PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

DEVELOPMENT AND ASSESSMENT OF COLLABORATIVE PROBLEM-SOLVING SKILLS THROUGH TECHNOLOGY

MATIAS IGNACIO ROJAS MIRANDA

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences

Advisor:

MIGUEL NUSSBAUM VOEHL

Santiago de Chile, January, 2022

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

DEVELOPMENT AND ASSESSMENT OF COLLABORATIVE PROBLEM-SOLVING SKILLS THROUGH TECHNOLOGY

MATIAS IGNACIO ROJAS MIRANDA

Members of the Committee:

MIGUEL NUSSBAUM — DocuSigned by: Miguel Nussbaum V — 8FAB758C7BD541B...

PABLO CHIUMINATTO — DocuSigned by: Pablo Chiuminatto M — 541147FDDB334F6...

JORGE BAIER — DocuSigned by: Jorge Baier A. — 931881C8FA9348E...

ALEJANDRO ECHEVERRÍA — DocuSigned by: Alejandro Echeverria — 8FBE914629F9474...

SAMUEL GREIFF — DocuSigned by: Samuel Greiff — 2CAC983362174FC...

IGNACIO LIRA — DocuSigned by: IGNACIO LIRA — 4747E55F02D9401...

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree Doctor in Engineering Sciences

Santiago de Chile, January 2022

To César, my parents, my sisters, and my nephews, for their support, and for being the ones who inspire me to keep moving forward.

ACKNOWLEDGEMENTS

A complex and extensive process such as a doctorate involves the academic and emotional support of many people, without whom I would not have been able to write this thesis. First of all, I would like to thank Professor Miguel Nussbaum, who since my undergraduate studies and until today, has seen in me the interest and the necessary ability to face this challenge, and has allowed me to develop my academic goals successfully and with pride.

Secondly, I would like to thank Professors Pablo Chiuminatto, Jorge Baier, and Samuel Greiff for their recommendations and their contribution to the project ideation and publication of the scientific articles related to this thesis.

Similarly, the effort and work of Orlando Guerrero, Cristián Sáez, and Jaime González, among other colleagues and undergraduate students from the Computer Science Department, who accompanied the design, implementation, and analysis of different phases of the more general process involved in this thesis, was fundamental.

On the other hand, I would like to thank my family, my parents, Jaqueline and Ricardo, for their support and backing in difficult times, and for having allowed me to build this journey even before I was aware of how far I could go. I thank my sisters, Elisa, Eva, and Camila, who share the passion for education with me, and from their different professional areas, contribute day by day to my motivation and even supported me in developing my thesis at times when I lacked support.

Finally, I thank César Moreno, who not only accompanied me in fundamental moments of this process but also inspired me to continue with my goals every day even when it seemed like an uphill battle.

INDEX

PAGE

INDEX OF TABLES

PAGE

INDEX OF FIGURES

PAGE

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

MEDICIÓN Y DESARROLLO DE HABILIDADES DE RESOLUCIÓN COLABORATIVA DE PROBLEMAS A TRAVÉS DE TECNOLOGÍA

Tesis enviada a la Dirección de Postgrado en cumplimiento parcial de los requisitos para el grado de Doctor en Ciencias de la Ingeniería.

Matias Ignacio Rojas Miranda

RESUMEN

En los últimos años, la resolución colaborativa de problemas (CPS) ha llegado a ser considera esencial en casi todas las áreas de la vida. En consecuencia, han surgido diferentes marcos teóricos e instrumentos de evaluación para medir estas habilidades. Sin embargo, se requieren más estudios aplicados sobre su implementación y evaluación en entornos educativos reales. Además, todavía hay poca investigación sobre el desarrollo de habilidades CPS entre estudiantes de primaria. En este sentido, los diseños experimentales pre-post son fundamentales para identificar nuevos métodos para desarrollar habilidades CPS.

En esta tesis se abordarán dos objetivos principales, por un lado, la evaluación de las habilidades CPS, y por otro lado, el desarrollo de las habilidades CPS de estudiantes a través de la tecnología. Para ello, se aporta evidencia cuantitativa y cualitativa de una serie de estudios de intervención y experimentos de campo.

