Article

Investigating Sequence Patterns of Collaborative Problem-Solving Behavior in Online Collaborative Discussion Activity

Yafeng Zheng [1] **, Haogang Bao** [2] **, Jun Shen** [3] **and Xuesong Zhai** [4,*]

[1] School of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046, China; zlzyf@126.com

[2] Faculty of Education, Beijing Normal University, Beijing 100875, China; haogangbao@126.com

[3] School of Computing and Information Technology, University of Wollongong, Wollongong, NSW 2522, Australia; jshen@uow.edu.au

[4] College of Education, Zhejiang University, Hangzhou 310058, China

[*] Correspondence: xszhai@zju.edu.cn

check for updates

Abstract: Collaborative problem solving (CPS) is an influential human behavior affecting working performance and well-being. Previous studies examined CPS behavior from the perspective of either social or cognitive dimensions, which leave a research gap from the interactive perspective. In addition, the traditional sequence analysis method failed to combine time sequences and sub-problem sequences together while analyzing behavioral patterns in CPS. This study proposes a developed schema for the multidimensional analysis of CPS. A combination sequential analysis approach that comprises time sequences and sub-problem sequences is also employed to explore CPS patterns. A total of 191 students were recruited and randomly grouped into 38 teams (four to six students per team) in the online collaborative discussion activity. Their discussion transcripts were coded while they conducted CPS, followed by the assessment of high- and low- performance groups according to the developed schema and sequential analysis. With the help of the new analysis method, the findings indicate that a deep exploratory discussion is generated from conflicting viewpoints, which promotes improved problem-solving outcomes and perceptions. In addition, evidence-based rationalization can motivate collaborative behavior effectively. The results demonstrated the potential power of automatic sequential analysis with multidimensional behavior and its ability to provide quantitative descriptions of group interactions in the investigated threaded discussions.

Keywords: behavioral sequence patterns; problem-solving task; online collaborative discussion; visualization

1. Introduction

Collaborative problem solving (CPS) has been listed as one of the core competencies in the 21st century [1]. The digital era has witnessed the crucial influence of online CPS on individuals' working performance and well-being [2,3], especially during the post-pandemic recovery. From the perspective of learning science, it has been also demonstrated that CPS can lead to deeper understanding of the topic in computer-supported collaborative learning (CSCL) settings [4]. CPS is defined as individuals' working or learning status when solving a problem by sharing their understanding and pooling their knowledge, skills, and efforts to reach a solution [5], which involves multiple dimensions of collaboration processes, such as knowledge building, process regulation, social interaction, and emotion expression [6]. Therefore, a successful CPS is supposed to be considered and conducted in cognitive and social dimensions [7,8]. However, prior studies on CPS selected either the cognitive or social

perspective as a theoretical foundation to code schemes for evaluating individuals' collaborative behaviors, including discourses and frequencies [9,10], thereby leaving a research gap that the dual dimensions, which combine cognitive and social perspectives in CPS, are still lacking.

Moreover, based on social constructivism theory, interactions and arguments among individuals have been confirmed as a significant factor to facilitate deep learning in CPS [4,11–13]. Understanding how CPS works in enhancing individuals' knowledge construction and improving their learning performance during their interactions is thus of great importance [14]. Sequential analysis is an adaptive approach widely used to analyze time-based CPS patterns through discourse [15,16]. However, CPS is not only affected by time sequences but also by generated sub-problem sequences. Different sub-problems may be raised randomly, which lead to the necessity to analyze CPS by combining time sequences and sub-problem sequences.

Based on the above assumptions, the study proposed to conduct developed CPS in an online learning context, by which the discourse schema was designed in consideration of cognitive and social dimensions. In addition, an automated sequential analysis method was developed to explore the combined new CPS patterns. A qualitative analysis was conducted to examine the CPS behavioral patterns of different collaboration group learners on the basis of the proposed method.

2. Literature Review

2.1. CPS

CPS encourages the ideology that learning is conducted by doing, by which problem-solving consciousness and capacity are inspired in a collaborative context [17]. The Programme for International Student Assessment (PISA) put forward that learners' core competence and well-being largely lean on CPS, which is a joint activity transforming a problem state into a desired goal state during cooperative information exchange [18]. This process relies on how well individuals can establish the ideology in common in terms of the features of the problem, solution plan, assessment, and tolerance to disagreements.

Previous studies confirmed cognitive and social factors as two crucial dimensions when CPS behavior was investigated [9,19]. From the perspective of the cognitive dimension, the joint problem space model was proposed, which comprises the sharing of goals, understanding of problems, and possible solutions [20]. One of the early coding schemata that analyzed the epistemic dimension in the CPS process was proposed by Weinberger and Fischer [21], in which learners' work on a knowledge construction task is analyzed. Based on the findings from cognitive science, Hesse et al. [19] provided a five-stage conceptual framework of a collaborative problem, including problem identification, problem representation, planning, executing, and inspection. The Organization for Economic Co-operation and Development (OECD) extends the individual problem-solving framework into CPS, which contains four cognitive processes: exploring and understanding; representing and formulating; planning and executing; monitoring and reflecting [5]. Sun et al. [22] proposed a generalized competency model for CPS, which contains three primary facets: constructing shared knowledge, negotiation/coordination, and maintaining team function.

Moreover, increasing research puts forward collaboration processes from the social dimension. It is important for students to extract more unshared information through elicitation, externalization, and giving specialized feedback. For example, Andrews-Todd et al. [9] explored the four social skills of CPS, namely, maintaining communication, sharing information, establishing shared understanding, and negotiating. Some studies also found that management behavior has a positive effect on CPS performance [23]. Other studies considered emotional factors as important in analyzing the CPS process. For example, Zheng and Huang [24] adopted an affective coding schema containing six categories, such as positive, negative, neutral, insightful, confused, and joke emotions. Ozcan and Gumus [25] explored the effects of self-efficacy, anxiety, and metacognitive experience on students'

mathematical problem solving. However, how social and cognitive factors work together in the CPS process is still less noticed.

2.2. Approach to Analyzing CPS in Computer-Supported Environments

Computer-supported environments can capture actions and discourses in the information technology-based learning system, which has revealed the mechanism of learners' interaction during task engagement [26–28]. This system offers promising artefacts in measuring complex constructs, such as CPS. Past researchers emphasized the significance of analyzing cognitive and social dimensions associated with discourses and actions in computer-supported environments. For example, Avouris et al. [29] proposed the Object-oriented Collaboration Analysis Framework to analyze students' CPS behavior through discourse and action.

