education sciences

IMPACT FACTOR 2.5

CITESCORE 4.8

Article

Designing Dialogic Peer Feedback in Collaborative Learning: The Role of Thinq Tank

Amber Van Hoe, Joel Wiebe, Jim Slotta, Tijs Rotsaert and Tammy Schellens

Special Issue
STEM Education: Creative Designs and Models

Edited by
Dr. Chunfang Zhou and Dr. Wei Zhang

MDPI

https://doi.org/10.3390/educsci14111231

education sciences

Article

Designing Dialogic Peer Feedback in Collaborative Learning: The Role of Thinq Tank

Amber Van Hoe [1,*], Joel Wiebe [2], Jim Slotta [2], Tijs Rotsaert [1] and Tammy Schellens [1]

[1] Department of Educational Studies, Ghent University, 9000 Ghent, Belgium; tijs.rotsaert@ugent.be (T.R.); tammy.schellens@ugent.be (T.S.)
[2] Department of Curriculum, Teaching and Learning, University of Toronto, Toronto, ON M5S 1V6, Canada; joel.wiebe@mail.utoronto.ca (J.W.); jim.slotta@utoronto.ca (J.S.)
* Correspondence: amber.vanhoe@ugent.be

Abstract: As technology continues to reshape education, the integration of peer feedback (PF) is increasingly recognized as a key element in promoting student collaboration and learning. This design study tracks the development and implementation of the Thinq Tank tool, which is aimed at facilitating dialogic PF within computer-supported collaborative-inquiry learning (CSCiL) environments. The research was conducted through three design iterations, comparing two key experimental conditions: one contrasting non-dialogic PF with dialogic PF, and another comparing dialogic PF with scripted dialogic PF featuring dynamic sentence-starters and defined roles. Utilizing a quasi-experimental design, we assessed the overall impact on scientific (collaborative) learning outcomes, students' perceptions of PF, and the composition and elements of PF dialogues. The findings suggest that the integration of dialogic PF within a CSCiL lesson series helps to boost scientific collaborative learning outcomes and increases student receptivity to PF. The study demonstrates that integrating structured support mechanisms within PF tools promotes PF dialogues in which students critically evaluate each other's work and collaboratively think about ways to improve quality. These insights illustrate that structured, technology-supported PF can significantly advance inquiry outcomes and foster deeper collaborative learning in secondary STEM education.

Keywords: peer feedback; dialogic feedback; collaborative inquiry; technology-supported learning; STEM education

Citation: Van Hoe, A.; Wiebe, J.; Slotta, J.; Rotsaert, T.; Schellens, T. Designing Dialogic Peer Feedback in Collaborative Learning: The Role of Thinq Tank. *Educ. Sci.* **2024**, *14*, 1231. https://doi.org/10.3390/educsci14111231

Academic Editor: Heng Luo

Received: 20 July 2024
Revised: 17 October 2024
Accepted: 6 November 2024
Published: 11 November 2024

1. Introduction

Formative peer assessment is recognized as a valuable instructional strategy for providing students with ongoing insights into their learning processes [1]. Using this approach, equal-status learners take on the role of assessors and evaluate each other's work based on specific standards [2]. Formative peer assessment offers the advantage of providing students with more rapid and extensive feedback [3], allowing them to gain new perspectives [4] and sharpen their evaluative and critical thinking skills [5,6], thereby enriching their learning experience.

However, while most actors in the educational field praise the potential of formative assessment in general, critics argue against its often transmission-based interpretation, an interpretation which dominates the current literature. In this view, feedback is perceived as information passed from the assessor to the assessee [7]. This unilateral flow characterizes a traditional feedback pattern, and this would challenge peer assessment's foundation in participatory learning cultures [8]. Studies emphasize the need to embrace the inherent social nature of peer feedback (PF), approaching it as a social exchange or dialogue between individuals [9,10]. Er and colleagues define dialogic PF as "students' collaborative meaning-making about feedback by engaging in dialogue with their peers (providing the feedback)" [11] (p. 587). Framing PF as a dialogue transforms evaluation into a collaborative process, fostering active communication and mutual perspective-sharing [12].

Redesigning formative PF as a dialogue rather than a transmission-based approach could be particularly advantageous in inquiry-based science learning. In this context, social interaction facilitates the co-construction of meaning around scientific concepts [13] and enables peers to serve as valuable structural supports for each other during inquiry [14,15]. Although PF within inquiry-based science learning has received some research attention, there is a notable lack of studies focusing on the implementation and support of formative PF dialogues within this context. Given the central role of technology in today's education, integrating web-based tools to support these practices is crucial. Web 2.0 technology has advanced computer-supported collaborative learning environments, making them highly interactive and capable of facilitating communication, collaboration, and information sharing among users [16]. These environments can provide robust platforms for implementing PF dialogues, enabling more effective and engaging learning experiences.

Therefore, this design study aims to explore the use of PF dialogues within formative feedback practices by designing, implementing, and evaluating a web-based PF tool. This tool is dedicated to facilitating and supporting online PF dialogues within a computer-supported lesson series for secondary science education and is grounded in the principles of collaborative inquiry learning. The goal is to gain insights into the effectiveness, usability, and impact of this designed tool or intervention, ultimately contributing to the advancement of educational practices in this field.

1.1. Formative Peer-Assessment

As the assessment culture has transitioned from evaluating learning outcomes (i.e., assessment of learning) to evaluating learning processes in order to enhance learning outcomes (i.e., assessment for learning; [17]), formative peer assessment has become a key component in many educational settings. Peer assessment can take various forms, such as peer scoring (e.g., grades) or feedback (e.g., comments) [18]. PF, a crucial element of peer assessment, provides mid-point evaluations that highlight strengths, weaknesses, and areas for improvement [19]. By consistently receiving feedback, students can make necessary adjustments, ultimately improving their academic performance [20,21].

Although it is often argued that teachers are the most reliable source of feedback due to their expertise, providing individualized feedback to all students is a demanding task [5]. Research shows that PF can be equally effective and, in some cases, even more beneficial. Li et al. found a strong overlap between peer and teacher evaluations, while meta-analyses by Double et al. and Li et al. indicate that peer assessment improves academic performance more than does having no assessment, and even outperforms teacher assessment [20–22]. Moreover, it was also found that PF often leads to more meaningful revisions, ones particularly focused on content, whereas teacher feedback tends to prompt more surface-level improvements [23].

Formative PF typically involves four phases: performing the task, providing feedback, receiving feedback, and revising the task [8]. Traditionally, the focus has been on providing high-quality feedback, as this is considered essential for improving the evaluated work. However, challenges persist, such as students often providing superficial or unjustified feedback [24,25] and interpersonal factors potentially compromising the objectivity of the feedback [26].

To address these challenges, one key factor in improving the reliability and validity of peer assessment, and consequently the quality of PF, is rater training [27,28]. The meta-analysis by Li et al. confirmed that peer assessment becomes significantly more effective when supported by training [21]. Liu and Li further found that rater training reduces discrepancies between student and instructor ratings, leading to higher-quality feedback and better revisions [29]. This highlights the essential role of training in enhancing both PF and learning outcomes. Additionally, Gennip et al. found that students trained in peer assessment experienced greater psychological safety during the process, compared to those assessed by teachers, adding another dimension to the benefits of rater training [30].

In addition to rater training, the use of explicit assessment criteria, typically through rubrics, plays a critical role in successful peer assessment [21]. Rubrics help to structure evaluations and clarify expectations [27], leading to several key benefits, such as increased transparency, reduced anxiety, and improved feedback quality, all contributing to better performance [31]. When combined with other forms of support, rubrics become even more effective. For example, training on rubric use has been shown to significantly improve feedback quality [32]. Furthermore, incorporating sentence-starters to guide students in formulating feedback, along with rubrics to set clear expectations, has significantly improved the quality of feedback. This demonstrates the value of structured guidance in the feedback process [33].

1.2. Dialogic Peer-Feedback

Amidst the exploration of optimal PF mechanisms and the consequent main focus on the peer assessor, the collaborative aspect of peer assessment, integral to its essence, has been somewhat neglected [10]. Typically, the interaction between assessor and assessee is confined to the exchange of scores and comments, representing a rather limited form of interaction [10,34]. In this conventional approach, the receiver assumes a passive role, merely accepting PF without their active engagement. This passivity may inadvertently hinder the full realization of the learning opportunities inherent in peer assessment [35]. Peer assessment is fundamentally rooted in student-centered pedagogy and emerges from a perspective in which interaction and collaboration among students lay the groundwork for learning [10]. Refocusing on interaction embraces peer assessment's collaborative nature and highlights assessment as a two-way conversation in which both assessors and assessees have an active role in feedback processes [35,36].

Direct PF dialogues are envisioned as being a means to encourage a more dynamic and engaging exchange of ideas and perspectives [37]. PF dialogues are particularly implemented in the context of higher education to support the feedback processes associated with writing assignments through the use of face-to-face interaction (e.g., [12,35,38]. The benefits of integrating direct PF dialogues are diverse. Firstly, it empowers assessees by actively involving them in the feedback process. Therefore, PF evolves into a more engaged and enriching learning experience, offering opportunities for clarification and negotiation [39,40], as well as collaboration between assessor and assessee, aiming to enhance the quality of the assessed work [35]. Secondly, a PF dialogue benefits both the assessee and the assessor [41]. While assessees reflect on their learning products based on the feedback they receive, assessors can also receive feedback on the feedback they provide. This helps them reflect on their evaluative skills and improve their ability to give constructive feedback [35,42]. Thirdly, dialogic PF has the potential to increase the uptake of PF, as demonstrated by the enhanced student engagement seen with dialogic, compared to written, PF [38,42].

Just like traditional PF, PF dialogues face challenges, such as students providing superficial and unjustified feedback. This issue is particularly relevant in PF dialogues, in which detailed, justified responses are crucial [43,44]. Additionally, organizing and executing these dialogues can be time-consuming for both teachers and students [11], requiring time to build familiarity with the process and confidence in giving feedback to their peers [12,36].

1.3. The Use of Technology to Facilitate Formative Peer-Feedback Dialogues

Technology shows promising solutions for optimizing the efficiency and effectiveness of the PF process [21,45], offering immediate feedback access and response options [16]. Various web-based tools already exist to implement dialogic PF into educational practices [41,43,46,47]. These tools have predominantly been designed to facilitate asynchronous PF dialogues, often resulting in a lack of active engagement between assessors and assessees. This lack of engagement could be due to the extended period between the provision of PF and the assessees' responses. This consequentially leads to conversations that fail to gain momentum and often stop after the assessees' reactions, due to time limitations [43,47]. Only Kim and Ryu have enabled online direct interaction, via their dialogic PF tool, enabling Korean preservice teachers to assess each other's designed extracurricular programs [41]. However, as the main focus of the research was the comparison of dialogic PF with traditional, unilateral peer- and self-assessment as to performance and metacognitive awareness, their research gives no insight into what the PF dialogues looked like.

Additionally, the aforementioned tools mainly focused on aiding only the assessors during these PF dialogues, rather than supporting both parties [41,46,47]. Only Alemdag and Yildrim gave attention to the assessees, designing a dialogic PF tool in order to support Turkish high school students in helping each other out with writing tasks in language courses. They provided prompts to assessees to help them to respond to feedback in PF dialogues while focusing on the clarity of the PF and their agreement with it. However, these prompts in the form of questions proved to be insufficient in eliciting positive and constructive responses when the received PF was deemed superficial or inaccurate [43].

Furthermore, it is noteworthy that all of these existing dialogic PF tools are tailored specifically to higher education and used primarily to scaffold writing assignments. This context represents the predominant implementation setting in the research on peer assessment, overall [21,45]. However, other domains could also greatly benefit from the utilization of dialogic PF [20]. One such example is computer-supported collaborative inquiry learning (CSCiL).

1.4. Dialogic Peer-Feedback During Computer-Supported Inquiry Learning

CSCiL is a pedagogical approach that centers around student exploration and investigation, and is mostly implemented within science education, in order to deepen the understanding of scientific concepts [48,49]. Peers play an essential role in providing support and guidance within CSCiL [50]. Recent studies have started to explore ways in which learners could provide support to one another through formative peer assessment (e.g., [51–53]). These studies focus on determining the best conditions for providing peer assessment and achieving optimal outcomes, such as whether it is better to provide assessment criteria to peer assessors or not (e.g., [51,53]). However, these studies follow a transmission-based approach, one in which there is no direct interaction between the assessor and assessee. Particularly relevant to this research are findings suggesting that students are often reluctant to accept and process PF. Although the PF was generally aligned with expert judgments and scientifically accurate, critical feedback and improvement suggestions were rare and, when provided, often lacked supporting arguments. These findings, along with the recommendations from previous studies, highlight the need for forms of PF which are more interactive and dialogic in order to improve its adoption and effectiveness [52,53].

