

The Effectiveness of Computer-Supported Collaborative Dialogue in Schools



Baruch Schwarz
The Hebrew University of
Jerusalem, Israel



Ziv Lison
The Hebrew University of
Jerusalem, Israel



Noa Brandel
The Open University of Israel,
Israel

Abstract

Effectiveness has received bad press in Dialogic Education as it generally points to improvements in a stable educational context, while dialogic pedagogies aim at educational change. The present paper examines effectiveness in senses that are compatible with both the educational system and the aims of Dialogic Education. Institutional and organizational constraints limit the implementation of practices such as peer-led computer-supported collaborative learning (CSCL) discussions in schools. In an intervention that handled these constraints, we compared the outcomes of small-group peer-led CSCL discussions around texts and teacher-led whole-class face-to-face discussions around the same texts. The outcomes compared were individual historical essays. We relied on a methodological tool developed by Monte-Sano (2016) to measure two qualitative aspects of students' essays: argumentation and historical reasoning. Two groups of high-school students ($N_1 = 24$; $N_2 = 21$) participated in discussions around historical texts. The second author, who was the group's teacher, facilitated a discussion with the first group. The second group was divided into small groups, which interacted through a CSCL tool. There was no significant difference in the quality of essays written after participation in a peer-led dialogue through a CSCL tool, compared to those written after participation in a teacher-led face-to-face discussion – in terms of either argumentation or historical reasoning. However, the students who participated in peer-led collaborative dialogue did reveal distinctive epistemological insights. These results have had implications for the Ministry of Education's policy, which recently recognized this measure of essay quality as part of the history matriculation exam. In addition, the analysis of the peer-led CSCL group discussions suggests that a dialogical space was created. The intervention was thus effective in terms of standardized learning outcomes, dissemination of new dialogic practices, and ontological-dialogical changes.

Baruch Schwarz is a Professor (Emeritus) at the School of Education, the Hebrew University in Jerusalem. His expertise ranges from educational psychology to the development of mathematical/scientific reasoning. He is a specialist in the role of argumentation in learning and development. He has led several R&D European projects on the use of technologies to boost deliberative argumentation, collaborative learning, mathematical problem solving, and “Learning to Learn” skills. He is also involved in research on the moderation of (computer-supported) collaborative learning. He is the director of the Israeli Center of Dialogic Education. He is active in the study of learning in Yeshivas (Schools of study in Jewish education).

Ziv Lison is a PhD candidate at the Hebrew University of Jerusalem's School of Education and a pedagogical leader at the Israeli Ministry of Education's Pedagogical Secretariat. His research focuses on applying Jürgen Habermas's principles of communicative action to dialogic pedagogy to mitigate social polarization and address structural power imbalances in intergroup discourse. Lison leads a nationwide initiative integrating Civics education with "Shared Life" education, designing (AI-assisted) alternative assessments based on synchronous dialogues between students from diverse sectors (e.g., secular and religious, Jewish and Arab). His work strives to transform "objectified" relations into inclusive, humanizing, argumentative dialogue within conflictual environments.

Noa Brandel is a postdoctoral fellow at the Research Center for Innovation in Learning Technologies at the Open University of Israel. She holds a PhD in Linguistics from Tel Aviv University, where she was awarded several scholarships, including the Rotenstreich Scholarship for outstanding PhD students in the Humanities. Noa previously completed a postdoctoral fellowship in Educational Psychology at the Center for Dialogic Education at the Hebrew University of Jerusalem and worked as a visiting scholar in the Section of Theoretical and Applied Linguistics at the University of Cambridge. Her research lies at the intersection of theoretical and applied linguistics, focusing on processes of language learning and instruction in the areas of computer-supported collaborative learning (CSCL) and (digital) dialogic pedagogy, with an emphasis on the assessment of learners' proficiency and interaction using AI-based tools. She aims to contribute to the design of effective instructional practices across diverse educational contexts.



Introduction

'Effectiveness' in educational policy is often synonymous with standardized learning outcomes. Accordingly, in a meta-analysis of the effects of CSCL in STEM education, based on 143 studies published between 2005 and 2014, Jeong et al. (2019) found that the overall effect of STEM CSCL was moderately positive. This effect was measured through individual tests administered after collaborative learning. Another sense of effectiveness concerns the extent to which CSCL practices change the educational system. Wise and Schwarz (2017) have noted the relative absence of *impact* of CSCL practices on educational systems, in spite of the positive results obtained in controlled studies. Indeed, one rarely hears about the massive implementation of CSCL practices in schools or universities. This rarity is explainable: The implementation of CSCL practices demands the replacement or modification of old practices, and these pose significant organizational and institutional challenges, designated as *systemic changes* (Arnseth & Ludvigsen, 2006) or *synergy of national policy, classroom practice, and design research* (Looi et al., 2011).

Our inquiry adopts the two senses of effectiveness just mentioned: It checks the quality of individual outcomes after collaborative work and shows that the results led policy makers to favor CSCL practices at a level of policy. The current study further relates to an additional sense of effectiveness, grounded in a broad ontological aim: the creation of a *dialogic space* (Wegerif, 2007). In this space, the goal is not merely the transfer of historical facts, but a 'collective search for truth' (Bakhtin, 1984), where differences in perspective are not obstacles to be overcome but resources for meaning-making. We posit that the 'unguided' nature of CSCL is not a deficit in instruction but a structural necessity that opens space for peer-led collaborative dialogue. By removing the teacher's immediate authority, the digital environment forces students to engage in 'chaining', i.e., linking their ideas directly to one another without the mediation of an authoritative third party. Thus, while we measure individual outcomes to demonstrate systemic feasibility and to convince policy makers to disseminate CSCL practices, our deeper pedagogical interest lies in checking whether the autonomous spaces created allow students to turn history – the domain of the current study – into a polyphonic dialogue rather than a monologic narrative. It is precisely this threefold conception of effectiveness that we examine in the present study.

The COVID-19 pandemic has opened opportunities for reconsidering this multifaceted view of effectiveness. Indeed, the pandemic led to several changes in secondary education, including greater integration of technology, a shift towards personalized learning approaches, reevaluated assessment practices, increased collaboration and community engagement, and enhanced professional development for educators (UNESCO, 2022). These changes seemed to create opportunities for integrating CSCL practices into educational systems. In a series of papers in the *International Journal of Computer-Supported Collaborative Learning*, Järvelä and Rosé (2020) introduced the editorial note entitled “Advocating for group interaction in the age of COVID-19”, recognizing these opportunities:

“we as a global society reached out online for family, friends, and colleagues, and rediscovered that *people are everything*. At the same time, we have entered into a worldwide process of mediated sensemaking and search for a path forward. In the midst of all this, we have taken the leap to a reality of online instruction shared by students, teachers, parents, administrators as well as researchers” (p. 143).

The four papers of the special issue focused on studies of regulation strategies and new methods for scaling up design and evaluation, “both of which are critical themes at this time in the history of our field” (Järvelä & Rosé, 2020: p. 144). Indeed, in a paper entitled “*Does collaborative learning design align with enactment? An innovative method for evaluating alignment in the CSCL context*,” Zheng et al. (2020) evaluated the alignment between an online collaborative learning design and its enactment. In the same special issue, Saqr (2020) offered a methodology for analyzing collaboration processes at the scale of a whole community using Social Network Analysis (SNA). The paper provided practical guidance for the application of SNA in terms of the selection of appropriate network representations as well as metrics. In a nutshell, the studies in the special issue had potential practical implications. However, the realization of these potentialities was beyond the scope of the studies.

The lack of research on the effectiveness of CSCL practices in educational institutions is understandable, as the obstacles to be surmounted are immense: The CSCL vision aims to replace old practices with new ones. This replacement demands, among other things, the frequent use of personal computers in the classroom, the design of collaborative-learning tasks, and the frequent grouping of students into small groups. The replacement of old practices also involves the radical change of the role of teachers in collaborative learning. Thus, the abundant literature on orchestration is important (see, for example, Matuk et al.'s 2019 special issue on real-time orchestrational technologies in CSCL). Moreover, Dillenbourg, Järvelä, and Fischer's (2009) summary of research on the integration of CSCL activities into larger pedagogical scenarios, which include activities that must be orchestrated in real time, is still relevant. Yet, this summary, which provides an interesting list of recommendations, does not address the challenges inherent in facing the multiple constraints involved in the actual integration of CSCL practices into educational structures.

We claim that one of the most difficult challenges in the integration of CSCL practices into educational structures is ideological. Educational institutions have always represented authority, and whole-class teacher-led instruction in educational institutions represents the authority of educational systems. In contrast, one of the tenets of the CSCL pedagogical approach is to leave the group to operate autonomously. In CSCL literature, group guidance, when necessary, must be adaptive. Comparing the effectiveness of CSCL settings with traditional settings is inappropriate for CSCL theorists, as their adherence to collaborative practices is ideological.

We provide here a study that handles the compatibility of CSCL practices with educational structures. We focus on an assessment tool acceptable by both the educational system and researchers committed to the CSCL agenda, which is also a dialogical agenda. This paper tells the story of the

acceptance of the assessment tool by the educational system, and of the effectiveness of a CSCL practice as measured by this tool.

The assessment tool evaluates the quality of written history essays. Monte-Sano (2016) developed this tool by defining criteria and levels of argumentation and historical reasoning in such essays. So far, this tool has been used for research purposes. The second author negotiated with the Israeli Ministry of Education regarding the use of essays written after participation in a dialogic activity as a tool for assessing achievement in history for the matriculation exam. The Ministry agreed that the score given for the written essay would form part of the overall score of the matriculation exam. The superintendent in history recognized that high-level historical essays are suitable for measuring historical knowledge as part of the matriculation exam. Her agreement originated from her conviction that the tool complies with the official directives included in the history syllabus, which stress the importance of reasoning and sourcing. The second author was also aware that the Ministry of Education favors guidance on high-quality talk across all disciplines. Fostering accountability to the other, to knowledge, and to reasoning, Michaels et al. (2008) called this talk *Accountable Talk* and articulated moves teachers should enact in classroom talk to foster it. Other researchers proposed variants of this quality talk (e.g., Hennessy et al., 2020) and showed that such talk leads to higher learning outcomes than outcomes following classroom talk in which the teacher has not been trained to facilitate high-quality talk (Howe et al., 2019). Previous research has indicated that high-level talk improves the quality of argumentative essays. For example, Kuhn et al. (2016) showed that dialogic argumentation leads to high-level writing (see also Hemberger et al., 2017). Decision-makers in the Ministry of Education are aware of some of these studies and currently instigate professional development programs for improving the quality of classroom talk among in-service teachers.

While the Ministry views dialogue instrumentally – as a means to produce correct historical reasoning, we view the CSCL environment as an ontological space where students negotiate identity and meaning. Our use of Monte-Sano's (2016) rubric aligns with the Ministry's instrumental goals for the purpose of systemic integration, yet we acknowledge that this rubric prioritizes 'finalized' arguments over the 'unfinalizable' and messy process of dialogue (Bakhtin, 1984) often found in the chat logs.

The present study was conducted during the COVID-19 pandemic, when many students remained at home. In this context, electronic channels enabled the assignment of learning tasks to small groups. This assignment, which responded to an organizational necessity, had intrinsic advantages as it provided a written record of the exchange that served as a reflection tool during and after interaction (Kuhn et al., 2008, 2013). Fostering disciplinary engagement and dialogic argumentation in small groups, peer-led collaborative dialogue turned into a natural systemic move. The literature about design principles for this fostering became highly relevant (Andriessen & Schwarz, 2009; Engle & Conant, 2002).

It is in this highly propitious context that the present study was conducted. Three opportunities were opened. First, the educational system recognized that the clear argumentative and reasoning aspects of history essays could be measured and that these measures could partly account for the matriculation exam grade. Secondly, the educational system recognized the importance of high-quality classroom talk and that teachers should be trained to facilitate it. Thirdly, CSCL tools provided a solution to the problem of student attendance during the pandemic, enabling the division of classes into small groups to complete learning tasks.

Description of the research

Research questions

In light of our goal to inspect the effectiveness of CSCL-based instruction in educational infrastructures, we first aim to compare learning outcomes following participation in small-group peer-led

CSCL discussions vs. a teacher-led whole-class face-to-face discussion around a controversial historical issue. Our first research question thus asks:

1. Is the quality of learning outcomes different for learners participating in small-group peer-led CSCL discussions around a controversial historical issue, compared to the quality of learning outcomes for learners participating in a teacher-led whole-class face-to-face discussion around the same issue?

Our second question explores epistemic components involved in learners' participation in such discussions – an exploration that may explain the results of the aforementioned comparison:

2. Do peer-led e-discussions lead to distinctive epistemological insights?

In order to answer the first question, which involves the comparison of learning outcomes, we assigned an individual writing task in both conditions: a post-discussion historical essay. We measured the quality of the essays in terms of argumentation and historical reasoning. To address the second question, we used a questionnaire in which students reflected on their discussions and on their subsequent writing.

We had no clear hypothesis about the answer to the first question. Of course, in a renowned article, Kirschner and colleagues (2006) argued that direct instruction is preferable to autonomous exploration for problem solving and conceptual learning, but many have disputed their claim. Non-comparative studies on well-designed student-led argumentation have yielded conceptual change (e.g., Asterhan & Schwarz, 2007; Schwarz et al., 2000). Moreover, the setting of CSCL group discussions provides indirect guidance and boosts agency, and may thus help students in small groups to improve the quality of their essays in terms of argumentation and historical reasoning. Research on the role of teachers in CSCL small-group argumentation has shown that it is not beneficial for conceptual change, in contrast with student-led CSCL group argumentation (e.g., Schwarz et al., 2011). As for the second research question, which involves asking students for feedback on their participation in the discussions, Bächtold et al. (2022) studied the importance of feedback on argumentative activities around socio-scientific issues for fostering epistemic cognition. We thus hypothesized that students participating in a group discussion would gain more from their reflection on their discussion.

Methodology

Population

The intervention was carried out in a high school in Israel that offers a special program for students interested in science or the arts, while providing a common humanistic basis for the students' specialization. High-level students are directed to one of four majors: music, humanities, life sciences, and art. The program in each of these majors extends beyond the material covered in the official matriculation exams. Although the school is highly selective, acceptance does not rely solely on tests and previous scores, but also considers the suitability of the candidate, their motivation, their curiosity, and their specific skills. The school is a boarding school (although 25% of the students do not live in the school), and the suitability of students as inmates in a boarding school is considered as well. Accordingly, each student is assigned to several groups in their major. Anchor classrooms include students who chose different majors but still attend the same general courses.

The experiment was conducted within the framework of the study of history as part of a general training in humanities, which is mandatory for students from all majors. Although the content covered was entirely determined by the school, it included a substantial part of the official matriculation of the Israeli Ministry of Education. Student evaluations were independently conducted by the teaching staff.

Forty-five 11th-grade students participated in the experiment, recruited from two classes: 24 from Class 1 (10 males, 14 females), and 21 from Class 2 (10 males, 11 females). Most of the students are secular Jews. Five students were observant Jews. Two were Muslims.

The study utilized a quasi-experimental design dictated by existing school structures. Students were not randomly assigned to conditions; instead, intact classes were assigned to either the face-to-face or CSCL condition. This ecological validity comes with the limitation of potential pre-existing differences, which we controlled for using the student dedication covariate.

The second author served as the teacher. He has been teaching history for 15 years and is knowledgeable about Dialogic Teaching. Although he did not undergo special training in facilitating whole-class discussions, his self-taught skills in facilitating discussions enabled him to design professional development programs to help teachers use different types of historical sources in their teaching, especially in discussions.

Program Development

The *Design-Based Research* paradigm was adopted to develop an educational program. Accordingly, the second author designed the learning environment and implemented it through iterative experimental cycles. He relied on the general ideas that CSCL tools may promote group learning, and that collaborative argumentation-based learning environments are propitious for conceptual learning and for forming reasoned opinions about controversial issues (Baker et al., 2019). However, the incorporation of such activities and environments into the school routine is a challenging enterprise (Wise & Schwarz, 2017). We do not describe here the cycles of development and experiments we undertook. The study took place in the last cycle. The second author convinced the principals of his school that CSCL tools and discussions can be used extensively for teaching the history of modern Israel, in which controversial issues are extremely common. He negotiated with the Ministry of Education the possibility of relying not only on exams, but also on essays in history to provide a reliable assessment of students' performance in the history class. To do so, he provided criteria by which the quality of essays in history can be assessed.

The program combined discussions around multiple historical sources, which took place on a CSCL tool. The teaching of history in Israel raises a great public interest. Teachers are expected to instill identity and belonging, to foster technological skills as well as textual literacy, and to turn learning into an emotional and intellectual experience. They are supposed to go deep into texts, but at the same time present them in an accessible way to weak students. A reform was planned to create a synthesis across disciplines and to encourage students to achieve learning outcomes. From an ethical vantage point, it aimed to foster intellectual flexibility and attentiveness to the opinions of others, as well as to promote social equity, tolerance, and solidarity. From an epistemological point of view, the reform aimed to develop students' autonomy of thought, critical thinking, and expression of arguments based on ethical values. These aims were articulated to face a problematic political situation, in which public discourse is not deliberative – a situation worsened by the common use of social networks in which alarming behaviors (shaming, bullying, cursing) are common.

Dialogym is a platform developed by the Center for Dialogic Education to create a secure digital space for learners' discussions. Before discussions, students are trained to comply with proper dialogic behaviors and values through individual learning units built into the platform.

Two teaching units were developed and taught in Grade 11. The first unit dealt with the memory of the Holocaust, while the second concerned the Warsaw Ghetto Uprising. In both units, students were provided with texts and were invited to analyze them. Throughout this study, we distinguish between *direct*

instruction and what we term *peer-led collaborative dialogue*. We reject the notion that the absence of a teacher equates to a lack of guidance. Instead, we view the specific constraints of the CSCL environment – the written instructions, the threaded interface of the Dialogym platform, and the presence of conflicting historical texts – as 'indirect guidance'. These elements provide a scaffold that delegates agency to the learners. Therefore, when we compare this to teacher-led settings, we are not comparing 'instruction' versus 'lack of instruction,' but rather centralized, synchronous authority versus distributed, asynchronous autonomy.

To understand the comparative context, it is necessary to examine the interactional dynamics of Class 1 (Teacher-Led). Although the teacher used Accountable Talk moves (e.g., asking students to revoice others' claims or to provide evidence), the discourse followed a 'hub-and-spoke' pattern. The teacher remained the central mediator, and students' responses were largely directed toward him rather than toward one another. While this ensured a high adherence to historical accuracy and minimized diversions from the subject matter, it limited the opportunity for the 'exploratory talk' (Mercer, 2000) observed in the CSCL group. In the face-to-face setting, the temporal pressure of the lesson meant that only a few students could verbally articulate a full argument, whereas the parallel processing of the CSCL environment allowed multiple 'voices' to be written simultaneously, creating a denser, albeit messier, web of peer-to-peer interaction.

The questionnaire

Our questionnaire followed the approach elaborated by Bächtold et al. (2022), which provides feedback on epistemic cognition among students. The questionnaire included the following four questions:

- (1) Have you noticed a link between the writing of the essay and the discussion?
- (2) What makes an argument solid?
- (3) What makes a historical argument solid?
- (4) How do I evaluate my arguments in the historical essay – the justifications, the evidence, and the examples?

Like the essay, the questionnaire was designed to achieve a specific educational objective: fostering epistemic cognition (Barzilai & Chinn, 2018). We should stress that our evaluation of the essays relied on Monte-Sano's (2016) terminology, which emphasizes *claims*, whereas in the questionnaire, we asked students about their *arguments* in order to encourage them to focus on the justifications they used to support their claims.

Procedure

In the unit on the Warsaw Ghetto Uprising, the students first learned about the chronology of the revolt and about the central events it involved. They then received two mandatory sources from a newspaper presenting conflicting views from researchers and writers about the role played by the right-wing Jewish Military Union (JMU)¹ in both the history and collective memory of the Warsaw Ghetto Uprising, Arens and Blatman. (The students also received several non-mandatory sources and could choose whether to use them). They read the sources alone or in small groups. Then, in Class 1, a whole-class face-to-face discussion was conducted under the guidance of the teacher. In the discussion, the teacher, who had been trained in the past in enacting dialogic moves in his teaching, interacted with students about the various motives of the participants in the Warsaw revolt. While interacting with his students, the teacher encouraged them to present their reasoned viewpoints on the motives of the participants in the Warsaw Ghetto Uprising

¹ The Warsaw Ghetto Uprising involved two main resistance groups. The JCO (Jewish Combat Organization) was a socialist-Zionist alliance. The JMU (Jewish Military Union) was associated with the right-wing Revisionist movement. Historically, the socialist narrative dominated the collective memory in Israel, often marginalizing the role of the right-wing JMU. The JMU's original name is in Polish: Żydowski Związek Wojskowy, and ŻZW is often the acronym used to call this organization. 'Jewish Military Union' is the translation of the Polish name.

and invited them to offer alternative arguments to those presented so far. In Class 2, students were divided into small groups of 3-5 and discussed similar issues related to the Warsaw Ghetto Uprising, but their discussions took place in a digital setting: the Dialogym platform. The groups were given the following written instructions: (a) Give your reasoned viewpoint about the motives of the participants in the Warsaw Ghetto Uprising, and (b) React to your group members by expressing an alternative argument to those expressed by them. In fact, the written instructions coincided with the instructional moves the teacher enacted in the face-to-face whole-class discussion in Class 1. However, unlike Class 1, which met in person in the classroom, Class 2 operated remotely due to COVID-19 restrictions. Students logged into the Dialogym platform from their homes. The platform mediated their interaction entirely; they did not use video conferencing (Zoom) for synchronous communication. This asynchronous, text-based environment meant that 'time on task' differed structurally: While Class 1's discussion was bounded by the 45-minute class period, Class 2's discussion extended over a 48-hour window, allowing for what Howe et al. (2005) term 'incubation' periods between turns.