Some studies used basic behavioral data to reflect the effectiveness of online CPS, such as the number of visits to online discussions [30] or the number of posts read by students [31,32]. Additionally, it is suggested that the time students spend viewing posts could be utilized as a predictor to learning achievement [33]. Some researchers are aware of the limitations of only using basic data for analysis. Therefore, they adopt fine-level granularity in data to meaningfully examine student behaviors in online problem-solving discussions. Quantitative content analysis (QCA) was employed to explore the frequency of various discussion behaviors among online learners [34]. This method employed a QCA to code different behavior types and counted the frequency to find the distribution of different behavior types. Such a method is simple and easy to implement, but it can only observe behavior types in the amount of distribution and cannot focus on an interaction behavior change in a collaborative process.

To address the above defects, finding behavioral patterns becomes a new research focus. Behavioral patterns shed light on the examination of the threaded messages offered, and the relationship between the message sequences and problem solving. Given the dynamic and hidden nature of CPS behavioral intention, sequence analysis is one of the most widely used methods to explore the effects of time on CPS patterns. Cheng et al. [35] combined a content analysis method and sequence analysis method to analyze online peer assessment activities, exploring learners' emotional, cognitive, and metacognitive situations in the process of peer assessment with technical support. Employing lag sequential analysis, Chang et al. [36] explored the influence of students' group discussions, problem solving activities, and feedback on their collaboration patterns and problem-solving strategies. Although progress on CPS behavioral pattern analysis has been made, challenges remain. Using one method that only considers time-related characteristics is not enough to reveal how students solve sub-problems (e.g., planning, answering specific questions) during CPS. Researchers argued that the event-centered view should instead be used to tightly link quantitative and qualitative methods [37,38].

2.3. Purpose of the Study

To tackle these challenges, the study proposes a new analysis schema of CPS from the perspectives of cognitive and social dimensions. In addition, an automated sequential analysis method is developed to explore CPS patterns, combining time sequences and sub-problem sequences. Then, the new analysis schema and method are used to analyze interactions among students in a synchronous computer-supported collaborative environment to detect the behavioral sequential patterns of students' CPS processes in different groups. The current study aims to address the following three questions:

1. What are the different behavioral types of CPS that are distributed during the online collaborative problem-solving discussion?
2. What are the differences in the distribution of behavioral types between high- and low-performance groups during the online collaborative problem-solving discussion?
3. What are the differences in the CPS behavioral sequence between high- and low-performance groups during the online collaborative problem-solving discussion?

3. Methods

3.1. Participants

A total of 191 second-year undergraduate students (average age = 19; SD = 1.86) were recruited from two comprehensive universities in northern China. There were 113 female students (59.16%) and 78 male students (40.84%). All the participants majored in educational technology and had basic knowledge of computer programming. They also claimed that they had experience of collaborative discussing on the Moodle platform and had received pre-class tool training. Consistent with an earlier study [39], students were assigned into groups of five by the instructor randomly. Some students quit after grouping because of personal reasons. Finally, there were 29 groups with five students, four groups with four students and five groups with six students.

3.2. Research Ethical

The students were recruited through public media (internet, online forum, and poster). All the students were over 18 years old. The students consented that their performance would be recorded for study during the activities. Students were compensated ¥200 for their participation in this study. Researchers in this study had passed ethical principles training of human subjects. The activity was performed on a private server, students were not required to share any personally identifiable data during the activity, any data linked to the students were only shared by the research team, and identifiable data were stored in a private, encrypted file.

3.3. Learning Activities

The learning activities were chosen from a compulsory course in educational technology called "Data Structure". In the selected experimental class, students were asked to design an algorithm for the collection management program "Collection Auction," which involved knowledge points related to stack and queue. The learning activity was conducted in a blended learning context including a two-hour lesson in a physical classroom and a two-hour online learning activity of programming in a computer laboratory. During the online learning activity, the Moodle platform was used as a collaborative learning tool, in which discussion activities were implemented until they came to a final solution. Instructors designed the problem-solving tasks, and students solved each of the problems in their own group discussion space. The collaborative online discussion lasted about two hours, and each group was assigned a separate space during discussion to avoid potential interference.

3.4. Procedure

Before carrying out the experimental study, a pilot test was conducted with 10 students to determine the feasibility of the study with respect to the learning task, materials, instruments, and the platform. The students in the pilot test were divided into two groups. This pilot study resulted in a slight modification of the description of the learning task. Additionally, the robustness of the platform was improved as well. The data from the pilot study were excluded in the final analysis.

In the formal experiment, the whole session took about 3.5 h and consisted of four main phases (see Table 1): (1) key concept introduction and learning phase, (2) platform training phase, (3) introduction and grouping phase, and (4) collaborative discussion phase.

Table 1. Overview of the procedure of the formal experiment.

Phase	Duration
(1) Key concept introduction and learning phase	60 min
Individual learning from electronic learning materials	20 min
Instruction by researcher	15 min
Open discussion	25 min
(2) Platform training phase	20 min
Instruction on the usage of the collaborative platform	10 min
Tried to use the platform	10 min
(3) Introduction and grouping phase	10 min
Discussion task introduction	5 min
Logging into groups	5 min
(4) Collaborative discussion phase	120 min
Total time	About 3.5 h

During (1) key concept introduction and learning phase, which took 1 h, students needed to learn the key theories and concepts that were required in the discussion task. The students studied the electronic learning materials (e-book and slides) for 20 min individually firstly. Then one researcher gave 15 min of instruction for the key point. In the last 25 min, the students started an open discussion about the learning materials with either classmates or a teacher (researcher). Students could keep the materials and their notes in the following experiment.

During the (2) platform training phase, which took 20 min, the researcher firstly gave instruction on the usage of the collaborative platform (10 min). The instruction included the methods to browse the learning materials, check the group members, post new a discussion, and reply to other posts. Then the students spent 10 min to check the availability of their account and tried to use the platform.

During the (3) introduction and grouping phase, which took 10 min, the researcher introduced the content and requirement of the discussion task (5 min). Then the students logged into their account and joined their groups based on the researcher's instructions (5 min).

