, the audience took two minutes, using the pair-and-share technique, to turn to a classmate and discuss the questions or comments that each of the two students had generated. The debate teams, in the interim, prepared themselves for the questions that were likely to follow. Debate team members had been encouraged to use the free-flowing third phase of question-and-answer to seize the opportunity both to work in any material they meant to get to but had not during the first two phases—which was especially useful in case a team misjudged its use of allotted time in the previous phases—and to interrogate or respond to the other debate team whenever they chose to. Following the two-minute exchanges between audience members, audience members then turned to interrogate the debate teams as a hostile jury, each member of which was equipped with two of his or her own insights along with the two of his or her erstwhile interlocutor from the pair-and-share exercise. For the benefit of those students generally prone to shyness or those students who were somewhat less quick to process their responses, Ellis reminded audience members that this phase of the debate offered a great opportunity to pose a question or make a comment that they had time to study in consideration and write down in advance. (In post-mortems of the debates, shy students often expressed appreciation for these opportunities to make an active contribution to classroom discussion.) In this third phase, then, audience members posed their questions and comments to the debate teams. On occasion, the instructor would follow up by asking the student who posed the question whether he or she was persuaded by the answer received, often leading to further discussion as the opposing debate team joined in. For the most part, however, he found it was seldom necessary to nudge the discussion forward. As a rule, students (each having, again, at least four questions or comments ready at hand) moved the discussions forward on their own until the instructor closed off the discussions, usually after about thirty to forty minutes, to move toward a post-mortem discussion of the debate.

Post-Mortem and Student Feedback

In the post-mortem, which usually lasted five to eight minutes, Ellis made it clear that the debate was officially over and that the debate team members were released from their obligations to promote or contest the debate resolution. He then asked a series of questions to get students to reflect on their experiences in the debate, to relate the debate to previously studied material, and to connect their insights to assignments that would follow (usually, again, an essay on a topic related to the debate). In the post-mortem, he typically began by asking the audience members (1) what they saw debate team members doing well that they could use in their own debates, (2) if they could identify any specific manipulations of the evidence by the debate teams, and (3) how the experience of the debate impacted their understanding of the principal concerns addressed in the course unit. Typical responses to the first question were that diligent preparation before the debate paid off in detailed, evidence-rich answers, in good clock management, in the ability to anticipate what the other team might argue, and in heightened confidence and persuasiveness. Responses to the second question varied widely, and responses to the third usually led students to note connections they had not seen before, to understand specific primary sources in new ways, and to understand more broadly how a critic might plausibly interpret sources in ways students had not previously appreciated.

Ellis then asked three questions for the debate teams, asking first if the experience of being forced to argue for or against the resolution impacted their actual views. In most cases, debate team members (who had been free to sign up for the pro or con side) stated that the experience had rendered more complex and specific ideas, but did not fundamentally change their pre-debate response to the resolution. However, audience members frequently noted that the experience had in fact changed their response. Such teachable moments were used to stress that historical evidence should always shape—and in some cases perhaps correct—our views of the past, making connections to the broader course goal of teaching historical thinking.

Debate team members were also asked two related questions—namely, what each debate team thought of as its weakest element (and whether the audience thought it as well), as well as whether the teams had deliberately manipulated evidence (and whether anyone

in the audience noticed). In practice, many teams confessed after the debate to manipulating the evidence to win, while other teams insisted they had not done so. Still, their differing choices helped keep the audience usefully engaged for the full length of the debate, constantly questioning whether teams were in fact manipulating the evidence, offering audience members chances to become more critical listeners. Ellis told students he hoped such a skeptical stance would become habitual not only in class, but also in the real world of political campaigns, job interviews, etc. In some cases, audience members challenged the assertion of some teams that claimed they had not manipulated the evidence, leading to a productive re-examination of the problem of how to know when a different interpretation of evidence merely reflects a different perspective, and when it might reflect a bias or even an attempt to shoehorn the evidence to fit a preferred conclusion.