Regarding assessment in Class 2, a distinction was made between the process and the product. Students were graded only for participation, not for the content of their interactions on the Dialogym platform. This decision preserved the digital space as a slightly more 'playful' zone (Wegerif, 2007), where tentative ideas could be tested without fear of evaluation. The platform served not merely to finalize thinking steps, but to mediate the interaction by slowing down the conversation, allowing students to read, re-read, and compose responses at a pace impossible in a verbal debate. Grading was reserved solely for the individual post-discussion essays, in alignment with the school's need for individual accountability.

Three weeks after the discussion, students in both classes were asked to write an essay about the Warsaw Ghetto Uprising. In the essay, the students were required to articulate two reasons for the differences in the purposes of the revolt across the two sources. Students from both classes were asked to capitalize on the discussions they participated in. After writing the essays, all students were asked to complete the questionnaire.

Data analysis

We adapted the methodology developed by Monte-Sano (2016) and De La Paz et al. (2017) for evaluating the quality of historical essays, and incorporated Seixas's (2006) ideas concerning the measurement of historical reasoning, as well as his methods for analyzing historical sources. These methods were chosen as they combine the assessment of the quality of argumentation and of historical reasoning. Each qualitative aspect (argumentation and historical reasoning) was evaluated based on five criteria with a scale ranging from 0 to 4. These methods were also chosen because they align with the goals of the Israeli Ministry of Education (Zohar & Boshrian, 2020).

The quality of argumentation was assessed based on five criteria:

- Quality of claim: Is the claim clear, reasonable, and relevant to the issue?
- Evidence: Does the student use convincing and relevant evidence to substantiate their claim?
- Causal chaining: Do the evidence, examples, and justifications form a causal chain supporting the claim?
- Absence of mistakes: Does the student include wrong elements in their claim?
- Coherence of writing: Is the writing clear and coherent?

Historical reasoning was also evaluated based on five criteria:

- Sourcing: Are the sources thoroughly analyzed, does the interpretation stick to the source, and is it reasonable? Are the sources well-triangulated?
- Historical significance: Is the student aware of the choices and meaning-making that are inherent to the study of history (for example, how grand narratives differ from inclusive history)? Does the student understand the historical significance and express it in their writing?
- Historical causality: Does the student understand the concept of historical causality and integrate it properly in their writing?
- Historical perspective: Does the student use historical knowledge in order to create links and understand the historical perspective of the actors?
- Continuity and change: Does the student understand the chronology, and is it included in the claim? Is the context created by choosing certain events on the time axis reasonably used and grounded?

In order to exemplify these criteria, we present here two essays, with their scores for argumentation and historical reasoning split into two tables for each essay. The essay appears in the first column for all tables. The second column presents the essay scores for argumentation and historical reasoning (reported separately), along with the excerpts on which they were based. The third column justifies the scores. Tables 1 and 2 display the essay of one of the students, Nala, and her scores for argumentation and historical reasoning, respectively.² Note that grading is a necessity dictated by rigid institutional constraints, while it fails to capture the full richness of the dialogic experience.

Table 1. The essay of student Nala and her scores in argumentation

Essay	Scores in Argumentation	Explanation
<p>There is a difference between the sources in the description of the goals of the rebellion. The first source presents the goal of the rebellion mainly as warfare, to help fight against the Nazis and kill the enemy. On the other hand, the second source shows the goal of the rebels is to choose how to die and prove that they are not weak. The main reason for the difference between the description of the goals of the rebellion in the two sources is the views of the people writing these passages. The first section was</p>	<p><u>Quality of claims: 4/4</u> <i>"There is a difference between the sources in the description of the goals of the rebellion. The first source presents the goal of the rebellion mainly as warfare, to help fight against the Nazis and kill the enemy. On the other hand, the second source shows the goal of the rebels is to choose how to die and prove that they are not weak."</i></p> <p><u>Evidence: 4/4</u> <i>"Armia Kariova and therefore their interest is to show that the right thing to do right now is to fight against the Nazis. To prove that even the Jews in the ghettos who are in very difficult conditions are trying to fight them. In</i></p>	<p>Uses previously learnt knowledge and skilled sourcing to base a solid claim</p> <p>Analyzes nuances in the text to base Claim + Properly uses periodization (place and time)</p>

² All names used in the paper are pseudonyms to protect participants' anonymity.

<p>written by a military underground, Armia Kariova, and therefore their interest is to show that the right thing to do right now is to fight against the Nazis. To prove that even the Jews in the ghettos who are in very difficult conditions are trying to fight them. In the passage, they refer to the JCO [Jewish Combat Organization], which they support. But these are not the interests of the JMU [Jewish Military Union], so the goals it presents as the goals of the rebels are slightly different. The revisionist party is in Israel and therefore it wants to show that the "young people of Israel" are the ones fighting in the ghetto, they want to emphasize that the Jews are not weak and that they can fight even in the worst conditions.</p> <p>Apart from the different views, two significant differences explain the difference between the goals that the two sources wrote. The first difference is the place where the sources were written, the first was written in Poland, near the place of the rebellion and they had little communication with the rebels in the ghetto. The second source was written in the Land of Israel and therefore the distance together with the ideology it reflects influenced the goals they present. Another difference between the two sources is the time when they were written, the first source was published on</p>	<p><i>the passage, they refer to the JCO, which they support. But these are not the interests of the JMU, so the goals it presents as the goals of the rebels are slightly different. The revisionist party is in Israel and therefore it wants to show that the "young people of Israel" are the ones fighting in the ghetto, they want to emphasize that the Jews are not weak and that they can fight even in the worst conditions."</i></p> <p><u>And:</u> <i>"Another difference between the two sources is the time when they were written, the first source was published on 29.4.1943 while the second source was published months later ... depending on the period in the war and the situation in the country has a great influence on the goals presented in the sources."</i></p> <p>Causal chaining: 4/4 <i>"Armia Kariova and therefore their interest is to show that the right thing to do right now is to fight against the Nazis. To prove that even the Jews in the ghettos who are in very difficult conditions are trying to fight them. In the passage, they refer to the JCO, which they support. But these are not the interests of the JMU, so the goals it presents as the goals of the rebels are slightly different. The revisionist party is in Israel and therefore it wants to show that the "young people of Israel" are the ones fighting in the ghetto, they want to emphasize that the Jews are not weak and that they can fight even in the worst conditions."</i></p> <p><u>And:</u> <i>"Another difference between the two sources is the time when they were written, the first source was published on 29.4.1943 while the second source was published months later ... has a great influence on the goals presented in the sources."</i></p> <p>Absence of mistakes: 4/4</p>	<p>for evidence; Ties the way the different sources name the fighters to their agenda: Refers to JCO - specifically because of their affiliation and to "young people of Israel" to include everyone accordingly</p> <p>Shows connection between the sources' agenda to nuances in text + Uses periodization forming causal relationship.</p>
---	---	---

<p>29.4.1943 while the second source was published months later on 8.10.1943. Time also has a great influence on what is written in these sources and depending on the period in the war and the situation in the country has a great influence on the goals presented in the sources.</p>	<p><u>Coherence of writing: 4/4</u></p>	<p>No mistakes</p> <p>Well written</p>
--	--	--

Table 2. The Essay of student Nala and her scores in historical reasoning

Essay	<u>Scores in historical reasoning</u>	<u>Explanation</u>
<p>There is a difference between the sources in the description of the goals of the rebellion. The first source presents the goal of the rebellion mainly as warfare, to help fight against the Nazis and kill the enemy. On the other hand, the second source shows the goal of the rebels is to choose how to die and prove that they are not weak. The main reason for the difference between the description of the goals of the rebellion in the two sources is the views of the people writing these passages. The first section was written by a military underground, Armia Kariova, and therefore their interest is to show that the right thing to do right now is to fight against the Nazis. To prove that even the Jews in the ghettos who are in very difficult conditions are trying to fight them. In the passage, they refer to the JCO, which they support. But these are not the interests of the JMU, so the goals it presents as the goals of the rebels are slightly different. The</p>	<p><u>Sourcing: 4/4</u> <i>"The first section was written by a military underground, Armia Kariova and therefore their interest is to show that the right thing to do right now is to fight against the Nazis. To prove that even the Jews in the ghettos who are in very difficult conditions are trying to fight them. In the passage, they refer to the JCO, which they support."</i> <u>And:</u> <i>"The revisionist party is in Israel and therefore it wants to show that the "young people of Israel" are the ones fighting in the ghetto, they want to emphasize that the Jews are not weak and that they can fight even in the worst conditions."</i></p> <p><u>Historical significance: 4/4</u> <i>"The first source presents the goal of the rebellion mainly as warfare, to help fight against the Nazis and kill the enemy. On the other hand, the second source shows the goal of the rebels is to choose how to die and prove that they are not weak."</i></p> <p><u>Historical causality: 4/4</u> <i>"The first difference is the place</i></p>	<p>Analyzes nuances between sources to base Claim; Uses previously learnt knowledge and skilled sourcing to base a solid claim; Ties the way the different sources name the fighters to their agenda: Refers to JCO - specifically because of their affiliation and to "young people of Israel" to include everyone accordingly</p> <p>Shows connection between the sources' agenda to nuances in text, highlighting the differences between the two groups, how in the rebels' goals reflect different messages as</p>