Al utilizar un enfoque de investigación basada en el diseño (DBR), nuestro equipo de investigación diseñó e implementó tecnologías para avanzar en el estudio de CPS en estudiantes primaria. Luego, se realizaron sucesivos estudios de pequeña y mediana escala (entre 69 y 719 estudiantes de entre 10 y 13 años por iteración). Por tanto, este trabajo incluye: (i) el diseño y validación de una herramienta de evaluación con dos versiones equivalentes basada en un marco propuesto por la OCDE y aplicado a la actividad

colaborativa. (ii) el estudio y desarrollo de un sistema de retroalimentación que utiliza técnicas de planificación automatizada para evidenciar las habilidades de CPS de estudiantes en un juego colaborativo. (iii) el diseño y estudio de un juego colaborativo que integra guiones colaborativos y herramientas de conciencia grupal que apoyan las habilidades de regulación y emociones de estudiantes de primaria.

En el primer estudio, participaron un total de 719 estudiantes. Los resultados mostraron que el instrumento propuesto mide efectivamente la dimensión de resolución de problemas de las habilidades CPS. Además, los resultados permiten concluir que los puntajes obtenidos en la evaluación fueron equivalentes para las dos formas creadas y para ambos géneros. Finalmente, no hubo diferencias significativas al evaluar CPS en grupos humano-humano versus grupos humano-agente usando el instrumento propuesto.

Luego, en el segundo estudio, describimos un estudio experimental en el que participaron 69 estudiantes, en el que exploramos la efectividad de la retroalimentación que se brindó en un juego colaborativo. Demostramos que la retroalimentación permitió a los estudiantes desempeñarse mejor en el juego, aumentar el diálogo de los estudiantes y de esta forma generar más instancias para demostrar sus habilidades CPS. También describimos un enfoque novedoso para monitorear planes de orden parcial de múltiples agentes que es más eficiente que los enfoques anteriores y, por lo tanto, requiere menos recursos computacionales en el aula.

Finalmente, en el tercer estudio se realizó una intervención con un grupo experimental y uno de control a una muestra de 223. Las actitudes de los estudiantes hacia la colaboración se evaluaron antes y después de la intervención con un juego colaborativo. Además, una semana después de la intervención se realizó un grupo focal a 32 estudiantes de ambos grupos. El análisis cuantitativo reveló que las actitudes hacia la colaboración mejoraron significativamente entre los estudiantes del grupo experimental. Esta diferencia se puede explicar por la intervención, las actitudes iniciales de los estudiantes y sus GPA. El análisis cualitativo proporcionó evidencia de los procesos de regulación y las emociones que surgen al combinar un guión colaborativo con la conciencia grupal durante las actividades CPS.

Esta tesis contribuye a la literatura de muchas formas. Primero, se proporcionó evidencia empírica de la validez de versiones equivalentes de un instrumento basado en el

marco de la OCDE. Esto nos permite explorar nuevas intervenciones que buscan desarrollar habilidades CPS, además, ampliamos el conocimiento de la medición de habilidades CPS en estudiantes de primaria y agregamos hallazgos a la discusión en torno al uso de agentes virtuales en evaluación. En segundo lugar, propusimos un algoritmo de seguimiento novedoso para planes de orden parcial que se adapta mejor a los entornos educativos y que funciona mejor que otros algoritmos presentados en la literatura. También demostramos que la retroalimentación extraída de un plan permite que las habilidades de CPS de los estudiantes se evidencien durante un juego colaborativo. Finalmente, este trabajo proporciona evidencia de cómo diseñar e implementar un juego colaborativo que andamia la conciencia del grupo utilizando un guión colaborativo para apoyar las habilidades de regulación y las emociones, promoviendo así el desarrollo de las habilidades de CPS.

Esta tesis contó con el apoyo de FONDECYT/CONICYT [FONDECYT 1180024].

Palabras Claves: Resolución Colaborativa de Problemas; Aprendizaje Colaborativo; Aprendizaje basado en juegos; Autorregulación; Retroalimentación Automatizada.

Miembros de la Comisión de Tesis Doctoral

Miguel Nussbaum
Pablo Chiuminatto
Jorge Baier
Alejandro Echeverria
Samuel Greiff
Ignacio Lira
Santiago, Enero, 2022

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

DEVELOPMENT AND ASSESSMENT OF COLLABORATIVE PROBLEM-SOLVING SKILLS THROUGH TECHNOLOGY

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences by

Matias Ignacio Rojas Miranda

ABSTRACT

In recent years, Collaborative Problem Solving (CPS) has become considered essential in almost all areas of life. Consequently, different theoretical frameworks and assessment instruments have emerged for measuring this skill. However, more applied studies on its implementation and evaluation in real-life educational settings are required. Moreover, there is still little research on developing CPS skills among elementary-school students. In this sense, pre-post experimental designs are essential for identifying new methods to develop CPS skills.

In this thesis, two main objectives will be addressed, on the one hand, the assessment of CPS skills, and on the other hand the development of students' CPS skills through technology. To this aim, quantitative and qualitative evidence of a series of intervention studies and field experiments is provided.

By using a Design-Based Research approach, our research team designed and implemented technologies to advance the study of CPS in elementary school students. Successive small and medium-scale studies were conducted (between 69 and 719 students aged between 10 and 13 years old per iteration). This work includes: (i) the design and validation of an assessment tool with two equivalent versions based on a framework proposed by the OECD and applied to the collaborative activity. (ii) the study and development of a feedback system that uses Automated Planning techniques to evidence

CPS skills among students in a collaborative game. (iii) the design and study of a collaborative game that integrates collaborative script and group awareness tools that support elementary school students' regulation skills and emotions.

In the first study, a total of 719 students participated in the study´s different stages. The results show that the proposed instrument effectively measures the problem-solving dimension of collaborative problem-solving skills among students of this age. Moreover, the results from the test were equivalent for both versions and across genders. Finally, there were no significant differences when assessing collaborative problem-solving in human-human groups versus human-agent groups using the proposed instrument.

Then, in the second study, we describe an experimental study involving 69 students, in which we explore the effectiveness of the feedback that was given. We show that the feedback allowed the students to perform better in the game, increase student dialogue and thus generate more instances to demonstrate their CPS skills. We also describe a novel approach to monitoring multi-agent partial-order plans that is more efficient than previous approaches and therefore requires fewer computational resources in the classroom.

Finally, in the third study, an intervention was carried out with an experimental group and a control group in a sample of 223 students. The students' attitudes towards collaboration were evaluated before and after the intervention. In addition to this, a focus group was held a week after the intervention, involving 32 students from both groups. The quantitative analysis revealed that attitudes towards collaboration improved significantly among students in the experimental group. This difference can be explained by the intervention, the students' initial attitudes, and their GPAs. The qualitative analysis provided evidence of the regulation processes and emotions that emerge when combining a collaborative script with group awareness during CPS activities.

This thesis contributes to the literature in many ways. First, empirical evidence of the validity of equivalent versions of an instrument based on the OECD framework was provided. This allows us to explore new interventions that seek to develop CPS skills, furthermore, we expand the knowledge of the CPS skills measurement in elementary school students and add findings to the discussion around the use of virtual agents in assessment. Secondly, we proposed a novel monitoring algorithm for partial-order plans that is better suited to educational settings and that performs better than other algorithms presented in the

literature. We also showed that feedback extracted from a plan allows students' CPS skills to be evidenced during a collaborative game. Finally, this work provides evidence of how to design and implement a collaborative game that scaffolds group awareness using a collaborative script to support regulation skills and emotions, thus promoting the development of CPS skills.

This work was supported by FONDECYT/CONICYT [FONDECYT 1180024].

Key words: Collaborative Problem Solving; Collaborative Learning; Game-based Learning; Selg-Regulation; Automated Feedback.

Members of the Doctoral Thesis Committee:

Miguel Nussbaum
Pablo Chiuminatto
Jorge Baier
Alejandro Echeverria
Samuel Greiff
Ignacio Lira
Santiago, January, 2022

1 INTRODUCTION

1.1. Theoretical Background

1.1.1. The Importance of Collaborative Problem Solving

Collaborative problem solving is a 21st-century skill that has attracted the attention of researchers and educators in different countries, with the emergence of technology-supported environments and the relevance that teamwork has today (Griffin et al., 2012). Despite the above, both the workforce and academia recognize deficiencies in the collaborative problem-solving skills of graduates entering the market (National Science and Technology Council, 2018; Hesse et al., 2015). (Assessing elementary students' collaborative problem-solving in makerspace activities). However, developing collaborative problem-solving skills in students is not only important for meeting workforce needs, but also for them to be productive learners in elementary, secondary, and college education. Accordingly, previous research has defined collaborative problem-solving skills as a teachable and measurable concept (Hesse et al., 2015; Rosen & Foltz, 2014; Ham, & Hwang, 2021).

Collaboration has clear advantages over individual problem solving because it allows: effective division of labor; incorporation of information from multiple perspectives, experiences, and sources of knowledge; