



UvA-DARE (Digital Academic Repository)

Promoting written historical reasoning among undergraduate L2 students

Sendur, K.A.

Publication date
2021

[Link to publication](#)

Citation for published version (APA):

Sendur, K. A. (2021). *Promoting written historical reasoning among undergraduate L2 students*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

CHAPTER 3

HISTORICAL REASONING IN AN UNDERGRADUATE CLIL COURSE: STUDENTS' PROGRESSION AND THE ROLE OF LANGUAGE PROFICIENCY

In a study of undergraduate L2 students participating in a Content and Language Integrated Learning historical reasoning course, we examined students' changing performance on historical reasoning and how this was affected by their English reading and writing proficiency. Students engaged in written historical reasoning when answering a historical question by using sources and heuristics such as historical contextualization and corroboration. The course was designed based on principles likely to enhance historical reasoning and second language acquisition, and included an overt focus on language form. We found that students were able to reason historically at an emerging level of proficiency. The markers of historical reasoning present in their writing were neither significantly flawed nor highly proficient. A latent growth curve analysis was used to investigate the effect of students' English language proficiency on their source-based writing and the changes in their reasoning over the duration of the course. We found that students' English proficiency level did not predict either their reasoning or changes in their performance during the course. Students at differing levels of English proficiency improved similarly.

Published as: Sendur, K.A., van Drie, J.P., van Boxtel, C. & Kan, K.J. (2020). Historical reasoning in an Undergraduate CLIL course: Students' progression and the role of English proficiency. *International Journal of Bilingualism and Bilingual Education*, 1-17,
Doi: <https://doi.org/10.1080/13670050.2020.1844136>

INTRODUCTION

The European Union has identified multilingualism as an educational goal (Eurydice, 2006). One way of achieving this goal is through Content and Language Integrated Learning (CLIL), in which students learn content through an L2 (Coyle et al., 2010). CLIL is offered in secondary schools in many European Union countries, particularly in the social sciences and sciences (Dallinger et al., 2016; Eurydice, 2006; Lorenzo, 2017) and can also be found in universities (Myskow & Ono, 2018).

The complexity of writing in history, particularly in an L2, may necessitate an explicit focus on language form and production, so students can fulfill the expectations of genres often used in history. Writing in history typically involves integrating multiple sources to construct an interpretation supported by the available evidence (Monte-Sano, 2010). Furthermore, it involves markers of historical reasoning, such as the use of historical contextualization (Van Drie & Van Boxtel, 2008). The language of historical reasoning, however, is challenging. The text structure, such as the use of nominalization, and archaic vocabulary in historical writing makes reading challenging (Berkeley et al., 2016; Martin, 1991; Wineburg & Martin, 2009), while source-based writing can be difficult because of the demands when citing (Shi, 2012) and integrating multiple sources (Cumming et al., 2016).

History education researchers have investigated middle and high school students' historical writing (De La Paz, 2005; Monte-Sano, 2010). Cognitive apprenticeship and explicit instruction are promising instructional approaches for teaching historical reasoning and writing at the middle (De La Paz et al., 2017) and secondary levels (Monte-Sano, 2008; Stoel et al., 2017). However, much of the research on students' writing from history education focuses on L1 students, and less is known about how L2 students perform with similar tasks and whether cognitive apprenticeship is an appropriate model for L2 students as well. This is

important to investigate since the complexity of writing in history may be likely to pose a challenge, particularly for students writing in an L2.

L2 students' history writing has also been studied from a language perspective. Studies have explored markers of historical literacy in students' writing (Lorenzo, 2017), as well as university students' biographical (Myskow & Ono, 2018) and argumentative writing (Mitchell & Pessoa, 2017). These studies provide insight into L2 students' use of language to write about history. Studies from the perspective of history education are also needed to better understand L2 students' performance and inform research into instructional approaches that promote both content and the academic language needed in a successful bilingual education program.

This descriptive study focuses on undergraduate Turkish students' performance when composing source-based writing in English during a CLIL historical reasoning course. We investigate changes in students' historical reasoning over the duration of the course and if performance is influenced by language proficiency.

