The Historical Account Differences survey: Enriching methods for assessing metahistorical understanding in complex school environments

Kevin O’Neill

Sheryl Guloy

Faculty of Education

Simon Fraser University

Presented to the Teaching History SIG at the Annual Meeting of the American Educational Research Association, April 2010
ABSTRACT

To equip students for participation in a pluralistic, democratic society, history curriculum should help them develop mature ideas about why multiple accounts of the same events exist. But how can we know if we have achieved this aim? The problem is pressing for researchers and teachers who are attempting to refine new curriculum that is enacted in varying ways, in differing school environments, with students who bring unique individual experiences and capabilities to the classroom. To find clarity amid this complexity, new tools of assessment are needed.

We describe the design and piloting of a paper-and-pencil survey instrument, the Historical Account Differences (HAD) survey, which is intended to add to the toolset that researchers and practitioners can use to evaluate students’ conceptions about multiple historical accounts. Items were constructed based on Shemilt’s (1987) developmental theory. Data are presented from a pilot of the instrument with 65 11th-grade students, who formed two groups. One group completed the survey once, while the other completed it prior to and after a curriculum unit designed to enhance their conceptions about multiple accounts. Analysis showed that while the two groups’ responses were not detectably different initially, students completing the two-week curriculum unit showed an increase in mature conceptions. The limitations of this work and plans for future research are discussed.
The Historical Account Differences survey

INTRODUCTION

Helping students to understand and appreciate historical accounts has always been difficult work. However, as societies increasingly recognize their diversities, the pressure increases to understand and make sense of the existence of multiple accounts of past events (Banks, 2008; Seixas, 2004; Takaki, 1993). Without an understanding of why two carefully researched historical accounts may disagree, students may hold either a cynical view that history is written by the victors, or an “anything goes” conception according to which all accounts are equally valid in principle. Such conceptions are inconsistent with participation in a pluralistic, democratic society.

To refine curricula to help students construct more sophisticated ideas about why historians have produced varying accounts of the past, educators need practical tools of assessment that will help them to understand how well their efforts work across the dramatically varied conditions that characterize public schooling. At the present time, general-purpose tools of this kind do not exist. Over the past several years, we have begun to fill this void by developing and piloting a practical survey instrument that can assist teachers and researchers in assessing students’ understandings of the reasons why historical accounts may differ.

Students’ understandings of differing accounts are part of a larger set of “metahistorical” conceptions that scholars have studied for many years (Lee, 2004; Seixas, 2006; Wineburg, 2001). Other important metahistorical conceptions include historical significance, causes and consequences, constancy and change (Peck & Seixas, 2008). We do not claim that students’ beliefs about differing accounts are more important than these other metahistorical conceptions, but have been drawn to studying beliefs about differing accounts because they have figured large in students’ responses to our own work in the classroom.
The Historical Account Differences survey

While metahistorical conceptions have been studied extensively, most studies have been conducted in an exclusively qualitative fashion that does not lend itself readily to the examination of curriculum innovations at large scale. As all educators know, a curriculum or teaching strategy that functions well in one setting may fail in another for a variety of reasons (Dede, 1998). As educational researchers, teachers rely upon us to help them understand (and if possible, create) the necessary conditions for teaching innovations to succeed in the complex ecologies of classrooms, schools, and districts. The development of the HAD survey is an effort to enhance the toolkit available to history education researchers and teachers for understanding where and why particular innovations work well. The survey is intended for use in conjunction with classroom observations, participant interviews, and other techniques to evaluate the efficacy of innovative history teaching.

The HAD survey questions are based largely on the work of Denis Shemilt, who conducted many hours of interviews with British adolescents to examine their conceptions about the nature of historical knowledge and historians’ work (Shemilt, 1987, 2000). An early version of the survey and associated data analysis were presented at a prior AERA conference (O’Neill & Sohbat, 2004). Below we discuss the limitations of that prior work, and recent efforts we have made to improve upon it. These efforts have been carried out in conjunction with the development of a curriculum unit designed to foster understanding of conflicting historical accounts (O’Neill, Sensoy, & Guloy, 2009), and the data presented below stem from that work. The larger dataset about our curriculum implementation includes classroom observations, student work and interviews; however this paper focuses on the HAD survey itself, and what the data provided by student respondents may tell us about the present-day relevance of Shemilt’s 4-stage theory in North American schools. At the conclusion of the paper, we describe further work we plan to undertake to validate and improve the HAD survey.
THEORETICAL FOUNDATIONS

For many years now, educational researchers from a variety of research traditions have explored the question of what it means to know and understand history — not only in the limited sense of knowledge about specific events and people from the past that parents and teachers consider important, but also in the larger sense of how the past comes to be understood by historians (Wineburg, 1991). Because historical thinking is an “unnatural act” (Wineburg, 2001) it is a nontrivial matter for secondary students to appreciate that events that have been studied thoroughly in the past can attain new significance and attract new curiosity as we learn more about “how things turned out.” For example, the questions that historians will ask about the events of September 11, 2001 will undoubtedly be different 50 or 100 years from now. However, while new curiosities may arise about past events, historians (unlike scientists) cannot generate new evidence to answer specific questions. They may only search diligently to uncover evidence not considered before, or examine known evidence in a new light.

This starts to reveal how it can be that honest, well-trained historians differ in their interpretations of past events, and why they find it necessary to write new accounts of events that were written about previously in an honest and methodical way. (Consider the massive number of volumes written about the U.S. Civil War since that conflict ended.) However, scholars suggest that many students (even those in university) do not have mature conceptions about conflicting accounts of the past. This lack of mature conceptions is believed to be due in part to the fact that textbook-driven instruction, common throughout the United States and Canada, shields students from varying perspectives on historical events and personalities (Barton, 1997). For the most part, history textbooks present a single, homogenized perspective on the events they cover, written in an impersonal “voice of History” (Wineburg, 1991) that obscures the decisions
The Historical Account Differences survey made by the historian in constructing the account. Even worse, textbook accounts are often badly out of date with respect to historical scholarship, and in some cases are based more on myth than evidence (Loewen, 1995).

This situation is unfortunate, since students in a democratic society will inevitably encounter varying historical accounts as the implicit or explicit basis for political arguments about present decision-making. Consider, for example, the suggestion of U.S. vice president Dick Cheney in 2003 that troops arriving in Iraq would be “greeted as liberators” (Rusert, 2003) – a clearly intentional allusion to the greetings that U.S. soldiers received in Europe during World War II.