During the (4) collaborative discussion phase, which took 2 h, students started their collaborative discussion to achieve the requirement of the task. Specifically, they were asked to analyze and discuss the learning task and give the joint problem solution plan. Only the discussion data in phase (4) were collected and analyzed in this study.

3.5. Online Discussion Environment

The online forum in Moodle was adopted as the discussion tool, where participants ask questions and share ideas. To analyze the process of online collaborative discussion automatically, the posted content of the discussion was required to be annotated on the basis of the behavior type of the behavior coding scheme. Considering the manual annotation method is laborious, this study adopted the method of marking by students' automated selection [40]. The behavior types in the behavior coding table were preset in the Moodle posting area ahead of time. Participants could choose the corresponding behavior types in the drop-down box of content label options while posting on the Moodle platform discussion area. The corresponding behavior coding of the selected behavior type was inserted into the database tables along with the content of the posts, which could be directly invoked by subsequent analysis tools. An example of behavior type annotation in the Moodle posting area is presented in Figure 1.

Figure 1. Behavior type annotation in Moodle posting area.

4. Data Collection and Analysis

4.1. Students' CPS Behavior Collection

To analyze the CPS behavior from the perspectives of cognitive and social dimensions, an analysis scheme was adapted from Meijden's work [41] that comprehensively considered cognitive and social aspects in CPS. The category "Statement" examined the cognitive aspect of the CPS which referred to students' construction of new knowledge through social interaction [42]. This category contained four subcategories, which were: propose opinions/solutions, further explain opinions, revise opinions/solutions, and summarize views/solutions. To pay further attention to behavioral strategy, the categories "Negotiation" and "Asking questions" were proposed. "Negotiation" referred to students' answers in the discussion. It was classified as "Agree," "Agree, give evidence/reference," "Disagree/question," and "Disagree, give evidence." The category "Asking questions" referred to students' behavior of asking a specific question, which was classified as "Asking questions" and "Asking for elaboration/follow-up questions." For the characteristics of synchronous discussions in the social aspect, the "management" category was added. It used "Organize/assign tasks" and "Coordinate/regulate" to indicate organizational and coordination behavior strategies [43]. Considering the importance of affective interaction in problem-solving activities, the category "Share feelings" was added. With regard to attaining reliability in the study, the inter-rater kappa coefficient was also examined. Two analysts coded and classified 755 online problem solving-based collaborative discussion corpora and merged similarities according to classification results to form a behavior classification system. Eventually, six categories and 14 sub-categories of behavioral coding classification were formed. The kappa results were 0.91, greater than 0.7, the acceptable level, which confirmed the reliability of the consistent test.

Table 2 shows the coding schema and an example for each item for behavior type classification based on CPS. Each coding item represents a discussion behavior, and the off-topic discussion, which is irrelevant to C1–C5, is grouped into C6, "others".

Table 2. Coding schema for behavior type classification based on collaborative problem solving (CPS).

Type	Category	Subcategory	Code	Explain	Examples
C1	Statement	Propose opinions/solutions	C11	Provide or introduce new ideas, opinions, recommendations, or plans to identify problems	• It can be stored using an adjacency matrix or an adjacency table. • We can also read data from a file to generate an adjacency table.
		Further explain opinions	C12	Provide a detailed explanation of the point of view and provide further additions, explanations, and descriptions	• We can select sequential stacks and queues as data types, and then check whether the number of data stored is greater than N. • "What is the meaning of both lines are stored at the same time?" "It is stored in an adjacent table."
		Revise opinions/solutions	C13	Revise or refine the views already expressed	• Uh-huh, I also think that we should develop two lines according to the depth-first search and the breadth-first search.
		Summarize views/solutions	C14	Synthesize viewpoints or discourses to make refinements and generalizations	• Summary: Adjacency matrix can store undirected graphs while traversal using the depth-first search and breadth-first search. • We transform the pictures into a directed network and use the adjacency matrix for storage, and then use the depth-first search traversal and breadth-first-search traversal to output two paths.
C2	Negotiation	Agree	C21	Support the views of others	• Well, I think it should be. • Agree!
		Agree and give evidence	C22	Support the views of others and give the reasons for such support	• I think so. The advantages of the adjacency matrix are … • Yes, I think we can easily be confused between directed and undirected storage and output. Storage should be undirected, whereas output should be in order.
		Disagree	C23	Challenge or oppose the views of others, give further statements, usually as a form of questioning	• The undirected map seems to have no direction, right? It cannot be considered a two-way kind, as no direction exists.
		Disagree and give evidence	C24	Object to the views of others and justify the opposition	• No, I think if you store a directed map, it can be separated … • Adjacency matrix takes up too much memory. Whether it is a directed graph or an undirected graph, the degree is less.

Table 2. *Cont.*

Type	Category	Subcategory	Code	Explain	Examples
C3	Asking questions	Ask questions	C31	Put forward doubts and express what is unclear	• How do you store the presentation information? • It is the undirected graph, right?
		Ask for further explanations	C32	Ask for further information to clarify interpretations, which are usually in the form of questions	• Yuanyuan, how do you determine the direction of a directed map? • I agree with you, but why is it stored by the adjacency matrix?
C4	Management	Organize/Assign tasks	C41	Organize implementation plans, the resource management, and the allocation management	• First, we should discuss the advantages and disadvantages of various storage methods. Second, we should determine the reasons for the use of the adjacency table.
		Coordinate/Remind	C42	Manage and remind about collaborations and time schedules	• Do not open a new post; we will not know where to discuss. • We can build a new floor in the forum, which concludes all of our final results and put the results there.
C5	Share feelings	Positive emotions	C51	Express greetings, mutual introductions, and emotional support	• Interesting idea! • Forgive me for ignoring!
		Other emotions	C52	Express anger, frustration, shock, and difficulty	• It's too difficult!
C6	Others	Other behaviors		Expressions irrelevant to tasks	• Shall we hold the meeting after class?

In order to make a comparison of the discussion activity between high- and low-performance groups, the discussion scripts from the participants were graded. Referring to previous research on collaborative project learning [44], two experts on data structure were selected to evaluate the quality of collaborative discussion. The evaluation criteria were based on the extent to which the discussion covered the target knowledge and contributed to the problem solving. The assessment criteria scoring from 1 to 10 was classified into five levels, and the detailed evaluation criteria are shown in Table 3. To ensure the inter-rater reliability, Spearman's rho correlation coefficient was tested, and the coefficient was 0.755, achieving the acceptable range. Finally, we took the mean value of the two raters' scores as the discussion quality score.

