

# Learning gains in Pyramid computer-supported collaboration scripts: factors and implications for design

Ishari Amarasinghe <sup>[0000-0003-2960-4804]</sup> , Davinia Hernández-Leo <sup>[0000-0003-0548-7455]</sup>,  
Emily Theophilou <sup>[0000-0001-8290-9944]</sup>, J. Roberto Sánchez Reina <sup>[0000-0002-6068-1229]</sup>, and  
René Alejandro Lobo Quintero <sup>[0000-0003-2989-5357]</sup>

ICT Department, Universitat Pompeu Fabra, Barcelona, Spain  
{ishari.amarasinghe, davinia.hernandez, emily.theophilou,  
roberto.sanchez, renealejandro.lobo}@upf.edu

**Abstract.** Computer-Supported Collaborative Learning (CSCL) scripts aim to structure the process of collaboration creating opportunities for productive social interaction and learning. Despite CSCL research has shown these benefits for some scripts in particular contexts, more evidence is needed about to what extent learning gains are actually achieved for more families of scripts and in different conditions of implementation. This paper studies how three CSCL scripts based on the Pyramid collaborative learning flow pattern facilitate students learning in online classes. Learning gains are measured in terms of precision and confusion assessment criteria. Students' behaviour in the learning process, regarding agreement in the knowledge exchange, is also analysed in relation to the learning gains. Results bring out several factors, and implications for the design of fruitful Pyramid scripts implementation, that related to the pedagogical envelope, the type of tasks, pyramid design elements, the need for epistemic orchestration, and debriefing.

**Keywords:** Computer-Supported Collaborative Learning, Scripts, Collaborative Learning Flow Patterns, Learning Gains, Pyramid Script.

## 1 Introduction

Free collaboration does not necessarily induce productive interactions among students and therefore does not offer a guarantee of leading to effective learning gains [2]. Directly influencing desired students' interactions by adhering to different strategies, such as using collaborative learning scripts, are seen to enhance students' interactions during collaboration. Collaborative learning scripts prescribe how and when learners may interact, how learner groups require to be formulated, and how learners are required to solve problems together [7, 8]. In other words, scripts structure collaborative learning processes and can be seen as a detailed and explicit contract between teachers and learners [2]. Such guidance becomes important as learners may have a limited understanding of how to interact in collaborative learning situations, to share useful information, to build, and engage in fruitful argumentation [9]. A computational representation of a collaborative learning script forms what is known as a Computer-

Supported Collaborative Learning (CSCL) script, that provides opportunities to mediate and guide the implementation of the collaborative learning activities both in in-person and online learning situations.

The notion of Collaborative Learning Flow Patterns (CLFPs) refers to an abstraction of broadly accepted collaboration scripts that structure the flow of collaboration [5]. Some of the well-known examples of CLFPs include Pyramid, Jigsaw, Think-Pair-Share (TPS), and Thinking Aloud Pair Problem Solving (TAPPS) [5]. Through their design rationale, these CLFPs aim to structure collaboration. For instance, the Pyramid CLFP is expected to support learners in reaching an agreed solution for a complex problem that does not usually have a concrete solution [5]. First learners will study a given problem individually to propose an initial solution. Learners then join in small groups, usually in pairs to discuss their solutions, and to propose a shared solution at the small group level. The discussion and negotiation will repeat in growing sizes of groups following a Pyramid structure until the whole group reaches a common solution to the given problem. The design rationale of the script promotes equal participation, knowledge co-construction, and improved negotiation. The Jigsaw CLFP is expected to encourage intensive interactions among students as the completion of the task is not possible without all partners' contributions. The design rationale of the script fosters positive interdependence.

Even though CLFPs and scripts, in general, provide guidance and structure for collaboration, empirical studies that provide evidence to support the claim that scripted collaboration is effective for domain learning are lagging behind [15]. Moreover, how different factors, e.g., the type of task, script design elements, teacher regulation, influence domain learning during scripted collaboration is not fully understood yet. More research is needed to provide evidence of the benefits that scripted collaboration entails in real-learning situations and how such structured collaboration impacts students' learning gains and the related design factors. To this end, the present study aims to generate more evidence in this field by studying how scripted CSCL activities lead to learning gains across three online learning situations scripted according to the Pyramid pattern. The study also contributes with two measurements namely: 1) level of precision and 2) level of confusion that can be used to evaluate learning gains in such CSCL scenarios. Moreover, students' behaviour in the learning process, regarding the agreement with the knowledge exchange, is also analysed in relation to the learning gains. The findings of the study are discussed to shed light on different factors that may have influenced students' learning across the cases and how different factors may lead to implications in the design and implementation of Pyramid scripts. The research question proposed in this study is *"How does the Pyramid script impact students' learning gains?"*

The rest of the paper is organized as follows. Section 2 explains the methodology followed. Section 3 presents the study results followed by section 4 which provides a discussion of the study findings. The final section provides concluding remarks and future research directions.