THEORETICAL FRAMEWORK

HISTORICAL REASONING IN SOURCE-BASED WRITING TASKS

History is an interpretative discipline rooted in evidence. Historical writing serves multiple functions, such as the novel interpretation of evidence or a response to others' interpretations (Nokes & De La Paz, 2018). In their writing, historians construct an argument based on their interpretation of available evidence (Carr, 1990; Stanford, 1986/1995). When considering which evidence to use and how to frame it, historians employ sourcing, contextualization and corroboration heuristics (Wineburg, 1991).

Coffin identified three genres in secondary school history writing: recording, explaining and argument (2006). Argumentation, the most cognitively

complex of these genres, is typically seen as the most privileged. In history education, students often use argumentation in the context of source-based writing.

In history, when students write a source-based argument they (may) engage in historical reasoning. Van Drie and Van Boxtel (2008, p. 89) define historical reasoning “an activity in which a person organizes information about the past in order to describe, compare, and/or explain historical phenomena.” Researchers have investigated individual components of historical reasoning in students’ writing, primarily in students’ L1s. We focus on the following components featured prominently in the historical reasoning course under study: the use of evidence from sources, historical contextualization, and heuristics such as sourcing, and corroboration.

Students’ conceptions of evidence in history differ, varying from those approach sources as verifiable information to others who understand that evidence must be contextualized (Lee & Shemilt, 2003). When writing, some students are able to use accurate and persuasive evidence, but other struggle with forming a correct interpretation, possibly because of problems when reading sources (Monte-Sano, 2010). In a comparison of eighth and eleventh grade student, older more competent writers included more sources than younger poorer writers (De La Paz et al., 2012). Writers with higher scores used more argumentative strategies in the writing, such as the use of examples. Nokes et al. (2007) reported that while many students used sources as evidence when writing, there was little change after instruction.

In historical contextualization, the writer considers when and where the source was produced (Wineburg, 1991). This appears to be a lesser-used heuristic in student writing, and not as prone to improvement compared to other features of historical reasoning. When writing, some students were able to incorporate information to help the reader better understand the historical time period (Monte-Sano, 2010). Others, however, included inaccurate evidence or an

interpretation contrary to the writer's intent. De la Paz et al (2012) also found that better writers were able to use contextualization, in contrast to poorer writers. Van Drie et al. (2015) found that students in general scored low on contextualization in writing an argumentative text. The use of contextualization across all intervention groups in another study was so low as to prevent analysis (Nokes et al., 2007).

When reasoning about history, historians consider who has created the source, which may take the form of assessments of reliability or usefulness (Wineburg, 1991). Some high school students' historical writing included limited levels of sourcing, including references to authors or documents (Monte-Sano, 2010). Others failed to note information about the source; none assessed reliability. The quantity and quality of sourcing in eighth grade students' writing varied greatly (Nokes, 2017). While most students do not spontaneously include sourcing in their essays, high school and university students can include more source-related information with instruction (Britt & Aglinskias, 2002).

Historians corroborate by comparing multiple sources for similarities and differences (Wineburg, 1991). Students can corroborate to some extent. High school students used evidence from multiple documents to support a claim (Monte-Sano, 2010). While students used corroboration less than sourcing, students in Nokes et al's (2007) study trained in a corroboration heuristic using multiple texts scored higher on the corroboration posttest than those who studied textbooks. Better writers in de la Paz et al's (2012) study also used corroboration in their writing.

These studies demonstrate that (primarily L1) students from middle school to university can use features of historical reasoning in their writing. Some features, such as historical contextualization, appear to develop less with instruction. Nokes and de la Paz (2018) conclude that writing in history benefits students in terms of knowledge transformation, the development of content knowledge and literacy skills.

THE POTENTIAL ROLE OF READING AND WRITING PROFICIENCY IN SOURCE-BASED WRITING

Reading sources may be challenging because of the abstract nature of historical writing. While history textbooks are not overly technical, features such as nominalization make comprehension challenging (Martin, 1991). Texts in history can also make interpretation difficult for students, for example, by hiding the author's interpretation (Unsworth, 1999). In addition to its abstract nature, organizational structures and genres in textbook writing also prove challenging. When history is presented as a list of facts, students find it difficult to understand how information forms a coherent narrative (Schleppegrell et al., 2004). Textbooks tend to include multiple genres, particularly recording and explaining genres, increasing the variety students encounter (Coffin, 2006). The diversity of genres and linguistic features in primary sources provides an even more significant challenge than secondary sources.