Table 3. The evaluation criteria of collaborative discussion quality.

Level	Evaluation Criteria
Level 5 (9–10)	Fully covers the target knowledge; finds the solution to the task
Level 4 (7–8)	Covers the majority of the target knowledge; finds the task solution, but some details are inaccurate
Level 3 (5–6)	Covers some of the target knowledge; presents the task solution with noticeable errors
Level 2 (3–4)	Covers only a limited range of the target knowledge; fails to find the task solution
Level 1 (1–2)	Fails to address the task; answer is completely unrelated to the task

4.2. Mining the Sequence Pattern of CPS Behavior

In computer science, association rule method is a suitable method to find repeated patterns and sequence relations in datasets. It can discover the association among discussion behavior occurrences, which can help us understand which behavior sequence occurs frequently. To identify the sequential patterns of CPS behaviors with time and sub-problems, an improved Apriori algorithm [45] was used in this study.

First, according to the structure of students' discussions, a tree structure was formed and formalized to represent different sub-problems in the discussions. Second, the behavior sequences in each sub-tree (sub-problems) were coded and analyzed to arrange the sequence pairs on the basis of the timestamp. The acquired behavior sequence pairs were stored in a set. Third, the "support rate" and "confidence rate" between different behavior sequence pairs were calculated via an improved Apriori algorithm. Last, based on these frequent behavioral patterns, the high frequency conversion mode of behavior state was visualized.

5. Results

5.1. Distribution of CPS Behavioral Types

The distribution of the 1577 coded annotated messages gathered from 31 groups in the Moodle database is shown in Table 3. Codes of C6 were not found and thus are not shown.

Table 4 presents that the most frequent behavior is "Statement" (C1, 46.78%), followed by "Asking questions" (C3, 21.84%), and "Negotiation" (C2, 17.52%). "Management" (C4, 9.38%) and "Sharing feelings" (C5, 4.48%) are less common.

Table 4. Descriptive statistical table of the first-level dimension behavior distribution.

	Statement (C1)	Negotiation (C2)	Asking Questions (C3)	Management (C4)	Sharing Feelings (C5)
Number	878	329	410	176	84
Percentage	46.78%	17.52%	21.84%	9.38%	4.48%
Standard Deviation (SD)	8.57	8.23	8.24	5.17	4.04
Minimum (Min)	1	0	0	0	0
Maximum (Max)	43	41	36	23	18

Furthermore, the distribution of second-level dimensions is shown in Table 5. The most frequent behavior is "Propose opinions/solutions" (C11), followed by "Ask questions" (C31), "Agree" (C21), "Organize/Assign tasks" (C41), "Coordinate/Remind" (C42), "Further explain opinions/solutions" (C12), and "Ask for elaboration/follow-up questions" (C32). Other behaviors are less common, especially "Summarize views/solutions" (C14) and "Agree and give evidence" (C22), which account for only 1.87% and 1.72% of all the behaviors, respectively.

Table 5. Descriptive statistical table of second-level dimension behavior distribution.

	C11	C12	C13	C14	C21	C22	C23	C24	C31	C32	C41	C42	C51	C52
Number	622	124	84	48	146	56	44	83	284	126	83	93	41	43
Percentage	0.31	0.07	0.05	0.03	0.08	0.03	0.02	0.04	0.15	0.07	0.04	0.05	0.02	0.02
SD	8.26	2.87	2.52	1.87	4.29	1.72	2.45	3.23	7.45	2.80	3.18	3.35	2.81	2.03
Min	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Max	43	15	11	8	21	8	10	12	32	13	16	14	11	7

5.2. Comparison of the Distributions of CPS Behavioral Types between High- and Low-Performance Groups

The research analyzed the distributions of the behavioral transformations of high- and low-performance groups to explore the differences in behavior conversions among different groups. To investigate the characteristics of the distributions of behavior types in high- and low-performance groups, 31 groups were ranked according to the evaluation results of their discussion outcome. The first 27% of groups were selected as high-performance groups, whereas the last 27% of groups were low-performance groups. The high- and low-performance groups consisted of eight groups.

Table 6 shows the distributions of high- and low-performance groups in the second-level dimension. In high-performance groups, the most frequent behavior was C11 (Propose opinions and solutions), followed by C31 (Ask questions), C24 (Disagree, give evidence), and C13 (Revise opinions/solutions). In low-performance groups, the most frequent behavior was also C11 (Propose opinions/solutions), followed by C31 (Ask questions), C21 (Agree), and C32 (Ask for elaboration/follow-up questions). C3 (Revise opinions/solutions) was the least frequent.

Table 6. Behavior distributions of high- and low-performance groups in the second-level dimension.

		C11	C12	C13	C14	C21	C22	C23	C24	C31	C32	C41	C42	C51	C52
High-quality group	Number of posts	168	19	35	13	45	26	19	37	51	29	18	31	16	14
	Percentage	32%	4%	7%	2%	8%	5%	4%	7%	9%	6%	4%	6%	3%	3%
Low-quality group	Number of posts	149	42	6	9	46	12	11	11	105	43	33	21	12	14
	Percentage	29%	8%	1%	2%	9%	2%	2%	2%	20%	9%	7%	4%	2%	3%

"Propose opinions/solutions" (C11) was the main behavior in high- and low-performance groups. However, significant differences were observed between both groups in the second dimension of the statement. "Disagree, give evidence" (C24) occupied a high proportion in high-performance groups (7%). On the contrary, the proportion of this behavior in low-quality groups was 2%. In addition, the "Agree" (C31) in low-performance groups was 20%, but in high-quality groups, it was less (9%). Finally, in high-performance groups, the "Revise opinions/solutions" (C13) occurred at a high frequency (7%), whereas in low-performance groups, this behavior was 1%. In high-performance groups, the total proportion of "Disagree, give evidence" (C24) and "Agree, give evidence" (C22) reached 12%. However, in low-performance groups, these two evidence-related behaviors together accounted for 4%.

5.3. Different Behavioral Sequences between High- and Low-Performance Groups

To further explore the differences in the behavioral patterns of different groups, the proposed behavioral pattern analysis method was used to conduct a sequential analysis of the coded operations of high- and low-performance groups. The results are shown in Tables 7 and 8. A total of 121 conversion

sequences were obtained in high-performance groups. Based on the extraction rules with a support ratio ranking in the top 10%, 12 behavior conversions were extracted, whereas one sequence of behavior with a confidence level of less than 10% was removed. Thus, 11 conversion sequences were obtained. Figure 2 illustrates the behavior conversion for high-performance groups.

Table 7. Frequency behavior conversion patterns of high-performance groups.

Type	Frequency	Support	Confidence	Type	Frequency	Support	Confidence
C11→C11	75	14.70%	21.92%	C32→C11	12	2.35%	24.49%
C11→C31	42	8.23%	12.28%	C21→C11	12	2.35%	17.91%
C42→C42	17	3.33%	21.79%	C22→C11	11	2.15%	25.58%
C31→C11	16	3.14%	12.50%	C21→C21	11	2.15%	16.14%
C31→C31	14	2.74%	10.93%	C13→C11	11	2.15%	17.74%
C24→C11	13	2.55%	21.67%				

Table 8. High-frequency behavior conversion for low-performance groups.

Type	Frequency	Support	Confidence	Type	Frequency	Support	Confidence
C11→C11	75	14.88%	22.06%	C32→C11	15	2.98%	20.55%
C11→C31	43	8.53%	12.65%	C41→C41	14	2.78%	20.89%
C31→C31	24	4.76%	12.12%	C12→C31	12	2.38%	17.65%
C21→C11	22	4.36%	32.35%	C12→C11	11	2.18%	16.18%
C31→C11	20	3.97%	10.10%				
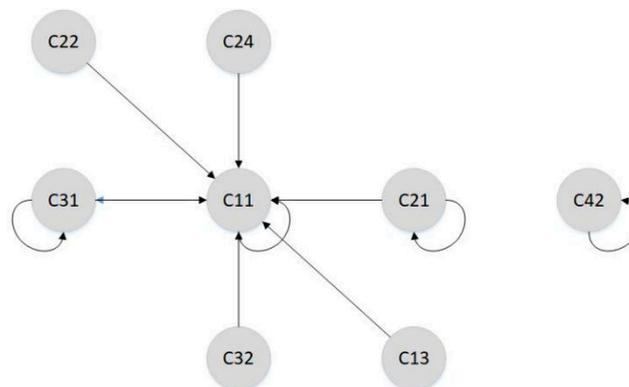
Figure 2. Behavior conversion for high-performance groups.

In low-performance groups, 107 conversion sequences were obtained. On the basis of the same extraction rule as high-performance groups, nine conversion sequences were obtained (shown in Table 8). Figure 3 presents the behavior conversion for low-performance groups.
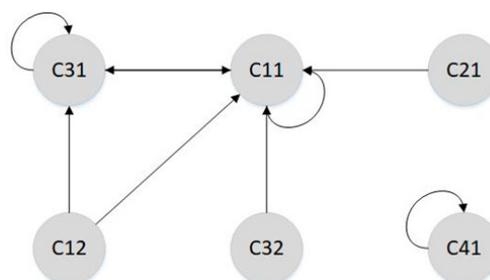
Figure 3. Behavior conversion for low-performance groups.

The automatic extraction of behavior conversion clearly shows that high-performance groups present a high proportion of "Propose opinions/solutions" (C11)→"Revise opinions/solutions" (C13) and "Revise opinions/solutions" (C13)→"Revise opinions/solutions" (C13), whereas low-performance groups do not have any behavior conversions associated with "Revise opinions/solutions" (C13). At the same time, in the high-frequency behavior transition, high-performance groups have a high proportion of "Disagree, give evidence" (C24)→"Propose opinions/solutions" (C11) and "Agree, give evidence/reference" (C22)→"Propose opinions/solutions" (C11). In low-performance groups, no conversion behavior is observed between "Disagree, give evidence" (C24) and "Agree, give evidence/reference" (C22). An obvious difference is that high-performance groups have a high proportion of "Coordinate/remind" (C42)→"Coordinate/remind" (C42); in low-performance groups, "Organize/assign tasks (C41)"→"Organize/assign tasks (C41)" is observed.

6. Discussion

This study reveals several implications for educational practice. First of all, the study reveals some overall characteristics of students' CPS behavioral types in the collaborative discussion. The results show that the "Statement (C1)" is most frequency behavior in the discussion. This result corresponds to the prior studies that found that the interactions about knowledge construction compose the main part of the CPS process [21]. A statement in CPS is a kind of interaction that enriches the learning material by additional information. This could promote knowledge acquisition and quality of the argumentation in the CSCL. Regarding the second-level indicators, C14 (Summarize views/solutions) is the least frequent behavior in the statement category. This corresponds to the earlier study that found thay integration consensus takes place rarely, as learners seem to hardly elaborate on the change of their perspectives in discourse [21].

The comparison between the group distributions of CPS behavioral types shows that during collaborative discussions, high-performance groups can revise and improve their views on the basis of the information shared by the group and finally converge on a common solution after continuously proposing, proving, opposing, and arguing the solution to the problem. This process continues to advance collaborative tasks and ultimately contribute to the success of collaborative activities. On the contrary, low-performance groups only provided options but did not summarize and revise them.

Additionally, our findings confirm that in collaborative discussions, if the knowledge construction level is only in sharing and comparing views, then it is not enough to promote the generation of new knowledge, which is also consistent with Shukor et al.'s research [46]. To reach a high-level knowledge building context, students tend to express their own ideas through debates, defenses, and decision-making. These attributes help students become critical and thus be able to build new knowledge [47]. Although a few studies also mentioned that revising others' viewpoints is an attribute of high-level knowledge building [48], they did not further explore whether the high proportion of revised viewpoints in the discussion can be used as an important observational indicator of high-level knowledge construction. In this study, through the comparison of statement coding items between high- and low-performance groups, the revise opinion behavior was used as an important externalized attribute of new knowledge construction to identify whether the groups reached high-level knowledge construction. Thus, the "Revise opinions/solutions" behavior can be taken as an important indicator of high-level knowledge construction.

Moreover, our data suggested that high-performance groups are more controversial during the collaborative discussion process, whereas low-performance groups are more expressive as an echo. Debate during the collaborative process is an important indication of the in-depth discussion of a group. After the high-performance group members put forward their opinions, they also questioned the stated statements and provided sufficient explanations. The final debate provoked the group members to reach a consensus. On the contrary, in low-performance groups, the lack of understanding of the problem made it easy for the members to agree with others' views, rather than identify the problems and question them. Collaborative learning is an important approach to trigger debates, which also

confirms Jonassen's findings that if argumentation-based teaching is implemented, problem-solving and critical thinking can be generated [49].