## 2 Methodology

### 2.1 Collaborative scripting tool: PyramidApp

In this study, a tool called PyramidApp was used to design and mediate interactions among students following the Pyramid pattern [12]. The tool implements the macro script structure employed by the Pyramid CLFP and has been used by teachers to implement CSCL activities in different learning situations [13]. The tool provides an authoring space for the teachers where they can design CSCL activities configuring several aspects such as the number of students expected, the number of levels in the Pyramid script, and the duration allocated to different phases of the script (details related to the design decisions of the tool are presented in [11]). Once done with the activity design teachers can generate a unique link to the activity which they can later share with students to enact Pyramid activities in their learning scenarios. The tool also provides a teacher-facing learning analytics dashboard through which teachers can view answers produced by individual students and agreed upon as well students' participation at the group levels [1].

In PyramidApp collaboration among students is structured within several phases. 1) First, students are required to login into the tool and provide an individual answer to the given task. 2) Then students are allocated into small groups ( $2 < n < 5$ ) randomly where they get the opportunity to see the answers submitted by their peers. The tool provides a rating mechanism and a discussion space for students to discuss and rate for the best answers (see Fig. 1). 3) At the end of the small group level, the small groups are merged to formulate larger groups automatically. Within larger groups, students can then view the answers agreed upon at the small group level. Students can further engage in discussions at the larger group level and rate for the best answers to reach a consensus at the end of the activity.

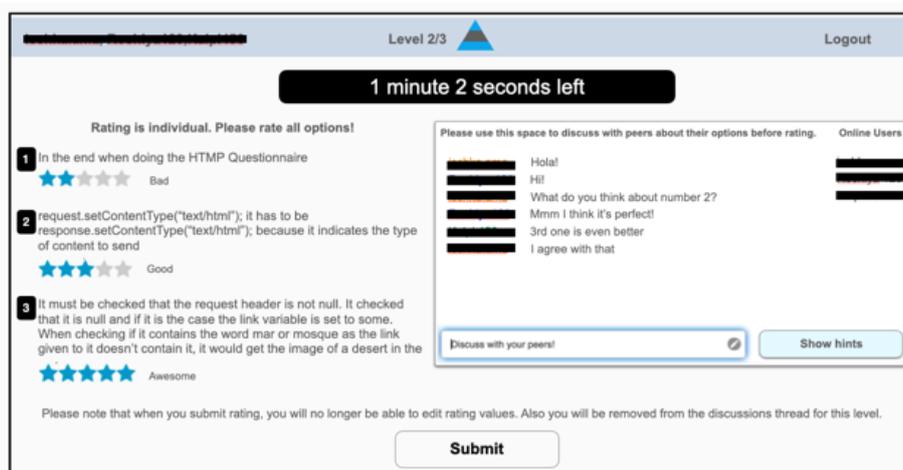


Fig. 1. User interface of the PyramidApp tool (rating space (left), discussion space (right)).

## 2.2 Contexts and Tasks

Two teachers from the Engineering School of a public university in Spain participated in our study. The design of each collaborative learning activity varied based on the teacher's requirements to conduct CSCL activities in their respective sessions and the time available (see Table 1).

Teacher A deployed two Pyramid scripts in her third-year undergraduate course named "Distributed Applications". One activity was focused on analysing a scenario to identify requirements and in total 25 students (4 female, mean age = 20) took part in these activities. The other activity was related to analysing errors in a computer program and a total of 28 students (4 female, mean age = 20) participated. Teacher A allocated in total 9 minutes for each Pyramid activity and Pyramids consist of three levels, i.e., option submission level, first rating level, and second rating level. As Pyramid activities were conducted as part of a full-fledged course, before the Pyramid scripts teacher A provided explanations to students about the concepts to be applied in the tasks and how the collaboration activity is connected with what has been taught.

Teacher B deployed a Pyramid activity within the context of an extracurricular workshop on social media. In Teacher B's sessions, 24 first-year undergraduate students (4 female, mean age = 17) participated in the Pyramid activity. As shown in Table 1 teacher B allocated in total 5 minutes for her Pyramid activity and Pyramids consisted of two levels, i.e., option submission level and a single rating level. The task focused on analysing the effects of digital footprints through a reflection activity based on a fictional scenario. Previous to the Pyramid activity, the students received information about three influencers and their behaviour in social media and had some time to read the information. Following this information they had to identify positive or negative aspects of each influencer and based on their personal judgement they had to select who would be more appropriate for a job position.

**Table 1.** A summary of collaborative learning activities conducted.

Teacher	Scenario	Activity Duration	Pyramid Levels	No. of students	Task	Educational Level
A	1 <sup>st</sup>	9 mins.	3	25	Determine the non-functional requirements for a scenario of a distributed application.	Third year undergraduate students.
	2 <sup>nd</sup>	9 mins.	3	28	Identify errors in software code.	Third year undergraduate students.
B	3 <sup>rd</sup>	5 mins.	2	24	Reflect about social media threats in digital footprint scenarios.	First year undergraduate students.

In both teacher A's and teacher B's sessions students were given training on how to use the PyramidApp for collaboration prior to the experimental sessions reported in this study. Students provided their informed consent before participating in the activities. A post-activity questionnaire asking students to rewrite an answer to the given task was administered immediately after the Pyramid activity.