1.5. Research Aim

In this design study, we develop and implement a web-based peer assessment tool to facilitate and support online synchronous PF dialogues. This tool, integrated into a CSCiL lesson series for secondary STEM education, encourages real-time interaction, fostering more dynamic exchanges between students. Our study stands out by focusing on synchronous PF dialogues, which contrasts with the predominantly asynchronous approaches used in prior research. Additionally, this work explores a unique context by implementing PF dialogues specifically within inquiry-based tasks in secondary science education; this is an area in which previous studies have largely centered their analyses on writing tasks in higher education. Furthermore, we go beyond the typical focus on the assessor by offering support to both assessors and assessees during PF dialogues, creating a more reciprocal and collaborative process. This dual support is facilitated by the tool, which uses structured scripting mechanisms to guide these interactions. Finally, this study is one of the first to implement dialogic PF in the CSCiL context, making a significant contribution to the fields of both educational technology and peer assessment.

The development of this tool progressed through an iterative design process, with each iteration involving the design, testing, and refinement of the tool based on classroom data. Each iteration included a quasi-experimental study in real classroom settings, providing empirical evidence to guide subsequent design improvements [54,55]. By intertwining design-based research with quasi-experimental classroom studies, our work contributes to the research field surrounding the design and efficacy of PF mechanisms [10,20]. Through this design research, we aim not only to enrich the theoretical understanding of its pedagogical underpinnings within the literature but also to refine the practical application of PF within CSCiL environments. Consequently, this research helps align theory, pedagogy, and educational technology [55]. The following research questions guided our study:

- RQ1. To what extent does a computer-supported collaborative inquiry lesson series with online formative peer assessment improve students' scientific conceptual knowledge, inquiry skills, and inquiry outputs?
- RQ2. How do users perceive and interact with the design elements of an online formative peer-assessment tool tailored for computer-supported collaborative inquiry learning?
- RQ3. What are the composition and elements of peer feedback dialogues within the online formative peer assessment tool in the context of computer-supported collaborative-inquiry learning?

2. Materials and Methods

2.1. Materials

This research article presents the design and evaluation of a peer assessment tool, Thinq Tank, which aims to provide students with an environment where they can collaboratively think about, assess, and discuss their inquiry efforts (Think + Inquiry = Thinq). This investigation is part of a larger research endeavor aiming to develop a series of lessons, called Climate colLab, designed for students in grades 9 and 10.

Climate colLab

The Climate colLab series consists of four 50 min lessons aligning with national educational standards in Flanders (Belgium). The lessons aim to enhance the scientific conceptual understanding of climate change and promote inquiry skills, following the principles of CSCiL.

To support this, student groups' scientific-inquiry processes are structured around the inquiry cycle of Pedaste et al. (see Figure 1) [48]. The first phase, the *orientation phase*, aims to capture students' interest and introduce them to the overarching topic of the inquiry. This is achieved by having them watch and read excerpts from documentaries or newspaper articles, explore simulations with real-world data, and complete accompanying questions and exercises. Moving on to the *conceptualization phase*, students are introduced to the research question they will explore, one for which they are subsequently prompted to formulate a research hypothesis. Next, in the *investigation phase*, students engage with simulations to experiment and address related questions. They seek to confirm or refute stated hypotheses by searching for, manipulating, and analysing relevant data. The *conclusion phase* comes next and tasks students with drawing conclusions based on their findings, formulating a response to the given research question. Finally, there is the *discussion phase*, which can be placed after any of the stages that are already in place. Discussions are facilitated as students work in pairs using one computer. During inquiry students are allowed to revisit stages as needed to refine their understanding.

Students go through four distinct inquiry cycles during Climate colLab. The first three inquiry cycles span two class periods and focus on providing students with a solid understanding of climate change. After these sessions, students are expected to be familiar with the process of navigating the inquiry cycle and key scientific concepts related to climate change. The fourth inquiry cycle takes place during the final two class periods, when student groups are given more freedom to conduct research on a sustainability topic of their choice from a list of five options. During these last two lessons, the *Thinq Tank* tool is introduced to offer additional peer support in the form of formative peer assessment as students engage in more independent inquiry.

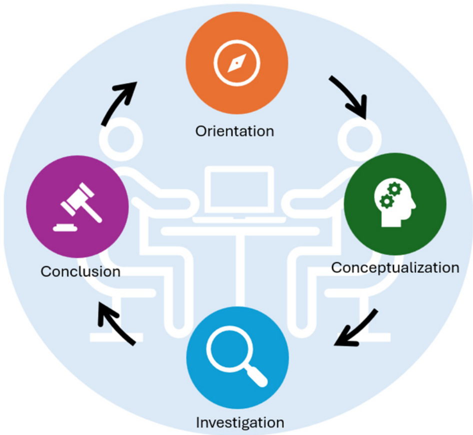

Figure 1. Simplified illustration of the inquiry cycle [48].

The Scripting and Orchestration Environment (SCORE) serves as the platform for Climate colLab (see Figure 2). Developed at the University of Toronto and the University of Berkeley, SCORE allows educators to design and implement CSCiL, enabling the creation of dynamic content using various formats (e.g., drag and drop, labelling) and tools (e.g., concept mapping, graphing, and simulations).

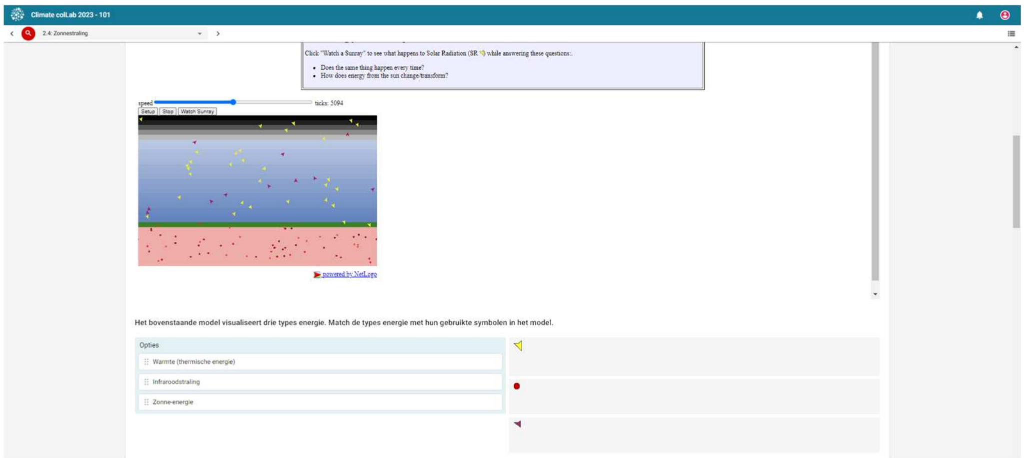

Figure 2. An example screenshot from SCORE.

Thinq Tank

As mentioned before, this research article focuses on the design and evaluation of the *Thinq Tank* peer assessment tool, which has specifically been developed to implement dialogic PF within a CSCiL platform. To design and refine *Thinq Tank*, an iterative and formative design approach was used, following three stages: Preparation and Design, Implementation, and Analysis [54]. In the Preparation phase, the core design issues were identified, relevant knowledge was gathered, and the scope of the tool was defined. The Implementation phase involved testing the tool in real classroom settings, collecting data, and selecting analysis methods. In the Analysis phase, the data were reviewed, making comparisons with the initial design goals and guiding further refinements [54].

Three design iterations were conducted, each with their distinct research focus, design, data collection methods, and analyses, to finetune the *Thinq Tank* tool.

The first iteration aimed to test the Climate ColLab lessons in order to identify student needs in a CSCiL environment and determine the optimal timing for peer assessment.

In the second iteration, the *Thinq Tank* tool (see Figure 3) was introduced during the fourth inquiry cycle to facilitate peer support. Peer assessment, including dialogic PF, was integrated at three crucial stages of the inquiry process: conceptualization, investigation, and conclusion (see Figures 5 and 6). At these stages, students, respectively, formulated a research question, collected data, and generated conclusions which became the foci of peer assessments. Each student pair worked collaboratively, both conducting their inquiry and engaging in reciprocal PF. Student pairs were randomly assigned and requested to provide feedback to each other throughout the study, ensuring consistency and fostering deeper engagement in the peer assessment process.

Figure 3. Screenshot of the PF dialogue component during the conceptualization phase.

The third iteration built upon this by introducing two new scripting features to *Thinq Tank* to further enhance the PF process: (1) the Role Assignment feature, which designated students as either assessors or assessees, and (2) the Dynamic Question Bank, which provided sentence-starters to guide meaningful PF interactions (see Figure 4). Both features were embedded within the peer chats to support structured and productive PF dialogues.

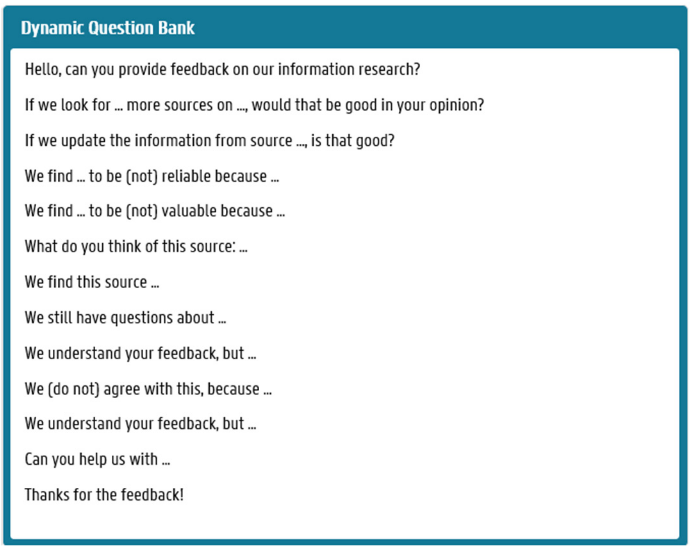

Figure 4. Dynamic question bank during the investigation phase, from the assessee's perspective.

2.2. Methods

In this section, we will provide an overview of each design iteration, outlining its research focus, design, participants, data collection methods, and analyses. Each iteration builds upon the previous one, addressing different aspects of CSCiL and peer assessment, with a particular focus on implementing and refining dialogic PF through *Thinq Tank*.

Design iteration 1: Identifying guidance needs in online CSCiL

Design and participants

In the first design iteration, the focus was on investigating the guidance needs of the students and identifying areas in which peer support could be most effectively integrated within a CSCiL environment. Due to the shift to online education during the COVID-19 pandemic, this iteration was conducted in an online setting. A one-group pre-test–post-test design was used to measure learning outcomes. A total of 38 students from grades 9 and 10 (Mage = 15.44, SD = 0.54) participated, grouped into 17 student groups for online collaboration. Pre-tests and post-tests, conducted over a two-week interval, were used to evaluate changes in students' scientific conceptual understanding and inquiry skills. Additionally, two online focus groups, each with 25 students, were held to explore their perceptions of online CSCiL and the guidance provided during the process.

Data collection and analysis

To answer **RQ1**, which focuses on the outcomes of the *Climate ColLab* intervention, pre-test and post-test data were collected to assess changes in students' scientific knowledge and inquiry skills. A one-way ANOVA with repeated measures was applied to evaluate the effect of the intervention.

For **RQ2**, which explores students' perceptions of the online CSCiL environment, two semi-structured focus groups were conducted, each with 25 participants. Each focus-group session lasted approximately 30 min, during which five propositions about online CSCiL and collaboration were presented using the Socrative tool. Participants rated their agreement with each proposition on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), and the anonymized responses were projected on a screen and discussed in real time. Descriptive statistics were used to analyse the Socrative responses, while the focus group discussions were transcribed and analysed qualitatively to provide deeper insights into students' perceptions and experiences.

Design iteration 2: Implementing dialogic PF support through Thinq Tank

Design and participants

In the second design iteration, a quasi-experimental pre-test–post-test design was employed, involving two experimental conditions and one control condition. A total of 382 students from grades 9 and 10 (Mage = 15.04, SD = 0.67) across 28 classes in 12 schools participated, divided into 187 student groups. The pre-test and post-test, taken two weeks apart, and between which the educational content was delivered, measured changes in students' scientific knowledge of climate change and their perceptions of the PF.

The two experimental conditions focused on different PF mechanisms implemented during the fourth inquiry cycle (see Figure 5). In the first peer assessment condition, students engaged in rubric-based grading accompanied by open-ended PF. In the second peer assessment + PF dialogue condition, an additional synchronous web-based PF dialogue was introduced after the exchange of peer assessment, allowing students to engage in PF dialogues via the *Thinq Tank* tool's chat feature. Lastly, the control condition involved students receiving traditional instruction without the *Climate colLab* intervention, allowing for comparison between the experimental groups and the control group.

Data collection and analysis

For **RQ1**, which examines how the lesson series with embedded formative peer assessment impacts students' scientific learning outcomes, two methods were used. First, changes in students' scientific knowledge were analysed using linear mixed models applied to the pre-test and post-test data to capture differences between the two experimental conditions and the control condition. Second, the inquiry outputs produced during the inquiry were evaluated to assess the quality of research questions, data, and conclusions. Considering the hierarchical structure of the dataset, (generalized) linear mixed models were used to examine differences in the quality of inquiry outputs between the experimental conditions.