Audience members often stated that they relished this element of cat-and-mouse, having playfully explored through posing critical questions involving whether a debate team had chosen the path of sophistry. When debate teams successfully fooled the audience with a bad historical approach, the teams took obvious pleasure in confessing their exploit after the debate, and, occasionally, dismay that no one had caught them. Common techniques of bad historical thinking included deliberately ignoring evidence that contradicted a debate team's position, developing specious reasoning to question the quality of such evidence, and quoting in a misleadingly selective way from a primary source, conveniently disregarding the nuances in the document and cherry-picking the parts of the document that agreed best with the team's position. Finally, the instructor used the post-mortem to remind students about the scaffolding of the assignments, having students make the connections between assignments that had preceded and would follow the debate. A few minutes toward the end of class were then reserved for discussing questions and problems connecting the debate experience to upcoming student essays.

Grading and Learning Achievements

In addition to discussing and evaluating the debate in the post-mortem, Ellis also gave debaters graded feedback using a scoring guide that had been made available since the first day of class. The

scoring guide considered student performance in five categories. Every debate team member was graded on how well the team explained why the balance of evidence in the readings assigned for the topic supported or did not support the team's thesis. Every debate team member was also graded on how well the team addressed the most serious problems with or limitations of the team's argument, evaluating the merits of the team's argument in comparison to other points of view about the topic and other readings of the evidence. Additionally, each team member was graded for their individual performance in each of the three phases of debate—the phase of presentation of a prepared case, the phase of extemporaneous response to the opposing team's prepared case, and the phase of questions and answers with the audience.

Each of the five categories was worth an almost risibly small amount—only 1% of the course grade, for a combined total of just 5% of the entire course grade. Three factors played a role in the choice to make the debate worth a relatively small amount of the grade. First, many students confessed to considerable anxiety about any kind of public debate, and the low number of points helped to reduce their concerns. Second, the low stakes of the assignment worked, perhaps paradoxically, to free students to explore new interpretations with greater engagement and enthusiasm. Students identified the debates in course evaluations as one of the most enjoyable elements of the course and one of the most productive learning moments. Third, the traditional essay that followed was expected to be a qualitatively more important achievement due to the design of the low-points, high-impact debate assignment.

In this style of structured controversy debate, then, students operating within rigidly structured roles as debaters or audience members still retained a wide range of free choices, keeping everyone guessing about exactly what they would say and whether they were arguing in good faith or arguing as bad historians. The principal sources of free choices for debaters involved the fairly large pool of assigned primary sources (and an unlimited pool of unassigned primary and secondary sources), the precise evidence that would serve their team best, and the invitation to manipulate the evidence. To compound the liminal element, students were told before the debate that the instructor would reserve the right to “spike” the debate by asking students in the audience to try to achieve goals that

he gave them before the debate began—goals that were inimical to one team or both teams. He might, for example, ask a student to try to get one team to admit to a particular reading of a telling primary source or to grant that there was a serious problem with the way a team defined a key term in the debate. Since student engagement was already high regarding the debates, Ellis rarely exercised this right in practice, but many debate teams that encountered difficult questions from the audience incorrectly suspected that he had planted them there to begin with.

Ellis' choice of such liminal elements is connected to his use, in some courses, of *Reacting to the Past*, a pedagogy pioneered by Mark Carnes, in which students engage in extended role-playing scenarios set amid past historical controversies, such as the debate over whether to adopt the French Constitution of 1791 or whether to award Charles Darwin the Copley Medal—issues connected to much bigger humanistic concerns.¹⁶ As we have heard many *Reacting* instructors say, *Reacting* students receive detailed character descriptions and secret objectives rather than scripts. To achieve their secret objectives, students typically must speak, write, and undertake various actions as their character, and to do so well, they need to master documents from the past that inform them not only about their own character's views, but also about those of other characters, whom they can decide (within certain bounds) to cooperate with, ignore, or work against. In other words, they must investigate the same kinds of sources often assigned in more traditional pedagogies, but the liminal, ludic, and competitive nature of *Reacting* typically results in greater student engagement with the sources.¹⁷ Because each character makes many choices in response to other characters' choices, students learn (among other things) a great deal about the promise and limits of agency. They also learn that they will never know exactly what will happen when they walk into the classroom, as there is no way to determine beforehand how each individual character will respond to the individual and collective choices of other characters. Familiar with *Reacting* and its brilliant use of liminality and playfulness to achieve powerful learning outcomes, Ellis sought to blend some of those elements into his structured controversy debates.¹⁸ In several of his courses, *Reacting* games are also pursued after the units involving class debates have been concluded.

Using this structured controversy style of debate involved some difficulties, as it required the use of a format that serves admirably well for developing rhetorical skills in the context of some disciplines (e.g., Speech Communication) in a very different context for Ellis' courses.¹⁹ The resolution style of debate was adopted at the turn of the nineteenth to the twentieth century as the standard format for developing rhetorical skills. In the context of the courses, the development of rhetorical skills, while important, was subordinate to the goal of developing skills of historical thinking. In this specific context, the "good" format of resolution debates was ironically transformed into a "bad example" of historical thinking, resulting—by design—in overly simplified analyses of complex problems and evidence, so that the superior merits of good historical thinking might shine more brightly.

In other words, the use of structured controversy was based on the paradox that good scholarly habits can develop through reflectively practicing bad habits. It is a commonplace observation in history, as in many other disciplines, that the evidence can be twisted to fit many potential conclusions, and that the thus mistreated evidence will bear the marks of that twisting, which a careful investigator can detect. Put differently, while debate team members forced the evidence to fit a preconceived (indeed, pre-assigned) thesis, each debate team and the student audience were invited to expose the invalid use of evidence in the course of the debate. Ellis stressed to the students that the kind of yes/no responses that were so effective in developing rhetorical skills often resulted in poor historical analysis. Knowing that that the debates were a kind of parlor game designed to demonstrate the importance of a more nuanced approach restored an element of liminality to students, enabling them to advocate potentially controversial positions at little personal risk. The use of flowcharts was also based on the presumption, probably a very safe one, that students who take detailed and structured notes will be well placed to offer significant critiques.

Anecdotal impressions of the debate indicated significant improvement. Student engagement seemed consistently higher both for debate team members and for audience members. In collaboration with Mark Vincent, Ellis designed a SoTL project to determine whether data could support such anecdotal impressions and, if so, yield more precise information about who benefited most from the changes, and why.

Data Collection

Ellis and Vincent developed a series of Likert measures that would allow students to express their evaluation of the debate's effects (student feedback prompts are included in the **Appendix**). All students responded to sixteen assertions about the debate, while debate team members responded to an additional six, and audience members responded to an additional and different set of assertions. In order to create a significant data pool, data were gathered in courses taught between the Fall 2007 and the Spring 2015, resulting in a total of 1,444 responses from students assessing the effectiveness of the debates. A total of fifty-seven different debates were assessed, with an average of 25.3 students assessing each debate's effectiveness. The fifty-seven debates were conducted across fourteen course sections ranging from introductory, intermediate, and advanced undergraduate levels (i.e., 100-, 200-, and 300-level courses). With as few as two debates and as many as six debates per course section assessed in the fourteen course sections, an average of 4.07 debates per course section were assessed. No substantial differences were found between the introductory and advanced level, which may be related to the fact that Ellis' advanced-level courses often attract many students (usually about one-third to one-half of the total class population) who may be beyond the first year of college, but are taking an advanced history course as their first college history course. They therefore mostly mirror the general student population. The control group was conducted in an introductory (100-level) history course with a general student population, resulting in sixty-five total student evaluations of a total of three debates, with an average of 21.67 responses per debate. One course section, in which three debates were assessed, was a gateway course for the history major, and therefore had only history majors and minors in the course. All fifty-seven debates concerned European history, although the format of the debate is readily transferable to other national, regional, or world history courses.