### 2.3 Variables

Together with the teachers who participated in this study we came up with two scales to measure the learning gains namely: 1) level of precision and 2) level of confusion (see Table 2 and Table 3). We define the level of precision as the degree of agreement between a particular measurement (student's response) with an accepted standard (teacher's response). The level of confusion is defined as the degree of misunderstanding present in a student's response (e.g., mistakes, misconceptions, incorrect ways of organizing facts and figures). Both precision and confusion become two important variables to be considered when evaluating learning gains in Pyramid CLFP based collaboration due to a number of reasons. First, in our study both teachers proposed open-ended tasks to students, i.e., there were no simple "yes / no" answers to a given problem but multiple correct solutions and elaborations were possible. In this regard, the level of precision indicates how far the answers produced by the students align with the teachers' solutions and it also demonstrates the students' knowledge of a particular subject domain [10]. Second, previous studies have shown that pedagogical approaches such as "productive failure" and "confusion" are conducive to learning, as it enables students to learn from their own mistakes [4, 6].

Therefore, it is possible to define the *expectation for CSCL scripts, in terms of learning gains, as an increase in the levels of precision and a decrease in the levels of confusion as a result of the social interaction promoted by the script*. In the case of Pyramid scripts, learning gains are expected to be achieved by a structure that enables students 1) knowledge exchange by contrasting their initial answer to a task with peers' answers leading to socio-cognitive conflicts [14] in contrasting opinions, and 2) knowledge revision by discussing and selecting together the best possible answer.

As a secondary variable, to further understand the learning process happening during the Pyramid script (while refining knowledge based on the knowledge exchange), we also examined the data tracked by the PyramidApp. In particular, we analyse the ratings provided to students' submissions by group members across the pyramid rating levels. It is known that disagreement in collaborative learning can provoke rich argumentation leading to learning [11]. Yet, in the context of factual knowledge tasks, it would be expected that the collaboration process facilitates reconciliation of approaches, and therefore a minimized disagreement by the end of the learning activity. In Pyramid scripts, the expectation would be that in pyramid groups, students would identify confusion and (the most possibly) precise answers in their own and others' submissions, rating accordingly the better submissions in each pyramid level, and therefore advancing in the pyramid flow towards more accurate knowledge. Thus, in this study, a possible expectation is that measures of (dis)agreement could be an indicator of

potential learning gains: the more the agreement, the more chances of reaching precise/non-confused answers; the more the disagreement, the more chances for reaching non-precise/confused answers.

Section 3 reports to what extent different designs and implementations of Pyramid scripts are achieving those expectations, in terms of learning gains (precision, confusion) and (dis)agreement measures.

**Table 2.** Level of Precision.

Definition	Measure
Student's response does not match the concepts presented in teacher's answers.	0 – Not Precise.
Student's response is oriented to the concepts presented in Teacher's response, but it is a lack of consistency.	1 – Slightly Precise.
Student's response mostly aligns with the the concepts presented in Teacher's response.	2 – Almost Precise.
Student's response matches with the the concepts presented in Teacher's response.	3 –Precise.

**Table 3.** Level of Confusion.

Definition	Measure
Student's response does not present any mistakes, biases, misunderstandings. It fully arranges the given information in his/her explanation or task performance.	0 – None.
Minor mistakes appear in student's response, but these do not compromise the clarity of his/her explanation or quality in performed tasks.	1 – Low.
Mid-size mistakes appear in student's response which may compromise his/her answers or performed tasks.	2 – Intermediate.
Critical mistakes are present in student's response which misguide his/her explanation or performed task.	3 – High.

## 2.4 Data Analysis

Statistical analysis was conducted using SPSS software (version 26). The paired-sample t-test or wilcoxon signed rank test was used to report the learning gains based on the distribution (normality) of precision and confusion measures.

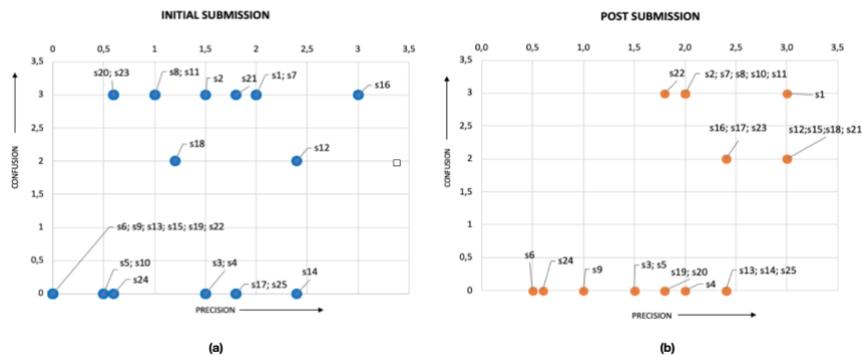
To measure disagreement of students' ratings we used standard deviation. For a given group a larger standard deviation indicates a higher disagreement, and a smaller standard deviation indicates a lower disagreement. As there were a number of groups formulated for a given Pyramid rating level, e.g., first rating level and second rating level, we calculated the average standard deviation to represent the disagreement as a whole for a given level.

### 3 Results

#### 3.1 Learning Gains

As shown in Table 1, the different types of activities considered were: 1) Determine the non-functional requirements for a scenario; 2) Identify errors in software code, and 3) Reflect about social media threats. First, Teacher A and Teacher B evaluated the students' answers for precision and confusion following the scales presented in Table 2 and Table 3. Then they also indicated the learning gains for each student considering the precision and confusion presented in initial and post submissions. In the following, we present the results regarding learning gains for each Pyramid activity conducted.