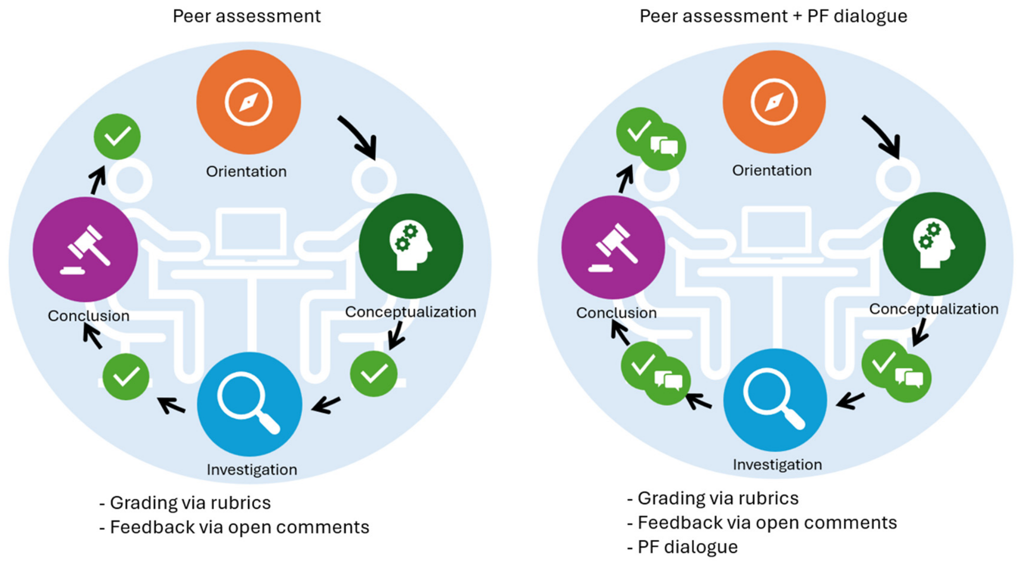

Figure 5. Formative peer assessment scenario according to experimental conditions.

To answer **RQ2**, which explores how users perceive and interact with the design elements of *Thinq Tank*, students' perceptions of PF were assessed at both pre-test and post-test timepoints. These perceptions were measured using validated instruments that assessed factors such as fairness, usefulness, acceptance, willingness to improve, and affect [56]. Linear mixed models were applied to analyse changes in these perceptions across the experimental conditions.

For **RQ3**, which investigates the composition and structure of PF dialogues, 113 dialogues were included in the qualitative dataset. From this, 24 dialogues (20%) were randomly selected, with eight being taken from each of the three inquiry phases. The qualitative analysis began by organizing dialogues into episodes, each marking the point when a new topic was introduced. Next, the PF dialogues were divided into dialogue acts—defined as independent sentences or discrete segments with a single communicative purpose. Using an adapted version of the categorization framework by Jones et al., dialogue acts were classified into statements, questions, and directives [57]. Additional categories were integrated to capture social interactions (e.g., salutations, camaraderie) and comments on the platform [58]. Dialogue responses were further categorized as either supportive (e.g., compliance, acceptance) or confronting (e.g., refusal, contradiction) [57]. The complete coding scheme can be found in Appendix A. After the coding scheme was finalized, the PF dialogue acts were coded and a descriptive analysis was conducted.

Design iteration 3: Scripting dialogic PF through Thinq Tank

Design and participants

In the third design iteration, a quasi-experimental pre-test–post-test design with two experimental conditions was employed to examine how the scripting of PF dialogues could enhance students' perceptions and interactions. A total of 230 students from grades 9 and 10 (Mage = 14.93, SD = 0.55) across 12 classes in 5 schools participated, divided into 108 groups. The *Climate ColLab* lesson series was implemented in both conditions, with the pre-test and post-test measuring changes in students' scientific-inquiry skills.

The two experimental conditions were structured as follows: In the PF dialogue condition, students engaged in PF dialogues via *Thinq Tank*, but the tool was not equipped with the new scripting features. Nonetheless, students received six sentence starters throughout the inquiry. In the scripted PF dialogue condition, two new features were introduced within *Thinq Tank*: (1) the Role Assignment feature, which assigned student groups to the role of assessor or assessee, and (2) the Dynamic Question Bank, which provided clickable sentence-starters adapted to the inquiry phase, student role, and identified areas for improvement. As in the previous iteration, *Thinq Tank* was used at the same points during the inquiry cycle (see Figure 6).

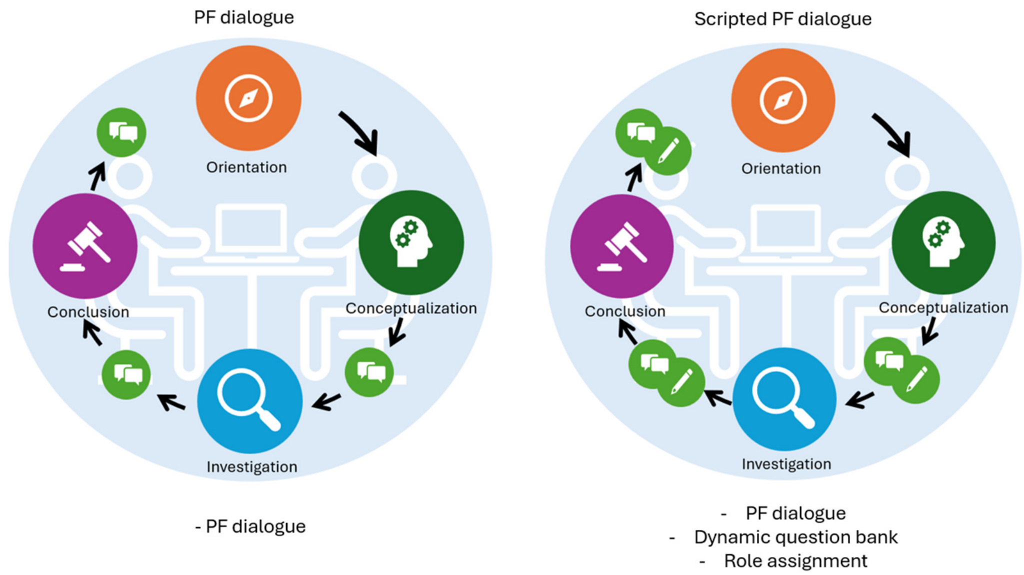

Figure 6. Formative PF scenario according to experimental conditions.

Data collection and analysis

For **RQ1**, which investigates how the lesson series with embedded formative peer assessment impacts students' learning outcomes, the scientific-inquiry skills of students were assessed by two independent raters (i.e., formulating a research question, setting up an investigation, and drawing a research conclusion; see Appendix B). The independent rater coded 30 percent of the answers to the pre-test to check the inter-rater reliability. The intra-class correlation coefficient (ICC) for the research question was 0.97; for the investigation, 0.87; and for the conclusion, 0.97, indicating good (0.75–0.90) to excellent (>0.90) inter-rater agreement [59]. As the data had a nested structure, linear mixed models were again used to analyse the effect on students' inquiry skills.

To answer **RQ2**, which explores how users perceive and interact with the design elements of *Thinq Tank*, students' perceptions of PF were measured through the pre-test and post-test. Factors such as the usefulness of PF [60], the engagement in providing PF [61], its utility, and the receptivity to PF [62] were assessed using validated questionnaires. Confirmatory factor analysis was conducted to confirm the subscales of the questionnaire for Flemish students, and the results are detailed in Appendix C. The perceived ease-of-use of *Thinq Tank* was also evaluated [63], with the survey instrument available in Appendix D. Linear mixed models were employed to analyse the differences in pre-test and post-test scores as well as conditions.

For **RQ3**, which examines the composition and structure of PF dialogues, 57 dialogues from the PF dialogue condition and 192 dialogues from the scripted PF dialogue condition were included in the qualitative dataset. Similar to design iteration 2, a random selection of 24 PF dialogues was drawn, compromising 8 dialogues from each inquiry phase. To allow for comparison with design iteration 2, the dialogues were likewise segmented into episodes and dialogue acts and coded using the same framework, which can be found in Appendix A. After coding the data, descriptive analysis was again conducted, and the results were compared with the descriptive analysis from the second design iteration.

2.3. Supportive Measures for Collaborative Inquiry and Peer Assessment

As none of the participants across the iterations had prior experience with CSCiL or peer assessment, several measures were implemented to support the collaborative inquiry as well as the peer assessment process. These measures addressed group collaboration and facilitator support, as well as the integrity of the peer assessment process.

To safeguard the quality of *collaborative inquiry,* particular attention was given to the formation of student groups and the roles of facilitators during the interventions. Teachers, who were familiar with their students, were responsible for forming collaboration groups to ensure compatibility and effective teamwork. In all iterations, two facilitators were present in each classroom to support the inquiry process. These facilitators were trained in the principles of CSCiL and actively monitored student progress, using the teacher dashboard tool to track responses and intervene when necessary. Students were also encouraged to ask questions at any time. When groups provided incorrect or insufficient answers, facilitators initiated a dialogue to help students correct their approach. In the first and third iterations, the first author and a scientific collaborator facilitated the sessions, providing guidance throughout the process. In the second iteration, due to the larger sample size, 54 rigorously trained master's degree students in Educational Sciences led the sessions, following strict protocols to scaffold the inquiry process, with two facilitators present per class to ensure effective support.

To support the *peer assessment process* during the second and third iterations, structured rubrics and targeted training were implemented. Rubrics, which were developed for each inquiry phase (i.e., research question, data collection, and conclusion), provided students with clear assessment goals and criteria, helping to structure their evaluations and clarify expectations [27]. These rubrics were developed in consultation with experts and pilot-tested in earlier phases, and then refined after each iteration. The rubrics used a 5-point Likert scale, helping students differentiate between performance levels. In the second iteration, students were also asked to provide PF via comments that were prompted to include strengths, weaknesses, and suggestions for improvement [64].

Additionally, students received training to ensure they could use the rubrics and give PF effectively. Research indicates that training enhances the reliability and validity of peer assessments [21,27,28]. Instructors explained the peer assessment process, its purpose, and the rubric in class, giving students the opportunity to ask questions. In the second iteration, a short instructional video demonstrated the peer assessment procedure, including working out examples to guide students through the process. In the third iteration, the training was embedded directly within SCORE, where students completed exercises based on example assessments and received feedback on their performance.

To further support the provision PF during the second iteration, sentence-starters were provided to guide the PF dialogue. These starters, such as "We do (dis)agree with the feedback about... because...", helped students engage in meaningful discussions and provide constructive feedback, rather than simply agreeing with one another's work [64]. Additionally, during the third iteration, the scripting condition itself served as a manipulation in order to enhance the PF process and collaborative learning by structuring interactions and guiding students through more productive exchanges.

Furthermore, through both iterations, facilitators played a key role in safeguarding the quality of PF. When PF lacked depth, was off-topic, or led to confusion or disagreement among students, facilitators intervened directly. They engaged with student groups to clarify misunderstandings, address concerns, and ensure that the PF process remained meaningful and aligned with the intended goals of the intervention.

3. Results

3.1. Design Iteration 1: Identifying Guidance Needs in Online CSCiL

The results of the first design iteration have been published in Van Hoe et al. [65], where the detailed findings for RQ1 can be found. While RQ2 data are also available there, the research primarily focuses on online collaboration. In this design article, we have chosen to provide a summarized presentation, with the detailed numerical data having been made available in the previously published version. What is new in this article is our focus on students' perceptions regarding receiving guidance—specifically, where and when they feel the need for support. We consider this to be the baseline for the further development of peer-based guidance in the subsequent design iterations. Revisiting these aspects is necessary to give a complete picture of the design process.

RQ1: Effects of CSCiL on scientific conceptual knowledge and inquiry skills

Quantitatively, the pre- and post-test results indicated that students exhibited significant improvements in their conceptual understanding of the climate change concept, as well as in their inquiry skills, indicating a positive learning effect of online CSCiL.

RQ2: Perceptions of CSCiL

The focus-group discussions provided insights into students' perceptions of CSCiL. Participants appreciated the interactive and engaging nature of CSCiL, noting that it encouraged greater activity and participation compared to traditional classroom settings. Participants highlighted the value of working together, which provided them with a richer understanding of scientific concepts and phenomena. However, the focus groups also shed light on the challenges associated with CSCiL. One key challenge is that about half of the students would have liked to have received more guidance from the teachers while going through the lessons. Students had to wait until teachers joined their group, sometimes resulting in long waiting periods, because the students had become stuck on some exercises and could not find a way out by themselves, with this ultimately resulting in the loss of a considerable amount of time.