<p>revisionist party is in Israel and therefore it wants to show that the "young people of Israel" are the ones fighting in the ghetto, they want to emphasize that the Jews are not weak and that they can fight even in the worst conditions. Apart from the different views, two significant differences explain the difference between the goals that the two sources wrote. The first difference is the place where the sources were written, the first was written in Poland, near the place of the rebellion and they also had little communication with the rebels in the ghetto. The second source was written in the Land of Israel and therefore the distance together with the ideology it reflects influenced the goals they present. Another difference between the two sources is the time when they were written, the first source was published on 29.4.1943 while the second source was published months later on 8.10.1943. Time also has a great influence on what is written in these sources and depending on the period in the war and the situation in the country has a great influence on the goals presented in the sources.</p>	<p><i>where the sources were written, the first was written in Poland, near the place of the rebellion and they had little communication with the rebels in the ghetto. The second source was written in the Land of Israel and therefore the distance together with the ideology it reflects influenced the goals they present."</i></p> <p>Historical perspective: 4/4 <i>"The first section was written by a military underground, Armia Kariova and therefore their interest is to show that the right thing to do right now is to fight against the Nazis. In the passage, they refer to the JCO, which they support. But these are not the interests of the JMU, so the goals it presents as the goals of the rebels are slightly different. The revisionist party is in Israel and therefore it wants to..."</i></p> <p>Continuity and change: 4/4 <i>"the first source was published on 29.4.1943 while the second source was published months later on 8.10.1943. Time also has a great influence on what is written in these sources and depending on the period in the war and the situation in the country has a great influence on the goals presented in the sources."</i></p>	<p>quoted above.</p> <p>Connects context, periodization, and agenda to form causal relationships</p> <p>Understands different group's agenda and analyzing text accordingly and uses it as evidence.</p> <p>Properly uses periodization for Evidence to base historical claim</p>
---	--	---

Tables 3 and 4 show the essay of another student, Andy, whose scores for argumentation and historical reasoning were lower than the scores Nala received

Table 3. They Essay of student Andy and her scores in argumentation

<u>Essay</u>	<u>Scores in argumentation</u>	<u>Explanations</u>
Armia Krajowa was a Polish underground movement that opposed	Quality of claims: 2/4	Satisfactory claim but ignores some

<p>the German occupation and wanted to restore Polish rule. This movement helped the Red Army in the war against Germany in the hope that it would assist in restoring Polish rule. According to the Armia Krajowa newspaper, the goals of the rebels in the Warsaw Ghetto Uprising were to weaken the German forces, try to smuggle people out of the ghetto, and to die after fighting and struggling.</p> <p>HaMashkif was an Israeli daily newspaper of the Revisionist Party that served as a platform for the [Jewish Military] Union. The Revisionist movement behind this newspaper believed that the Jews should declare their intention to establish a Jewish state on both sides of the Jordan River. According to HaMashkif, the rebels chose to revolt because they wanted to die on their own terms and were not willing to be humiliated. They chose to die as heroes rather than be oppressed by the Nazis.</p> <p>One possible reason for the difference in describing the rebellion is that Armia Krajowa was an active rebel movement in the turbulent war zone and therefore needed to provide an example of how to conduct a rebellion and justify its meager significance in order to inspire the rebels in the movement and enhance their willingness to rebel. Armia Krajowa chooses not to present the fact that the rebels knew they had no chance, but instead to present their bold goals with nobility and heroism. HaMashkif, on the other hand, was a newspaper published in Israel, and it did not need to arouse a desire to rebel in the same way as Armia Krajowa.</p> <p>Another reason could be that HaMashkif, as an Israeli newspaper aspiring to establish a Jewish state, wants to create an image that events in Europe and the rest of the world are not good in order</p>	<p><i>"One possible reason for the difference in describing the rebellion is that Armia Krajowa was an active rebel movement in the turbulent war zone and therefore ... HaMashkif, on the other hand, was a newspaper published in Israel, and it did not need to arouse a desire to rebel in the same way as Armia Krajowa."</i></p> <p>Evidence: 0/4 <i>"One possible reason for the difference in describing the rebellion is that Armia Krajowa was an active rebel movement in the turbulent war zone and therefore needed to provide an example of how to conduct a rebellion and justify ... and it did not need to arouse a desire to rebel in the same way as Armia Krajowa."</i></p> <p>And: <i>"Another reason could be that HaMashkif, as an Israeli newspaper aspiring to establish a Jewish state, wants to create an image that events in Europe and the rest of the world are not good in order to encourage Jews to remain in the country."</i></p> <p>Causal chaining: 0/4 <i>"HaMashkif, on the other hand, was a newspaper published in Israel, and it did not need to arouse a desire to rebel in the same way as Armia Krajowa."</i></p> <p>And: <i>"Another reason could be that HaMashkif, as an Israeli newspaper aspiring to establish a Jewish state, wants to create an image that events in Europe and the rest of the world are not good in order to encourage Jews to remain in the country."</i></p>	<p>of the meanings of the sources; Only partly answers the question</p> <p>Incorrect and unsound interpretation of evidence</p> <p>Unsound reasoning; Uses context wrongly and jumps to conclusions</p>
---	---	---

<p>to encourage Jews to remain in the country.</p>	<p>Absence of mistakes: 0/4 <i>Wants to create an image that events in Europe and the rest of the world are not good in order to encourage Jews to remain in the country.</i></p> <p><u>Coherence of writing: 2/4</u></p>	<p>There are factual errors or unbased/irrelevant facts.</p> <p>Satisfactorily written</p>
--	--	--

Table 4. The essay of student Andy and her scores in historical reasoning

<u>Essay</u>	<u>Scores in historical reasoning</u>	<u>Explanations</u>
<p>Armia Krajowa was a Polish underground movement that opposed the German occupation and wanted to restore Polish rule. This movement helped the Red Army in the war against Germany in the hope that it would assist in restoring Polish rule. According to the Armia Krajowa newspaper, the goals of the rebels in the Warsaw Ghetto Uprising were to weaken the German forces, try to smuggle people out of the ghetto, and to die after fighting and struggling.</p> <p>HaMashkif was an Israeli daily newspaper of the Revisionist Party that served as a platform for the Union. The Revisionist movement behind this newspaper believed that the Jews should declare their intention to establish a Jewish state on both sides of the Jordan River. According to HaMashkif, the rebels chose to revolt because they wanted to die on their own terms and were not willing to be humiliated. They chose to die</p>	<p><u>Sourcing: 2/4</u> <i>"According to the Armia Krajowa newspaper, the goals of the rebels in the Warsaw Ghetto Uprising were to weaken the German forces, try to smuggle people out of the ghetto, and to die after fighting and struggling."</i> <u>And:</u> <i>"HaMashkif was an Israeli daily newspaper of the Revisionist Party that served as a platform for the Union... They chose to die as heroes rather than oppressed by the Nazis."</i></p> <p><u>Historical significance: 0/4</u> <i>"One possible reason for the difference in describing the rebellion is that Armia Krajowa was an active rebel movement in the turbulent war zone ... HaMashkif, on the other hand, was a newspaper published in Israel, and it did not need to arouse a desire to rebel in the same way as Armia Krajowa."</i> <u>And:</u> <i>"Another reason could be that HaMashkif, as an Israeli newspaper aspiring to establish a Jewish state, wants to create an image that events in Europe and the rest of the world are not good in order to encourage Jews to remain in the country."</i></p> <p><u>Historical causality: 0/4</u> <i>"One possible reason for the difference in describing the rebellion is that Armia</i></p>	<p>Weak analysis of sources; Partially correct interpretation of the evidence</p> <p>Presents the basic differences between the goals of the rebels but without explaining their meaning</p> <p>Wrongly interprets the significance of the sources</p>

<p>as heroes rather than oppressed by the Nazis.</p> <p>One possible reason for the difference in describing the rebellion is that Armia Krajowa was an active rebel movement in the turbulent war zone and therefore needed to provide an example of how to conduct a rebellion and justify its meager significance in order to inspire the rebels in the movement and enhance their willingness to rebel. Armia Krajowa chooses not to present the fact that the rebels knew they had no chance, but instead to present their bold goals with nobility and heroism. HaMashkif, on the other hand, was a newspaper published in Israel, and it did not need to arouse a desire to rebel in the same way as Armia Krajowa.</p> <p>Another reason could be that HaMashkif, as an Israeli newspaper aspiring to establish a Jewish state, wants to create an image that events in Europe and the rest of the world are not good in order to encourage Jews to remain in the country.</p>	<p><i>Krajowa was an active rebel movement in the turbulent war zone and therefore ... HaMashkif, on the other hand, was a newspaper published in Israel, and it did not need to arouse a desire to rebel in the same way as Armia Krajowa."</i></p> <p><u>And:</u> <i>"Another reason could be that HaMashkif, as an Israeli newspaper aspiring to establish a Jewish state, wants to create an image that events in Europe and the rest of the world are not good in order to encourage Jews to remain in the country."</i></p> <p><u>Historical perspective: 2/4</u> <i>"Armia Krajowa was a Polish underground movement that opposed the German occupation and wanted to restore Polish rule ... and to die after fighting and struggling."</i></p> <p><u>And</u> <i>"HaMashkif was an Israeli daily newspaper of the Revisionist Party that ... They chose to die as heroes rather than oppressed by the Nazis."</i></p> <p><u>Continuity and change (periodization): 2/4</u> <i>"Armia Krajowa was a Polish underground movement that opposed the German occupation and wanted to restore Polish rule ... HaMashkif was an Israeli daily newspaper of the Revisionist Party that served as a platform for the Union."</i></p>	<p>Misses the correct causal relationship between sources and agenda</p> <p>Recognizes the different views of stakeholders but fails in the analysis</p> <p>Uses place but not time for evidence; Misses the relevance of source B's time frame</p>
--	---	---

Two raters independently scored the essays on each qualitative aspect (namely, argumentation and historical thinking). Inter-rater reliability was assessed using Cohen's Kappa. The results indicated substantial agreement for argumentation ($\kappa = 0.71$) and near-perfect agreement for historical thinking ($\kappa = 0.85$).

The statistical analysis was performed using R Statistical Software (version 4.4.0; R Core Team, 2021) in the RStudio environment (RStudio Team, 2024).

To address our first research question, that is, whether the learning outcomes were different for learners participating in peer-led collaborative group discussions around a controversial historical issue, compared to learners participating in a teacher-led whole-class discussion around the same issue, we conducted an analysis of covariance (ANCOVA). The covariate concerned students' dedication during history class and was based on their grades in three written assignments. It was necessary since each class's heterogeneity in terms of students' dedication could not be controlled for, as the schools' division into classes had to be adopted.

The sample size was determined based on the necessary statistical tests. Since we were looking for pairwise comparisons, a large sample was required (at least 100 data points per condition, see e.g., Brysbaert, 2019). Observations were thus collected from as many students as possible, given the limitations imposed by resource constraints. Since the data were not normally distributed, we applied the Aligned Rank Transform procedure (e.g., Salter & Fawcett, 1995) using the R package ARTool (Kay et al., 2021) to examine main effects and interactions (Wobbrock et al., 2011) and to conduct *post hoc* pairwise comparisons (Elkin et al., 2021).

Partial eta squared (η_p^2) was calculated for each main effect and interaction to assess the magnitude of the effects. Effect sizes were interpreted according to Cohen's guidelines, with η_p^2 values of 0.01, 0.06, and 0.14 considered small, medium, and large effects, respectively.

In what follows, only significant effects ($p < .05$) are reported.

Our second research question was qualitative in nature. We analyzed the feedback provided by the students who completed the questionnaire and identified leading principles (Braun & Clarke, 2006).

We were aware of a potential conflict of interest: The second author served as both the teacher for the face-to-face group and the designer of the CSCL intervention. As a proponent of dialogic education, this dual role introduces the risk of experimenter bias. It is possible that the teacher's enthusiasm for the CSCL tool subtly influenced the classroom atmosphere. However, it is equally plausible that his expertise in face-to-face dialogic instruction provided the 'control' group with a higher-than-average standard of instruction, making the 'null result' of the CSCL group even more robust. To mitigate this bias, grading was conducted using a rubric with high inter-rater reliability ($\kappa > 0.7$), and the teacher was not involved in the statistical analysis.

Findings: Effectiveness of small-group peer-led collaborative CSCL discussions through the analysis of individual outcomes

Table 5 and Figure 1 present students' scores (means, medians, and standard deviations) in argumentation and historical reasoning – the two qualitative aspects of the historical essay written by the students. Scores are divided according to the discussion type the students participated in prior to writing the essay: teacher-led whole-class face-to-face discussion versus small-group peer-led collaborative CSCL discussion. Recall that each qualitative aspect was evaluated based on five criteria, with each criterion worth up to 4 points – totaling a maximum score of 20 points per aspect. Students' scores are thus presented as percentages out of 20 points.

Table 5. Scores in argumentation and historical reasoning of historical essays of students participating in unguided small-group CSCL and teacher-led whole-class FTF discussions

Discussion Type	Unguided Small-group CSCL (N = 21)		Teacher-Led Whole-Class FTF (N = 24)	
	Argumentation	Historical Reasoning	Argumentation	Historical Reasoning
Mean	68.3	60.6	60.4	54
Median	75	60	60	52.5
Standard Deviation	27.9	28.1	26.4	25.8

Note. CSCL = computer-supported collaborative learning; FTF = face-to-face.

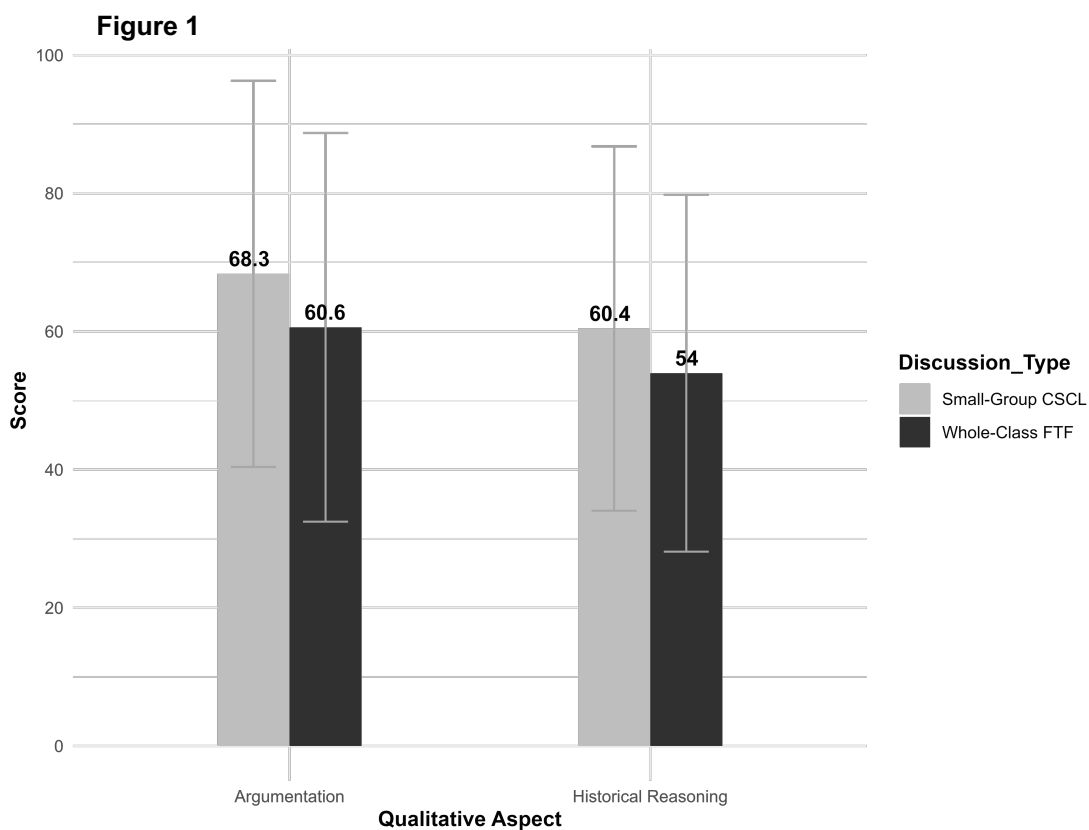


Table 6 and Figure 2 present students' scores (means, medians, and standard deviations) in argumentation and historical reasoning, divided according to the discussion group these students partook in – eight groups in total.³

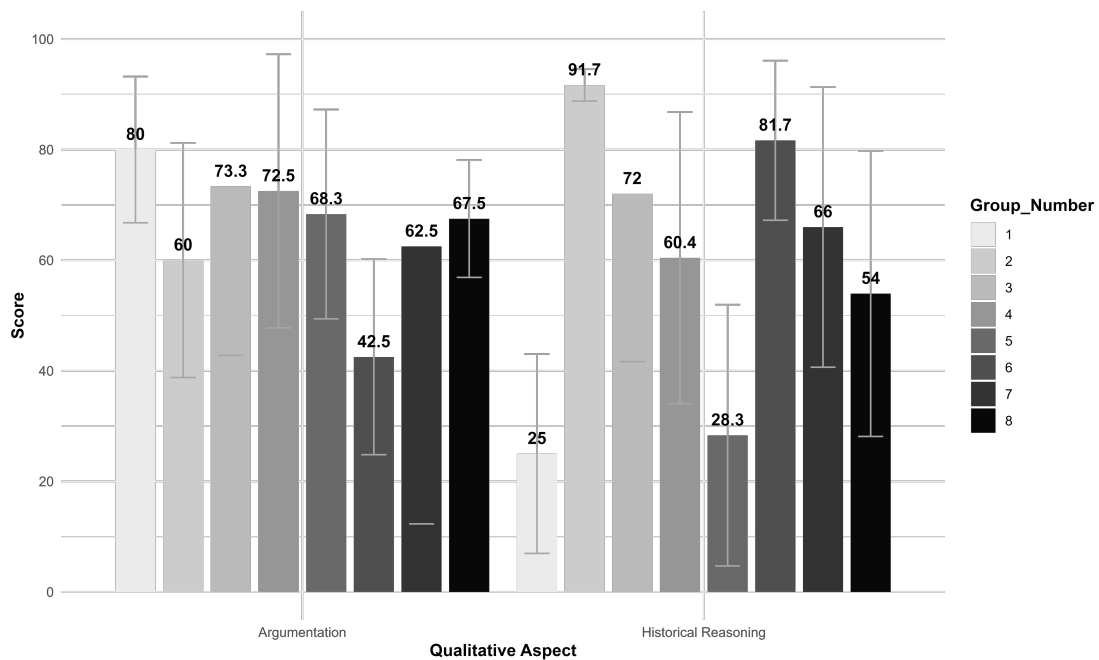
³ Recall that only CSCL discussions involved different discussion groups. Those were seven different groups, while the whole-class discussion was considered as the eighth group for the purpose of this analysis.

Table 6. Scores in argumentation and historical reasoning of historical essays of students, divided according to discussion groups

Group Number	Qualitative Aspect	Argumentation	Historical Reasoning
1	Mean	80	68.3
	Median	75	60
	Standard Deviation	13.2	18.9
2	Mean	25	28.3
	Median	20	20
	Standard Deviation	18.03	23.6
3	Mean	60	42.5
	Median	60	42.5
	Standard Deviation	21.2	17.7
4	Mean	91.7	81.7
	Median	90	90
	Standard Deviation	30.3	14.4
5	Mean	73.3	62.5
	Median	80	85
	Standard Deviation	30.6	50.2
6	Mean	72	66
	Median	90	65
	Standard Deviation	30.3	25.3
7	Mean	72.5	67.5
	Median	72.5	67.5
	Standard Deviation	24.8	10.6
8	Mean	60.4	54
	Median	60	52.5
	Standard Deviation	26.4	25.8

Note. CSCL = computer-supported collaborative learning

Figure 2



In order to address our first research question, i.e., whether the learning outcomes of students participating in small-group peer-led collaborative CSCL discussions were different from those of students participating in a teacher-led whole-class face-to-face discussion, we conducted an ANCOVA involving Qualitative Aspect as a 2-level within-subject factor (*Argumentation, Historical Reasoning*), Discussion Type a 2-level between-subject factor (*CSCL, FTF*), Group Number as an 8-level between-subject factor (1-8), and Student Dedication as a covariate. The analysis revealed no main effects, but gave rise to three significant interactions:

- Qualitative Aspect * Group Number: $F(6, 13) = 2.93, p = 0.04, \eta^2_p = 0.575$, indicating that the effect of qualitative aspect on students' scores was influenced by their discussion group (large effect).
- Qualitative Aspect * Student Dedication: $F(21, 13) = 4.62, p = 0.003, \eta^2_p = 0.882$, indicating that the effect of qualitative aspect on students' scores varied substantially with their level of dedication (very large effect).
- Qualitative Aspect * Discussion Type * Student Dedication: $F(3, 13) = 4.51, p = 0.022, \eta^2_p = 0.51$, suggesting that the combined effects of qualitative aspect and discussion type on students' scores were moderated by their dedication (large effect).

However, *post hoc* comparisons did not reveal any significant pairwise differences between specific levels of the factors, and some contrasts could not be estimated, likely due to sparse data and empty cells in certain factor combinations. Hence, although scores tilted towards the students participating in small-group peer-led collaborative CSCL discussions in both argumentation (68.3 versus 60.4) and historical thinking (60.6 versus 54), no significant differences were found between the discussion types.

At first glance, the findings are quite surprising. As already mentioned, Kirschner et al. (2006) showed that direct instruction is preferable to autonomous exploration for problem solving and conceptual

learning. Consequently, one might expect that a teacher-led discussion would help students write higher-quality argumentative essays than a peer-led CSCL discussion. We will discuss this point in the discussion section. However, in the next section, we will analyze a discussion in which student Nala participated. Nala is a student whose essay, which was presented and analyzed above (see Tables 1-2), received top scores for both argumentation and historical reasoning. Through the analysis of the discussion in which she participated, we will suggest a possible explanation for the findings.