**1<sup>st</sup> Scenario: Software Engineering problem: Determine non-functional requirements.** In Fig. 2(a) and Fig. 2(b) each data point represents students based on the precision and confusion level scores assigned by Teacher A to their initial and post submissions. As shown in Fig. 2(a) the initial answers submitted by 44% of the students were precise (in different levels) but confusion was also present. Data points positioned on the positive X-axis indicates that 32% of the students submitted initial answers with levels of precision and without levels of confusion. Data points in (0,0) position indicate that initial answers submitted by 24% of students were not precise but confusion related elements were also not detected. As shown in Fig. 2(b) when considering the post submissions, it can be seen that 56% of students submitted post submissions that were in general more precise than in the initial submission, but confusion was still also present. 44% of students submitted post submissions with positive levels of precision without levels of confusion.



**Fig. 2.** Precision and Confusion: Initial Submission vs. Post Submission in the 1<sup>st</sup> scenario.

In the following we provide a summary of the students' learning gains in terms of precision and confusion (difference between the post submission and the initial submission). In Fig. 3 a data point in red indicates cases without learning gains and green indicate cases with learning gains. The data points in the Q1 quadrant show that 16% of students' answers have an increased precision but also present an increase in

confusion. Data points in Q4 quadrant indicate increased precision and decreased confusion in 12% of students' answers. Data points positioned on the positive X-axis also indicate increased precision and zero confusion in 52% of students' answers. Such scenarios, i.e., increased precision and decreased confusion and increased precision and zero confusion in students' answers indicate a clear learning gain (being a total of 64%). A data point in (0,0) position indicates that 16% of students' answers did not indicate a change in precision and confusion. Q3 quadrant indicates decreased precision but also decreased confusion in one students' answer. In terms of levels of precision, post submissions reached a higher average value ( $M = 2.076$ ;  $SD = 0.69$ ) when compared to the initial submissions ( $M = 1.11$ ;  $SD = 0.89$ ). A paired sample t-test indicated statistically significant differences;  $t(24) = 5.8$ ,  $p < .01$ . Regarding the average levels of confusion, post submissions reached higher average values ( $M = 1.4$ ;  $SD = 1.32$ ) when compared to the initial submissions ( $M = 1.24$ ;  $SD = 1.45$ ). A wilcoxon signed rank test indicated the difference was not significant;  $T = 13$ ,  $z = 0.71$ ,  $p > .05$ .

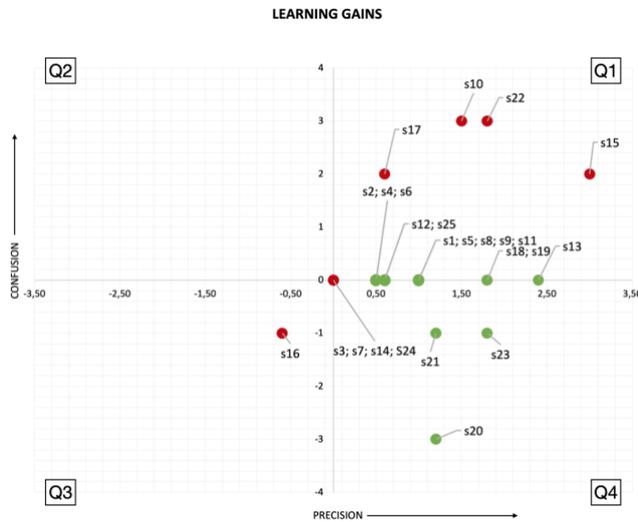
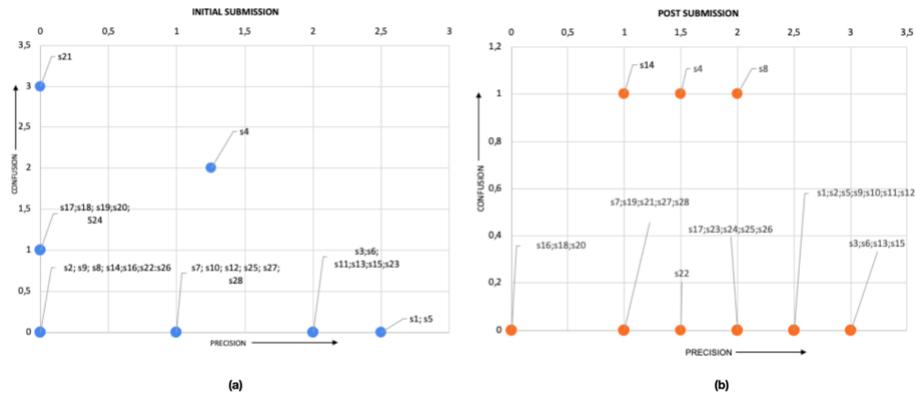


Fig. 3. Learning Gains in the 1<sup>st</sup> scenario.

**2<sup>nd</sup> scenario: Programming problem: Identifying errors in code.** In Fig. 4(a) and Fig. 4(b) each data point represents students based on the precision and confusion scores assigned to their initial and post submissions by Teacher A for a programming problem. As shown in Fig. 4(a) the initial answers submitted by a student (3.6%) consisted of both precision and confusion related elements. Data points positioned on the positive X-axis indicate that 50% of students submitted initial answers with levels of precision and without presenting confusion. Data points in (0,0) position indicate that initial answers submitted by 25% of students were not precise but confusion related elements were also not detected. Data points along the Y-axis indicate that 21.4% of students submitted initial answers that were not precise and included confusion-related elements. As shown in Fig. 4(b), when considering the post submissions, it can be seen

that 10.72% of students submitted post submissions that were precise, but confusion was also present. 78.57% of students submitted post submissions that were precise and in those submissions, confusion was not detected. 10.71% students' submitted post submissions that did not indicate a change in precision and confusion.



**Fig. 4.** Precision and Confusion: Initial Submission vs. Post Submission in the 2<sup>nd</sup> scenario.

In the following we present a summary of students' learning gains in the 2<sup>nd</sup> scenario in terms of precision and confusion. In Fig. 5 a data point in red indicates cases without learning gains and green indicates cases with learning gains. The data points in the Q1 quadrant show that 7.14% of students' answers have an increased precision but also present an increase in confusion. The assessment of the teacher clearly indicated that this was a case of confusion propagation in the knowledge revision, where students adopted incorrect knowledge exchanged by another student (about a supposed error in the code that was not such). Data points in the Q4 quadrant indicate increased precision and decreased confusion in 17.86% of students' answers. Data points positioned on the positive X-axis also indicate increased precision and no modification in the level of confusion in 42.86% of students' answers. A data point along the Y-axis indicates that 7.14% of students' answers show a case of no increase in precision but a decreased confusion. Cases including increased precision and decreased confusion, increased precision and no modification in the level of confusion, and no increase in precision but decreased confusion indicate a clear learning gain (being a total of 67.86%). A data point in (0,0) position indicates that 25% of students' answers did not indicate a change in precision and confusion. No data points fall into Q2 and Q3 quadrants. In terms of average levels of precision, the post submissions were more precise ( $M = 1.80$ ;  $SD = 0.93$ ) when compared to the initial submissions ( $M = 0.87$ ;  $SD = 0.92$ ). A wilcoxon signed rank test indicated that this difference was significant;  $T = 0$ ,  $z = 3.86$ ,  $p < .01$ . In terms of levels of confusion, post submissions consisted of less confusion ( $M = 0.11$ ;  $SD = 0.31$ ) when compared to the initial submissions ( $M = 0.36$ ;  $SD = 0.73$ ). A wilcoxon signed rank test indicated the difference was not significant;  $T = 9$ ,  $z = 1.73$ ,  $p > .05$ .

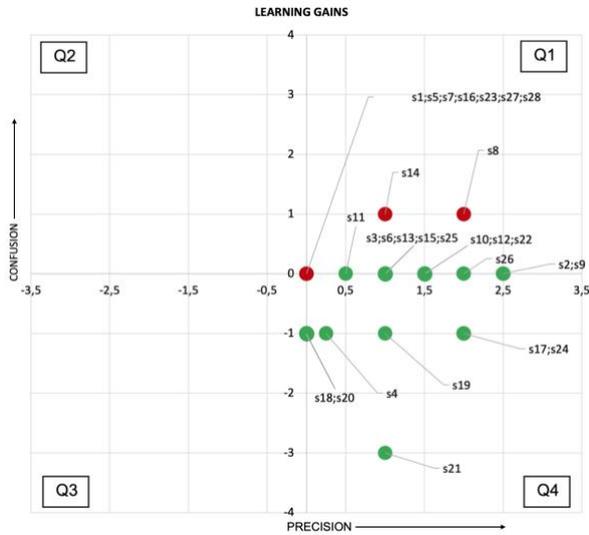


Fig. 5. Learning Gains in the 2<sup>nd</sup> scenario.

**3<sup>rd</sup> scenario: Reflect about social media threats in digital footprint scenarios.** For the third learning scenario, in Fig. 6(a) and Fig. 6(b) each data point represents students based on the precision and confusion scores assigned to their initial and post submissions by Teacher B. As shown in Fig. 6(a) the initial answers submitted by 79.16% of the students had some level of precision but confusion was also present. A data point positioned on the positive X-axis indicates that a student (4.17%) submitted an initial answer that was slightly precise and in the answer, confusion was not detected. Data points along the Y-axis indicate that initial answers submitted by 16.67% of students were not precise and confusion related elements were detected.

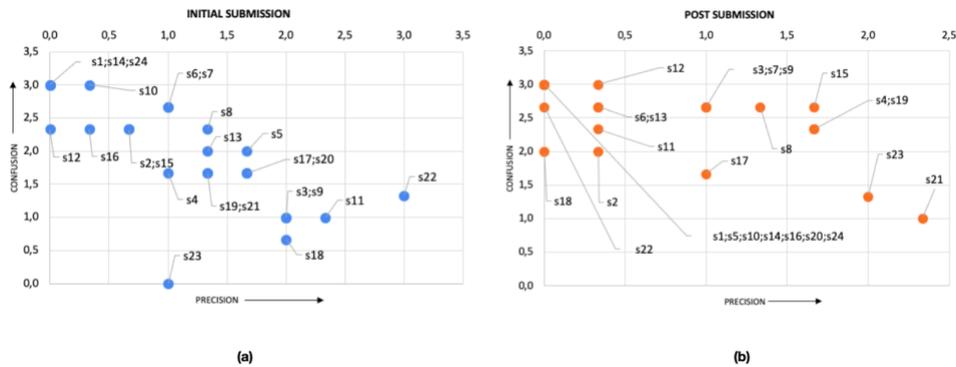


Fig. 6. Precision and Confusion: Initial Submission vs. Post Submission in the 3<sup>rd</sup> scenario.