Source-based writing is a cognitively demanding task, but the role of language proficiency is not entirely clear. Some features of students' writing may be independent of language, such as the evidence students choose (Keck, 2014) and overall text quality in source-based argumentative writing (van Weijen et al., 2019). This has implications for bilingual education if source-based writing instruction in an L2 can impact similar writing tasks in an L1. Students' proficiency may depend partly on their prior knowledge in the discipline and their experience in writing (Cumming et al., 2016). Novice L1 and L2 undergraduate writers both paraphrased more extensively when writing a summary than experienced L1 and L2 students (Keck, 2014).

It is possible that writing may be affected by students' L2 proficiency. When focusing on features such as grammar uses up working memory, the overall message can be affected (Schoonen et al., 2009; Weigle, 2002). Language proficiency may also affect how students write with sources. L2 students may be

more likely to stay close to the original text, possibly because of difficulty in comprehension (Cumming et al., 2018; van Weijen et al., 2019). This can be problematic in history, in which original interpretation is valued. Despite these challenges, recent studies have shown that L2 students at the secondary and university level can produce writing with features of historical reasoning, such as the use of evidence and elements of historical contextualization (Lorenzo, 2017; Miller et al., 2014; Myskow & Ono, 2018).

Instruction appears to influence students' historical reasoning, but there have been inconsistent findings regarding the role of students' reading and writing proficiency. All students, including the small percentage of L2 students, showed improvement from one disciplinary writing intervention, but proficient readers benefitted more (De La Paz et al., 2017). In contrast, students in another writing-focused study improved their historical reasoning after instruction, but there was no difference between students with high and low writing proficiency scores (Van Drie et al., 2015) or between students with differing levels of reading proficiency (De La Paz et al., 2014). In a reading-focused intervention, all students improved, particularly struggling readers (Reisman, 2012a).

Given the complexity of reading and writing with historical sources, and the conflicting findings regarding the influence of reading and writing proficiency on students' historical reasoning, it is important to consider the potential influence of students' reading and writing proficiency when studying how students write about history with sources.

CLIL IN THE HISTORY CLASSROOM

Because of its focus on writing, interpretation and argumentation, history is well-suited for CLIL. With additional instructional time, students in a CLIL history class were able to make similar gains in content knowledge, and comparatively larger gains in listening compared to their non-CLIL peers (Dallinger et al., 2016).

Humanities courses, such as history, also tend to involve more speaking opportunities for students (Lo, 2014), building both content and language skills. Because of the complexity of reading and writing in history, however, a CLIL history course may need an explicit focus on language.

In their 4Cs Framework, Coyle et al. (2010) emphasize the integration of *content* and *communication* together with *cognition* and *culture* as principles for teaching a CLIL course. Based on Westhoff's (2004) formulation of best practices in second language acquisition, de Graaff et al. (2007) developed an observational tool for CLIL teachers that focuses on the use of meaningful and comprehensible input, an overt focus on language form, pushed output and strategic learning. While these works consider language an important aspect of CLIL pedagogy, de Graaff et al. (2007) found that teachers rarely focused on language form. In other studies as well, CLIL teachers also had a greater emphasis on content than language (Morton, 2010; Oattes et al., 2018). It would therefore be valuable to study how students progress in their historical reasoning in a CLIL course with a more overt focus on language.

Written historical reasoning in history education is well-studied among various age groups, but has rarely focused on L2 students or CLIL contexts. When included, like in Nokes et al (2007) and de la Paz et al (2012), the percentage of the population is often low and proficiency levels are not included. It is therefore difficult to draw conclusions about the impact of language proficiency on students' written historical reasoning, and whether instruction works similarly. Language may be a factor in L2 students' historical reasoning and writing, though. In an earlier study (in press) of undergraduate L2 students, we found that when language interfered with reading comprehension of a primary source, their reasoning also had errors. We therefore believe it would be valuable to investigate the written historical reasoning of L2 students, whether their historical reasoning changes with instruction and the role of language proficiency.

RESEARCH QUESTIONS

Our research question is: Does historical reasoning in L2 students' writing improve over the duration of a CLIL historical reasoning course and is the level and improvement in historical reasoning influenced by reading and writing proficiency?

We would expect students' historical reasoning to improve over the duration of the course. Students at higher levels of English proficiency may demonstrate higher levels of historical reasoning and more change in their writing than those at lower levels.

METHODOLOGY

PARTICIPANTS

Sixty-three non-native English speaking undergraduate students in a pre-university intensive English program at a small English-medium university in Istanbul, Turkey participated in this study. Eight students were excluded because of an incomplete dataset or because they did not complete the course. See Table 1 for student demographics. Most students complete one to two semesters of intensive English before beginning undergraduate coursework, and are around 18 to 19 years old. Approximately 45% of students major in engineering. Students were at the B2 level of the Common European Framework of Reference for Languages. Data collection took place during a CLIL historical reasoning course that serves as a bridge between the intensive English program and a series of required history courses.

Table 1

Student Gender and Intended Area of Study (N=55)

Gender		Intended Area of Study		
Male	Female	Arts and Social Sciences	Engineering	Management
35	20	5	46	4

Note. The university does not offer an undergraduate degree in history.

HISTORICAL REASONING COURSE

This CLIL course had the dual purposes of teaching historical reasoning and English. Over eight weeks, (four hours weekly, 32 hours total), students were incrementally introduced to the historical reasoning components source evaluation, argumentation, corroboration and historical contextualization (Van Drie & Van Boxtel, 2008; Wineburg, 1991). Argumentation focused on claim and evidence as features of disciplinary writing (Monte-Sano, 2010). See Appendix D for a lesson overview. Students learned historical reasoning through explicit instruction using a cognitive apprenticeship model (Britt & Aglinskas, 2002; Collins et al., 1991; De La Paz et al., 2017; Stoel et al., 2017). Cognitive apprenticeship is a model of explicit teaching in which students learn from expert models with substantial support in the form of coaching and scaffolding, which is withdrawn as students demonstrate competent performance (Collins et al., 1991).

Students studied gladiators during the late Roman Republic and early Empire. Units included socioeconomic class, politics and cultural values. During the lessons, students read and discussed sources produced during the Roman Empire (primary sources) and recent textbook and journal articles (secondary sources) about gladiators and their time. Students wrote a document-based question (DBQ) using the provided sources after each unit. In a DBQ, students answer a historical question using a set of sources, typically both primary and

secondary. This genre is used in history classes as well as in history education research (McCarthy Young & Leinhardt, 1998; Monte-Sano, 2010).

The course focused on language identified as best practices in CLIL (Coyle et al., 2010; de Graaff et al., 2007; Westhoff, 2004). To help focus on meaning, secondary source readings were modified, and students were taught to annotate them to improve comprehension and identify questions. Primary sources were modified using Wineburg and Martin (2009, p. 214)'s guidelines. Specifically, we shortened sources, simplified the presentation with white space around the primary source, and simplified the vocabulary and structure to a level appropriate for B2 learners. When reading primary sources, students completed a graphic organizer and a series of guiding questions intended to promote comprehension and the following aspects of historical reasoning: source evaluation, historical contextualization and corroboration. For example, before reading an excerpt of Suetonius, students completed a matrix that scaffolded the process of source evaluation. Next, students labelled an amphitheater diagram to illustrate where different socioeconomic classes sat during gladiator games, and identified background information necessary to help them contextualize the primary source. Later, students discussed the extent of corroboration between primary sources. To facilitate English production, classes were conducted in English with frequent opportunities for speaking, writing and feedback on language production.

A focus on form included two aspects: the concept of an argument as a genre in history (Morton, 2010) and language for writing an argument. When learning to write an argument, for example, students learned about claims using a set of criteria. First, they practiced distinguishing between claims and statements using the criteria, and later revised their own claims based on the same criteria. Students also learned about grammatical aspects helpful in argumentation, such as citation language and connectors to help structure the argument (Coffin, 2006; Lorenzo, 2017). In our instruction we distinguished between writing about

primary and secondary sources. For example, when writing about artifacts such as portraits of gladiators, students learned to first describe features of the art, and then use hedging language to make a tentative interpretation of its significance. Finally, we also provided language models for aspects of historical reasoning, including claim, evidence, corroboration, source evaluation and historical contextualization. For example, “this evidence seems to suggest that...”

The course was designed by the first author, but was taught by six instructors from the intensive English program. Instructors must meet minimum education and experience requirements, and use a highly standardized curriculum. Since CLIL courses are typically taught by content specialists, the English language instructors used highly scripted lesson plans to ensure that the course was carried out as intended. The course coordinator tracked instructor’s weekly self-reports and reported that all required lesson components were completed as expected. Extensive support materials were available to instructors, including annotated sample DBQ answers.

DATA SOURCES AND ANALYSIS

ASSESSMENT OF READING AND WRITING PROFICIENCY IN ENGLISH

Students’ proficiency in reading and writing in English was assessed with an integrated task completed during the first historical reasoning lesson but before instruction took place. This task type is increasingly used to measure reading and writing proficiency, partly because it is more representative of how university students read and write (Leki & Carson, 1997; Weigle et al., 2013) and because it is an appropriate measure of both reading and writing (Shin & Ewert, 2015; Weigle, 2004; Weigle et al., 2013). The task was closely based on a previously validated reading and writing proficiency test for university-level L2 students (Weigle, 2004; Weigle et al., 2013). The task, on adolescent competency in digital literacy, was developed by the first author. Question types were identical to those

in Weigle (2004)'s study. Students' English instructors evaluated the reading passages and questions to ensure they were an appropriate measure of reading and writing for B2 level students. Students completed two separate tasks: a reading comprehension task based on two readings and a writing task that required the use of the information in both readings using a question similar in structure to those in the course.

The tasks were assessed using the rubrics in Weigle's study (2004). The writing rubric was slightly simplified in consultation with an English language instructor because of the brevity of the tasks. Students' answers were scored as in Weigle's (2004) study. Each answer set was scored by two trained raters using the rubrics and a set of scored samples with scoring explanations. The sample scores were determined by the first author and an English language instructor. If the scores were more than 1 point apart, a third rater was used. Interrater agreement within 1 point was 98% for reading, 86% for writing (rhetoric) and 91% for writing (language). The final reported score is the average of the two closest raters' scores.

One reading score and two writing scores are reported: rhetoric and language. The rhetoric score is the average of the content and organization scores from Weigle's (2004) rubric. The language score is the average of the rubric's accuracy, and range and complexity scores.

DOCUMENT-BASED QUESTION WRITING

Students completed an in-class DBQ after finishing each of the three units. See Table 2 for the questions and Appendix E for an example of a primary source. All DBQ questions fell within the explaining genre (Coffin, 2006), which can be used to explain complex causes and consequences. Students could consult the sources and language models while writing. Students had 50 minutes to write each essay with 10-20 minutes of required planning. Instructors provided feedback on

the historical reasoning and academic English of the first drafts of DBQ1 and DBQ2 using a simplified version of the analytical rubric (described below). After this, students completed a final draft, which was also assessed using the rubric. No feedback was given for DBQ3. We analyzed the first drafts of each DBQ.

Table 2

Document-based Questions, Available Sources, and Timing

Document-based Question prompt	Assigned essay length	Available Sources	When the essay was completed
DBQ1: Describe Rome’s social hierarchy and explain one way it affected people’s lives.	120-150 words	2 primary sources 3 secondary sources	Lesson 3
	250-300 words	3 primary sources 3 secondary	Lesson 5
DBQ2: Why did rich Romans sponsor gladiator games?			
DBQ3: How did Roman society view gladiators? Explain two views.	250-300 words	6 primary sources 3 secondary sources	Post-Lesson 8

HISTORICAL REASONING ANALYTICAL RUBRIC

A 5-point analytical rubric was developed to assess the extent to which students incorporated the following features of historical reasoning in their writing: claim, evidence, source evaluation, historical contextualization and corroboration. In this rubric, a score of 1 or less indicates that students either did not include these features or that there were significant errors. A 2 indicates an attempt with minor errors or of limited support. Scores in the 3 to 4 bands indicate appropriate historical reasoning, with those scoring a 4 showing greater proficiency. See Appendix F for the rubric. These features were a part of the course and are consistent with features of argumentation and historical reasoning found in student writing in history (Monte-Sano, 2010; Van Drie & Van Boxtel, 2008). An analytical rubric was chosen over a holistic rubric since students’ historical reasoning abilities develop separately (Lee et al., 2001; Monte-Sano, 2010). An

analytical rubric better conveys differences in students' abilities and tracks changes over the three DBQs.

We developed the rubric by first consulting how others have measured historical reasoning and argumentation in student writing (Monte-Sano & De La Paz, 2012; Nokes, 2017; Van Drie et al., 2015). Using a structure similar to Monte-Sano and De La Paz (2012), we added the categories of claim and corroboration based on the aforementioned literature, and adapted the features within categories to correspond to how we taught students historical argumentation. The first and second authors then scored a random sample of students essays with differing levels of performance in historical reasoning, after which they discussed differences in scoring and amended the rubric to both conform to the literature and the sample dataset.

The first and second authors coded two rounds of nine students' three DBQs, 27 essays per round, each round of which was approximately 15% of the total sample. The primary difference in coding was determining the extent of the claim. During the second round, Pearson's r was determined as between .74 and .95 for the five categories. Evidence and source evaluation corresponded to the lowest and highest correlations, respectively. The first author coded the remaining essays.

To analyze the progression students made between DBQ1 and DBQ3, we used paired samples t-tests. To analyze the role of language proficiency on historical reasoning and progression throughout the course, a latent growth curve analysis was conducted (Raykov & Marcoulides, 2012). We fitted an intercept slope model on the DBQ scores (DBQ1, DBQ2, and DBQ3) with Reading, Writing Rhetoric, and Writing Language as indicators of a single latent covariate (predictor of both intercepts and slopes), interpreted as proficiency in English.

RESULTS

First we present the extent to which students include features of historical reasoning in their DBQs and whether these change during the course. We provide further analysis of students' performance for the historical reasoning component source evaluation, since in a previous study (in press) sourcing remained at an emerging level of proficiency, with some students including sourcing that undermined their argument. We therefore wanted to investigate how the quality of students' sourcing changed over three DBQs. Finally, we analyze the role of students' language proficiency on their historical reasoning and progression throughout the course.

HISTORICAL REASONING IN STUDENTS' DBQ WRITING

In DBQ1, mean scores fall below 2 and in DBQ2 and DBQ3, mean scores fall below 3. See Table 3 for students' scores and Appendix G for a sample DBQ. There is individual variation between scores. Claim was the highest score in DBQ1; source evaluation and historical contextualization were either absent or flawed. Students made progress in both source evaluation and historical contextualization, and there was less of a difference with the other categories in DBQ3. However, these two categories remained the lowest-scoring features in DBQ3. DBQ1 scores in claim, evidence and corroboration began as flawed or with emerging proficiency, but DBQ3 scores approached appropriate historical reasoning. There were statistically significant differences between students reasoning in DBQ1 and DBQ3 in all areas we measured.

Table 3

Results of t-test and Descriptive Statistics for Aspects of Historical Reasoning in Students' Document-based (DBQ) Writing (N=55)

	DBQ1		DBQ2		DBQ3		DBQ1-DBQ3 <i>t</i>
	M	SD	M	SD	M	SD	
Claim	1.85	1.16	2.33	1.14	2.51	1.23	-3.31 *
Evidence	1.51	.54	2.49	.77	2.58	.90	-8.29 *
Source Evaluation	.87	1.32	1.18	1.35	2.11	1.20	-5.66 *
Historical Contextualization	.64	1.04	1.93	1.12	2.38	1.23	-8.89 *
Corroboration	1.35	1.34	1.84	1.34	2.78	1.41	-5.98 *
<i>Total</i>	<i>6.22</i>	<i>2.67</i>	<i>9.76</i>	<i>3.01</i>	<i>12.36</i>	<i>4.01</i>	<i>-11.88 *</i>

Note: For each DBQ, $df=54$, * $p < .01$

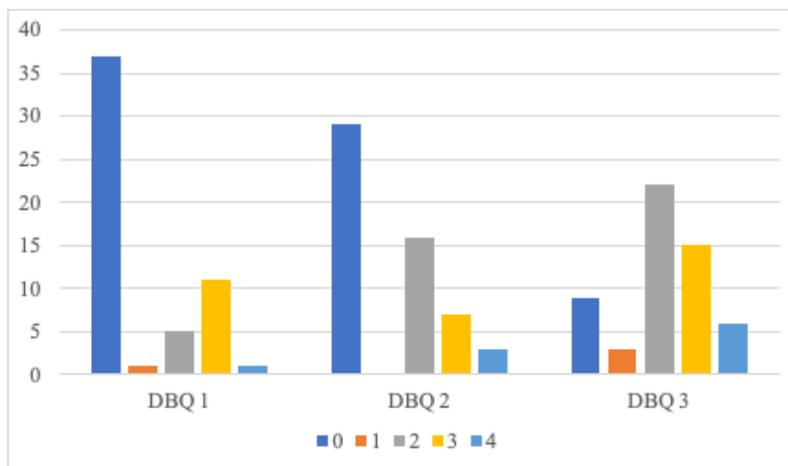
USE OF SOURCE EVALUATION ACROSS DBQs

Figure 1 shows a small number of students included an evaluation of a primary source in DBQ1. In contrast, most students scored a 2 or 3 in DBQ3; few scored a 4. In most of the 2 scores, students included information about the source as well as an evaluation of reliability based on the information, which is in line with our definition of proficient sourcing. As in our previous study, although these students assessed the source as unreliable they continued to use the source as evidence, thereby undermining their argument. For example: “Politicians spent their money to gladiator games. So they may probably want to show positive effects of gladiators. For this reason this source (Cicero) may not be reliable (Student 6).” The number of students who included this type of an evaluation rose both in numbers and proportion over the course. In DBQ1, only two students included this type of an assessment, but in DBQ3 17 of 22 students with a score of 2 included a negative assessment. This type of an assessment seems to

indicate that students are taking the first steps towards a more proficient source evaluation.

Figure 1

Frequency of Scores 0-4 for the Historical Reasoning Characteristic, Source Evaluation, for Each DBQ (N=55)



ENGLISH PROFICIENCY

Table 4 shows that students scored relatively poorly on both reading and writing. The test was scored on each aspect from 1 (low) to 5 (high). In this assessment, a score of 2 or 3 indicates that students are able to partially or minimally fulfill the criteria for the assigned tasks, but with errors. The difference between students is low, indicating that students were at similar levels. This is expected since all students had been placed at the B2 level.

Table 4

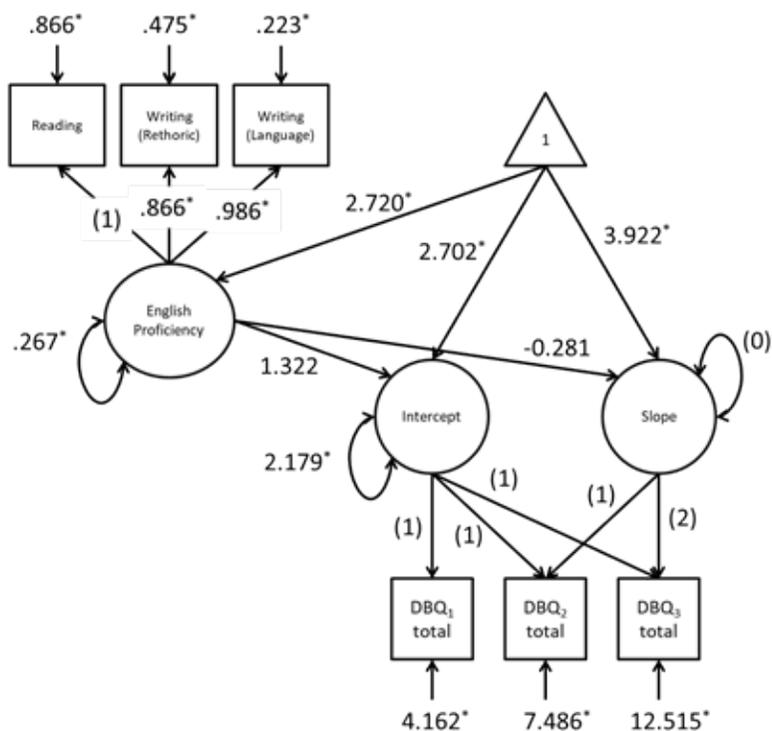
Means and Standard Deviations for Students' Reading and Writing Proficiency Assessment (N=55)

	Mean (SD)
Reading	2.75 (.98)
Writing: Rhetoric	2.33 (.89)
Writing: Language	2.68 (.70)

The results of the latent growth curve analysis on students' historical reasoning are provided. As, initially, the variance in the slope was estimated negative, we fixed this variance to 0. We thus assumed, eventually, that individuals have different initial scores possibly due to differences in English proficiency, but show the same increase in DBQ score. Because the variance in the slope was 0, the covariance between intercept and slope was also fixed to 0. This adapted model fit the data well ($\chi^2 = 16.4524967$, $df = 12$, $P = 0.17$). The parameter estimates and their significance are presented in Figure 2.

Figure 2

A Graphical Representation of the Intercept-slope Model Fitted on the Three DBQ Scores. Fixed Parameters in Parentheses. Significant Effects are Denoted with an Asterisk ()*



From Figure 2, one can obtain that English Proficiency did not have any significant effect on historical reasoning (intercepts) or the growth herein (slope). Therefore, we concluded that English Proficiency could be left out the model, so we ended up with a standard intercept slope model. This model demonstrated good fit as well ($\chi^2 = 0.881$, $df = 3$, $P = 0.83$). According to this model, the initial historical reasoning score was estimated as 6.308 on average ($CI_{95} = [5.638, 6.979]$) and increased with 3.156 points ($CI_{95} = [2.636, 3.675]$) over every new assessment. This increase turned out to be significant ($t = 11.900$, $p < .001$).

DISCUSSION AND CONCLUSIONS

We examined the extent to which undergraduate L2 students are able to use historical reasoning in their writing and whether students' reasoning changes over the duration of a CLIL historical reasoning course that explicitly integrates language and content knowledge. We also investigated the relationship between students' historical reasoning in a DBQ task and their level of English proficiency.

As expected, we found that undergraduate L2 students are able to include features of historical reasoning when writing DBQs with primary and secondary sources. These findings support studies with students in middle and secondary school contexts (De La Paz et al., 2017; Monte-Sano, 2010), the secondary school CLIL context (Lorenzo, 2017) and add undergraduate L2 students to the literature on historical reasoning and writing. While students made significant gains in all areas, there is room for improvement in their historical reasoning. Students struggled particularly with source evaluation and historical contextualization, similar to that found by others (Nokes, 2017; Nokes et al., 2007). These findings demonstrate that students appear to need significant practice in these areas in order to use them in disciplinarily-appropriate ways.

Students made early gains in claim and evidence, and then less progress later in the course. In contrast, students showed stronger gains in corroboration, source evaluation and historical contextualization between DBQ2 and DBQ3. This may stem from the order of teaching, as claim and evidence were taught first, giving students more opportunities to practice. The individual primary sources may have played a role in students' historical reasoning as well, but the improvement in students' scores combined with the decrease in instructor support make this less likely. It is also possible that since these are features of both general argumentative and history-related writing, that they were easier for students. This is at least partly supported by students' higher scores in these areas in DBQ1. In contrast, domain-specific features may have posed a greater

challenge for these students. It is also possible that students needed additional instructional time, similar to that found by Dallinger et al. (2016).

We found that English proficiency has no predictive value on students' historical reasoning in their writing or on students' progress in the course. Students at differing levels of English proficiency benefitted from the course similarly. This is similar to others who did not find a difference between those with different levels of writing (Van Drie et al., 2015) or reading proficiency (De La Paz et al., 2014), but in contrast to a study in which struggling readers particularly benefited (Reisman, 2012a). We propose three primary reasons for this finding. First, as B2 level students, their English proficiency was relatively similar. Studies with diverse levels of English proficiency may have resulted in different findings. Secondly, we focused on scoring historical reasoning, which may also have affected the extent to which English proficiency influenced students' scores.

Most importantly, our instructional decisions may have enabled students at different levels of English proficiency to make similar gains in historical reasoning. The use of a cognitive apprenticeship model has been effective when teaching historical reasoning and writing to diverse groups of learners, including L2 students, (De La Paz et al., 2017), and may also have been effective for our students. In our instruction, we focused on meaning through modified readings and graphic organizers, and engaged students in speaking and writing. Notably, we included an explicit focus on form to help students write arguments and used English language specialists as CLIL instructors. While there may have been variations between these instructors, our scripted instruction makes this less likely. This seems to indicate that the combination of the cognitive apprenticeship and CLIL models may show promise as an instructional approach for L2 students.

In sum, this descriptive study demonstrates that undergraduate L2 students can reason historically when writing, and that their reasoning improves during a CLIL course. This is an important contribution given the lack of studies

that focus on content in the CLIL context (Cenoz et al., 2013). Our finding that students' English proficiency does not predict historical reasoning scores seems to indicate that our instructional choices, notably a cognitive apprenticeship and CLIL model, may enable students with different levels of English proficiency to benefit similarly from the course, and should be considered as an instructional approach for CLIL history courses.

The descriptive nature of this study limits the conclusions we can draw about the impact of the instructional model. Future research in which different approaches are compared could provide further insight into how this model affects L2 students' historical reasoning and writing.