Description of a small-group discussion and of its impact on essay quality

We analyze here one CSCL-mediated discussion that took place in a small group of three students. In the discussion, students were asked to provide their reasoned viewpoint about the motives of the partakers in the Warsaw Ghetto Uprising, after having read two written sources. In the essays, students were asked to explain the differences between motives conveyed by these sources. Three students participated in the discussion: Nala, Morin, and Gil. Nala's essay was written three weeks after the discussion. It was presented and analyzed in the methodology section. We decided to wait three weeks for the writing task because we wanted to evaluate middle-term gains from the discussion. We present the discussion here. The numbers indicate the chronological order of the turns. Indentation marks the threads in the CSCL-mediated discussion: An indented turn signifies a direct response to a previous turn.

1 Nala

I didn't understand what I'm supposed to ask :(

2 Morin

Haha I think Arens or Blatman,⁴ JMU [Jewish Military Union] or JCO [Jewish Combat Organization]

3 Nala

I agree with Blatman's position that claims Arens's research is not reliable and is influenced by his personal views on the Bund⁵ movement and on JCO. On the other hand, I agree that there is an underrepresentation of JMU's significant contribution to the rebellion.

8 Morin – Responds to 3 Nala:

I agree with the unreliability of Arens's research and, combined with the lack of information and diminished research on the uprising, mentions of the JMU movement are decreasing and even being erased

9 Gil – Responds to 3 Nala:

Nala, You're absolutely right, but I think we need to emphasize the lack of representation of JMU in the mention of the Warsaw Ghetto Uprising. It's really really important in my opinion...

4 Gil

I agree that the representation of JMU in the Warsaw Ghetto Uprising in political memory is unfair. The state's choice to mainly mention JCO stems from their political views. The lack of documentary material could have affected the level of recognition and the mentioning of JMU as part of the uprising, but if they wanted to, they could mention everyone who took part in the uprising, they would have made an effort and taken testimonies.

6 Nala– Responds to 4 Gil:

I support your position. Collective memory is important when it comes to historical events. The collective memory of the Warsaw Ghetto Uprising and its central

⁴ Arens and Blatman are the names of the authors of the sources provided to the students.

⁵ The Bund was a secular Jewish socialist party in Eastern Europe that opposed Zionism, focusing instead on Jewish cultural autonomy in the diaspora.

perception were created by Yitzhak Zuckerman, a member of the Dror movement,⁶ who greatly contributed to the documentation but was a member of the Dror movement and also wrote in a subjective way.

5 Morin

I agree with Blatman's position that says JMU had a major part in the Warsaw Ghetto Uprising. The main historical argument for this is Muranowski Square, where they raised the symbols of the uprising, the Israeli flag and the Polish flag. There are no mentions of flags in the reports during the war period, but there is a mention after the war. Blatman's explanation for this is that during the war, the SS⁷ commander couldn't admit his defeat. The organization responsible for Muranowski Square is JMU. Another argument is Arens's unreliable research.

7 Nala– Responds to 5 Morin:

Dear Morin, I was interested to read what you wrote. These are indeed convincing historical arguments. Was it the Israeli flag or the JMU Flag?

10 Gil

Additionally, Ofer Aderet⁸ mentions that JMU fought almost like a military and more as a professional fighting force, unlike JCO who fought like a youth movement. This detail is important and adds to the unfairness towards JMU in the memory of the uprising.

It appears that Nala immediately mentions “*an underrepresentation of JMU's significant contribution to the rebellion*” (turn 3), but she expresses neither her appraisal nor the causes of this underrepresentation. Gil claims in response that it is “*political*” and “*unfair*” (turn 4). Nala expands Gil's claims by stressing that “*The collective memory of the Warsaw Ghetto Uprising and its central perception were created by Yitzhak Zuckerman, who greatly contributed to the documentation but was a member of the Dror movement [a socialist movement] and also wrote in a subjective way*” (turn 6). Morin adds that the first historical source provides evidence for the pivotal role of JMU (“*The main historical argument for this is Muranowski Square, where they raised the symbols of the uprising, the Israeli flag and the Polish flag*”; turn 5). Nala clearly admits that what Morin says is new to her (“*Dear Morin, I was interested to read what you wrote. These are indeed convincing historical arguments. Was it the Israeli flag or the JMU Flag?*”; turn 7). Gil adds that JMU was a military force, in contrast with JCO (“*Ofer Aderet mentions that JMU fought almost like a military and more as a professional fighting force, unlike JCO who fought like a youth movement*”; turn 10).

Nala's essay takes into consideration the aspects that her peers contributed to the discussion. The fact that the difference between the three sources relies on political and ideological grounds, as expressed by Gil, is incorporated into her essay: “There is a difference between the sources in the description of the goals of the rebellion. The first source presents the goal of the rebellion mainly as warfare, to help fight against the Nazis and kill the enemy. On the other hand, the second source shows the goal of the rebels is to choose how to die and prove that they are not weak” (see Tables 1-2). The fact that JMU acted as a professional movement, which was stressed by Morin in the discussion, appears in her essay in the following paragraph: “*The first section was written by a military underground, Armia Krajowa,⁹ and therefore*

⁶ The Dror movement was a pioneering socialist-Zionist youth movement that took part in the revolt, Yitzhak 'Antek' Zuckerman was one of its leaders.

⁷ The SS (German for *Schutzstaffel*; lit. 'Protection Squadron') was a major Nazi paramilitary organization that controlled the German police forces and the concentration camp system.

⁸ Ofer Aderet is the author of another source that was at the students' disposal during the discussion and the writing of the essay.

⁹ Armia Krajowa (Polish for 'Home Army') was the largest resistance movement in German-occupied Poland during World War II.

their interest is to show that the right thing to do right now is to fight against the Nazis. To prove that even the Jews in the ghettos who are in very difficult conditions are trying to fight them" (again, see Tables 1-2).

The most salient characteristic of Nala's written essay is that it integrates the different views expressed in the discussion, but not verbatim. It shows a remarkable synthesis, like in the following part of the essay: *"In the passage, they refer to the JCO, which they support. But these are not the interests of the JMU, so the goals it presents as the goals of the rebels are slightly different. The revisionist party is in Israel and therefore it wants to show that the "young people of Israel" are the ones fighting in the ghetto, they want to emphasize that the Jews are not weak and that they can fight even in the worst conditions."*

We characterize the exchange between Nala, Morin, and Gil not merely as information exchange, but as a 'dialogic gap' leading to new understanding (Wegerif, 2008). In the chat, Nala does not simply accept Morin's correction regarding the JMU; she aims at refining it (*"Was it the Israeli flag or the JMU Flag?"*; turn 7). This search for refinement marks a moment of 'unfinalizability' (Bakhtin, 1984): Nala is struggling to reconcile her prior knowledge with the new voice introduced by Morin. Her final essay does not merely copy Morin's fact; it synthesizes Morin's emphasis on the 'professional' nature of the fighters with her own understanding of political memory. This demonstrates that the peer-led group succeeded in moving from cumulative talk to *constructive conflict*, resulting in an essay that was multivocal rather than a recitation of a single textbook narrative. Three weeks after the discussion, the views of the three discussants intertwine in a sophisticated way in Nala's historical essay.

In this case 'collaboration' was manifest in the engagement of different historical views on a CSCL platform, without necessarily leading to consensus or agreement. On the contrary, our design aligns with the view of 'constructive conflict' (Baker et al., 2019). The instructions explicitly required students to 'react to your group members by expressing an alternative argument.' We view the friction between differing interpretations not as an obstacle to be smoothed over, but as the essential engine of 'inter-thinking.' In the peer-led collaborative dialogue CSCL environment, the lack of a teacher – who often acts as a peacemaker or consensus-builder – allowed these disagreements to remain unresolved longer, forcing students to struggle with the 'unfinalizability' of the historical narrative. The small-group discussion shows, then, the creation of a dialogic space. The other peer-led discussions shared this feature. For our research team, this was a sign of effectiveness, in terms of dialogic aims.

Findings: Epistemological insights following the discussions

The questionnaire served to both boost epistemological insights and compare the responses of the students from the two classes. Twenty-two students from Class 1 (who attended the teacher-led whole-class face-to-face discussion) and 19 students from Class 2 (who participated in a peer-led collaborative small-group CSCL discussion) answered the questionnaire.

In general, the questionnaire prompted students to articulate epistemological insights about the discussion, which might not have emerged otherwise. In the analysis of the feedback, we identified leading principles (Braun & Clarke, 2006), including time pressure, reflectiveness, and the way students refer to various sources in their writing.

At a surface level, the students of the two classes responded in the same way concerning the link between the discussion and the writing of the essay: 10 out of the 22 students in Class 1 (45%), and 11 out of the 19 students in Class 2 (58%), mentioned that the discussion was helpful. Delving more deeply into the answers nevertheless revealed some inter-class differences. In Class 1, the response in support of a link between the discussion and the writing task was a laconic "yes" or the like for 5 out of the 10 students acknowledging the link. For others, it concerned factual knowledge, like in the case of Rob:

The Effectiveness of Computer-Supported Collaborative Dialogue in Schools

Baruch Schwarz, Ziv Lison, Noa Brandel

The discussion helped me understand the events in the upheaval in the Warsaw Ghetto and what the consequences of this uprising were.

In other words, the teacher-guided whole-class face-to-face discussion helped Rob understand the deployment of the events during the Warsaw Ghetto Uprising. Other students who participated in the teacher-led discussion and whose responses about the link between the discussion and the writing of the essay were more elaborate, provided similar feedback.

Six out of the 11 students in Class 2 that identified a link between the discussion and the essay writing, provided explanations to support their claim. Their explanations were more reflective than the explanations of the students in Class 1. For example, Terez wrote:

During the writing of the essay, I identified links between the question and the discussion. I remember that things from the discussion popped up in my head as I was writing. In the discussion, we brought several pieces of evidence and we responded to our mutual justifications. In my opinion, this taught me how to justify and strengthen our claims in a better way.

Louis, also from Class 2, affirmed that he detected a link between the discussion and the writing of the essay, but added the following reservation:

It was then that I identified the "dry" link between JCO and JMU and the articles, the entities, and interests behind them, but I did not delve deeply into the implications of these links.

In other words, Louis stressed the importance of personal reflection in order to internalize the role of these entities and interests, but reported that this internalization failed in his case. At any rate, the request to report on the contribution of the peer-led collaborative dialogue to the writing of the essay led to extended reflections – whether positive or not. We will come back later to reflectiveness: the reflective nature of the responses of students from Class 2.

As for the question "What makes an argument or a historical argument solid?", as noted earlier, these questions were not asked to measure the argumentative skills of the students, but were aimed at fostering epistemological insights. For example, Perl, who participated in the peer-led collaborative CSCL discussion, responded to the question "What makes a historical argument solid?" in the following way:

First validity: To look for events relevant to the claim. Secondly, to use primary sources. Third, to think according to a historical perspective and to avoid presentism. Fourth, to understand what makes historical events significant, and finally, causality, to link different events.