If we believe that a meta-level understanding of historical knowledge should be part of students’ preparation for life in democratic society, what sort of understanding should students reach? Furthermore, given the proven importance of students’ prior conceptions in learning new ideas (Bransford, Brown, & Cocking, 2000), how can we characterize the various preconceptions from which students begin to develop the understanding we desire for them?

While there are many possible answers to these questions, one useful framework for addressing them comes from the work of Denis Shemilt (1987, 2000, 2003). As part of the National School Councils History 13-16 project in the United Kingdom, Shemilt conducted research into the development of adolescents’ thinking about historical evidence and methodology. Shemilt’s 4-stage developmental theory, based on many hours of interviews with British adolescents, presents a rich and nuanced description of adolescents’ thinking about historical evidence and research methods. Since this scheme is important to understanding our work, we will review it in some detail here.
The Historical Account Differences survey

At the lowest level of understanding, Stage\(^1\) of Shemilt’s model, students take knowledge of the past as given. The only difficulty they associate with history is the difficulty of reading the stories and remembering them. By extension, they think of historians as no more than good memorizers of stories about the past. Students move on to Stage 2 when they realize that the past does not speak with a single voice: knowledgeable people do disagree in their accounts of the past. However, the only explanations a Stage 2 thinker can muster for this disagreement are that the evidence may not be reliable (e.g., some people saw the events in question while others did not) or that some reporters may be biased. At this stage, students think of historians as people who are somehow able to sniff out false or biased stories about the past, “read off” the truth from sources, and piece together the one true account.

Students at Stage 3 are distinguished by the understanding that historical knowledge can never be absolutely certain. At best, we can use the evidence available about the past (some of which is other peoples’ stories, but which also includes relics and non-narrative records) to reduce the uncertainty of our knowledge. In line with this view, Stage 3 thinkers understand historical scholarship to involve reasoning methodically with evidence to come up with an account that represents the most likely and believable reconstruction of events. Stage 3 thinkers understand historians as people who know how to do this methodical work.

Finally, in the 4\(^{th}\) and most advanced stage of Shemilt’s model, historical knowledge is viewed as kaleidoscopic. Students recognize that it is possible to have several equally defensible accounts of the past – particularly if they have been constructed to address different questions. In

\(^1\) Readers may associate the word “stage” with the work of developmental psychologist Jean Piaget, whose work postulated fixed, age-dependent stages of psychological development (Piaget, 1972). However, Shemilt did not observe any firm age dependency in students’ progressions from one set of ideas to another. Quite the contrary, he found students at nearly every stage in each age cohort in his sample.
The Historical Account Differences survey

this view, a historical account is understood to be a creature of its time. It is constrained by the available evidence, but is also shaped by the questions it seeks to answer. Like a kaleidoscope, history’s “patterns are ordered and determinate, but do not yield a single stable picture” (Shemilt, 2000).

This developmental account has been influential in the literature, garnering a large number of citations. Yet to our knowledge, the work discussed in this paper remains the only attempt to operationalize Shemilt’s 4-stage model for large-scale data collection. As a result, there has not yet been large-scale examination of whether the metahistorical ideas described by Shemilt actually cluster into the coherent stages he describes.

Lee & Shemilt (2003) revisited their prior research on the development of students’ metahistorical conceptions, urging researchers (especially in North America) to examine how the constructs they had formulated in the 1980s corresponded to students’ current metahistorical conceptions. They argued that metahistorical conceptions are likely to vary both from place to place, based on the nature of the curriculum students are exposed to, and over time, based on changes in culture. For this reason, both the HAD instrument and the analysis that follows were designed to build upon Shemilt’s previous findings about adolescents’ metahistorical conceptions, while avoiding assumptions about whether these conceptions clustered in the specific coherent stages previously formulated.

Lee and Shemilt are not alone in suggesting that stage models of development may deserve careful examination. In the years since Shemilt’s developmental theory was first published, there has been growing scholarly interest in the ideas that students hold about the nature of knowledge (Hofer & Pintrich, 2002; Schommer-Aikins & Easter, 2006), and some published critique of the ability of tidy stage models to represent the complexity of epistemological beliefs (Chandler, Hallett, & Sokol, 2002). While developmental models of
epistemological conceptions, often stemming from the work of Perry (1970), continue to draw scholars’ interest, current models such as Schommer-Aikins’ (Schommer-Aikins, 2004) represent epistemological beliefs using several interrelated but independent scales, rather than discrete stages (Schraw, Bendixen, & Dunkle, 2002).

Notwithstanding legitimate questions about whether students actually approach an appreciation of the “kaleidoscopic” nature of historical knowledge in discrete stages, the work we present below is premised on the belief that it is an appropriate long-term goal for all students in a pluralistic, democratic society to grasp the ideas Shemilt describes as Stage 4. If students appreciate that honestly constructed historical accounts can differ because they address different questions, they will be better equipped to negotiate life in a multicultural society without surrendering to the impoverished notion that everyone is entitled to his own opinion about the past.

CONTEXT OF THE STUDY

The Compassionate Canada unit

The work discussed here took place as part of a larger project sponsored by the Canadian Council on Learning, aimed at designing a technology-intensive curriculum unit that would help secondary social studies teachers to teach more sophisticated metahistorical conceptions while covering mandated curriculum. The unit, which came to be called “Compassionate Canada?” was developed by a design team that included three experienced social studies teachers and two university professors working in the areas of social studies and educational technology. Over a series of design sessions, the team agreed on the following broad parameters for the new unit:

1) Students would pursue the question “Has Canada become a more compassionate country in the last 100 years?” This question was chosen because it covered the entire scope of
The Historical Account Differences survey

the British Columbia Social Studies 11 curriculum, and addressed what experienced teachers considered to be a central theme of the course.

2) The materials provided to students should include a set of primary source evidence that could be used to address this question.

3) Groups of students in each class would pursue the overarching question using evidence about different historical events (cases) during the 100-year period. For example, two of the cases involved the internment of Japanese-Canadians during World War II and the exclusion of Chinese immigrants in the late 1800s and early 1900s.

4) Building upon prior work by O’Neill and his collaborators (O’Neill, 2004; O’Neill et al., 2003, April), the unit would involve volunteer historians serving as “telementors” to help students understand historians’ thinking and work practices. These volunteers would provide advice and guidance to students online as their work developed, taking the opportunity to model how a trained historian would pursue the students’ assignment.

5) The unit would take no longer than two weeks to implement. Given the overall scope of the prescribed curriculum for the year, teachers believed that this was as much time as could reasonably be dedicated to the unit.

Students bring a wide array of prior conceptions to bear on their studies in school – some of which are in conflict with curriculum goals (Bransford et al., 2000). History is no exception. Because every student enters the classroom with preexisting ideas about the nature of historical knowledge, improving students’ metahistorical conceptions is necessarily a matter of conceptual change (Vosniadou, 2007). Researchers have proposed that when one’s goal is conceptual change, learners must experience dissatisfaction with their existing understanding of a phenomenon in order to be open to developing a new one (Limon, 2001; Posner, Strike, Hewson, & Gertzog, 1982). Thus, whether a student began the unit with conceptions that Shemilt would
The Historical Account Differences survey describe as Stage 1, 2, 3, or 4, we aimed for *Compassionate Canada* to provide challenges to understanding that our volunteer telementors and teachers could help students to resolve.

To meet this goal, we developed a set of resources and activities that would arouse cognitive conflict about appropriate answers to the overarching question of the unit, and how these answers can be arrived at. Students in each participating class were divided into small groups, with each group studying a different event relating to the overarching question. These events were deliberately chosen by the design team to promote different opinions about the question of whether Canada had become more compassionate over time.

A range of documents pertaining to each historical case, including period photographs, newspaper articles and legislation, were arranged on a clickable timeline of events in Knowledge Forum, a commercially available online discussion environment (see Figure 1). A “backgrounder” document for each case, written by the design team in language intended to be readily understandable to students, was also attached to this timeline.

For students beginning the unit with Stage 1- or Stage 2-like conceptions, we expected that working with primary sources to address a historical question would bring about cognitive conflict in itself. For students with Stage 3-like conceptions, we aimed to arouse cognitive conflict through dialogue with peers who were attempting to address the overarching question of the unit using different evidence. At the conclusion of the unit, a culminating debate would drive the student teams to confront differences in their thoughts about the overarching question of the unit, if they had not already done so. Our expectation was that the juxtaposed views of students with in-depth knowledge of different cases would push those students already at Stage 3 to Stage 4 conceptions.

As students worked, volunteer telementors posted responses to their efforts, helping to ensure that they were being responsible to all the evidence provided, and were considering all
The Historical Account Differences survey

reasonable interpretations. A total of seven volunteer telementors were recruited to support the implementation, primarily via e-mail listservs for Canadian graduate students in history. Four of the volunteers were male, three female. Two had completed Ph.D.s in history, three were Ph.D. candidates in history programs, and the remaining two were completing MA degrees in history. Volunteers ranged in age from 25 to 36 years, averaging 30.33 years in age. All participated without pay, and submitted to requested criminal records and personal reference checks. For support, recruits had access to a web-based telementor’s guide, and could contact the first author via e-mail and telephone for advice.
The Historical Account Differences survey

Figure 1: A timeline view in Knowledge Forum, with source documents attached

Figure 2: A “take a stance” view for the Komagata Maru case. Each blue box represents a student or mentor posting. Notes that are responses to previous ones are linked with a blue arrow. Usernames are obscured for participant anonymity.
The Historical Account Differences survey

THE DESIGN OF THE HAD SURVEY

As described above, the HAD survey was intended to provide researchers and practitioners with an additional means to inform themselves about the effectiveness of their efforts to improve students’ understandings about differing historical accounts. The first version of HAD (O’Neill & Sohbat, 2004) was a step in this direction, but proved limited in some important ways. After consulting with a psychometrician, it was decided that three flaws needed to be addressed:

a) the survey provided few student reactions relative to the amount of time students were required to spend completing it. This was because the survey contained several long passages of text, which were challenging for many students to read with understanding during the brief time typically available

b) the questions were too few in number to support statistical tests of reliability; and

c) the survey was based largely on multiple-choice questions, which essentially assumed the existence of discrete stages in students’ thinking about historical evidence and methodology. These questions did not afford the opportunity to empirically examine whether students’ metahistorical conceptions corresponded to Shemilt’s theorized stages.

The revised HAD

Given the fact that most history instruction is based on textbooks that present a single unified account (Wineburg, 2001), no instrument that intends to evaluate students’ understandings of varying historical accounts can rely upon students having previously encountered this phenomenon. For this reason, HAD includes a set of three brief published narratives, chosen to present substantially different accounts of the same event.
The event selected for this study was the 1994 crisis in Rwanda, since a) it is an event about which widely divergent opinions have been published, b) it involves Canada and falls within the time period considered by the Grade 11 BC Social Studies curriculum addressed in *Compassionate Canada*, but c) it was not included as a case in the *Compassionate Canada* unit. (Thus, the instrument itself could not bias the educational outcomes of the unit.) To adapt the instrument for use in another setting, the Rwanda narratives should be replaced with age-appropriate, discrepant published narratives relating to the curriculum being studied.

Following the three narratives, the survey poses a series of comprehension questions designed to ensure that students process the narratives deeply enough to appreciate that historical accounts do, in fact, differ. After these comprehension questions come several rating-scale questions, on which our analysis below will focus. These questions include:

- What makes somebody a historian?
- How do historians know what caused an event in the past?
- Why do historians write new books about events that were already written about before?
- If a historian is learning about the events of a period and finds two stories about them that disagree, what should she do?

Students respond to each question by rating their agreement with four or more non-exclusive statements. Each statement was written by the first author to reflect the conceptions of students at one of the four stages hypothesized by Shemilt. These statements were examined and revised prior to the data collection by a colleague specializing in social studies scholarship, to help ensure correspondence with Shemilt’s constructs.

As an example, the question “What makes somebody a historian?” asked students to rate the following four statements on a scale from 1 (disagree strongly) to 7 (agree strongly):
The Historical Account Differences survey

- They have studied a lot about the truth of what happened in the past, and remember it (Stage 1)
- They can figure out which stories about the past are biased or untrue and put together the one true story (Stage 2)
- They don’t necessarily know exactly what happened, but can use evidence to figure out what probably did (Stage 3)
- They investigate new questions that people ask about what happened in the past and what it means today (Stage 4)

In this and the other questions, an effort was made to formulate prompts that were of roughly equal length, and free of obscure vocabulary. Most important, the new rating-scale questions afford an important kind of analysis not supported by the previous version of HAD. Rather than assuming a priori the coherence of particular ideas in particular stages, these questions allowed us to empirically examine how students’ commitments to particular ideas grouped together.

DATA COLLECTION

Student data collection involved two distinct groups, which we will refer to as the “intervention” and “comparison” groups. The intervention group was offered the Compassionate Canada curriculum unit, and completed pre- and post-unit surveys including the HAD questions. The comparison group received only the HAD survey.

Comparison group

The comparison group was composed of Grade 11 Social Studies students from two public schools: one in Vancouver, BC and the other in Mission, BC. The Vancouver students came from a school in a relatively affluent area and were largely university-bound, while many of the
The Historical Account Differences survey

Mission students were bound for vocational schools or the workforce. Together, these two schools enabled us to construct a sample that was comparable to the intervention group. No statistically significant differences were detected between the two groups with respect to parents’ education, academic self-concept, or plans for future schooling. See Table 1 for details on the two groups’ characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Comparison Group</th>
<th>Intervention Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students</td>
<td>26</td>
<td>39</td>
</tr>
<tr>
<td>Percent of females</td>
<td>46.2</td>
<td>43.6</td>
</tr>
<tr>
<td>Average hours spent on homework per day</td>
<td>3.1</td>
<td>1.9</td>
</tr>
<tr>
<td>Average hours per week spent working</td>
<td>7.4</td>
<td>8.4</td>
</tr>
<tr>
<td>Average Grade (Median on the rating of categorical statements)</td>
<td>“About half A and half B”</td>
<td>“About half A and half B”</td>
</tr>
<tr>
<td>Mother’s highest level of education (Percent Frequency)</td>
<td>High School</td>
<td>26.9</td>
</tr>
<tr>
<td></td>
<td>Bachelor’s Degree</td>
<td>11.5</td>
</tr>
<tr>
<td></td>
<td>Graduate Degree</td>
<td>3.8</td>
</tr>
<tr>
<td>Father’s highest level of education (Percent Frequency)</td>
<td>High School</td>
<td>15.4</td>
</tr>
<tr>
<td></td>
<td>Bachelor’s Degree</td>
<td>23.1</td>
</tr>
<tr>
<td></td>
<td>Graduate Degree</td>
<td>7.7</td>
</tr>
<tr>
<td>Percentage of students bound for university</td>
<td>69.2</td>
<td>66.7</td>
</tr>
</tbody>
</table>

Table 1: Demographic data for the two student groups

After securing ethical approval from each school district, teachers at comparison group schools were contacted through colleagues on the design team. At the invitation of teachers and administrators in these schools, the principal investigators distributed research permission forms. Ultimately, 26 of 78 potential comparison group participants contacted (approximately 33%)
The Historical Account Differences survey provided the required forms. Though the original research design had called for comparison group students to complete the survey twice, with the same time interval as the intervention group, access to participants proved to be limited in both of the comparison group schools. Ultimately, it was only possible to have the comparison group students complete the HAD survey once.

**Intervention group**

The intervention group, which participated in the *Compassionate Canada* unit, was made up of students from a single secondary school in Vancouver (different from that in the comparison group). As it is a relatively small school (approximately 500 students), all sections of the Grade 11 Social Studies course were recruited for participation. Eighty-seven students, all taught by the same teacher, were approached for research consent, of whom 41 provided both required forms. The overall consent rate for the intervention group was 47.13%. A handful of these participants did not provide all the necessary data for the analyses presented below, due to absence on survey administration days.

During implementation of the *Compassionate Canada* unit, each class session was observed by at least two members of the research team. Detailed analysis of the classroom observation notes lies outside the scope of this paper, but some initial analysis is provided elsewhere (O'Neill & Guloy, in press). More germane to our analysis here, students’ online activities in Knowledge Forum were logged automatically by the server computer, providing tallies of the numbers of notes (postings) written and opened by each student. At the conclusion of the unit, students completed a post-survey that included the HAD items, several questions about their interest in and satisfaction with the *Compassionate Canada* unit, and questions about the support they believed they had received from their assigned telementors.
RESEARCH QUESTIONS

In the sections that follow, we present a series of analyses aimed at answering three questions:

1) What do the HAD responses suggest about students’ conceptions regarding differing historical accounts?

2) Does HAD provide evidence of change in students’ metahistorical conceptions as a result of instruction?

3) Can HAD help us tease out which aspects of instruction influence students’ conceptions regarding differing historical accounts?

In what follows, we use data from the *Compassionate Canada* unit to illustrate the potential uses of HAD. To an extent, *Compassionate Canada* also provides opportunities for validating the HAD instrument. If students’ responses change significantly in the implementation group over time, and if these changes are associated with the opportunities for learning that theory suggests they should be associated with, it reinforces our sense that HAD is measuring what it is intended to measure.

DATA ANALYSIS AND RESULTS

In this section we examine the evidence with regard to each of our research questions, incorporating data from the implementation and comparison schools as appropriate. We used a significance level of alpha = .05 for our statistical tests.

1. **What do the HAD responses suggest about students’ conceptions regarding differing historical accounts?**

Both the HAD survey and the analysis that follows were designed to build upon Shemilt’s (1987) previous findings about adolescents’ metahistorical conceptions, while avoiding assumptions
The Historical Account Differences survey about whether these conceptions clustered in the coherent stages previously formulated. The first step in analyzing students’ responses to the survey was to perform a factor analysis on the rating scale items. Factor analysis is a statistical data-reduction technique based on examining the common variance among a set of variables (Williams, 1992). Common variance can be thought of as the apparent overlap between what two different measures are measuring. When applied to a large set of variables, factor analysis can produce a smaller set of abstract “factors” that reflect the underlying relationships in the data (Kachigan, 1991). In this case, we used factor analysis to examine whether students’ responses to 17 scaled-response questions on HAD cohered in the ways that would be predicted by Shemilt’s developmental theory.

The factor analysis was performed with SPSS 16 using Principal Axis Factoring and Varimax rotation with Kaiser normalization. We included in this analysis all 89 available student responses to the scalar items – both the comparison group’s responses and the intervention group’s pre- and post-unit responses. Both pre- and post-unit responses were included on the assumption that regardless of whether the participants had completed the survey once or twice, the underlying factor structure should be the same.