As shown in Fig. 6(b) when considering the post submissions it can be seen that 62.5% of students submitted post submissions that had levels of precision but confusion was also present. Data points along the Y-axis indicate that 37.5% of students submitted post submissions that were not precise and in those submissions, confusion was detected.

In the following a summary of students' learning gains in the 3<sup>rd</sup> scenario in terms of precision and confusion are presented. In Fig. 7 a data point in red indicates cases without learning gains and green indicate cases with learning gains. The data points in the Q1 quadrant show that 20.83% of students' answers have an increased precision but also present an increase in (propagated) confusion. This scenario presents cases (data points) in the Q2 quadrant, indicating that 37.5% of the students' revised answers presented a decreased precision and increased confusion. Q3 quadrant indicates decreased precision but with decreased confusion in one students' answer (4.16%). Data points in the Q4 quadrant indicate increased precision and decreased confusion in one students' answers (4.16%). Only such cases (in Q4) indicate a clear learning gain (i.e. 4.16%). Data points in (0,0) position indicate that there were no changes in precision nor in confusion in 16.66% of students' answers. Data points positioned on the negative X-axis indicate decreased precision and no changes in confusion in 12.5% of students' answers. A data point positioned on the positive Y-axis indicates no changes in precision but increased levels of confusion in one students' answer (4.16%). Finally, in terms of average levels of precision the post submissions were less precise ( $M = 0.68$ ;  $SD = 0.75$ ) when compared to the initial submissions ( $M = 1.15$ ;  $SD = 0.81$ ). A paired sample t-test indicated significant differences;  $t(23) = -2.2$ ,  $p < .05$ . In terms of confusion, post submissions presented more average levels of confusion ( $M = 2.52$ ;  $SD = 0.56$ ) when compared to the confusion presented in initial submissions ( $M = 1.93$ ;  $SD = 0.81$ ). A paired sample t-test indicated that this difference was significant;  $t(23) = 4.3$ ,  $p < .01$ .

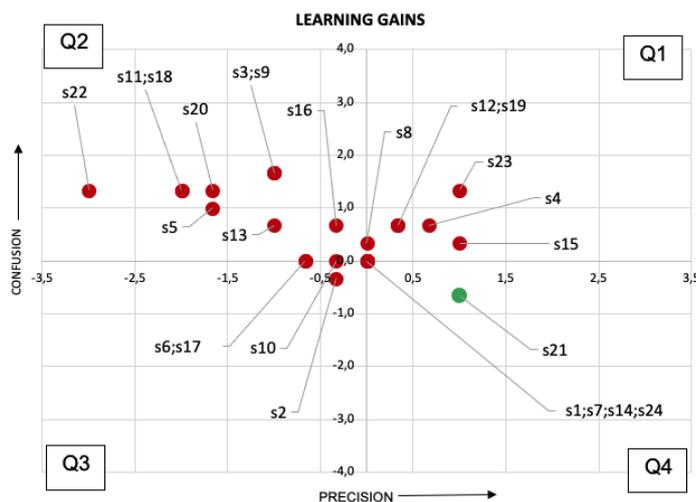


Fig. 7. Learning Gains in the 3<sup>rd</sup> scenario.

### 3.2 (Dis)agreement measures in ratings across the collaborative learning flow

As mentioned before (see Table 1) teacher A conducted Pyramid activities with two rating levels and Teacher B conducted a Pyramid activity with one rating level. As shown in Fig. 8 in teacher A's class in general the disagreement seems to decrease in the second rating level showing that students are reaching an agreement as the Pyramids evolve in time. In teacher B's class disagreement was observed for the first rating level of the Pyramid activity. As there was no second rating level in this class, we couldn't compare the differences of the level of disagreement across Pyramid levels in this case.

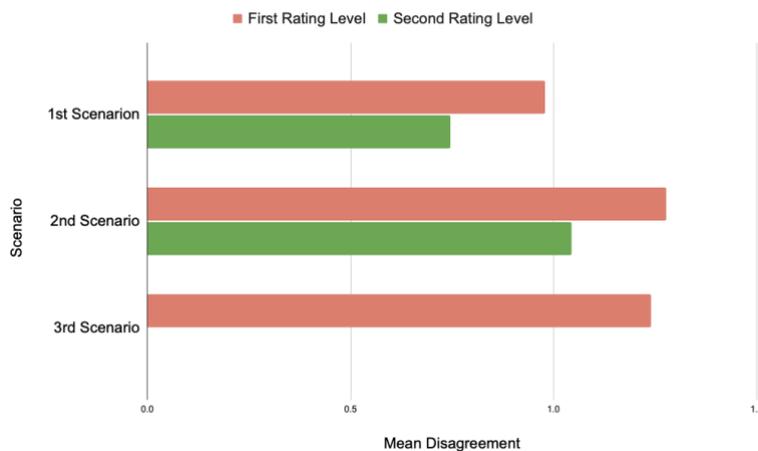


Fig. 8. Disagreement during Pyramid activities.

### 3.3 On how results align with the expectations

The results provide interesting findings regarding the stated expectations (section 2.3).

In terms of learning gains, the expectation was that “Pyramid scripts can lead to *an increase in the levels of precision and a decrease in the levels of confusion*”. The results show that it is actually the case in two out of the three learning scenarios analysed. In particular, in the 1<sup>st</sup> and the 2<sup>nd</sup> scenarios, clear learning gains in terms of precision are observed. Yet, the levels of precision achieved are not maximized for most students and some (but limited) propagation of confusion is observed. On the contrary, in the third scenario, the situation was that the knowledge exchange and the knowledge revision guided by the Pyramid script did not lead to learning gains in terms of maximizing precision and minimizing confusion. Moreover, the activity contributed to the creation and propagation of confusion.

In terms of measures of *(dis)agreement* and to what extent it be an indicator of potential learning gains, the expectation was that “*the more the agreement, more chances reaching precise/non-confused answers; the more the disagreement, the more chances for reaching non-precise/confused answers*”. As summarised in Fig. 8 during the 1<sup>st</sup> scenario and 2<sup>nd</sup> scenario the disagreement became low in the second rating level of the Pyramid script. Moreover, as commented above in the 1<sup>st</sup> and the 2<sup>nd</sup> scenarios

clear learning gains in terms of precision are observed. These results suggest that the study findings are in alignment with our expectations that a high agreement creates more chances to reach precise/non-confused answers. Moreover, the disagreement observed within the first rating level of the Pyramid script in the 3<sup>rd</sup> scenario can be interpreted as a lower level of agreement creates chances to reach non-precise/confused answers.

The next section discusses how these results suggest that several factors may affect the potential learning gains facilitated by these scripts and implications for the design of Pyramid scripts implementation.

## **4 Discussion: Factors and implications for design and implementation**

The present study aimed to analyse how scripted Pyramid CSCL activities could lead to learning gains across three different online learning scenarios. Overall findings suggest that scripted CSCL activities can enhance learning outcomes although the precision and confusion in learning gains may depend on teaching scenarios and mediating factors. In our consideration, factors such as the pedagogical envelope, the type of tasks, pyramid design elements, the need for epistemic orchestration and debriefing may have influenced the design and implementation of the examined Pyramid activities. The following section discusses how factors may have affected the results of the study.

### **4.1 The pedagogical envelope: prior activities**

As observed in the data analysis, learning outcomes in the designed activities varied from class to class. Results obtained from Teacher B's class activity may suggest contrary expected outputs. Nevertheless, it is important to observe the nature of the activities and the prior experience of students with the explored topic. Teacher A's activities were part of a formal lesson and therefore students were familiar with the class and the teaching materials. Conversely, Teacher B's activity occurred in an extracurricular environment. Enrolled students were not asked for any prior preparation, and the activity was used as an ice breaker activity to encourage students to meet each other.

### **4.2 The type of task**

The results of the study indicated that the type of task can have an impact on the students' learning gains in these scripted learning situations. Precision and confusion in post submissions varied in all developed tasks (1<sup>st</sup> scenario and 2<sup>nd</sup> scenario showed more precision and less confusion while 3<sup>rd</sup> scenario showed less precision and more confusion). Although confusion is propagated in all tasks, results show that openness of the task might lead to major confusion. Teaching activities in 1<sup>st</sup> scenario and 2<sup>nd</sup> scenario presented more answer-oriented exercises in contrast to 3<sup>rd</sup> scenario which demanded dealing with assumptions based on previous experiences and personal

opinions. Although open-ended tasks can be given to students with the expectation that tasks will create an opportunity for students to negotiate and co-construct a solution, current study findings show that these types of tasks require careful consideration and intervention of teachers to achieve the intended learning outcomes.

### **4.3 Pyramid design elements: pyramid levels**

Pyramid flow design elements may have also had an impact on the findings. In comparison to the two rating level activities, single rating level activity may have not provided opportunities for students to revise their knowledge in the process of reaching an agreement. As observed in the analysis, the level of disagreement between students tended to decrease with the increasing levels of group phases of the Pyramid script. Post submissions in the 1<sup>st</sup> scenario and 2<sup>nd</sup> scenarios, where Pyramid activities consisted of two rating levels showed increased precision whereas the 3<sup>rd</sup> scenario in which Pyramid activity consisted of a single rating level did not produce significant learning gains. These findings are aligned with the findings of previous similar research [14].

### **4.4 The need for teacher intervention: epistemic orchestration**

Another possible explanation for the results can be connected to the role of the teacher in monitoring the scenario. Teachers' intervention is essential to mediate cases where top levels of precision are not achieved, and confusion occurs so that confusion related elements do not propagate to the subsequent levels of the activity. Commonly referred to as *orchestration* in these types of learning situations teachers are required to manage collaboration in real-time [3]. Teachers not only have to regulate students' participation in groups, but they also need to constantly evaluate the content produced by students that can also be described as epistemic orchestration of the learning activity. This could be supported, for example, by learning analytics tools such as teacher-facing dashboards, through which they can monitor a large number of groups simultaneously, and embedded warning mechanisms that alert teachers automatically to the relevant indicators [1]. These indicators could include the presence of confusion, decreased precision in students' answers, and increased disagreement that can facilitate their epistemic orchestration, overall contributing to establish a fruitful collaborative learning climate.