Our results also reveal that during the negotiation process, high-performance groups used more evidence-based behavioral strategies, whereas this was not found in low-performance groups. Evidence data in collaboration can be used to measure its quality. High-performance groups consulted many times and often pushed forward the problems' solutions in the form of questioning and giving important evidence, whereas low-performance group members preferred to accept the opinions of others directly. A previous study suggested that using evidence is considered one of the most important dimensions in improving the knowledge understanding of students and in developing their argumentation ability [50]. Our results indicated that using evidence in the process of supporting, evaluating, questioning, or refuting a view reflects the cognitive context of students [51]. The way students use data and evidence can show how they interpret and evaluate information fragments and how they transform information into a part of their knowledge. In high-performance collaborative interaction, individuals can examine each other's ideas and rationalize them with claims and evidence, thereby allowing group discussions to focus on the target effectively for enhancing the effect of learning. Therefore, reasoning, evaluating alternatives, presenting evidence, and weighing the reliability of evidence are suggested to promote the depth and quality of collaboration [52]. From a practical point of view, teachers should pay attention to group members' usage of evidence and encourage students to use solid evidence to support their own ideas and opinions in discussion. That is, students should explain their opinions to one another, rather than simply check the views and answers.

The sequential analysis revealed that the high-performing and low-performing groups applied different process patterns of CPS during the collaboration. When comparing the high frequency behavior conversions of high- and low-performance groups, the high frequency conversion of "Coordinate/Regulate" (C42) to itself existed in high-performance groups, that is, after the management behavior related to the task was performed, it was directly transferred into the discussion activity; in low-performance groups, high-frequency conversion was observed in "Organize/Assign tasks" (C41) to itself, indicating that they were continuously falling into management problems during the collaborative discussion. Therefore, high-performance groups have a better management status than low-performance groups.

Students' management of task plans, learning resources, task processes, and time has a positive significance for achieving high-performance group collaboration. Jahng and Nielsen [23] highlighted the importance of management status to group collaboration in their collaborative learning analysis framework and found that management status has a positive significance for achieving high-quality group collaboration. Our study also confirmed this finding. High-performance groups can coordinate the team collaboration process in terms of time planning, conflict resolution, and technical problem support, thus achieving a good collaboration quality. Due to the lack of group management, low-performance groups have problems in conducting meaningful negotiations, narrowing the gap between views, and overcoming the conflict of personal opinions, which may lead group members to give up complex discussions and keep dialogues at a superficial level. Recent studies on regulated learning in the collaborative learning context also showed that students' learning plans, time, and behavior management can lead to an improved performance in group emotion, confidence, motivation, and task interest. These efforts made an improved CPS possible, resulting in improved learning outcomes in the collaborative setting [53,54].

7. Conclusions

On the basis of the cognitive and social perspective, this study provided a quantitative study to analyze the CPS behavioral patterns of students between high- and low-performance discussion teams with a developed scheme and a sub-problem-related sequence analysis method. The revision behavior in the collaboration process can be used as an important indicator of high-level knowledge construction. Arguments involving joint decision-making are beneficial to discussion tasks. In addition,

high-performance group members can provide sufficient evidence whether they agree or disagree with others' opinions during discussions. These behavioral strategies help high-performance groups engage in clear goal-based consultations, which make the problem-solving process continuously develop in depth. Another finding is that students' management of learning processes can promote groups' academic discussions and learning performance.

Our results also revealed that by using the multidimensional behavior feature schema and sub-problem-related sequence analysis, we can build accurate mining models that can automatically identify CPS patterns for collaborative discussions. Our analysis developed an automatic method, including the identification of behavior classification, and presented tools that can be used to provide a quick, accurate group behavioral pattern. The results demonstrated the potential power of automatic sequential analysis and its ability to provide quantitative descriptions of group interactions in the investigated threaded discussions. CPS is a process of knowledge construction based on certain interaction modes of collaborative team members facing complex problems, which inevitably imply specific behavioral strategies. The adoption of an automated method to extract interaction behavior sequences in collaboration groups can help educators and learning designers to further understand the knowledge construction process of online collaborative discussion. It can provide guidance and suggestions for teachers in online collaborative problem-solving activities and provide a basis for designing improved collaborative teaching strategies.

However, a limitation of this study is that the proposed data mining method is verified only in one collaborative discussion course. Additional studies can replicate and extend the results of this research by examining other collaborative discussion activities. In this study, the epistemic dimension of CPS is not focused upon. Thus, the study did not explore how students respond to the learning challenges in the discussion. A future study could give insight into the epistemic aspect in CPS to find out how students construct new knowledge while they are solving problems. The other limitation of this study is the effect of the group composition and group size on the CPS process. In the future, more studies should be designed to explore different CPS processes with different group situations [55].

Further research may also focus on investigating behavioral transformation modes at different stages of activities, which can be extracted and compared, to gain a deep understanding of behavioral strategies in online collaborative discussion activities. Such research can also help teachers further understand the interactive process of online collaborative discussion and provide suggestions and a basis for teaching feedback. Another future research direction may focus on how to use these results to implement appropriate intervention strategies and timing that can support students' problem solving and knowledge construction.

References

1. Neubert, J.C.; Mainert, J.; Kretzschmar, A.; Greiff, S. The Assessment of 21st Century Skills in Industrial and Organizational Psychology: Complex and Collaborative Problem Solving. *Ind. Organ. Psychol.* **2015**, *8*, 238–268. [CrossRef]

2. Risopoulos-Pichler, F.; Daghofer, F.; Steiner, G. Competences for Solving Complex Problems: A Cross-Sectional Survey on Higher Education for Sustainability Learning and Transdisciplinarity. *Sustainability* **2020**, *12*, 6016. [CrossRef]

3. Kim, A.Y.; Sim, I.O. Communication Skills, Problem-Solving Ability, Understanding of Patients' Conditions, and Nurse's Perception of Professionalism among Clinical Nurses: A Structural Equation Model Analysis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4896. [CrossRef]

4.	Noroozi, O.; Biemans, H.J.A.; Busstra, M.C.; Mulder, M.; Chizari, M. Differences in Learning Processes Between Successful and Less Successful Students in Computer-Supported Collaborative Learning in the Field of Human Nutrition and Health. *Comput. Hum. Behav.* **2011**, *27*, 309–318. [CrossRef]

5.	OECD. *PISA 2015 Assessment and Analytical Framework*; OECD Publishing: Paris, France, 2017. [CrossRef]

6.	Care, E.; Scoular, C.; Griffin, P. Assessment of Collaborative Problem Solving in Education Environments. *Appl. Meas. Educ.* **2016**, *29*, 250–264. [CrossRef]

7.	Baker, K.; Greenberg, S.; Gutwin, C. Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration. In *Engineering for Human-Computer Interaction*; Little, M.R., Nigay, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2001; Volume 2254, pp. 123–139. [CrossRef]

8.	Hadwin, A.F.; Järvelä, S.; Miller, M. Self-Regulation, Co-Regulation, and Shared Regulation in Collaborative Learning Environments. In *Handbook of Self-Regulation of Learning and Performance*, 2nd ed.; Schunk, D., Greene, J., Eds.; Routledge Custom: New York, NY, USA, 2018; pp. 83–106. [CrossRef]

9.	Andrews-Todd, J.; Forsyth, C.M. Exploring Social and Cognitive Dimensions of Collaborative Problem Solving in An Open Online Simulation-based Task. *Comput. Hum. Behav.* **2020**, *104*, 105759. [CrossRef]

10.	De Wever, B.; Schellens, T.; Valcke, M.; Van Keer, H. Content Analysis Schemes to Analyze Transcripts of Online Asynchronous Discussion Groups: A Review. *Comput. Educ.* **2006**, *46*, 6–28. [CrossRef]

11.	Jermann, P.; Dillenbourg, P. Group Mirrors to Support Interaction Regulation in Collaborative Problem Solving. *Comput. Educ.* **2008**, *51*, 279–296. [CrossRef]

12.	Pear, J.J.; Crone-Todd, D.E. A Social Constructivist Approach to Computer-Mediated Instruction. *Comput. Educ.* **2002**, *38*, 221–231. [CrossRef]

13.	Oxenswärdh, A.; Persson-Fischier, U. Mapping Master Students' Processes of Problem Solving and Learning in Groups in Sustainability Education. *Sustainability* **2020**, *12*, 5299. [CrossRef]

14.	Hmelo-Silver, C.E.; Jordan, R.; Liu, L.; Chernobilsky, E. Representational Tools for Understanding Complex Computer-Supported Collaborative Learning Environments. In *Analyzing Interactions in CSCL*; Puntambekar, S., Erkens, G., Hmelo-Silver, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 12. [CrossRef]

15.	Lin, P.; Hou, H.; Wu, S.; Chang, K. Exploring College Students' Cognitive Processing Patterns During a Collaborative Problem-Solving Teaching Activity Integrating Facebook Discussion and Simulation Tools. *Int. High. Educ.* **2014**, *22*, 51–56. [CrossRef]

16.	Lin, P.; Hou, H.; Chang, K. The Development of a Collaborative Problem Solving Environment that Integrates a Scaffolding Mind Tool and Simulation-based Learning: An Analysis of Learners' Performance and Their Cognitive Process in Discussion. *Interact. Learn. Environ.* **2020**. [CrossRef]

17.	Nelson, L.M. Collaborative Problem Solving. In *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory*; Reigeluth, C.M., Ed.; Erlbaum Associates: Mahwah, NJ, USA, 1999; pp. 241–267.

18.	Programme for International Student Assessment Survey. 2012. Available online: http://www.oecd.org/pisa/aboutpisa/ (accessed on 13 October 2020).

19.	Hesse, F.; Care, E.; Buder, J.; Sassenberg, K.; Griffin, P. A Framework for Teachable Collaborative Problem Solving Skills. In *Assessment and Teaching of 21st Century Skills*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 37–56.

20.	Roschelle, J.; Teasley, S.D. The Construction of Shared Knowledge in Collaborative Problem Solving. *Int. J. Comput. Support. Collab. Learn.* **1995**, *128*, 69–97. [CrossRef]

21.	Weinberger, A.; Fischer, F. A Framework to Analyse Argumentative Knowledge Construction in Computer-Supported Collaborative Learning. *Comput. Educ.* **2006**, *46*, 71–95. [CrossRef]

22.	Sun, C.; Shute, V.J.; Stewart, A.; Yonehiro, J.; Duran, N.; D'Mello, S. Towards a Generalized Competency Model of Collaborative Problem Solving. *Comput. Educ.* **2020**, *143*, 103672. [CrossRef]

23.	Hadwin, A.F.; Järvelä, S.; Miller, M. Self-regulated, Co-Regulated, and Socially Shared Regulation of Learning. In *Handbook of Self-Regulation of Learning and Performance*; Zimmerman, B.J., Schunk, D.H., Eds.; Routledge Custom: New York, NY, USA, 2011; pp. 65–84. [CrossRef]

24.	Zheng, L.; Huang, R. The Effects of Sentiments and Co-regulation on Group Performance in Computer Supported Collaborative Learning. *Int. High. Educ.* **2015**, *28*, 59–67. [CrossRef]

25.	Ozcan, Z.C.; Gumus, A.E. A Modeling Study to Explain Mathematical Problem-Solving Performance through Metacognition, Self-efficacy, Motivation, and Anxiety. *Aust. J. Educ.* **2019**, *63*, 116–134. [CrossRef]

26. Sun, G.; Shen, J. Facilitating Social Collaboration in Mobile Cloud-Based Learning: A Teamwork as A Service (TaaS) Approach. *IEEE Trans. Learn. Technol.* **2014**, *7*, 207–220. [CrossRef]

27. Honey, M.A.E.; Hilton, M.E. *Learning Science through Computer Games and Simulations*; National Academies Press: Washington, DC, USA, 2011.

28. Quellmalz, E.S.; Pellegrino, J.W. Technology and Testing. *Science* **2009**, *323*, 75–79. [CrossRef]

29. Avouris, N.; Dimitracopoulou, A.; Komis, V. On Analysis of Collaborative Problem Solving: An Object-Oriented Approach. *Comput. Hum. Behav.* **2003**, *19*, 147–167. [CrossRef]

30. Webb, E.; Jones, A.; Barker, P.; van Schaik, P. Using E-learning Dialogues in Higher Education. *Innov. Educ. Teach. Int.* **2004**, *41*, 93–103. [CrossRef]

31. Hamann, K.; Pollock, P.H.; Wilson, B.M. Learning from "Listening" to Peers in Online Political Science Classes. *J. Polit. Sci. Educ.* **2009**, *5*, 1–11. [CrossRef]

32. Palmer, S.; Holt, D.; Bray, S. Does The Discussion Help? The Impact of A Formally Assessed Online Discussion on Final Student Results. *Br. J. Educ. Technol.* **2008**, *39*, 847–858. [CrossRef]

33. Morris, L.V.; Finnegan, C.; Wu, S. Tracking Student Behavior, Persistence, and Achievement in Online Courses. *Int. High. Educ.* **2005**, *8*, 221–231. [CrossRef]

34. Lin, P.; Hou, H.; Wang, S.; Chang, K. Analyzing Knowledge Dimensions and Cognitive Process of a Project-Based Online Discussion Instructional Activity Using Facebook in an Adult and Continuing Education Course. *Comput. Educ.* **2013**, *60*, 110–121. [CrossRef]

35. Cheng, K.H.; Liang, J.C.; Tsai, C.C. Examining the Role of Feedback Messages in Undergraduate Students' Writing Performance during an Online Peer Assessment Activity. *Int. High. Educ.* **2015**, *25*, 78–84. [CrossRef]

36. Chang, C.; Chang, M.; Chiu, B.; Liu, C.; Fan Chiang, S.; Wen, C.; Chen, W. An Analysis of Student Collaborative Problem Solving Activities Mediated by Collaborative Simulations. *Comput. Educ.* **2017**, *114*, 222–235. [CrossRef]

37. Reimann, P. Time Is Precious: Variable- and Event-Centred Approaches to Process Analysis in CSCL Research. *Int. J. Comput. Support. Collab. Learn.* **2009**, *4*, 239–257. [CrossRef]

38. Molenaar, I. Advances in Temporal Analysis in Learning and Instruction. *Front. Learn. Res.* **2014**, *2*, 15–24. [CrossRef]

39. Su, Y.; Li, Y.; Hu, H.; Rosé, P.C. Exploring College English Language Learners' Self and Social Regulation of Learning during Wiki-Supported Collaborative Reading Activities. *Int. J. Comput. Support Collab. Learn.* **2018**, *13*, 35–60. [CrossRef]

40. Jeong, A.C. The Sequential Analysis of Group Interaction and Critical Thinking in Online. *Am. J. Distance Educ.* **2003**, *17*, 25–43. [CrossRef]

41. Van der Meijden, H. Knowledge Construction through CSCL: Student Elaborations in Synchronous, Asynchronous and Three-Dimensional Learning Environments. Ph.D. Thesis, Radboud University, Nijmegen, The Netherlands, 2005. Unpublished.

42. Isaacs, E.; Walendowski, A.; Ranganathan, D. Hubbub: A Sound-Enhanced Mobile Instant Messenger that Supports Awareness and Opportunistic Interactions. In Proceedings of the CHI 2002, Minneapolis, MN, USA, 20–25 April 2002; ACM Press: New York, NY, USA, 2002. [CrossRef]

43. Nardi, B.A.; Whittaker, S.; Bradner, E. Interaction and Outeraction: Instant Messaging in Action. In Proceedings of the CSCW 2000, Philadelphia, PA, USA, 2–6 December 2000; ACM Press: New York, NY, USA, 2000. [CrossRef]

44. Hou, H.; Wu, S. Analyzing The Social Knowledge Construction Behavioral Patterns of An Online Synchronous Collaborative Discussion Instructional Activity Using An Instant Messaging Tool: A Case Study. *Comput. Educ.* **2011**, *57*, 1459–1468. [CrossRef]

45. Wu, X.; Kumar, V.; Ross Quinlan, J.; Ghosh, J.; Yang, Q.; Motoda, H.; Steinberg, D. Top 10 Algorithms in Data Mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]

46. Shukor, N.A.; Tasir, Z.; Van der Meijden, H.; Harun, J. Exploring Students' Knowledge Construction Strategies in Computer-Supported Collaborative Learning Discussions Using Sequential Analysis. *J. Educ. Technol. Soc.* **2014**, *17*, 216–228.

47. McLoughlin, C.; Luca, J. Cognitive Engagement and Higher Order Thinking through Computer Conferencing: We Know Why but Do We Know How? In *Flexible Futures in Tertiary Teaching, Proceedings of the 9th Annual Teaching Learning Forum, Perth, Australia, 2–4 February 2000*; Herrmann, A., Kulski, M.M., Eds.; Curtin University of Technology: Perth, Australia, 2000.

48. Zhu, E. Meaning Negotiation, Knowledge Construction, and Mentoring in a Distance Learning Course. *Class. Commun.* **1996**, *25*, 821–844.

49. Jonassen, D.H.; Kim, B. Arguing to Learn and Learning to Argue: Design Justifications and Guidelines. *Educ. Technol. Res. Dev.* **2010**, *58*, 439–457. [CrossRef]

50. Manz, E.; Renga, I.P. Understanding How Teachers Guide Evidence Construction Conversations. *Sci. Educ.* **2017**, *101*, 584–615. [CrossRef]

51. Sandoval, W.A.; Millwood, K.A. The Quality of Students' Use of Evidence in Written Scientific Explanations. *Cogn. Instr.* **2005**, *23*, 23–55. [CrossRef]

52. Maloney, J.; Simon, S. Mapping Children's Discussions of Evidence in Science to Assess Collaboration and Argumentation. *Int. J. Sci. Educ.* **2007**, *28*, 1817–1841. [CrossRef]

53. Järvelä, S.; Hadwin, A.F. New Frontiers: Regulating Learning in CSCL. *Educ. Psychol.* **2013**, *48*, 25–39. [CrossRef]

54. Dindar, M.; Järvelä, S.; Järvenoja, H. Interplay of Metacognitive Experiences and Performance in Collaborative Problem Solving. *Comput. Educ.* **2020**, *154*, 103922. [CrossRef]

55. Noroozi, O.; Weinberger, A.; Biemans, H.J.A.; Mulder, M.; Chizari, M. Argumentation-Based Computer Supported Collaborative Learning (abcscl): A Synthesis of 15 Years of Research. *Rev. Educ. Res.* **2012**, *7*, 79–106. [CrossRef]