Many students in both groups provided similar appraisals about what makes a historical argument solid. Students from both groups distinguished between solid arguments in general and solid historical arguments in particular. For example, Sunny from Class 1 wrote about what makes arguments solid:

As I understand it, a solid argument is the combination of a clear claim, a number of justifications (including examples and data) and refutation of opposing arguments.

As for what makes historical arguments solid, he wrote:

As I understand it, a solid argument in history has to meet the criteria that I wrote about arguments in general, and to meet other criteria like the historical link, historical perspective, and various sources (primary and secondary) that support the conclusions.

Such discernments were common in both classes. They attest to the high pedagogical standards according to which the teacher facilitated/designed activities in both classes.

We observed little to no differences between the two classes' responses to the first three questions. However, some differences arose between Class 1 and Class 2 concerning the fourth question: "How do I evaluate my arguments in the historical essay?". The first difference is that *time pressure* was more common in Class 1 (7 out of 22 responses; 32%) than in Class 2 (2 out of 19 responses; 10.5%). For example, Tony from Class 1, wrote:

I justified my arguments but not in a deep way, especially at the end because of the time pressure. The arguments were good but not elaborate enough. I hope that next time, they will be better.

For Tony and for other students in Class 1, the time limit for writing adds a lot of pressure because articulating clear and grounded arguments demands high precision and reflection, as well as argumentative writing, and these require time. The following response, written by Don, also from Class 1, makes the point clear:

I think that my arguments in the essay were good and were well justified, except for two places. When I referred to the picture [...] but I missed it and I forgot to add [...]. Thus, I regret that I didn't come back to the question (I didn't have time) to fix this. As for the question regarding which I needed to add JMU, when I read about the revisionists, they referred to the JMU movement, and I don't know why I did not add a sentence about it. I hope that next time, I will not miss details like those.

As already mentioned, the issue of time pressure was rarely mentioned by students in Class 2.

In Class 1, the evaluation of the quality of the historical essay often concentrated on students' compliance with what was expected from them, as in the two evaluations cited above. Bill's reflection on his own work conveys the same character of compliance:

The arguments were partially substantiated with historical claims, partially with general arguments, and some of them were not adequately substantiated either way.

At the same time, Bill praises the questionnaire:

I enjoyed the reflection because I understood what I can improve. I am not sure that I will be able to apply what I learned in real time, but it helps.

For 14 out of the 19 students in Class 2 (74%), the evaluation was critical and sincere, like in the examples we brought from Class 1, but they were somewhat more personal. They involved general justifications in addition to the justifications that can be extracted from the sources they had at their disposal. For example, Pete wrote:

In the essay, the claims are not very well-justified and most of them rely on evidence, for example, I brought evidence to support my claim but it did not provide enough facts to explain it. I also used general justifications, and did not use historical justifications like causality and the creation of meaning through the use of primary sources.

The personal nature of the appraisal is even more salient in the case of Suki:

In my opinion, I should have used more historical justifications that were mentioned above. In particular, I have to think about the historical perspective and to disconnect from my personal viewpoint and to avoid presentism. In addition, I should have taken more care of small details in the historical sources and given weight to each word.

The example we bring here stresses even more that, in their response about the appraisal of their argumentative essay, students from Class 2 did not refer to their compliance with the demands of the task, but to their own position with respect to the issue at stake. Indeed, the student AI criticized the position of the historian Segev (whose source was at the students' disposal during the writing of the essay comparing it to the discussion) and claimed:

In the question in which there is a contradiction between Segev's claim and the picture, I explained in great detail what appears in the picture and how this contradicts Segev's claim. I also expressed some reservations about my own claim, and I also presented facts that support the position claimed in the picture. However, the answer misses a clear-cut affirmation, the goal – the contradiction. In addition, I should have possibly thought about what is known and to what extent their contradiction is strong. The evidence of the number of people in the photo [of a protest in support of Jewish immigration] in relation to their [general] number [i.e., the number of Jews in the population at the time] strengthens the basis of the argument. There is something to improve.

In such a reflection, we can see deep epistemic considerations about the quality of arguments in general and about the difficulties that AI encountered while writing the essay in particular.

AI went on to write the following:

I did not understand what makes an argument solid and this activity helped me understand how to construct an argument in history and in the humanities.

The following appraisal, written by Gil, shows the same personal nature of the response:

In my justifications, there are no primary sources, but only secondary sources. The justifications are historical as they relate to events that happened, they take into account chronology, [they] are validated. I generally refer to causality and to historical importance. Sometimes, I used general justifications. Most of my justifications are good, although they can be improved, for example by inserting primary sources. Anyway, it was fun and I learned a lot about the world, history, my people, and myself. Thank you, Ziv [Ziv is the name of the teacher – the second author].

In a nutshell, the appraisals in Class 2 were more reflective and conveyed richer considerations of the sources they used in their essays (especially general personal considerations). It also seems that they were more relaxed in their writing than the students in Class 1, as they rarely referred to time pressure they experienced during their writing. We will suggest some explanations for these differences in the general discussion.

Discussion

The present study is important on three very different scales that correspond to the three meanings of effectiveness outlined in this article: the quality of learning outcomes at the individual level, changes to the educational system at the policy level, and the creation of a dialogic space at the ontological level. First, as far as our first research question is concerned, we found that in a particular context, gains from small-group peer-led CSCL discussions do not differ from gains from high-quality teacher-led whole-class discussions. In other words, the quality of small-group peer-led collaborative CSCL dialogue is comparable to the quality of guided adaptive instruction. It might be higher for peer-led CSCL discussions, but a larger

sample would be required to reveal such differences. The fact that the essays in history were even higher quality (though not statistically significant) for students who discussed the issue in small groups suggests that small-group collaborative dialogue provides opportunities for learning insights that teacher-led whole-class discussions do not.

Our findings revealed no statistically significant difference in essay quality between the teacher-led and peer-led groups. Following Rogoff (1990), this resonates with the historical debate between Piagetian (peer-based) and Vygotskian (adult-guided) paradigms, suggesting that well-scaffolded peer interactions can rival expert instruction. However, we must consider alternative explanations for this 'null result.'

First, it is possible that the 'ceiling effect' (high performance in both groups) masked differences; the teacher's expertise in dialogic instruction likely elevated the face-to-face group's performance, preventing the CSCL group from showing a distinct advantage. Second, we must consider the context of COVID-19. The stress of the pandemic and the physical constraints of the classroom (masks, social distancing) may have hampered the natural flow of the face-to-face discussion, leveling the playing field with the remote CSCL group. Finally, while some might argue that the teacher was 'ineffective,' his established track record in dialogic pedagogy suggests otherwise. We argue that the lack of difference is actually a positive finding for systemic scalability: It proves that peer-led digital dialogue can sustain high-level historical reasoning without the constant, resource-heavy intervention of a teacher.

We are aware that the discussion groups were not homogeneous. In future implementations, students should not be randomly divided into groups; rather, the teacher should ensure that groups are balanced (for instance, that each discussion group comprises one high-level student, one med-level student, and one low-level student). We did not check students' pre-existing abilities in history reasoning and argumentation skills. We explained that the study was not set as a controlled experiment. However, the experiment helped change the Ministry of Education's policy: The second author persuaded Ministry officials to adopt the quality of peer-led collaborative dialogue as a measure of compliance with the standards set for students in secondary education.

The quite positive results achieved in the absence of teacher guidance during small-group discussions should be understood in context. The students were enculturated into dialogic practices and worked with CSCL tools. In addition, the small groups were given written instructions ((a) Give your reasoned viewpoint about ..., and (b) React to your group members by expressing an alternative argument to those expressed by them). These instructions are collaborative and argumentative scripts. They guided students' moves during the discussion, and at the same time, they gave them authority. The students were also provided with two historical sources. Such a design echoes the four principles articulated by Engle and Conant (2002) to foster disciplinary engagement: problematization, authority, accountability to disciplinary norms, and resources. Still, it is remarkable that this meticulous design yielded such high-quality essays among students who participated in small-group discussions. As mentioned before, this design aligns with the view of 'constructive conflict' (Baker et al., 2019).

The contribution of the CSCL tool to the high quality of the historical essays has been referred to sporadically. First, as electronic channels whose traces remain visible, they invite reflection during and after the discussions, as already claimed by Kuhn et al. (2008, 2013). The fact that the discussions were threaded and that students were encouraged to explicitly refer to previous turns in new turns led to the activation of meta-argumentation. As shown by Bächtold et al. (2022), this activation helps in productive argumentation. Moreover, the fact that the CSCL tool concretized a material *dialogic space* (Wegerif, 2007) in which the texts that the students discussed were visible also contributed to this productivity.

The epistemological prompts provided in the questionnaire led the students participating in small-group peer-led collaborative discussions to express epistemological insights: How the discussion impinged on the writing of the essay, what constitutes a solid (historical) argument, and how to evaluate personal arguments in the historical essay, spanning the justifications, the evidence, and the examples. First of all, students who participated in small-group discussions were more relaxed – they hardly mentioned time pressure. Their insights were more elaborate than those of students who participated in a teacher-led whole-class face-to-face discussion: Quite often, students participating in peer-led collaborative dialogue actually internalized moves expressed by others in their essay. We have already mentioned that the epistemological prompts not only measure, but also boost, epistemological insights (Bächtold et al., 2022). It is possible, then, that students who had participated in peer-led collaborative discussions became aware of these insights while answering the questionnaire. Anyway, from an epistemological point of view, this reflective activity suggested that these students were superior to those who attended the teacher-guided whole-class discussion.

To a large extent, the findings concerning our second research question explain why peer-led collaborative dialogue led to high-quality essays. The students participating in the small-group discussions positioned themselves with regard to the historical issue under discussion. Yet, they took different positions into consideration. We first attribute their unusually vivid memory of the discussions three weeks after they took place to the CSCL tool, which materialized the turns and preserved their threaded structure. There is, however, a more general reason for this phenomenon. Durable memory long after collaborative learning in small groups is well-known (Howe et al., 2005). Howe et al. explain this phenomenon as the result of an *incubation* after collaboration. Asterhan and Schwarz (2007) replicated this phenomenon when conceptual learning was at stake. The example of the student discussion that we presented above exemplifies interactions in which autonomous students interacted towards the co-elaboration of reasoned viewpoints on a historical issue. These precious moments were processed – *incubated* – by the students until they wrote their final essays. It stresses the long-term effects of the design we elaborated. At any rate, we showed in this study that peer-led collaborative dialogue in CSCL settings can be effective in schools.