Following a preliminary analysis of results from the first course, Ellis and Vincent collected additional data on subsequent courses. Specifically, all students—both those in the “unstructured” (control group) and structured debates—were asked to provide their name, sex, class level (first-year, sophomore, etc.), the number of college-level history courses taken previously, their intended major(s) and minor(s),

Assertion	Mean	Standard Deviation
Debate made me more interested in today's material.	1.9847	1.04191
Debate improved my understanding of today's material.	2.0320	1.09697
Debate made me more interested in unit material.	2.1639	.97196
Debate improved my understanding of unit material.	2.1873	1.02103
I think historical knowledge can change.	1.8949	2.32853
I think knowledge is something produced through an exchange of scholarly views and through criticism.	1.7996	1.05898
I now have a more complex answer to the guiding question for the week than I had before the unit.	1.9119	1.03202
I have a good sense of how information from today fit with information discussed in earlier classes.	1.9890	.98297
Because I had written a journal, I showed up to class having already thought carefully about the material debated.	1.9677	1.13254
On the whole, students connected to each other in a productive way today.	1.9577	.99581
Today's class meeting was driven by student's analysis.	1.7790	1.01953
Most of the audience consisted of active rather than passive listeners.	2.2073	1.34787
I am more curious about the question than when I first encountered the guiding question.	2.0445	1.00955
I think other perspectives are valuable in understanding history.	1.6405	1.01137
Studying this unit has made me more aware of the strangeness of the past.	1.9861	2.94674

Figure 1: Students' perceptions of the impact of structured controversy debates on learning and interest (based on a Likert scale of 1 to 5, in which a lower number indicates agreement).

the number of courses in which team debates were organized and required of all students, and the number of students on their debate team. Additionally, students were asked to rank in order of importance the reasons why they took the course, including the fact that the course satisfied a general education requirement, the time of the day in which the course was offered, the reputation of the instructor, and interest in the specific content (European history) of the course. Students were also asked, more broadly, to add anything else they thought the instructor should know about their experience of the debate.

To help contextualize the data, it may be helpful to note some characteristics of Augustana students in general and of the students in these courses in particular. Augustana College is a selective, private liberal arts school with a residential campus and about 2,500 undergraduates, drawn from many parts of the United States and several foreign countries. Enrollments for the courses ranged from 19 to 30, and was typically about 25. In most courses, the students met three times per week, for 75 minutes per meeting, for ten weeks. In a few courses, students met two times per week, for 110 minutes per meeting, for ten weeks. Using data from an “unstructured” control group and a structured group, which were typical of the data pool overall, we found that most students in introductory courses were first-year (74.6%), with little or no experience in college-level history courses, and that upper-level courses typically had a more even mixture of beginning and advanced students in history. The largest group of students (42.7%) reported they took the course to satisfy a general education requirement—or as they often put it, to “get rid of a requirement.” 47.8% had declared no major. 27.5% were History and History Education majors. 53.6% reported no significant debate experience in other classes.

Evaluation of the Data

We found that the evaluation of the data confirmed our impression that the structured controversy debates were working well. In fact, on *every* measure, the numbers showed that students perceived the debates to have deepened their learning and piqued their interest (**Figure 1**). Using the Likert scale of 1 to 5, with 1 indicating strong agreement and 5 strong disagreement with the statements listed above (meaning that a lower number showed agreement with

Student Grade Level	Males	Females
First-Year	2.004	2.064
Sophomore	1.671	2.403
Junior	1.867	1.949
Senior	1.783	1.754

Figure 2: Students' perceptions of debate effectiveness, categorized by grade level for male and female students (based on a Likert scale of 1 to 5, in which a lower number indicates agreement).

the statements), we found that the mean response to all claims was 1.936, a clear indication that students found the debates generally effective in advancing their learning. We further found that the structured controversy debate offered a significant improvement over unstructured debates. Finally, we found that while adding structure to the debates worked slightly better for some students than others, the differences had little explanatory value. In other words, very similar results for different subsets of students support the claim that the structured controversy model works quite well for all student groups.

We found that the items were strongly correlated, and used factor analysis to confirm they were dealing with a single factor. This factor consists of all items except "historical knowledge can change." The reliability of the resulting fourteen-item scale was classified as "good" (Cronbach's alpha = 0.94).

Additionally, we wanted to determine if adding structure to the debates had gendered implications for student learning. The answer to this question is complex. On the one hand, the debates generally affected self-reported men and women differently ($F(1, 1170) = 8.781, p = .003, \eta_p^2 = .007$). Males reported having a more positive response than females, averaged across the two formats, $M_s = 1.838$ vs. 2.056 for male and female students, respectively. More importantly, this difference reached significance only for those in the *unstructured* debate condition, $M_s = 1.752$ vs. 2.177 , for males and females, respectively; $F(1, 1170) = 12.755, p < .01$. This difference between males and females was *not* significant when structured debates were used ($M_s = 1.093$ and 1.965 for male and female students, respectively).

Debate Number (and Topic)	Mean
Debate 1 (on the origins of WWI)	2.9514
Debate 2 (on the rise of National Socialism)	1.9560
Debate 3 (on the origins of the Cold War)	1.8478
Debate 4 (on post-WWII immigration)	1.7681
Debate 5 (on progress/regress in European civilization/s)	1.8825

Figure 3: Students’ perceptions of debate effectiveness, over multiple debates within a sample course section (based on a Likert scale of 1 to 5, in which a lower number indicates agreement).

We interpret the data as meaning that while the more traditional (unstructured) format clearly benefited men more than women in a statistically significant fashion, the structured controversy debates were beneficial for both males and females, moving toward (without quite reaching) gender parity in outcomes. This result is informed by what some other researchers have found. For example, Joe Bellon observed that “Colbert (1994) notes that ‘debating may be an effective method of assertiveness training,’ especially for women.”²⁰ In our study, a *t*-test comparing men versus women on the scale totals, ignoring the structured/unstructured distinction, was not significant. However, a difference did re-emerge when correlated with the students’ year in college. We found that sophomore and senior men reported that the debate worked especially well for them (meaning they had significantly lower means on the scale), and sophomore women found the debates worked less well than women at other levels of education (**Figure 2**).

If we were to gather further data, we would like to ask the students more information about their backgrounds, especially asking them how they would identify their race, ethnicity, and nationality, as well as whether students were the first in their families to attend college. We would also like to inquire in a more open-ended way about their sex and/or gender (in part so as to encompass those who might not self-identify as male or female). Having such data would enable a richer and more detailed interpretation of the debates’ effectiveness for various groups.

Motivation for Taking the Course	Mean
Satisfies a requirement	1.902
Time of day in which the course is offered	1.767
Interest in content of the course	1.920
Reputation of the instructor	1.943
History major or minor	2.213

Figure 4: Students' perceptions of debate effectiveness, categorized by motivation for taking the course (based on a Likert scale of 1 to 5, in which a lower number indicates agreement).

In many courses, multiple debates were conducted, with the pattern revealing that the first debate was the least effective, while subsequent debates were usually increasingly more effective, as students became more accustomed to the format. In a typical course on twentieth-century Europe, for example, the results for one section are displayed in **Figure 3**.

Unsurprisingly, we found that students who were members of the classroom debate teams reported higher levels of learning (i.e., lower means) than audience members (2.0156 for audience members vs. 1.8415 for debaters). However, the audience, too, clearly found the debates enhanced their learning.

When we examined the students' reported motivations for signing up for the course, post hoc means revealed that history majors and minors did find the debates to be effective, but other motivations were more closely connected with perceptions of debate effectiveness (**Figure 4**). We interpreted the results as indicating that students who likely had lower levels of motivation to take the course (e.g., students who took the course mainly because it happened to fit their time schedule) found the debates especially effective for learning, while the history majors and minors who presumably entered the course with higher levels of motivation found other means of learning (including class discussion and lecture) to be nearly as effective for their learning. The implication is that the structured debate format is valuable for both novices and students more familiar with historical thinking.

Meanwhile, we found that both the reason for taking course and type of debate (structured and unstructured) had a marginally significant effect on student perceptions. However, the interaction between the two variables was not significant. Finally, we found that the correlation between mean rating and the number of other courses that had used formal debates was statistically significant, but extremely weak (-.073). In other words, while previous debate experience is related, it is of little explanatory value.

Conclusion

The adoption of a new debate format based on structured controversy with liminal and ludic elements proved effective for *all* measured student groups, and notably moved closer to parity in self-reported gains for men and women than a more traditional debate format. Students reported gains in learning across all of the measured areas. Paradoxically, when debate teams were encouraged make the bad historical move of forcing the evidence to fit a pro or con response to a debate resolution, both the debaters and audience members developed more sophisticated skills of historical analysis, as actively engaged audience members used critical questions to expose nuances in the evidence, developing an evidence-driven, complex analysis.

Notes

1. Grant P. Wiggins and Jay McTighe, *Understanding by Design*, second ed. (Upper Saddle River, NJ: Pearson, 2005), 14.
2. L. Dee Fink, *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses*, Jossey-Bass Higher and Adult Education Series (San Francisco, CA: Jossey-Bass, 2003).
3. Lendol Calder, "Uncoverage: Toward a Signature Pedagogy for the History Survey," *The Journal of American History* 92, no. 4 (March 2006): 1358-1370.
4. In some cases, students could write a question on one of two units covered.
5. Cf. Janell M. Mead and Lawrence C. Scharmann, "Enhancing Critical Thinking through Structured Academic Controversy," *The American Biology Teacher* 45, no. 7 (October 1994): 416-419.
6. Elizabeth Green Musselman, "Using Structured Debate to Achieve Autonomous Student Discussion," *The History Teacher* 37, no. 3 (May 2004): 336.
7. Samuel S. Wineburg, *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past* (Philadelphia, PA: Temple University Press, 2001), 6-7.
8. Gerald Graff and Cathy Birkenstein, *They Say, I Say: The Moves that Matter in Academic Writing*, third ed. (New York: W. W. Norton, 2014).
9. Thomas Andrews and Flannery Burke, "What Does It Mean to Think Historically?" *Perspectives on History* 45, no. 1 (January 2007), <<https://www.historians.org/publications-and-directories/perspectives-on-history/january-2007/what-does-it-mean-to-think-historically>>.
10. Lendol Calder and Tracy Steffes, "Measuring College Learning in History," in *Improving Quality in American Higher Education: Learning Outcomes and Assessments for the 21st Century*, ed. Richard Arum, Josipa Roksa, and Amanda Cook (San Francisco, CA: Jossey-Bass/Wiley, 2016), 55-56.
11. Calder and Steffes, "Measuring College Learning," 57-60.
12. Calder and Steffes, "Measuring College Learning," 60-63.
13. Joel M. Sipress and David J. Voelker, "From Learning History to Doing History: Beyond the Coverage Model," in *Exploring Signature Pedagogies: Approaches to Teaching Disciplinary Habits of Mind*, ed. Regan A. R. Gurung, Nancy L. Chick, and Aeron Haynie (Sterling, VA: Stylus, 2009), 27.
14. Joel M. Sipress and David J. Voelker, "The End of the History Survey Course: The Rise and Fall of the Coverage Model," *The Journal of American History* 97, no. 4 (March 2011): 1064.
15. Frances A. Maher and Mary Kay Thompson Tetrault, *The Feminist Classroom* (New York: Basic Books, 1994), 203 and 208.
16. To examine the student manual for these games, see Jennifer J. Popiel, Mark C. Carnes, and Gary Kates, *Rousseau, Burke, and Revolution in France, 1791*, second ed. (New York: W. W. Norton, 2015); and Marsha Driscoll, *Charles Darwin, the Copley Medal, and the Rise of Naturalism, 1861-1864* (New York: W. W. Norton, 2010). A full list of published Reacting games can be found at Barnard College's "Reacting to the Past" page, <<https://reacting.barnard.edu/games>>.

17. See, for example, Russell Olwell and Azibo Stevens, “‘I had to check my thoughts’: How the Reacting to the Past Methodology Impacts First-Year College Student Engagement, Retention, and Historical Thinking,” *The History Teacher* 48, no. 3 (May 2015): 561-572.

18. For a general introduction to Reacting and to its impact on student engagement, see the Teagle Foundation White Paper, “Reacting to the Past: A New Approach to Student Engagement and to Enhancing General Education,” by Richard Gid Powers, John M. Burney, and Mark C. Carnes, 2010, <https://reacting.barnard.edu/sites/default/files/inline-files/reacting_white_paper_teaglefoundation_0.pdf>; and Matthew C. Weidenfeld and Kenneth E. Fernandez, “Does *Reacting to the Past* Increase Student Engagement? An Empirical Evaluation of the Use of Historical Simulations in Teaching Political Theory,” *Journal of Political Science Education* 13, no. 1 (2017): 46-61. The Teagle White Paper found (p. 14) that in a 2009 survey of fifty-three instructors who used Reacting in their classrooms, more than 90% reported “very effective” or “effective” in “producing student learning of content or skills” in the following areas: “inquiry and analysis” (96.2%), “critical thinking” (96.1%), “oral communication” (96.1%), “integrative learning” (92.4%), and “teamwork” (90.6%).

19. Perhaps for this reason, it has proven somewhat difficult to find scholarly literature that examines how this style of debate plucked from one discipline serves the learning goals in courses developed in another.

20. Joe Bellon, “A Research-Based Justification for Debate across the Curriculum,” *Argumentation and Advocacy* 36, no. 3 (Winter 2000): 169. Here, Bellon is citing the work of Kent R. Colbert, “Replicating the Effects of Debate Participation on Argumentativeness and Verbal Aggression,” *The Forensic of Pi Kappa Delta* 79, no. 3 (Spring 1994): 7.

Appendix: Student Feedback Questions

Students were provided with these prompts following the debates, and were asked to respond using a Likert scale of 1 to 5, with 1 indicating strong agreement and 5 strong disagreement with the statement.

All students responded to these statements:

- The debate made me more *interested* in today's material.
- The debate improved my *understanding* of today's material.
- The debate made me more *interested* in the other material assigned for the *unit*.
- The debate improved my *understanding* of the other material assigned for the *unit*.
- I think *historical knowledge can change*.
- I think knowledge is something produced through an *exchange* of scholarly views and through criticism.
- I now have a more *complex answer* to the guiding question(s) for the week than I had at the start of the unit.
- I have a good sense of how information from today *fits* with information discussed in earlier classes.
- Because I had already written a journal entry, I showed up to class having already *thought carefully* about the material that was debated.
- On the whole, students *connected to each other* in a productive way today.
- Today's class meeting was driven by *students' analysis*.
- Most of the audience consisted of *active rather than passive listeners*.
- I am more *curious* about the question than when I first encountered it.
- I think *other perspectives* are valuable in understanding history.
- Studying this unit has made me more aware of the *strangeness* of the past.

Debate team members (only) were asked to respond to these statements:

- I feel that I have improved my *public speaking skills* today.
- There was a healthy sense of *competition* between the debate teams.
- Today's debate made me more *critical of the position* my team advocated in the debate.
- *Being an audience member in previous classes helped to prepare me* to be a critical and active part of today's debate team.

- My experience as a debater today makes me more likely to *participate actively in this course*.
- My experience as a debater today makes me more likely to *participate actively in other courses*.

Audience members (only) were asked to respond to these statements:

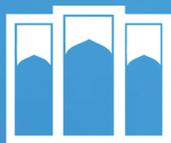
- The flowcharts made me a more *active and critical* audience participant.
- I was an *active rather than a passive listener*.
- The flowcharts helped me *develop good questions and/or criticisms* for the debate team.
- I felt a sense of *connection and community* among audience members.
- *Being a debate team member in a previous class helped to prepare me* to be a more critical and active member of the audience today.
- My experience as an audience member today makes me more likely to *participate actively in other courses*.

In a control group using an “unstructured” debate format (with more open-ended questions than the debate resolutions, and with the charge merely to take notes but not necessarily keep a flowchart of the debate), audience members were asked to respond to the following questions:

- Taking notes made me a more *active and critical* audience participant.
- I was an *active rather than a passive listener*.
- Taking notes helped me *develop good questions and/or criticisms* for the debate team.
- I felt a sense of *connection and community* among audience members.
- *Being a debate team member in a previous class helped to prepare me* to be a more critical and active member of the audience today.
- My experience as an audience member today makes me more likely to *participate actively in other courses*.



INDIAN OCEAN IN WORLD HISTORY



Sultan Qaboos
Cultural Center

www.IndianOceanHistory.org