Need for guidance during CSCiL

In conclusion, this study suggests that Climate colLab can effectively support students in their acquiring of scientific knowledge and the application of scientific-inquiry skills. However, finding the right balance of guidance for each student pair is a challenging task. Students often turn to the teacher for additional guidance at points when their peers and the online learning environment itself could also provide valuable support [50]. Leveraging these resources is a key strength of CSCiL, as it promotes a more collaborative and autonomous learning experience. Encouraging students to utilize peer support and digital tools can enhance their learning processes, fostering a more participatory and interactive educational environment [8,49].

3.2. Design Iteration 2: Implementing Dialogic PF Support Through Thinq Tank

3.2.1. Redesign and Implementation

Based on feedback from the previous iteration, in which students expressed a need for additional support, this second iteration addresses that need by incorporating extra support within Climate colLab. This is achieved by focusing on peer support, specifically, by involving different pairs of students in each other's inquiries through dialogic PF. This approach emphasizes one of the key aspects of CSCiL, namely, that peers support each other's learning through collaboration and interaction [16,66]. In this study, pairs of students provide formative peer assessment to one another during their inquiry, as it is an effective form of learning support.

To accomplish this, Thinq Tank was designed to enhance peer-to-peer interactions by allowing instructors to pair student groups for the provision of reciprocal peer assessments. Thinq Tank enables students to provide peer assessments on various activity components within SCORE, using both scoring (i.e., via rubrics) and feedback (i.e., via open comments). Additionally, Thinq Tank includes a chat function that facilitates PF dialogue, encouraging elaboration, negotiation, and deeper understanding of feedback (see Figure 3). These features enable students to give and receive peer assessment, and then revise their initial responses based on the feedback received. Collectively, these functions create a structured process for peer assessment, as illustrated in Figure 7.

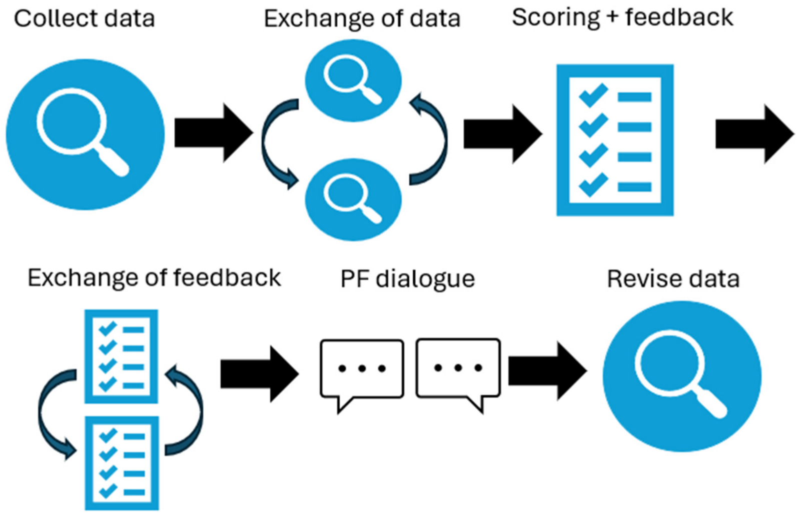

Figure 7. Peer assessment procedure in Thinq Tank, applied to the investigation phase.

To evaluate the implementation of the Thinq Tank tool within Climate ColLab and address the research questions, a quasi-experimental study was conducted. This study compared a control condition involving direct teaching with two experimental conditions: Thinq Tank with peer assessment and Thinq Tank with a complementing PF dialogue (see Figure 5).

3.2.2. Results of the Analysis

We take a similar approach for the second design iteration, for which the detailed figures and analyses for both RQ1 and RQ2 can be found in Van Hoe et al. [15]. As with design iteration 1, we provide a summarized version here, as we believe it is important to present the full picture of the results on which the redesign for the next iteration was based. What is new in this design article, however, relates to the data for RQ3. These data are crucial, as they directly informed the redesign process for the subsequent iteration.

RQ1: Effects of CSCiL with formative peer assessment on scientific conceptual knowledge and inquiry output

The study confirmed that Climate colLab, this time embedded with formative peer assessment, significantly enhanced students' scientific conceptual knowledge of climate change, aligning with the previous research findings of design iteration one. Across the experimental conditions, about half of the student pairs made adjustments to the inquiry output after receiving peer assessment. The study revealed varying frequencies of adjustments to inquiry output, with research questions and data receiving more adjustments, compared to conclusions, after engaging in (dialogic) peer assessment. Interestingly, the inclusion of PF dialogues did not significantly impact the number of adjustments made. When students did choose to make adjustments following peer assessment, there was a significant improvement in the quality of all types of inquiry output, with the most substantial improvements being observed in the quality of research conclusions. No significant difference in the size of the quality improvement was found between the experimental conditions.

RQ2: Effects of (dialogic) PF on PF perceptions

Students' perceptions of peer assessments were generally positive. They agreed that both forms of peer assessment were fair and useful, and they indicated that they accepted feedback from their peers, with no significant differences in perceptions found between the two experimental conditions. Additionally, students in both groups reported that peer assessment did not evoke as many negative emotions as initially expected, indicating an overall positive experience with (dialogic) peer assessment.

RQ3: Exploration of PF dialogues

In the 24 PF dialogues analysed, 52 episodes and 220 dialogue acts were identified, averaging 2.17 (SD = 0.87) episodes and 9.17 (SD = 7.51) dialogue acts per dialogue. On average, each episode comprised 4.42 (SD = 2.96) dialogue acts.

Our qualitative analysis revealed several key insights (see Table 5 below). First, statements formed the largest proportion of the dialogue acts, constituting 39.6% (n = 87). Notably, evaluations represented a significant portion of these statements, amounting to 20.9% (n = 46). These evaluative statements predominantly expressed positive feedback, affirming various aspects of the inquiry outputs. Interestingly, while the positive tone was prevalent, there was also a smaller but important fraction of evaluations that expressed criticism as to the inquiry output. Explanations constituted the smallest category of statements, making up 6.4% (n = 16) of the total. These explanations aimed at elucidating or defending one's inquiry products and the rationale behind them, as well as interpreting or justifying the PF given.

Second, social interaction comprised 18.6% (n = 41) of the dialogue acts, including greetings, expressions of social support, and instances of social resistance. Positive social support, such as jokes, were much more prevalent (7.3%, n = 16), contrasting sharply with the fewer derogatory remarks found in the social resistance acts (2.7%, n = 6).

Third, questions posed by students represented 16.8% (n = 37) of the dialogue acts, with a notable portion seeking explanations to clarify the feedback received (7.7%, n = 17). This indicates a genuine interest in understanding PF, although questions seeking additional evaluations (5.0%, n = 11) or further information were less common (0.9%, n = 2).

Fourth, directives accounted for 9.5% (n = 21) of the dialogue acts, including commands, suggestions, and requests, all aimed at improving the inquiry outputs. The presence of such directives underscores the students' engagement in providing actionable feedback. However, the number of directives was limited, which implies that students think that their main role is to point out the good and bad aspects of the inquiry output, and not collaborative thinking about how the inquiry output can be improved.

Lastly, supporting answers, such as agreements, acknowledgments, and compliance, made up 7.8% (n = 17) of the dialogue acts, whereas confrontational answers, including challenges and contradictions, accounted for 6.3% (n = 14). While the willingness to challenge and contradict peers indicates a level of critical engagement crucial for meaningful PF, it is evident that agreeing and complying were more common.

In conclusion, these results show that PF dialogues often lacked depth, were brief, and consisted of frequently changing topics. Additionally, a significant amount of time was spent discussing the PF itself rather than focusing on how to collaboratively improve the inquiry outputs. Moreover, the predominance of positive feedback overshadowed necessary constructive criticism, which is crucial with regard to motivating significant changes to the inquiry output. Our analysis also highlighted that explanations of PF were limited, possibly hindering students' understanding and their ability to enhance their inquiry outputs. Although some critical engagement was present in responses, it was less common than agreements and compliance, indicating a passive attitude towards receiving feedback. While PF dialogues included valuable feedback and positive social interactions, they often lacked the depth and critical collaboration needed for meaningful improvement. Effective PF dialogues should balance positive reinforcement with constructive criticism, provide detailed justifications for feedback, and foster a collaborative environment to truly enhance inquiry outputs.

3.3. Design Iteration 3: Scripting Dialogic PF Through Thinq Tank

3.3.1. Redesign and Implementation

Based on the results of second iteration, we made substantial updates to Thinq Tank. In order to promote the collaborative character of dialogic PF, we decided it would be useful to also redesign the inquiry assignment in terms of collaboration promotion. Therefore, we decided to adhere the core principles of collaborative learning as described by Johnson and Johnson—namely, positive interdependence, individual accountability, active interaction, support of social skills, and group processing [67]. This transformed the inquiry task into a collaborative research project between student pairs, in which they conducted investigations together and provided feedback to each other. By together choosing a topic from a sustainability documentary and dividing research areas between them, these student pairs aim to achieve a thorough understanding of the subject through shared inquiry efforts. This setup ensures that, for each student pair, their success is directly linked to their peers' input, fostering a supportive environment that still highlights personal responsibility through the PF dialogues. Interactions within student pairs are face-to-face, while those between pairs take place online through Thinq Tank.

Next, to support the PF dialogues, and responding to the shortcomings found in the previous design iteration, such as the lack of critical feedback or thinking together about how the inquiry output can be improved, it was decided to adopt some scripting methods within Thinq Tank. Scripting in CSCL involves creating a series of instructions or prompts that scaffold learners' interactions to enhance the quality of collaboration and learning outcomes [68]. Students themselves also express the need for guidance during dialogic PF in other educational contexts [42]. This desire for guidance is supported by the prior meta-analysis of [45], which discovered that supporting strategies during peer assessment improve the latter's potential. Given the pivotal role of interaction in PF dialogues, it is imperative to explore the necessary support that an online PF tool can offer to both assessors and assessees. Both parties, and not only the assessor, should be empowered to express their perspectives constructively and clearly [69,70] and should receive adequate support in doing so [25,43,44].

To this end two new scripting features were introduced to Thinq Tank: (1) the Role Assignment feature, which assigns student groups to the assessor or assessee role; and (2) the Dynamic Question Bank, which offers sentence-starters that support student pairs during the PF dialogues. Both features are embedded within the peer chat of Thinq Tank.

The *Role Assignment* feature introduces an automated system used for designating roles among the student pairs who engage in PF dialogues with one another. Specifically, it randomly assigns students to act as either assessors or the ones being assessed in the PF dialogues. This role is dynamically switched after each PF dialogue, ensuring that student groups experience both sides of the PF dialogue within one inquiry phase.

The *Dynamic Question Bank* is designed to script the PF dialogue by offering structured yet flexible sentence starters to students (see Figure 4). Bürgermeister et al. introduced sentence starters to aid student teachers in generating PF, resulting in improved feedback quality [33]. By narrowing the range of options, sentence starters provide learners with structure and concrete suggestions for formulating feedback, thereby potentially steering both parties toward a constructive and supportive learning environment [33,69]. Sentence starters could address various challenges across cognitive dimensions (e.g., enhancing the articulation and expression of feedback and back-feedback, promoting ongoing dialogue among peers, stimulating clarification and justification of feedback) and socio-affective dimensions (e.g., fostering trust among peers, enabling them to see each other as reliable and trustworthy assessors) of peer assessment, while also promoting high-level thinking skills (e.g., the formulation of critical feedback) [25,44]. In the context of this research, these sentence starters are tailored based on the inquiry phases and the specific role (i.e., assessor or assessee) the students in the pairs are designated in the PF process. For those in the role of assessor, the question bank provides prompts that help articulate focused constructive feedback, both positive and critical, based on identified areas for improvement in their peers' work. Conversely, when students find themselves in the roles of the ones being assessed, the sentence starters are designed to elicit comprehensive, reflective responses that engage with the feedback received. Thus, each set of sentence starters that Thinq Tank provides to student pairs during PF dialogues is unique and personalized, promoting a PF dialogue that is relevant to the participants' specific learning processes. Moreover, these sentence starters are clickable, meaning that when they are clicked, they are interjected within the peer chat and students can adjust them according their needs. All of the sentence starters can be found in Appendix E.

The addition of these two features to Thinq Thank introduces a new cyclic process of dialogic PF, illustrated in Figure 8.

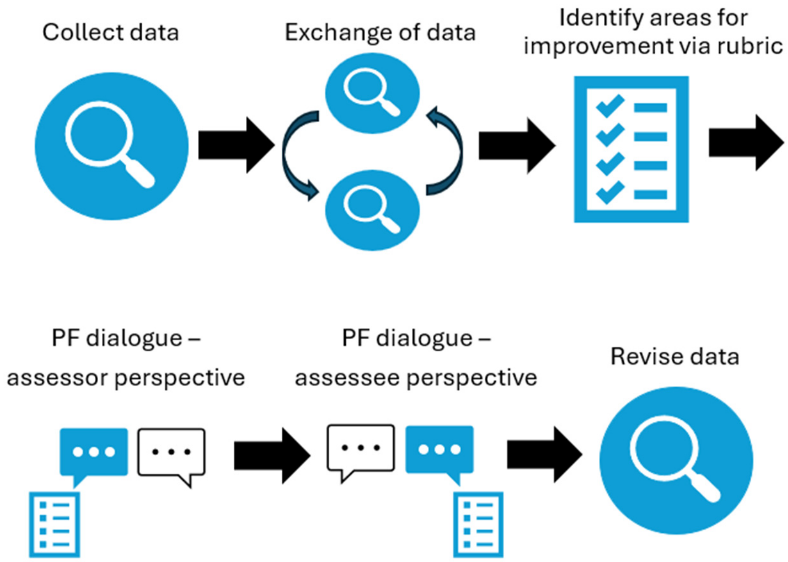

Figure 8. PF procedure in Thinq Tank, applied to the investigation phase.

To answer the research questions regarding the effect of scripting features within Thinq Tank, a quasi-experimental study was conducted. This study compared two experimental conditions: one with the scripting features (role assignment and dynamic question bank) and one without them (see Figure 6). The goal was to evaluate how these features influenced the PF dialogues and supported students' engagement and learning outcomes.

3.3.2. Results of Analysis

For the third design iteration, all data for each research question are provided as originally processed for this article. This provides the final insights into the development of the Thinq Tank tool. The results presented here represent the culmination of the design process, offering a comprehensive understanding of the tool's impact and its potential for further implementation.

RQ1: Effects of CSCiL with formative PF dialogues on scientific inquiry skills

Table 1 presents descriptive statistics of students' scores on the pre- and post-tests, categorized by inquiry output. The initial levels of proficiency among students varied across different inquiry skills, as evidenced by the pre-test scores. Prior to the intervention, students demonstrated a relatively high level of competence in formulating research questions, and proficiency in conducting investigations and gathering relevant information, as indicated with the mean scores of, respectively, 3.99 (SD = 0.91) and 3.81 (SD = 0.76). This suggests that students possessed a solid foundation in these two areas but had room for improvement in refining their skills further. The lowest initial proficiency level was observed in the conclusion skill, where students obtained a mean pre-test score of 2.77 (SD = 1.09). This indicates that students faced greater challenges in drawing meaningful conclusions based on their inquiry findings, highlighting an area requiring substantial development.

Table 1. Descriptive analysis of students' pre- and post-test scores across inquiry skills.

	Pre (M, SD)	Post (M, SD)
Research Question	3.99 (0.91)	4.45 (0.65)
Investigation	3.81 (0.76)	4.14 (0.73)
Conclusion	2.77 (1.09)	3.80 (1.06)

Note: scores should be interpreted based on a total of 5 points.

Table 2 presents a two-level modelling analysis depicting the progress made by students between the two test moments. In Model 0, representing the intercept-only model, the intercept signifies the average progression across all three inquiry skills, with a measurement of 0.57 (SD = 0.08). This indicates substantial overall improvement in students' scores from the pre- to the post-test across all inquiry skills. In Model 1, the predictor 'inquiry skill' is incorporated, facilitating a comparison of improvement sizes among the inquiry skills. Here, the intercept reflects the average enhancement, specifically, in the inquiry skill of formulating research questions. Notably, the Tukey test for post hoc comparisons reveals no significant difference between the progress made in formulating research questions (M = 0.46, SD = 0.10) and that in gathering relevant information during investigation (M = 0.29, SD = 0.12). However, a significant disparity emerges between the progress made in formulating research questions and investigation, and that in drawing research conclusions, indicating a notably greater advancement (M = 1.00, SD = 0.12) in the latter. This could possibly be explained by the lower starting level associated with drawing conclusions, as noted earlier, which leaves more room for improvement in this skill.

Table 2. Summary of the model estimates for the two-level analysis of students' difference scores for inquiry skills, as measured in the pre- and post-tests.

	Model 0		Model 1	
	Coefficient	**SE**	**Coefficient**	**SE**
Fixed effect				
Intercept	0.57 ***	0.08	0.46 ***	0.10
Data			−0.17	0.12
Conclusion			0.54 ***	0.12
Random effect				
Class variance	0.04	0.19	0.04	0.20
Student variance	1.25	1.12	1.17	1.08
Model fit				
Loglikelihood	−751.67		−733.72	
χ^2, df = 2			35.893	
p			$p < 0.001$	

Note. Scores should be interpreted based on a total of 5 points.; *** $p < 0.001$.

This analysis confirmed that Climate colLab significantly enhanced students' scientific-inquiry skills in both conditions, consistent with the findings from iteration one, demonstrating that Climate colLab's effectiveness remained intact after the third iteration redesign.

RQ2: Effects of (dialogic) PF on PF perceptions

The multilevel models in Table 3 provide insights into students' perceptions of PF dialogues. In each of the four models, the intercept of the null model represents the average score on the scale. Model 1 introduces the experimental condition as a predictor to examine whether perceptions significantly differ based on whether students received support during peer dialogue through scripting. Overall, students indicated they were open to receiving PF, found it valuable for their learning process, and put effort into delivering high-quality PF. The only difference between the two conditions was in receptivity towards PF: students in the scripted dialogue condition reported being more receptive to the PF compared to those in the control condition, without scripting.

Table 3. Summary of the model estimates for the two-level analysis of students' scores regarding PF perceptions.

	Receptivity (/5)				Utility (/5)			
	Model 0		**Model 1**		**Model 0**		**Model 1**	
	Coefficient	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**
Fixed effect								
Intercept	4.04 ***	0.06	3.88 ***	0.08	3.93 ***	0.07	3.86 ***	0.13
Scripting			0.25 *	0.10			0.11	0.16
Random effect								
Class variance	0.01	0.11	0.00	0.00	0.03	0.16	0.00	0.00
Student variance	0.38	0.62	0.38	0.62	0.51	0.72	0.38	0.62
Model fit								
Loglikelihood	−168.74		−165.81		−196.84		−196.59	
χ^2, df = 1			5.8525				0.5018	
p			0.01555 *				0.4787	
	Usefulness of PF (/7)				**Providing PF (/6)**			
	Model 0		**Model 1**		**Model 0**		**Model 1**	
	Coefficient	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**
Fixed effect								
Intercept	5.24 ***	0.15	5.04 ***	0.27	4.68 **	0.07	4.59 ***	0.12
Scripting			0.30	0.33			0.13	0.15
Random effect								
Class variance	0.18	0.43	0.19	0.43	0.02	0.14	0.02	0.14
Student variance	1.52	1.23	1.52	1.23	0.55	0.74	0.55	0.74
Model fit								
Loglikelihood	−331.62		−331.42		−201.78		−201.32	
χ^2, df = 1			0.9652				0.9276	
p			0.3259				0.3355	

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Finally, multilevel models were constructed for each of the 14 items of the Perceived Ease-of-Use Questionnaire [63]. Detailed descriptive statistics for each item are provided in Appendix D. The multilevel analyses, found in Table 4, indicate significant differences between the two conditions across three items of the questionnaire. Students under the scripting condition more strongly indicated that Thinq Tank with the embedded scripting was easier to use, in comparison to students under the control condition, who experienced Thinq Tank without the added scripting. Students under the scripting condition also reported that Thinq Tank, with its scripting features, provided better support for giving feedback and made it easier for them to correct errors.

Table 4. Summary of the model estimates for the two-level analysis of students' perceived ease-of-use of Thinq Tank.

	I Find it Easy to Recover From Errors Encountered While Using the Feedback Tool.				The Feedback Tool Provides Helpful Guidance in Giving Feedback.				Overall, I Find the Feedback Tool Easy to Use.			
	Model 0		**Model 1**		**Model 0**		**Model 1**		**Model 0**		**Model 1**	
	Coefficient	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**
Fixed effect												
Intercept	4.93 **	0.16	4.42 ***	0.19	5.06 ***	0.25	4.29 ***	0.35	5.07 ***	0.19	4.52 ***	0.28
Script			0.80 ***	0.24			1.18 **	0.43			0.86 *	0.34
Random effect												
Class variance	0.13	0.36	0.00	0.00	0.59	0.77	0.32	0.56	0.27	0.52	0.15	0.38
Student variance	2.14	1.46	2.12	1.46	2.02	1.42	2.02	1.42	2.12	1.46	2.10	1.45
Model fit												
Loglikelihood	−296.27		−291.66		−292.14		−288.62		−292.30		−289.23	
χ^2, df = 1			9.2125				7.0491				6.1353	
p			0.002 **				0.00793 **				0.01325 *	

Note: scores should be interpreted based on a total of 7 points. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

RQ3: Exploration of PF dialogues

The qualitative analysis revealed that the 24 PF dialogues could be divided into 38 episodes and 248 dialogue acts. This indicates that an average PF dialogue consisted of 1.58 (SD = 1.41) episodes and 10.33 (SD = 7.52) dialogue acts. On average, each episode comprised 7.43 (SD = 4.00) dialogue acts.

In the detailed analysis of dialogue acts, an overview of which is provided in Table 5, several patterns emerged. The largest proportion of dialogue acts, 40.7%, (n = 101), were statements, with a significant portion, 29.8% (n = 74), being dedicated to evaluations. While a mix of positive affirmations and critical comments was present, the majority were critical, highlighting areas for improvement in the inquiry outputs. Another surprising finding was that despite the use of scripting, the fraction of statements devoted to explanations was small, namely 7.7% (n = 19); these were almost exclusively about clarifying the PF provided.

Table 5. Summary of the qualitative analysis of the PF dialogues.

	Iteration 2	Iteration 3
Episodes	**52**	**38**
Dialogue acts	**220**	**248**
Statements	**39.6%**	**40.7%**
Evaluation	*20.9%*	*29.8%*
Explanation	*6.4%*	*7.7%*
Information	*12.3%*	*3.2%*
Social interaction	**18.6%**	**13.7%**
Greetings	*8.6%*	*9.3%*
Social support	*7.3%*	*3.2%*
Social resistance	*2.7%*	*1.2%*
Questions	**16.8%**	**12.9%**
Explanation	*7.7%*	*2.8%*
Evaluation	*5.0%*	*10.1%*
Information	*0.9%*	-
Directives	**9.5%**	**13.7%**
Commands	*6.8%*	*3.2%*
Suggestions	*1.8%*	*10.5%*
Requests	*0.9%*	-
Answers	**14.1%**	**17.6%**
Supporting	*7.8%*	*11.2%*
Confrontational	*6.3%*	*6.4%*
Online learning environment	**1.4%**	**1.4%**

Next, directives accounted for 13.7% (n = 34) of the dialogue acts, with suggestions (10.5%; n = 26) being more common than commands (3.2%; n = 8). This indicates that students were actively engaged in providing constructive advice to improve the quality of the inquiry outputs. Student groups actively collaborated to enhance the quality of the inquiry outputs, viewing it as a shared responsibility.

Social interactions also played a notable role, comprising 13.7% (n = 34) of the dialogue acts. Most social remarks were greetings (9.3%; n = 23), which helped open discussions, followed by social support (3.2%; n = 8), the latter often manifesting as jokes that fostered a positive atmosphere. However, social resistance (1.2%; n = 3) included negative remarks and swearing, a phenomenon which could disrupt the collaborative spirit.

Questions made up 12.9% (n = 32) of the dialogue acts, with a majority (10.1%; n = 25) seeking evaluation of the inquiry output and approval for its revision, while only a minority (2.8%; n = 7) sought additional explanations about the PF, suggesting that students were keen on improving their work. One possible explanation could be that the scripting methods supported the students in clearly communicating their PF, for which no additional explanations were necessary.

Supporting acts, such as acknowledgments, compliance, and agreement, accounted for 11.2% (n = 28) of the interactions, indicating a generally cooperative environment. However, 6.4% (n = 16) were confrontational acts, including challenges and contradictions, showing that there was some critical engagement, though infrequent.

The qualitative analysis of the PF dialogues reveals that these interactions were deeply focused on specific pieces of feedback, encouraging students to engage critically with each other's work. The dialogues were characterized by a mix of positive affirmations and constructive criticism, with students often offering suggestions for improvement and viewing the quality of inquiry outputs as a shared responsibility. Social interactions, including supportive comments and jokes, helped create a positive and collaborative atmosphere, although occasionally, negative remarks were noted. Questions primarily sought evaluations and approval for revisions, indicating students' eagerness to improve their work. Overall, the scripted peer dialogues facilitated active engagement, critical but constructive feedback, and a collaborative learning environment, though there is potential for further enhancing the depth and explanatory nature of these interactions.

4. Discussion

The principal objective of this research was to explore how PF can be designed as a dialogue and to validate the Thinq Tank Tool as an instrument for formative PF within the framework of CSCiL. Through a rigorous iterative design process, the goal was to refine the tool itself and enhance the collaborative learning experience by fostering meaningful and constructive PF dialogues. Grounded in the theoretical foundations of CSCiL [48], formative PF, and dialogic PF, our study emphasized the shift from traditional, unilateral feedback models to collaborative, dialogue-based interactions [12,37]. The incorporation of Thinq Tank into the CSCiL platform aimed to actualize these modern feedback paradigms by enabling real-time, reciprocal feedback dialogues among peers. Throughout the iterations of our study, enhancements to Thinq Tank were informed by both quantitative and qualitative data analyses. Each iteration focused on distinct aspects of the tool and its deployment within Climate colLab: the first iteration focused on identifying needs of students during CSCiL; the second tested PF modalities with or without PF dialogue; and the third explored the effect of adding additional scripting via dynamic question banks and explicit role assignments in order to enhance formative PF dialogues. This structured approach to iterative improvement allowed for targeted refinements based on specific feedback, aligning closely with the principles of design-based research [54].

The results from the first research question across the three design iterations indicate that CSCiL, when integrated with formative peer assessment, particularly dialogic PF, can enhance learning outcomes such as students' scientific conceptual understanding and inquiry skills. These findings align with the existing literature emphasizing the positive effects of CSCiL on conceptual learning and inquiry [66,71]. Specifically, in design iteration 2, the inquiry outputs—such as the quality of students' research questions, data analysis, and conclusions—demonstrated notable improvement after revisions based on peer assessment, both with and without PF dialogue. This suggests that integrating dialogic PF at key phases of the inquiry cycle can support students in refining their inquiry output and producing more sophisticated work. These findings are in line with studies highlighting the importance of formative peer assessment in enhancing inquiry-based learning outcomes, confirming previous findings that peer assessment can exacerbate positive influences on learning and skill development within STEM subjects [5,20].

Our findings for the second research question, which examines students' perceptions and interactions with the Thinq Tank tool, show that students consistently view PF as a valuable guide during CSCiL, one addressing their need for additional guidance identified in the first design iteration. During the second iteration, it was found that students viewed (dialogic) peer assessment as fair and useful and accepted it. Moreover, the incidence of negative emotions was low, indicating a predominance of positive interactions with Thinq Tank. This positive perception of formative peer assessment within CSCiL aligns with previous findings, in which students also related that they enjoyed the process and would be willing to use it again in the future [53]. During the third iteration, findings showed that students recognized the positive impact of Thinq Tank on their inquiry processes, appreciating the providing and receiving of thoughtful PF. Notably, students in the scripted condition reported a higher receptiveness to PF. This marks a positive shift, as previous research in CSCiL contexts found that students often ignored PF [52,53]. With students now more open to engaging with PF, this could potentially lead to a greater uptake and integration of PF in their work. This increased receptiveness possibly can be attributed to the clear guidance and structured support provided by Thinq Tank. Features like the role assignments and dynamic question bank were particularly valued for enhancing the clarity and support of the feedback process, aiding students in articulating and substantiating their feedback more effectively. The role assignment feature, which alternated roles between assessors and assessees, was instrumental in promoting equitable participation and deepening students' understanding of and appreciation for the feedback process. Additionally, these scripting methods contributed to the reasons why students in the scripted condition found Thinq Tank easier to use and more effective in correcting mistakes, compared to those in unscripted settings. Despite Thinq Tank's complexity due to its additional features, the prolonged exposure and the opportunity to interact from multiple perspectives facilitated ease-of-use and likely enhanced students' overall experience and engagement with the tool.

To answer the third research question regarding the facilitation of PF dialogues within Thinq Tank, a comparison between unscripted (i.e., iteration two) and scripted PF dialogues (i.e., iteration three) reveals significant differences in how students navigate PF dialogues within a CSCiL context. In the unscripted condition, dialogues displayed notable spontaneity, with episodes and dialogue acts varying widely in content and length, illustrating the inherent flexibility that allows students to tailor communication to their preferences. However, non-scripted dialogues tended to be brief and often shifted topics. With a straightforward approach, they quickly reached the point of the feedback, but this sometimes left student inquiries unanswered. Both parties focused on fulfilling the task of providing feedback, often leaving the conversation without truly addressing each other's comments. Conversely, the introduction of scripting not only reduced the number of episodes but also increased the number of dialogue acts per dialogue, suggesting more deep and focused exchanges. In this condition, sentence-starters encouraged more sustained and meaningful conversations, promoting the in-depth discussions which were less present in the unscripted condition. In contrast to earlier research in which PF dialogues often lacked depth and struggled to gain momentum, the use of sentence-starters in this study succeeded in facilitating and engaged discussions [43,47]. While social interaction remained consistent across conditions, scripted dialogues replaced informal banter with formal openings and closings, enhancing conversational structure and mirroring real-life dynamics more closely. This shift likely fostered a more welcoming and respectful dialogue atmosphere, increasing openness and receptivity among participants. Furthermore, in contrast to earlier findings where guiding questions led to some negativity, sentence-starters succeeded in keeping the conversation going and maintaining a positive atmosphere [43].

Both conditions featured positive feedback, reflecting the supportive nature of the PF dialogues facilitated by Thinq Tank. In the unscripted condition, students predominantly provided positive evaluative statements, mostly affirming their peers' work. This lack of critical engagement suggests that students felt more comfortable reinforcing each other's work rather than challenging it. This aligns with previous research on formative peer assessment within CSCiL, for which students similarly gave little critical feedback and mainly confirmed each other's work, resulting in few revisions and missed opportunities to improve the quality of their inquiry products [52,53]. In contrast, the scripted condition prompted a significant shift toward more critical feedback, directing dialogues toward specific areas for improvement and making the feedback more actionable. This was evident in the fact that scripted groups generated almost twice as many directives as those under the unscripted condition. By encouraging a suggestion-oriented communication style, Thinq Tank facilitated a more collaborative interaction in which students are partners in the inquiry process, collaboratively exploring improvements that may optimize the inquiry output. This interaction mode not only aligns with constructivist learning theories, which emphasize active engagement and co-construction of knowledge, but also framed the feedback process as a collaborative endeavour rather than a unilateral assessment [7,72] By encouraging a suggestion-oriented communication style, Thinq Tank minimized resistance to feedback [27] and promoted more collaborative and exploratory interactions.

4.1. Implications for Theory and Practice

The results from the iterative design process offer valuable theoretical implications that contribute to the existing research on CSCiL and formative peer assessment. Specifically, the findings support the idea that integrating dialogic PF into CSCiL environments can not only enhance students' conceptual understanding and inquiry skills but also foster a more collaborative and reflective learning process. This aligns with theories that emphasize the importance of social interaction and co-construction of knowledge, a proposition which is further strengthened by the structured support provided by tools, like Thinq Tank, that lead to enhanced learning outcomes [9,10]. Moreover, the results show that the scripted nature of dialogic PF helps to establish a more collaborative, reciprocal dynamic. This suggests that scripted PF can shift students from passive receivers to active participants in the feedback process, challenging traditional approaches that rely on one-directional feedback, such as scoring and comments given without interaction. It also highlights the potential of PF to foster meaningful collaborative learning experiences, encouraging deeper engagement and reflection among students. These implications offer a new direction for theoretical development in the areas of peer assessment, collaboration, and inquiry-based learning, suggesting that future studies should continue to explore the intersection of these domains with structured technological supports to maximize the educational impact.

Based on the findings from our design research on the Thinq Tank tool within a CSCiL context, we have identified several key design guidelines that can inform the development and implementation of online formative dialogic peer assessment tools. These guidelines are essential for facilitating dialogic PF and enhancing the overall collaborative learning experience. Additionally, they offer practical insights for educators and instructional designers looking to effectively integrate peer assessment into collaborative learning environments. One key design guideline is the facilitation of structured dialogue through dynamic tools. Our findings highlighted the significant benefits of structured dialogues in peer assessment, particularly when supported by features like the Role Assignment and Dynamic Question Bank. This tool proved instrumental in providing users with tailored sentence-starters, which not only focused the feedback but also guided students toward constructive and critical evaluations. This structure supports the development of critical thinking and evaluative skills, ensuring that feedback is not only frequent but also substantive and targeted toward specific improvement areas. Another important design guideline is promoting empathy and understanding through role-switching. The Role Assignment feature implemented in Thinq Tank allowed students to alternate between giving and receiving feedback, fostering a deeper understanding of the peer assessment process. This feature was effective in building empathy and a more balanced perspective on the peer assessment dynamics, as students experienced both sides of the feedback exchange. This role-switching encourages a more equitable and comprehensive engagement in the learning process, helping students to appreciate the challenges and responsibilities of both roles. We also emphasize the design guideline of embedding collaborative learning principles. The integration of collaborative learning principles into the design of Climate colLab facilitated a supportive learning environment that emphasized mutual support, shared goals, and individual accountability [67]. By aligning the tool's functionalities with these principles, the PF process became a cooperative venture that focused not only on individual contributions, but also on collective progress and understanding. Prioritizing usability and accessibility is another crucial design guideline. Usability and accessibility are essential for ensuring that all students can effectively engage with the tool without unnecessary complications. Thinq Tank was designed to be intuitive and user-friendly, with clear instructions and a supportive interface that made navigation and interaction straightforward. This ease-of-use is vital for maintaining student engagement and ensuring that the technology enhances rather than hinders the learning experience. Finally, the guideline aiming to ensure real-time interaction where possible was also highlighted in our study. Thinq Tank facilitated real-time interaction which kept the dialogue dynamic and more closely simulated face-to-face interactions. This immediacy in communication was crucial for maintaining the flow of discussion and ensuring that feedback was timely and relevant, thereby enhancing the responsiveness and engagement of the feedback process. These guidelines, rooted in empirical findings from the Thinq Tank iterations, suggest that the design of PF tools should not only focus on the functional aspects of feedback provision but also on fostering a supportive learning environment. By structuring interactions, facilitating empathetic exchanges, embedding collaborative principles, ensuring ease-of-use, and maintaining real-time engagement, online formative PF tools can significantly enhance the quality and effectiveness of PF in CSCiL settings.

4.2. Limitations and Suggestions for Future Research

The research conducted with the Thinq Tank tool has provided substantial insights into the application of PF in a CSCiL context. However, certain limitations and areas for future research have been identified based on the iterative design and findings of this study. One notable limitation is the structured nature of the Thinq Tank tool, which, while facilitating focused and constructive dialogue, may restrict spontaneous student interactions which can also yield valuable insights. The use of sentence-starters and role assignments, although beneficial in maintaining dialogue structure, could potentially constrain natural discourse dynamics. Future research could explore the balance between scripted and unscripted dialogue elements to optimize engagement without sacrificing the quality of the feedback.

Additionally, the study's dependence on technological interfaces may affect the generalizability of findings to contexts where digital tools are less integrated. The interaction dynamics observed in the Thinq Tank platform might not directly translate to traditional classroom settings, places where communication norms and educational practices vary. In future research, it would be valuable to explore how these tools perform across diverse educational settings and to examine the longitudinal impact of structured peer assessment tools on students' learning outcomes. Further studies could also investigate the scalability of such tools and their adaptability to various subject matters or larger group sizes. Moreover, the integration of real-time feedback mechanisms could be examined more closely to determine how immediacy affects the quality and reception of peer assessments. These suggestions aim to expand on the current research by addressing its limitations and exploring new dimensions of peer assessment within CSCiL, ultimately enhancing both the theoretical understanding and the practical applications of these educational tools.

5. Conclusions

In conclusion, the iterative development and testing of the Thinq Tank tool within the Climate colLab series have significantly enhanced our understanding of formative PF dialogues in CSCiL. The findings demonstrate that CSCiL, integrated with formative peer assessment, contributes to improvements in students' scientific conceptual understanding and inquiry skills, as well as their inquiry outputs. Moreover, students perceived the Thinq Tank tool positively, finding it easy to use. Additionally, students indicated that the PF dialogues were useful and fair, and the students were open to receiving PF. Next, the findings demonstrate that scripting can facilitate PF dialogues. Features such as role-switching and dynamic question banks lead to deeper, more constructive interactions, enabling students to critically evaluate each other's work and collaboratively discuss ways to improve it. This study highlights the importance of continually refining PF tools to ensure that they support effective dialogic exchanges, which are vital for CSCiL. Future research should investigate the scalability of such tools across various educational levels and subjects to ensure adaptability and effectiveness in diverse learning settings.

Author Contributions: Conceptualization, A.V.H. and J.W.; methodology, A.V.H.; software, J.W. and J.S.; formal analysis, A.V.H.; investigation, A.V.H.; resources, A.V.H.; data curation, A.V.H., J.W. and J.S.; writing—original draft preparation, A.V.H.; writing—review and editing, A.V.H., J.W., J.S., T.R. and T.S.; visualization, A.V.H.; supervision, J.S., T.R. and T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Special Research Fund of Ghent University.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Ghent University (protocol code 022/004; 4 March 2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Dataset available on request from the authors.

Acknowledgments: We would like to thank the ENCORE lab of the University of Toronto for programming the peer assessment tool in SCORE to make this study possible.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Coding scheme for dialogue acts—initiating acts.

INITIATING ACTS	**Offer**	Wish Offer
	Directive	Request Suggestion Command
	Statement	Information Explanation Evaluation
	Purpose of question	Explanation Evaluation Information
	Social activities	Greetings Support Resistance
	Online learning environment	

Table A2. Coding scheme for dialogue acts—responding acts.

	Question		Clarification Confirmation
	Response to offer		Acceptance Rejection
	Response to directive		Acknowledgement Compliance Challenge Refusal
	Response to statement		Acknowledgement Agreement Contradiction
RESPONDING ACTS		Offer	
	Response to question	Directive	Request Command Suggestion
		Statement	Information Explanation Evaluation
	Social activities		Greetings Support Resistance
	Online learning environment		

Appendix B

There are two versions of this task measuring scientific-inquiry skills; both are focused on sustainability, but they cover different topics: **e-waste** and **fast fashion**. Although the subjects are different, the structure and assignments are identical. Each student completed one version in the pre-test and the other in the post-test, meaning that if a student worked on e-waste in the pre-test, they worked on fast fashion in the post-test, and vice versa

From Electronics to Waste

Electronic devices have become a part of our daily lives. As a result, the production and consumption of these devices have skyrocketed in recent years, leading to a growing problem of electronic waste, or e-waste.

It is estimated that the world generates about **53.6 million tons** of e-waste each year. This is worrying because the production of electronic devices relies on valuable natural resources, and the fast disposal of these devices causes these resources to run out quickly.

Recycling e-waste also produces a significant amount of greenhouse gases. Moreover, many electronic devices contain **dangerous substances** that, when not recycled properly, can leak into the environment and be harmful to both nature and humans.

Your Task as a Researcher

You are a researcher, and you've been asked to investigate one of the consequences of electronic waste mentioned above. Follow these steps to complete your task:

1. Formulate a research question.
2. Create a hypothesis.
3. Find three online sources to help answer your research question.
    a. Copy and paste the link of each online source you plan to use into the provided space.
    b. Copy and paste the information from the online source that you will use to answer your research question into the provided space.
4. Formulate your answer

Appendix C

Table A3. Confirmatory factor analyses.

<table>
<tr><th>Scales</th><th>Items</th><th>Likert</th><th>Example</th><th>Test</th><th>p-Value</th><th>CFI</th><th>RMSEA</th><th>SRMR</th><th>Bentler's ρ</th></tr>
<tr><td>Usefulness</td><td>4</td><td>1–7</td><td>I found it valuable to consider classmates' feedback on our research.</td><td>YB χ^2 (6) = 261.837</td><td>$p < 0.001$</td><td>0.96</td><td>0.23</td><td>0.03</td><td>0.90</td></tr>
<tr><td>Provide</td><td>5</td><td>1–6</td><td>When I gave feedback to my classmates, I referred to the given feedback criteria.</td><td>YB χ^2 (10) = 189.70</td><td>$p < 0.001$</td><td>1.00</td><td>0.02</td><td>0.02</td><td>0.81</td></tr>
<tr><td>Utility</td><td>3</td><td>1–5</td><td>When I received feedback from my classmates, I found the feedback more useful because we could discuss it.</td><td rowspan="2">YB χ^2 (15) = 268.39</td><td rowspan="2">$p < 0.001$</td><td rowspan="2">0.96</td><td rowspan="2">0.08</td><td rowspan="2">0.04</td><td>0.78</td></tr>
<tr><td>Receptivity</td><td>4</td><td>1–5</td><td>When I received feedback from my classmates, I was convinced that even when my classmates had a different point of view, their feedback could still be valuable.</td><td>0.76</td></tr>
</table>

Appendix D

Table A4. Overview of perceived ease-of-use questionnaire with descriptive statistics.

Item	M (SD)
I often become confused when I use the feedback tool.	3.31 (1.76)
I make errors frequently when using the feedback tool.	2.86 (1.47)
Interacting with the feedback tool is often frustrating.	3.10 (1.89)
I need to consult the teacher often when using the feedback tool.	2.68 (1.55)
Interacting with the feedback tool requires a lot of my mental effort.	2.66 (1.73)
I find it easy to recover from errors encountered while using the feedback tool.	4.94 (1.50)
The feedback tool is inflexible to interact with.	3.52 (1.63)
I find it easy to get the feedback tool to do what I want it to do.	4.69 (1.46)
The feedback tool often behaves in unexpected ways.	3.13 (1.65)
I find it cumbersome to use the feedback tool.	3.72 (1.55)
My interaction with the feedback tool is easy for me to understand.	5.27 (1.35)
It is easy for me to remember how to use the feedback tool.	5.02 (1.54)
The feedback tool provides helpful guidance in giving feedback.	5.12 (1.59)
Overall, I find the feedback tool easy to use.	5.13 (1.53)

Note: Results should be interpreted on a 7-point Likert scale, encompassing values from 1 (strongly disagree) to 7 (strongly agree).

Appendix E

Table A5. Overview of sentence-starters, by role and inquiry phase.

	Assessor	Assessee
General	- Hello, we have reviewed your *RQ/IR/RC* and think you should work on the following criteria. . . - Your *RQ/IR/RC* would be even better if... - Maybe you can adjust your *RQ/IR/RC* so that... - You did... well, but... - If we were you, we would... - We suggest... - We (do not) think this is a good idea, because... - We (dis)agree with this, because... - Good job! - Keep it up!	- Hello, can you provide feedback on our *RQ/IR/RC*? - We understand your feedback, but... - We still have questions about... - We (dis)agree with this, because . . . - Can you help us with... - We understand your feedback, but... - Thanks for the feedback!
Conceptualization	<u>Focus</u> - You can improve your RQ by focusing more on..., because... - We suggest narrowing down your RQ by... - Your RQ is currently too... formulated. Perhaps you can... <u>Clarity</u> - You can make your research question clearer by..., because... - We have read your research question and currently find... a bit unclear. - Try to clarify what you want to investigate with... <u>Relevance</u> - We think your research question could be more relevant if you focused more on... - We believe we would learn more from your research if it focused on... <u>Feasibility</u> - We fear that your research question is not feasible within the time frame, because... - We suggest making your research more feasible by... <u>Spelling</u> - We have found some spelling errors in your research question:... - We believe that... is spelled incorrectly. <u>Punctuation</u> - We think you may have forgotten some punctuation marks or capital letters, such as... - We think you have used some punctuation marks incorrectly, for example...	- If we adjust our RQ by..., do you think it's better? - If we formulate our RQ as follows, would it be better in your opinion:... - Maybe we can adjust our RQ by... - Can you explain... more? - We (do not) think this is a good idea, because...

Table A5. *Cont.*

	Assessor	Assessee
Investigation	Number of Sources - We currently think that you have researched too few sources because... - We believe source . . . does not really fit your topic because... - Perhaps you could look for an additional source on.... Diversity of Sources - Currently, we do not think your sources are diverse enough because... - We suggest searching for information in this area to make your sources more diverse:... - To make your sources more diverse, you could look for a source written by... Reliability of Sources - We do not find source . . . to be very reliable because... - Perhaps this source is more reliable than yours:... Usability of Information - We do not find the information from source . . . particularly valuable for your research question because... - We suggest focusing more on the main points in the information from source..., because...	- If we look for . . . more sources on..., would that be good in your opinion? - If we update the information from source . . ., is that good? - We find... to be (not) reliable because... - We find... to be (not) valuable because... - What do you think of this source:... - We find this source...
Conclusion	Specificity - We find the answer to your RQ somewhat vague because... - We suggest mentioning . . . in your RC. - Perhaps you could expand your RC by... Coherence - We found it somewhat challenging to read your RC because... - The transition between... and... is a bit rough and could perhaps be made smoother by... 323. In some places, we find your RC not so clear, for example.... Wording - The wording of your RC at this place is... somewhat... - We think that . . . could be worded more smoothly, for example by... - In our opinion, you have literally taken... from a site. Spelling - We have found some spelling mistakes in your RC:... - We believe that... is spelled incorrectly. Punctuation - We think you may have forgotten some punctuation marks or capital letters, such as... - We think you have used some punctuation marks incorrectly, for example.	- If we adjust our RC by..., do you think that's better? - If we phrase our RC as follows, would that be better in your opinion:... - Maybe we can expand our RC by... - We (don't) think this is a good idea because...

References

1. Shute, V.J. Focus on Formative Feedback. *Rev. Educ. Res.* **2008**, *78*, 153–189. [CrossRef]
2. Topping, K.J. Peer Assessment: Learning by Judging and Discussing the Work of Other Learners. *Interdiscip. Educ. Psychol.* **2017**, *1*, 1–17. [CrossRef]
3. Topping, K.J. Digital Peer Assessment in School Teacher Education and Development: A Systematic Review. *Res. Pap. Educ.* **2021**, *38*, 472–498. [CrossRef]
4. Nicol, D.; Thomson, A.; Breslin, C. Rethinking Feedback Practices in Higher Education: A Peer Review Perspective. *Assess. Eval. High. Educ.* **2014**, *39*, 102–122. [CrossRef]
5. Tenório, T.; Bittencourt, I.I.; Isotani, S.; Silva, A.P. Does Peer Assessment in Online Learning Environments Work? A Systematic Review of the Literature. *Comput. Hum. Behav.* **2016**, *64*, 94–107. [CrossRef]
6. Barahona, C.; Nussbaum, M.; Martin, V.; Meneses, A.; Arriagada, S.; Di Serio, A.; Hilliger, I. Technology-Scaffolded Peer Assessment for Developing Critical Thinking in Pre-Service Teacher Training: The Importance of Giving Feedback. *Educ. Technol. Res. Dev.* **2023**, *71*, 667–688. [CrossRef]
7. Carless, D. Feedback as Dialogue. *Encycl. Educ. Philos. Theory* **2016**, 1–6. [CrossRef]
8. Kollar, I.; Fischer, F. Peer Assessment as Collaborative Learning: A Cognitive Perspective. *Learn. Instr.* **2010**, *20*, 344–348. [CrossRef]
9. Ajjawi, R.; Boud, D. Researching Feedback Dialogue: An Interactional Analysis Approach. *Assess. Eval. High. Educ.* **2017**, *42*, 252–265. [CrossRef]
10. Strijbos, J.W.; Ochoa, T.A.; Sluijsmans, D.M.; Segers, M.S.; Tillema, H.H. Fostering Interactivity through Formative Peer Assessment in (Web-Based) Collaborative Learning Environments. In *Cognitive and Emotional Processes in Web-Based Education: Integrating Human Factors and Personalization*; Mourlas, C., Tsianos, N., Panagiotis, G., Eds.; IGI Global: Hershey, PA, USA, 2009; pp. 375–395. [CrossRef]
11. Er, E.; Dimitriadis, Y.; Gašević, D. A Collaborative Learning Approach to Dialogic Peer Feedback: A Theoretical Framework. *Assess. Eval. High. Educ.* **2021**, *46*, 586–600. [CrossRef]
12. Nicol, D. From Monologue to Dialogue: Improving Written Feedback Processes in Mass Higher Education. *Assess. Eval. High. Educ.* **2010**, *35*, 501–517. [CrossRef]
13. Zuiker, S.J.; Anderson, K.T. Fostering Peer Dialogic Engagement in Science Classrooms with an Educational Videogame. *Res. Sci. Educ.* **2021**, *51*, 865–889. [CrossRef]
14. Lazonder, A.W.; Harmsen, R. Meta-Analysis of Inquiry-Based Learning: Effects of Guidance. *Rev. Educ. Res.* **2016**, *86*, 681–718. [CrossRef]
15. Van Hoe, A.; Wiebe, J.; Rotsaert, T.; Schellens, T. The Implementation of Peer Assessment as a Scaffold During Computer-Supported Collaborative Inquiry Learning in Secondary STEM Education. *Int. J. STEM Educ.* **2024**, *11*, 1–21. [CrossRef]
16. Chen, T. Technology-Supported Peer Feedback in ESL/EFL Writing Classes: A Research Synthesis. *Comput. Assist. Lang. Learn.* **2016**, *29*, 365–397. [CrossRef]
17. Black, P.; Wiliam, D. Assessment and Classroom Learning. *Assess. Educ.* **1998**, *5*, 7–74. [CrossRef]
18. Liu, N.F.; Carless, D. Peer Feedback: The Learning Element of Peer Assessment. *Teach. High. Educ.* **2006**, *11*, 279–290. [CrossRef]
19. Falchikov, N. Peer Feedback Marking: Developing Peer Assessment. *Innov. Educ. Train. Int.* **1995**, *32*, 175–187. [CrossRef]
20. Double, K.S.; McGrane, J.A.; Hopfenbeck, T.N. The Impact of Peer Assessment on Academic Performance: A Meta-Analysis of Control Group Studies. *Educ. Psychol. Rev.* **2020**, *32*, 481–509. [CrossRef]
21. Li, H.; Xiong, Y.; Hunter, C.V.; Guo, X.; Tywoniw, R. Does Peer Assessment Promote Student Learning? A Meta-Analysis. *Assess. Eval. High. Educ.* **2020**, *45*, 193–211. [CrossRef]
22. Li, H.; Xiong, Y.; Mindy, K.; Chung, K.S. Peer Assessment in the Digital Age: A Meta-Analysis Comparing Peer and Teacher Ratings. *Assess. Eval. High. Educ.* **2016**, *41*, 245–264. [CrossRef]
23. Gielen, S.; Tops, L.; Dochy, F.; Onghena, P.; Smeets, S. A Comparative Study of Peer and Teacher Feedback and of Various Peer Feedback Forms in a Secondary School Writing Curriculum. *Br. Educ. Res. J.* **2010**, *36*, 143–162. [CrossRef]
24. Latifi, S.; Noroozi, O.; Hatami, J.; Biemans, H.J.A. How Does Online Peer Feedback Improve Argumentative Essay Writing and Learning? *Innov. Educ. Teach. Int.* **2021**, *58*, 195–206. [CrossRef]
25. Tsivitanidou, O.; Zacharia, Z.C.; Hovardas, T.; Nicolaou, A. Peer Assessment among Secondary School Students: Introducing a Peer Feedback Tool in the Context of a Computer-Supported Inquiry Learning Environment in Science. *J. Comput. Math. Sci. Teach.* **2012**, *31*, 433–465.
26. Vanderhoven, E.; Raes, A.; Montrieux, H.; Rotsaert, T.; Schellens, T. What if Pupils Can Assess Their Peers Anony-mously? A Quasi-Experimental Study. *Comput. Educ.* **2015**, *81*, 123–132. [CrossRef]
27. Heitink, M.C.; Van der Kleij, F.M.; Veldkamp, B.P.; Schildkamp, K.; Kippers, W.B. A Systematic Review of Prerequisites for Implementing Assessment for Learning in Classroom Practice. *Educ. Res. Rev.* **2016**, *17*, 50–62. [CrossRef]
28. Van Zundert, M.; Sluijsmans, D.; Van Merriënboer, J. Effective Peer Assessment Processes: Research Findings and Future Directions. *Learn. Instr.* **2010**, *20*, 270–279. [CrossRef]
29. Liu, X.; Li, L. Assessment Training Effects on Student Assessment Skills and Task Performance in a Technology-Facilitated Peer Assessment. *Assess. Eval. High. Educ.* **2014**, *39*, 275–292. [CrossRef]

30. Van Gennip, N.A.E.; Segers, M.S.R.; Tillema, H.H. Peer Assessment as a Collaborative Learning Activity: The Role of Interpersonal Variables and Conceptions. *Learn. Instr.* **2010**, *20*, 280–290. [CrossRef]
31. Panadero, E.; Jonsson, A. The Use of Scoring Rubrics for Formative Assessment Purposes Revisited: A Review. *Educ. Res. Rev.* **2013**, *9*, 129–144. [CrossRef]
32. Camarata, T.; Slieman, T.A. Improving Student Feedback Quality: A Simple Model Using Peer Review and Feedback Rubrics. *J. Med. Educ. Curric. Dev.* **2020**, *7*, 238212052093660. [CrossRef] [PubMed]
33. Bürgermeister, A.; Glogger-Frey, I.; Saalbach, H. Supporting Peer Feedback on Learning Strategies: Effects on Self-Efficacy and Feedback Quality. *Psychol. Learn. Teach.* **2021**, *20*, 383–404. [CrossRef]
34. Deiglmayr, A. Instructional Scaffolds for Learning from Formative Peer Assessment: Effects of Core Task, Peer Feedback, and Dialogue. *Eur. J. Psychol. Educ.* **2018**, *33*, 185–198. [CrossRef]
35. Zhu, Q.; To, J. Proactive Receiver Roles in Peer Feedback Dialogue: Facilitating Receivers' Self-Regulation and Co-Regulating Providers' Learning. *Assess. Eval. High. Educ.* **2022**, *47*, 1200–1212. [CrossRef]
36. Topping, K.J. Elaborated Dialogic Feedback and Negotiated Action in Peer Assessment: Metacognitive Benefits for Assessor and Assessee. In *Fostering Inclusion in Education: Alternative Approaches to Progressive Educational Practices*; Postiglione, E., Ed.; Palgrave Macmillan: London, UK; New York, NY, USA, 2022; pp. 79–103. [CrossRef]
37. Carless, D.; Salter, D.; Yang, M.; Lam, J. Developing Sustainable Feedback Practices. *Stud. High. Educ.* **2011**, *36*, 395–407. [CrossRef]
38. Schillings, M.; Roebertsen, H.; Savelberg, H.; Van Dijk, A.; Dolmans, D. Improving the Understanding of Written Peer Feedback through Face-to-Face Peer Dialogue: Students' Perspective. *High. Educ. Res. Dev.* **2021**, *40*, 1100–1116. [CrossRef]
39. Kim, M. The Impact of an Elaborated Assessee's Role in Peer Assessment. *Assess. Eval. High. Educ.* **2009**, *34*, 105–114. [CrossRef]
40. Prins, F.J.; Sluijsmans, D.M.A.; Kirschner, P.A.; Strijbos, J.W. Formative Peer Assessment in a CSCL Environment: A Case Study. *Assess. Eval. High. Educ.* **2005**, *30*, 417–444. [CrossRef]
41. Kim, M.; Ryu, J. The Development and Implementation of a Web-Based Formative Peer Assessment System for Enhancing Students' Metacognitive Awareness and Performance in Ill-Structured Tasks. *Educ. Technol. Res. Dev.* **2013**, *61*, 549–561. [CrossRef]
42. Zhu, Q.; Carless, D. Dialogue within Peer Feedback Processes: Clarification and Negotiation of Meaning. *High. Educ. Res. Dev.* **2018**, *37*, 883–897. [CrossRef]
43. Alemdag, E.; Yildirim, Z. Design and Development of an Online Formative Peer Assessment Environment with Instructional Scaffolds. *Educ. Technol. Res. Dev.* **2022**, *70*, 1359–1389. [CrossRef]
44. Cheng, L.; Li, Y.; Su, Y.; Gao, L. Effect of Regulation Scripts for Dialogic Peer Assessment on Feedback Quality, Critical Thinking, and Climate of Trust. *Assess. Eval. High. Educ.* **2023**, *48*, 451–463. [CrossRef]
45. Zheng, L.; Zhang, X.; Cui, P. The Role of Technology-Facilitated Peer Assessment and Supporting Strategies: A Meta-Analysis. *Assess. Eval. High. Educ.* **2020**, *45*, 372–386. [CrossRef]
46. Wood, J. Making Peer Feedback Work: The Contribution of Technology-Mediated Dialogic Peer Feedback to Feedback Uptake and Literacy. *Assess. Eval. High. Educ.* **2022**, *47*, 327–346. [CrossRef]
47. Filius, R.M.; De Kleijn, R.A.M.; Uijl, S.G.; Prins, F.J.; Van Rijen, H.V.M.; Grobbee, D.E. Strengthening Dialogic Peer Feedback Aiming for Deep Learning in SPOCs. *Comput. Educ.* **2018**, *125*, 86–100. [CrossRef]
48. Pedaste, M.; Mäeots, M.; Siiman, L.A.; De Jong, T.; Van Riesen, S.A.; Kamp, E.T.; Manoli, C.C.; Zacharia, Z.C.; Tsourlidaki, E. Phases of Inquiry-Based Learning: Definitions and the Inquiry Cycle. *Educ. Res. Rev.* **2015**, *14*, 47–61. [CrossRef]
49. De Jong, T. Moving towards Engaged Learning in STEM Domains: There Is No Simple Answer, but Clearly a Road Ahead. *J. Comput. Assist. Learn.* **2019**, *35*, 153–167. [CrossRef]
50. Kim, M.C.; Hannafin, M.J. Scaffolding Problem Solving in Technology-Enhanced Learning Environments (TELEs): Bridging Research and Theory with Practice. *Comput. Educ.* **2011**, *56*, 403–417. [CrossRef]
51. Dmoshinskaia, N.; Gijlers, H.; De Jong, T. Giving Feedback on Peers' Concept Maps in an Inquiry Learning Context: The Effect of Providing Assessment Criteria. *J. Sci. Educ. Technol.* **2020**, *30*, 420–430. [CrossRef]
52. Hovardas, T.; Tsivitanidou, O.E.; Zacharia, Z.C. Peer versus Expert Feedback: An Investigation of the Quality of Peer Feedback among Secondary School Students. *Comput. Educ.* **2014**, *71*, 133–152. [CrossRef]
53. Tsivitanidou, O.E.; Zacharia, Z.C.; Hovardas, T. Investigating Secondary School Students' Unmediated Peer Assessment Skills. *Learn. Instr.* **2011**, *21*, 506–519. [CrossRef]
54. Bakker, A. *Design Research in Education: A Practical Guide for Early Career Researchers*; Routledge: London, UK, 2018. [CrossRef]
55. Chan, C.K.K.; Yang, Y. Developing Scientific Inquiry in Technology-Enhanced Learning Environments. In *Second Handbook of Information Technology in Primary and Secondary Education*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 161–180. [CrossRef]
56. Strijbos, J.W.; Pat-El, R.; Narciss, S. Validation of a (Peer) Feedback Perceptions Questionnaire. In *Proceedings of the 7th International Conference on Networked Learning 2010*; Dirckinck-Holmfeld, L., Hodgson, V., Jones, C., De Laat, M., McConnell, D., Ryberg, T., Eds.; Aalborg University: Aalborg, Denmark, 2010.
57. Jones, R.H.; Garralda, A.; Li, D.C.S.; Lock, G. Interactional Dynamics in Online and Face-to-Face Peer-Tutoring Sessions for Second Language Writers. *J. Second Lang. Writ.* **2006**, *15*, 1–23. [CrossRef]
58. Janssen, J.; Erkens, G.; Kanselaar, G. Visualization of Agreement and Discussion Processes during Computer-Supported Collaborative Learning. *Comput. Hum. Behav.* **2007**, *23*, 1105–1125. [CrossRef]
59. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef]

60. Shroff, R.H.; Ting, F.S.T.; Chan, C.L.; Garcia, R.C.C.; Tsang, W.K.; Lam, W.H. Conceptualisation, Measurement and Preliminary Validation of Learners' Problem-Based Learning and Peer Assessment Strategies in a Technology-Enabled Context. *Australas. J. Educ. Technol.* **2023**, *39*, 1–18. [CrossRef]
61. Dawson, P.; Yan, Z.; Lipnevich, A.; Tai, J.; Boud, D.; Mahoney, P. Measuring What Learners Do in Feedback: The Feedback Literacy Behaviour Scale. *Assess. Eval. High. Educ.* **2023**, *49*, 348–362. [CrossRef]
62. Kasch, J.; Van Rosmalen, P.; Henderikx, M.; Kalz, M. The Factor Structure of the Peer-Feedback Orientation Scale (PFOS): Toward a Measure for Assessing Students' Peer-Feedback Dispositions. *Assess. Eval. High. Educ.* **2022**, *47*, 15–28. [CrossRef]
63. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–339. [CrossRef]
64. Gan, M.J.S.; Hattie, J. Prompting Secondary Students' Use of Criteria, Feedback Specificity and Feedback Levels during an Investigative Task. *Instr. Sci.* **2014**, *42*, 861–878. [CrossRef]
65. Van Hoe, A.; Rotsaert, T.; Schellens, T. Online computer-supported inquiry learning in STEM education during the COVID-19 pandemic. In Proceedings of the 14th Annual International Conference of Education, Research and Innovation, Seville, Spain, 8–10 November 2021; pp. 782–791.
66. Sun, D.; Looi, C.K.; Xie, W. Learning with Collaborative Inquiry: A Science Learning Environment for Secondary Students. *Technol. Pedagog. Educ.* **2017**, *26*, 241–263. [CrossRef]
67. Johnson, D.W.; Johnson, R.T. Making Cooperative Learning Work. *Theory Pract.* **1999**, *38*, 67–73. [CrossRef]
68. Fischer, F.; Kollar, I.; Stegmann, K.; Wecker, C. Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educ. Psychol.* **2013**, *48*, 56–66. [CrossRef] [PubMed]
69. Nicol, D.J.; Dick, D.M. Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Stud. High. Educ.* **2007**, *32*, 199–218. [CrossRef]
70. Reinholz, D. The Assessment Cycle: A Model for Learning through Peer Assessment. *Assess. Eval. High. Educ.* **2016**, *41*, 301–315. [CrossRef]
71. Lindgren, R.; Tscholl, M.; Wang, S.; Johnson, E. Enhancing Learning and Engagement through Embodied Interaction within a Mixed Reality Simulation. *Comput. Educ.* **2016**, *95*, 174–187. [CrossRef]
72. Ajjawi, R.; Boud, D. Examining the Nature and Effects of Feedback Dialogue. *Assess. Eval. High. Educ.* **2018**, *43*, 1106–1119. [CrossRef]