A major finding of this study concerns effectiveness in its systemic sense. It is in this highly propitious context that the present study was conducted. As noted earlier, the Israeli educational system recognized that clear argumentative and reasoning aspects of essays in history could be measured and that these measures could partly account for the matriculation exam grades. Moreover, the educational system recognized the importance of high-quality classroom talk and supported training programs for in-service teachers to facilitate high-quality classroom talk. Finally, the Dialogym platform provided a possible solution for the problem of student attendance in schools during the COVID-19 pandemic, as it enabled the division of students into small groups to complete learning tasks. The study shows that these opportunities were capitalized on. Since the pandemic ended, the use of CSCL tools should have stopped, and the impact of CSCL practices would have appeared to be circumstantial. However, the system learned from this experience, and in the tormented geopolitical situation in which Israel stands, the Ministry's policy is to enable dialogues among schools enrolling students of different ethnicities or viewpoints. These dialogues are carried out through electronic platforms to handle issues of geographical distance, mainly through Zoom. They are considered vital by the Ministry, which has recognized that high-quality talk among students of different backgrounds is in itself an important achievement, and that promoting such talk should be an aim in civic education. When the second author informed officials in the Ministry of Education that clear criteria could be proposed to assess the quality of talk between students from different communities, and that tools empowered by artificial intelligence (AI) could measure this quality online, the Ministry was open to this proposition. This was probably because officials had already recognized that the system itself could foster effective practices within peer-led collaborative dialogue in CSCL environments.

Rather than viewing these results as a 'victory' for CSCL over traditional teaching, we frame them as evidence of systemic feasibility. The Ministry of Education's adoption of these assessment tools does not signal an ideological shift but rather a pragmatic recognition that 'dialogic spaces' can be assessed with the same rigor as monologic exams. The study demonstrates that the 'messiness' of peer-led student dialogue does not compromise academic standards (accountability), allowing such practices to survive within the hierarchical constraints of the school system.

Although we were satisfied with the results of the study, which attested to the effectiveness both in terms of individual gains and in the systemic impact of dialogic practices in history classes, we kept our pedagogical aim in mind. While our method is instrumental – measuring 'effectiveness' to satisfy the gatekeepers of the educational system – our pedagogical objective is ontological. We utilized these instrumental measures to carve out a protected space within the curriculum where students can engage in 'becoming' through dialogue. By proving that such spaces do not harm academic outcomes, we secure students' right to engage in meaning-making that is not limited to instrumental purposes. For us, this modest study points toward a humanistic essence of education in a digital age. The results suggest that the 'Thou' of the learning partner (Buber, 1923) can be encountered even through the mediation of a screen and without the intervention of an authority figure. By trusting students to organize their own 'dialogic space,' we affirm that education is not merely the transmission of content from teacher to student, but a horizontal act of shared creation. As educational systems become increasingly digitized, ensuring that these tools are used to foster autonomy rather than surveillance is the critical challenge for the future of dialogic pedagogy.

Looking forward, the integration of AI-empowered tools offers a solution to the methodological tension discussed above. While this study relied on post-discussion essays, future implementations can utilize Natural Language Processing (NLP) to assess the dialogic process in real-time. By training AI models such as large language models (LLMs) to recognize markers of 'exploratory talk' (such as tentative language, questioning, and building on others' ideas) rather than just 'final form' arguments, we can move beyond the 'essay' as the sole arbiter of quality. This would allow the educational system to value and accredit the act of dialogue itself, not just its written byproducts.

References

- Andriessen, J. & Schwarz, B. (2009). Argumentative design. In N. Muller Mirza & A.-N. Perret-Clermont (Eds.), *Argumentation and Education: Theoretical Foundations and Practices* (pp. 145–174). New York: Springer.
- Arnseth, H. C. & Ludvigsen, S. (2006). Approaching institutional contexts: Systemic versus dialogic research in CSCL. *The International Journal of Computer-Supported Collaborative Learning* 1, 167–185 DOI 10.1007/s11412-006-8874-3
- Asterhan C. S. C. & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *The Journal of Educational Psychology*, 99(3), 626-639.
- Bakhtin, M. M. (1984). *Problems of Dostoevsky's poetics*. Minneapolis: University of Minnesota Press.
- Bächtold, M., Pallarès, G., De Checchi, K., & Munier, V. (2022). Combining debates and reflective activities to develop students' argumentation on socioscientific issues. *Journal of Research in Science Teaching*, (1), 1–46.
- Baker, M. J., Andriessen, J., & Schwarz, B. B. (2019). Collaborative argumentation-based learning. In R. Wegerif, N. Mercer and L. Major (Eds.), *The Routledge International Handbook of Research on Dialogic Education* (pp. 76–88). Routledge.

- Barzilai, S., & Chinn, C. A. (2018). On the goals of epistemic education: Promoting apt epistemic performance. *Journal of the Learning Sciences*, 27(3), 353-389.
- Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77–101.
- Brysbart, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2(1), 1–38.
<https://doi.org/10.5334/joc.72>
- Buber, M. (1923). *Ich und Du*. Leipzig: Insel-Verlag.
- De La Paz, S., Monte-Sano, C., Felton, M., Croninger, R., Jackson, C., & Piantedosi, K. W. (2017). A historical writing apprenticeship for adolescents: Integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly*, 52(1), 31–52.
- Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). [The evolution of research on computer-supported collaborative learning: From design to orchestration](#). Technology-enhanced learning: Principles and products, 3–19.
- Elkin, L. A., Kay, M., Higgins, J. J., & Wobbrock, J. O. (2021). An aligned rank transform procedure for multifactor contrast tests. *Proceedings of the ACM Symposium on User Interface Software and Technology* (UIST '21) (pp. 754–768). <https://doi.org/10.1145/3472749.3474784>
- Engle, R. A. & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, 20(4), 399–483.
- Hemberger, L., Kuhn, D., Matos, F., & Shi, Y. (2017). A Dialogic Path to Evidence-Based Argumentative Writing. *The Journal of the Learning Sciences*, 26(4), 575–607.
<https://doi.org/10.1080/10508406.2017.1336714>
- Hennessey, S., Howe, C., Mercer, N., & Vrikki, M. (2020). Coding classroom dialogue: Methodological considerations for researchers. *Learning, Culture and Social Interaction*, 25.
<https://doi.org/10.1016/j.lcsi.2020.100404>
- Howe, C., Hennessey, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher-student dialogue during classroom teaching: Does it really impact on student outcomes? *Journal of the Learning Sciences*, 28, 1–51.
- [Howe](#), C., [McWilliam](#), D., & [Cross](#), G. (2005). Chance favours only the prepared mind: Incubation and the delayed effects of peer collaboration. *British Journal of Psychology*, 96(1), 67-93.
- Järvelä, S. & Rosé, C. P. (2020). [Advocating for group interaction in the age of COVID-19](#). *The International Journal of Computer-Supported Collaborative Learning*, 15(2), 143-147.
- Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review* 28, 100284.
- Kay, M., Elkin, L.A., Higgins, J.J., & Wobbrock, J.O. (2021). ARTool: Aligned rank transform. [R] package with documentation and two vignettes.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2), 75-86.

- Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argumentation skills in a computer-supported environment. *Child Development, 79*, 1310–1328.
- Kuhn, D., Hemberger, L., & Khait, V. (2016). *Argue with Me. Argument as a Path to Developing Students' Thinking and Writing*. Routledge. Taylor and Francis.
- Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentative competence. *Cognition & Instruction, 31*, 456–496.
- Looi, C.-K., S. Hyo-Jeong So, H.-J., Toh, Y., & Chen, W. (2011). The Singapore experience: Synergy of national policy, classroom practice and design research. *The International Journal of Computer-Supported Collaborative Learning, 6*, 9–37. <https://doi.org/10.1007/s11412-010-9102-8>
- Matuk, C., Tissenbaum, M., & Schneider, B. (2019). [Real-time orchestrational technologies in computer-supported collaborative learning: an introduction to the special issue](#). *The International Journal of Computer-Supported Collaborative Learning, 14*(3), 251–260.
- Michaels, S., O'Connor C., & Resnick, L. B. (2008). Deliberative Discourse Idealized and Realized: Accountable Talk in the Classroom and in Civic Life. *Studies in Philosophy and Education, 27*, 283–297.
- Monte-Sano, C. (2016). Argumentation in history classrooms: a key path to understanding the discipline and preparing citizens. *Theory into practice, 55*(4), 311–319.
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York: Oxford University Press.
- Salter, K. C. & Fawcett, R. F. (1993). The ART test of interaction: A robust and powerful rank test of interaction in factorial models. *Communications in Statistics: Simulation and Computation, 22*(1), 137–153. <https://doi.org/10.1080/03610919308813085>
- Saqr, M., Viberg, O., & Vartiainen, H. (2020). [Capturing the participation and social dimensions of computer-supported collaborative learning through social network analysis: which method and measures matter?](#) *International Journal of Computer-Supported Collaborative Learning, 15*(2), 227–248.
- Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two “wrongs” may make a right...If they argue together! *Cognition & Instruction, 18*(4), 461–494.
- Schwarz, B. B., Schur, Y., Pensso, H., & Tayer, N. (2011). Perspective taking and argumentation for learning the day/night cycle. *The International Journal of Computer Supported Collaborative Learning, 6*(1), 113–138.
- Seixas, P. (2006). Benchmarks of historical thinking: A framework for assessment in Canada. Vancouver, CA: Center for the Study of Historical Consciousness. Retrieved from archive.historybenchmarks.ca/documents/benchmarks-historical-thinking-frameworkassessment-Canada
- [UNESCO](#) (2022). The impact of the COVID-19 pandemic on education: international evidence from the Responses to Educational Disruption Survey (REDS). [International Association for the Evaluation of Educational Achievement](#). ISBN: 978-92-3-100502-2

- Wegerif, R. (2007). *Dialogic education and technology: Expanding the space of learning*. Springer Science & Business Media.
- Wegerif, R. (2008). Dialogic or dialectic? The significance of ontological assumptions in research on educational dialogue. *British Educational Research Journal*, 34(3), 347–361.
- Wise, A. F. & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12, 423–467.
- Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)* (pp. 143–146). Association for Computing Machinery. <https://doi.org/10.1145/1978942.1978963>
- Zheng, L. Cui, P., & Zhang, X. (2020). [Does collaborative learning design align with enactment? An innovative method of evaluating the alignment in the CSCL context](#). *The International Journal of Computer-Supported Collaborative Learning*, 15(2), 193–226.
- Zohar, A. & Boshrian, A. (Eds.) (2020). Adapting curricula and teaching materials to the 21st century — Summary of the work of the expert committee, status report and recommendations. Center for Knowledge and Research in Education, Israel National Academy of Sciences. Jerusalem.
-



New articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by [Pitt Open Library Publishing](#).