### **4.5 Pedagogical envelope: the need for teacher-led debriefing**

Connected to the preceding idea, a suitable intervention by the teacher during the learning scenarios have a potential influence on the learning gains. Teacher-led debriefing activities at the end of learning activities can contribute to reflect on the conflicting points with the whole class. Noted confusing elements in students' answers can be discussed at the end of the activity with the whole class by the teacher [4]. Teacher-facing dashboards can assist teachers in flagging answers consist of confusing elements that require discussion at the individual, group, or whole class level.

## 5 Conclusions and Future Work

This paper aimed to explore how Pyramid pattern-based learning scripts impact student's learning gains. The findings of the study have shown that the observed differences in the learning gains at three different cases can be interpreted by referring to a number of different factors: the pedagogical envelope; the type of task; Pyramid design elements; the need for teacher intervention and need for teacher-led debriefing. Moreover, based on the improved precision observed in students' post submissions it is evident that the particular Pyramid pattern-based CSCL activities conducted using the PyramidApp facilitate learning. However, the aforementioned factors require careful consideration.

There are several limitations of this study. First, the number of cases we considered is low and with each case representing a different task for collaboration. The number of students participated in each activity is relatively low and the prior activities they engaged in are different across the cases. Although teachers who participated in the study agreed precision and confusion as valid measurements to evaluate learning gains, the precision and confusion assigned to students' answers are strictly dependent on the teacher's personal judgement. The above limitations can have an impact on the obtained learning gains. However, the exploration of these factors needs to be acknowledged as an important contribution of our work. In the future, we are planning to explore further how some of these design elements, specifically the teachers' aspects related to epistemic orchestration and teacher-led debriefing, can influence the precision and confusion presented in students' answers and therefore learning gains. Moreover, in the future, it is also important to extend a more detailed analysis of the student's behaviour during the learning process (e.g., chat discussion) to further understand relevant indicators (such as (dis)agreement measures) of potential learning gains.

**Acknowledgements** This work has been partially funded by the Volkswagen Foundation (COURAGE project, no. 95567). TIDE-UPF also acknowledges the support by FEDER, the National Research Agency of the Spanish Ministry, TIN2017-85179-C3-3-R, PID2020-112584RB-C33, MDM-2015-0502 and by ICREA under the ICREA Academia programme (D. Hernández-Leo, Serra Hunter).

## References

1. Amarasinghe, I., Hernández-Leo, D., Michos, K., & Vujovic, M.: An actionable orchestration dashboard to enhance collaboration in the classroom. *IEEE Transactions on Learning Technologies* **13**(4), 662–675 (2020).
2. Dillenbourg, P.: Over-Scripting CSCL: The Risks of Blending Collaborative Learning with Instructional Design. In: Kirschner, P.A. (ed.) *Inaugural Address, Three Worlds of CSCL. Can We Support CSCL?*, pp. 61–91. Heerlen, Open Universiteit Nederland (2002).
3. Dillenbourg, P.: Design for classroom orchestration. *Computers & Education* **69**(1), 485–492 (2013).
4. D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A.: Confusion can be beneficial for learning. *Learning and Instruction* **29**, 153–170 (2014).

5. Hernández-Leo, D., Villasclaras-Fernández, E. D., Asensio-Pérez, J. I., Dimitriadis, Y. A., & Retalis, S.: CSCL scripting patterns: hierarchical relationships and applicability. In: 6th IEEE International Conference on Advanced Learning Technologies, pp. 388–392. (2006).
6. Kapur, M.: Productive failure in learning math. *Cognitive science* **38**(5), 1008–1022 (2014).
7. Kollar, I., Fischer, F., & Hesse, F.W.: Collaboration scripts - a conceptual analysis. *Educational Psychology Review* **18**(2), 159–185 (2006).
8. Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F.: Specifying computer-supported collaboration scripts. *International Journal of Computer Supported Collaborative Learning* **2**(2–3), 211–224 (2007).
9. Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, J. D.: A tough nut to crack: Measuring collaborative problem solving. *Handbook of research on technology tools for real-world skill development* (2016).
10. Loibl, K., Roll, I., & Rummel, N.: Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review* **29**(4), 693–715 (2017).
11. Malzahn, N., Aprin, F., Hoppe, U., Eimler, S., & Moder, S.: Measures of Disagreement in Learning Groups as a Basis for Identifying and Discussing Controversial Judgements. In: *Proceedings of the 14<sup>th</sup> International Conference on Computer-Supported Collaborative Learning*, pp. 233–236. (2021).
12. Manathunga, K., & Hernández-Leo, D.: Authoring and enactment of mobile pyramid-based collaborative learning activities. *British Journal of Educational Technology* **49**(2), 262–275 (2018).
13. Manathunga, K., & Hernández-Leo, D.: Flexible CSCL orchestration technology: Mechanisms for elasticity and dynamism in pyramid script flows. In: *Proceedings of the 13<sup>th</sup> International Conference on Computer Supported Collaborative Learning*, pp. 248–255. (2019).
14. Mugny, G., & Doise, W.: Socio-cognitive conflict and structure of individual and collective performances. *European Journal of Social Psychology* **8**(2), 181–192 (1978).
15. Radkowsch, A., Vogel, F. & Fischer, F.: Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer Supported Collaborative Learning* **15**, 5–47 (2020).