Results from this analysis are presented in Table 2. As the table shows, 13 of the 17 items included in the survey loaded on one of the four factors that emerged from our analysis. These factors did not map cleanly onto the stages described by Shemilt based on his interviews.
The Historical Account Differences survey

Reading the output of a factor analysis and labelling the factors is an interpretive process, not a formulaic one (Williams, 1992). The output tells the researcher that according to the input data, a set of variables have an underlying relationship (based on shared variance); but it does not say what this relationship is. Thus, interpreting a factor structure is a matter of theorizing about what the set of variables loading on each factor has in common, and how each set is distinct from the others. In naming a factor, it is common practice to emphasize the variable with the strongest factor loading. In this case, since we are working with students’ responses to a survey, we then theorize about how students may understand the whole set of variables to be related.

Table 2 provides details on the four factors that emerged from our analysis. Factor 1 loads students’ responses to four survey items – two of which were written to appeal to Stage 2 conceptions, one of which was written to appeal to Stage 1 conceptions, and one (the weakest loading) that was written to appeal to Stage 4 conceptions. The item with the strongest loading on this factor is “historians can figure out which stories about the past are biased or untrue, and put together the one true story” (emphasis added). Following this, we note that the second- and third-strongest loadings are on items that similarly imply a single, true account. One item mentions “the truth of what happened in the past,” while the other talks about evidence that might change “the story” (emphases added). The fourth, weakest-loading item, is more challenging to interpret together with the first three. We speculate that while it was written to capture the Stage 4 idea that questions about the past inevitably change, students interpreted the item to suggest that the historian’s job is to generate a new, single true story when others ask for it.

Factor 2 loads three variables, including one (the strongest loading) that was written to appeal to Stage 3 conceptions, and two that were written to appeal to Stage 4 conceptions. It appears that what unites these items from a student’s perspective are the ideas of investigation
The Historical Account Differences survey

and evidence. The strongest-loading item describes historians as “using evidence to figure out” what probably happened. The next strongest item mentions historians “investigating new questions,” and the weakest-loading item mentions historians “learning about” and “trying to understand” how authors looked at and felt about events. Based on these commonalities, we named this factor “investigation using evidence.”

Factor 3 loads three items – two of which were written to appeal to Stage 3 ideas, and one of which was written to appeal to Stage 2 ideas. In naming this factor, we noted that the two strongest-loading items both include the phrase “educated guess.” The third item talks about deciding on “the most believable” storyteller. Considering it together with the others, it seems that students interpreted this item to imply a kind of informed wager. Based on these interpretations, we decided that an appropriate name for this factor was “educated guess.”

Factor 4 loaded three items, two of which were written to appeal to Stage 1 ideas and one of which was written to appeal to Stage 2 ideas. The two strongest-loading items assert (for different reasons) that historians “should not write new books,” and the third asserts that “it does not make sense for historians to disagree.” Taken together, these items seemed to voice a frustrated assertion (in the face of the conflicting Rwanda accounts) that the past should be simple.

It is important to note that in none of the four factors was the strongest-loading item one that was written to appeal to Stage 4 conceptions. Shemilt’s most advanced “kaleidoscopic” conception of history is thus not visible in our factor structure. Since in Shemilt’s original work Stage 4 thinkers were found to be rare, it is possible that our sample simply did not contain enough Stage 4 thinkers for this idea to emerge in our factor structure. A substantially different or larger sample may have yielded different results.
The Historical Account Differences survey

The next step in our analysis was to examine the reliability (internal consistency) of the items loading on each factor. If a set of survey items is not sufficiently consistent, measurement specialists consider that these items cannot be interpreted as measuring the same underlying construct. A traditional measure of reliability, Cronbach’s alpha, was computed for the set of variables loading on each of the four factors, to evaluate the internal consistency of the respective items (see Table 2). Factor 4, while interesting, did not meet the generally-accepted threshold for reliability of .60, so this set of variables was excluded from further analysis.
### The Historical Account Differences survey

#### Table 2: Factors identified in the HAD survey responses

<table>
<thead>
<tr>
<th>Factor</th>
<th>Survey items</th>
<th>Mean Rating</th>
<th>Factor Loading</th>
<th>Researchers’ Interpretation</th>
<th>Reliability (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Historians can figure out which stories about the past are biased or untrue, and put together the one true story [Q2B] (Stage 2)</td>
<td>4.88</td>
<td>.689</td>
<td>One true story</td>
<td>.680</td>
</tr>
<tr>
<td></td>
<td>Historians have studied a lot about the truth of what happened in the past, and remember it [Q2A] (Stage 1)</td>
<td>4.83</td>
<td>.625</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Historians write new books because] new evidence is discovered that might change the story. [Q7A] (Stage 2)</td>
<td>5.66</td>
<td>.571</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Historians write new books because] as time goes on, new questions come up that historians need to answer [Q7D] (Stage 4)</td>
<td>6.00</td>
<td>.416</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Historians don’t necessarily know exactly what happened in the past, but can use evidence to figure out what probably did. [Q2C] (Stage 3)</td>
<td>5.08</td>
<td>.731</td>
<td>Investigation using evidence</td>
<td>.636</td>
</tr>
<tr>
<td></td>
<td>Historians investigate new questions that people ask about what happened in the past, and what it means today [Q2D] (Stage 4)</td>
<td>5.48</td>
<td>.630</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[If two stories disagree historians should] learn about the authors and try to understand how they looked at and felt about the events [Q8B] (Stage 4)</td>
<td>5.58</td>
<td>.462</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[If two stories disagree, historians should] make an educated guess about what likely happened, based on other evidence [Q8C] (Stage 3)</td>
<td>4.15</td>
<td>.679</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>[Historians know what caused an event in the past] by considering several possibilities and making an educated guess based on the evidence available for each one. [Q3B] (Stage 3)</td>
<td>4.54</td>
<td>.612</td>
<td>Educated guess</td>
<td>.639</td>
</tr>
<tr>
<td></td>
<td>[Historians know what caused an event in the past] by looking at what several people have said about what caused the event, and deciding who is the most believable [Q3C] (Stage 2)</td>
<td>3.54</td>
<td>.511</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Historians should not write new books about the same people and events that have been written about before, because the older ones are better. [Q7C] (Stage 2)</td>
<td>2.37</td>
<td>.618</td>
<td>Frustrated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Historians should not write new books about the same people and events that have been written about before. It makes no sense because there should only need to be one book. [Q7B] (Stage 1)</td>
<td>2.02</td>
<td>.606</td>
<td>assertion of simplicity</td>
<td>.581</td>
</tr>
<tr>
<td></td>
<td>[If two stories disagree, historians should] ignore both stories. It does not make sense for historians to disagree. [Q8A] (Stage 1)</td>
<td>2.12</td>
<td>.425</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Historical Account Differences survey

For reasons discussed earlier, we wanted to avoid simplifying students’ metahistorical conceptions artificially into a single stage. Thus, we continued our analysis using the notion of an “epistemic profile.” A student’s profile, in this analysis, is made up of three separate scores that can be thought of as reflecting the salience that the “one true story,” “evidence-based investigation,” and “educated guess” explanations have for her when asked to explain the reasons for different stories about past events. We computed an epistemic profile for each participant by averaging her responses to the items loading on each of the factors produced in our earlier analysis.

Epistemic profiles were computed for the implementation group’s pre-survey and post-survey responses, as well as for the comparison group’s responses. Figures 1-3 depict the implementation and comparison groups’ epistemic profiles prior to the Compassionate Canada unit.

![Histogram of students’ epistemic profiles on the “one true story” Factor](image)

Figure 1) Histogram of students’ epistemic profiles on the “one true story” Factor

As Figure 1 shows, both the implementation and comparison groups initially favoured the idea that there should be one true story about any past event. Mean scores on this factor were not significantly different between the two groups (comparison group mean = 5.38, intervention
The Historical Account Differences survey

group mean = 5.28) as determined by an independent samples t-test, where t= .412 and p= .662.
Two-tailed t-tests were used for this as well as the subsequent t-tests.

Figure 2) Histogram students’ epistemic profile on the “evidence-based investigation” Factor

As Figure 2 shows, both the implementation and comparison groups initially favoured the idea that historical accounts were developed based on evidence. The mean score on this factor was identical across the two groups (5.38), where t= .036 and p=.972.
Figure 3) Histogram of students’ epistemic profiles on the “educated guess” Factor

Figure 3 shows that both groups were less committed to what Shemilt would regard as the more sophisticated notion (relative to “one true story”) that historical accounts represent an educated guess about what most likely happened in the past, based on evidence that is never entirely complete. Again, no significant difference was detected between the test and comparison groups’ scores on this factor (comparison group mean = 4.19, intervention group mean = 4.03), where t=.485 and p=.629.

Thus, prior to the Compassionate Canada unit, the implementation group’s responses were similar enough to the comparison group that any change in the comparison group should be indicative of learning during the unit.

2. Does HAD provide evidence of change in students’ metahistorical conceptions as a result of instruction?

There is debate in the research literature as to how amenable students’ metahistorical ideas are to improvement through instruction. Gabella describes childrens’ naïve metahistorical ideas as “stubborn” and “resilient” (Gabella, 1994), while Lee (2004) contends that metahistorical ideas
The Historical Account Differences survey are not by nature more difficult to grasp than ideas dealt with in the “content” of history (e.g., feudalism). The present study cannot resolve this ongoing debate, but may be able to inform it.

As discussed in the previous section, three epistemic profile scores were computed for each of the participants in the Compassionate Canada implementation group, pre and post. Tests of significance were then performed between the pre and post scores for each of the profile variables.

Two-tailed tests on the three profile variables indicated a significant pre-post difference for Factor 3, “Educated guess” ($t(32)=-2.157$, $p=.039$). Recall that of the three factors resulting from our factor analysis of the HAD responses, this factor reflects the most sophisticated metahistorical conceptions. The difference between the pre and post means was 0.44 in the positive direction, meaning that students’ scores increased significantly on this variable over the course of the implementation. We interpret this to mean that the “educated guess” explanation for differences among historical accounts increased in salience for students over the course of the unit.

Since statistical tests can identify changes that are significant (most likely non-random) but small, it has become traditional in analyses like this one to compute an “effect size” as well, using a formula such as Cohen’s $d$. The effect size for Factor 3 was computed as 0.334, using Morris and Deshon’s (as cited in Bender, Kim, & Springer, 2007) calculation for $d$ adapted for paired sample t-tests. According to the typical interpretation of Cohen’s $d$ this would be considered a small effect; though in education, effect sizes are typically small even for highly successful interventions (Light, Singer, & Willett, 1990).

Subject to limitations we will discuss in our closing section, we interpret these results as evidence that HAD can be useful in measuring change in students’ ideas about varying historical accounts.
3. Does HAD help us tease out the aspects of instruction that influence students’ conceptions regarding differing historical accounts?

While the *Compassionate Canada* unit seems to have produced a positive outcome (especially for a short-term intervention), it is not sufficient in our view to demonstrate that an innovation leads to increased sophistication in students’ epistemic profiles overall. Like almost any curriculum unit, *Compassionate Canada* was a complex event involving many different potential learning experiences. Which among these actually mattered with respect to our educational goals? Answers to this question could lead to refinements in the design of the unit, in preparation for broader-scale implementation. Equally important, keeping in mind the possibility of Matthew effects (in which the most able students become more able, while the least able languish) (Stanovich, 1986), did the unit work better for students from privileged backgrounds than for other students?

To address the last question, we examined correlations between the pre-post changes in students’ epistemic profile scores and several survey variables reflecting their academic self-concept, their plans for future schooling, and their parents’ education. No significant correlations were found, suggesting that the outcomes of the unit were not biased by any of these attributes.

Our next set of analyses attempted to connect the changes in students’ epistemic profiles to measures of their engagement, and some of the learning opportunities offered them in the unit. We hypothesized that in the implementation group, student engagement would lead to reductions in the salience of naïve ideas like those clustered in Factor 1, and increases in the salience of relatively mature ideas clustered in Factor 3. Further, we expected that some types of guidance from mentors (such as alternate interpretations of source materials) would correlate significantly with reductions in Factor 1 scores and increases in Factor 3 scores. On the other hand, we did not
The Historical Account Differences survey expect to see positive change in metahistorical conceptions correlate with students’ reports that their mentors provided basic information about the historical cases they were studying.

To test these conjectures, we examined correlations between the implementation group’s pre-post change in epistemic profile scores and a wide array of participation variables, including the number of Knowledge Forum notes students wrote and opened, and students’ reports of the telementoring functions (forms of advice and guidance) they had received. No significant correlations were observed between these participation variables and pre-post changes on Factor 3, “educated guess,” but several negative correlations were observed between individual students’ change on the most naïve of the epistemic scales, “one true story” (Factor 1) and specific forms of advice and guidance they reported receiving from their telementors.

On the post-unit survey, implementation group students were asked to report the degree to which their mentors had been helpful to them in each of eleven different ways, by rating their mentors’ helpfulness on a scale from 1 (not helpful) to 7 (very helpful). The “telementoring function” questions from our survey, and students’ average ratings of them, are provided in Table 3.
Table 3: Telementoring functions asked about on the post-survey, and students’ average ratings of their mentors’ helpfulness with respect to each

<table>
<thead>
<tr>
<th>Mentoring Function</th>
<th>Average Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gave me background information on my topic</td>
<td>5.14</td>
</tr>
<tr>
<td>Reviewed my work as I went along</td>
<td>5.11</td>
</tr>
<tr>
<td>Offered other interpretations of source materials</td>
<td>5.00</td>
</tr>
<tr>
<td>Asked me questions to help me think about my work</td>
<td>4.96</td>
</tr>
<tr>
<td>Answered questions I had about specific people, events or ideas in history</td>
<td>4.91</td>
</tr>
<tr>
<td>Suggested challenging things for me to do that could improve my work</td>
<td>4.74</td>
</tr>
<tr>
<td>Suggested specific strategies that would help me get my work done</td>
<td>4.56</td>
</tr>
<tr>
<td>Helped me to understand material I read about my topic</td>
<td>4.52</td>
</tr>
<tr>
<td>Helped me to understand what historians do each day</td>
<td>3.93</td>
</tr>
<tr>
<td>Suggested books or library resources that I should examine</td>
<td>3.82</td>
</tr>
<tr>
<td>Gave me pointers to web sites with relevant materials</td>
<td>3.62</td>
</tr>
</tbody>
</table>

Four mentoring functions correlated with changes on Factor 1 (see Table 4). In line with our conjectures, these types of guidance are ones that might be expected to facilitate a reduction in naïve conceptions about historical knowledge. The first and second functions, “suggest sources” and “web pointers,” involve the telementor (a supposed authority) implying that one true story is not evident – that it is necessary to examine a variety of accounts. The third function, “offer other interpretations,” does the same even more forcefully. Finally, the “review work” function implies that there is a methodical element to understanding the past – that oversights occur or mistakes are made in constructing accounts, but that they can be corrected. Note that as expected, two functions involving the provision of basic information, “Provided background information on my topic” and “Answered questions I had about specific people, events or ideas in history” did not correlate with changes in naïve conceptions.
The Historical Account Differences survey

<table>
<thead>
<tr>
<th>Survey item</th>
<th>Spearman’s rho (one tailed)</th>
<th>Alpha level</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Suggested books or library resources that I should examine”</td>
<td>-.793</td>
<td>.01</td>
</tr>
<tr>
<td>“Gave me pointers to web sites with relevant materials”</td>
<td>-.784</td>
<td>.01</td>
</tr>
<tr>
<td>“Offered other interpretations of source materials”</td>
<td>-.514</td>
<td>.01</td>
</tr>
<tr>
<td>“Reviewed my work as I went along”</td>
<td>-.362</td>
<td>.05</td>
</tr>
</tbody>
</table>

Table 4: Significant correlations between reported mentoring functions and reduction of naïve conceptions

Given our understanding of conceptual change and metahistorical conceptions, we had expected students’ level of engagement with telementors (in terms of numbers of notes posted and read), or engagement of particular kinds (such as mentors offering alternative interpretations of source materials) to track reliably with the pre-post change on Factor 3. While we cannot offer a definitive explanation as to why this was not the case, we can speculate based on classroom observations. During sessions in the computer lab, students in the implementation group often wrote their online notes collaboratively under a single login name. On occasion they also read source documents and mentor notes collaboratively. Thus, the automated records of students’ note reading and posting may not reflect students’ online activity accurately enough to produce significant correlations.

DISCUSSION

Scholars of history education have developed a broad consensus that mature metahistorical conceptions – including sophisticated understandings of the nature of historical knowledge and where it comes from – are important. Substantial effort has been invested in the development of teaching innovations to address this need (e.g. Sandwell, 2005; Saye & Brush, 2002). To this point, however, there have been few attempts to gauge change in students’ metahistorical
The Historical Account Differences survey conceptions quantitatively. Especially in North America, the majority of scholarly innovations in history teaching appear to have been evaluated using qualitative methods exclusively.

We would argue that the lack of quantitative instruments for gauging students’ metahistorical conceptions makes it difficult for researchers and curriculum developers to refine large-scale innovations for history teaching in an evidence-based fashion. A curriculum of any sophistication contains many elements, some of which may be more important to informing students’ metahistorical conceptions than others. Even if a teaching innovation is demonstrably effective, it is useful to understand what its “active ingredients” are (Clark, 2009). Without this understanding, we dare not change it for fear of spoiling its efficacy.

To complicate matters further, some elements of a teaching innovation may have greater efficacy in one place than another, or with one part of the student population than another. Without effective means to gauge which particulars of the innovation actually help to address the curriculum designer’s goals, and with whom, we are hard pressed to understand how we can make our work equitable.

The HAD survey appears to hold some promise for addressing these problems, and we hope to further refine and disseminate the fruits of this effort. As discussed above, several of the new survey items discussed here were found to be reliable (internally consistent), and correlated with variables reflecting participation in our pilot curriculum. These correlations are suggestive of the validity of the instrument; but further validation work is needed on the individual survey items. In a future study, we plan to have practicing historians categorize each of the HAD questionnaire items according to the stages postulated in Shemilt’s developmental theory, blind of our own interpretation of them.

There are other important limitations to the assertions that can be made based on this study, due either to the study design or the fact that some of the planned data could not be
The Historical Account Differences survey collected. For example, given better access to the comparison group participants, we would have collected a second set of HAD responses at the same time interval as the implementation group. These data could potentially enhance the credibility of the instrument by indicating its “test-retest” reliability. In the absence of an instructional intervention, each student’s responses to the HAD questions should remain largely unchanged over time.

In a similar vein, it would have been ideal to collect re-test HAD data from the implementation group students a year or so after the implementation of the Compassionate Canada unit, to gauge the longevity of the effect reported on Factor 3. This was not possible for the present study, but may be in future work.

The HAD survey was conceived initially as a tool to help teachers and researchers examine and refine the efficacy of history curriculum; but beyond its potential value in this respect, data derived from HAD may have theoretical implications as well. The present study provided a small-scale opportunity to empirically evaluate whether the metahistorical ideas discussed by Shemilt (1987) occurred in our sample in the discrete stages originally postulated, or as Lee & Shemilt (2003) suggested may be the case, fell into a different pattern. Analysis of data from our sample of 65 11th-grade students did not entirely coincide with Shemilt’s original theory; however the robustness of the three-factor analysis presented here should be tested with a larger and more diverse sample of students. We hope to collaborate with like-minded colleagues at other institutions to carry out this work.
The Historical Account Differences survey

REFERENCES


The Historical Account Differences survey


The Historical Account Differences survey


The Historical Account Differences survey


APPENDIX A: HISTORICAL ACCOUNT DIFFERENCES SURVEY

This set of questions is designed to help us understand how you think about History and historians.

*There are no wrong answers. We are interested in your ideas.*

**Question 1**

What do you think historical accounts (books, documentaries, etc.) are useful for? (Rate your agreement with each statement.)

<table>
<thead>
<tr>
<th></th>
<th>Disagree Strongly</th>
<th>Agree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appreciating how good life is</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>today compared with the past</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keeping people together</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>by making them proud of the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>same things</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding how the world got</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>to be the way it is today</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teaching us the mistakes of the</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>past, so we don’t repeat them</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nothing – they’re boring and</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>don’t matter today</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question 2**

What makes somebody a historian? (Rate your agreement with each statement)

<table>
<thead>
<tr>
<th></th>
<th>Disagree Strongly</th>
<th>Agree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>They have studied a lot about the</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>truth of what happened in the past,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and remember it</td>
<td></td>
<td></td>
</tr>
<tr>
<td>They can figure out which stories</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>about the past are biased or untrue,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and put together the one true story</td>
<td></td>
<td></td>
</tr>
<tr>
<td>They don’t necessarily know exactly</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>what happened, but can use evidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>to figure out what probably did</td>
<td></td>
<td></td>
</tr>
<tr>
<td>They investigate new questions that</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>people ask about what happened the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>past, and what it means today</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Question 3**

How do historians know what caused an event in the past (for example, the internment of Japanese Canadians during the Second World War)?

<table>
<thead>
<tr>
<th></th>
<th>Disagree Strongly</th>
<th>Agree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>They consider who had a motive to make it happen and examine evidence as far back as they’ve decided the story needs to go</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>They consider several possibilities and make an educated guess based on the evidence available for each one</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>They look at what several other people have said about what caused the event, and decide who is the most believable</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>They read only what has been written or recorded about the event by the people who were actually there when it happened.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
</tbody>
</table>
Question 4

Below are three published accounts about the crisis in Rwanda, Africa in 1994, in which thousands of civilians were killed by their neighbours. Please read each story and answer the questions.

Story A: Published in 2004

In Rwanda the United Nations is doing its utmost to help people recover and reconcile. …In Tanzania, a United Nations criminal tribunal continues to pursue the main perpetrators of the genocide. The tribunal has handed down pioneering verdicts: the first conviction for genocide by an international court; the first to hold a former head of government responsible for genocide; …the first to find journalists guilty of genocide -- because they helped create the state of mind in which thousands of people could set aside the most fundamental moral instincts of all human society, and embark on the wholesale massacre of fellow human beings. With these and other steps, the United Nations is doing what it can to help Rwanda find a path to lasting security and peace, with itself and its neighbours.

According to this story, who was to blame for the deaths in Rwanda?:

________________________________________

Story B: Published in 2000

For 150 years, the outside world played a central part in carving out the building blocks that built to the genocide. This role extended way back: to the racism of the first European explorers, to Belgian colonial policy; to Catholic church support for “demographic democracy” under a Hutu military dictatorship; to the Structural Adjustment Programme imposed by western financial institutions; and to the legitimizing of an ethnic dictatorship by France, the US, and many international development aid agencies. In our very strong view, the world carried a heavy responsibility for the events in Rwanda. There was an honourable and useful way in which the world might have discharged that responsibility. Human rights groups and a small number of UN officials tried frantically to get it to do so. Instead, world leaders chose to play politics and to pinch pennies as hundreds of thousands of innocent Rwandans needlessly died.

According to this story, who was to blame for the deaths in Rwanda?:

________________________________________

Story C: Published in 2003

The United States knew exactly what was happening on the ground in Rwanda. Once in a blue moon, Americans would even share some information. …I received information that a friendly foreign power (we later confirmed that it was the United States through the American ambassador in Nairobi) had received information that I was to be assassinated in the next few days and recommended that I bring adequate security with me when I had to leave the compound. The assigned troops were the biggest, baddest, meanest-looking gentlemen I had ever seen, especially when their eyes were hidden by sunglasses, which was more or less all the time. I guess I should have been grateful for the tip from the Americans, but my larger reaction was that if delicate intelligence like this could be gathered by surveillance, how could the United States not be recording evidence of the genocide occurring in Rwanda?

According to this story, who was to blame for the deaths in Rwanda?:

________________________________________
The Historical Account Differences survey

**Question 5**

Which of the stories you just read do you think is the most believable one? (Check one box.)

- [ ] Story A
- [ ] Story B
- [ ] Story C

**Question 6**

Why is this story the most believable one?
The Historical Account Differences survey

Question 7
Over time, historians have written many books on the same people and events (for example, Canadian confederation, the War of 1812). Why do you think historians write new books about events that were already written about before? (Rate your agreement with each statement)

<table>
<thead>
<tr>
<th>Statement</th>
<th>Disagree Strongly</th>
<th>Agree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>New evidence is discovered that might change the story</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>They shouldn’t because there only needs to be one book</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Historians should not write new books because older ones are better</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>As time goes on, new questions come up that historians need to answer</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Older historians made mistakes in weighing the evidence. These need to be corrected</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
</tbody>
</table>

Question 8
If a historian is learning about the events of a period and finds two stories about them that disagree, what should she do? (Rate your agreement with each statement.)

<table>
<thead>
<tr>
<th>Statement</th>
<th>Disagree Strongly</th>
<th>Agree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>She should ignore both stories. It does not make sense for two accounts to disagree</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Learn about the authors and try to understand how they looked at and felt about the events</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Make an educated guess about what most likely happened, based on other evidence</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Try to figure out which author is less biased or was closer to the events, and use only that one</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
</tbody>
</table>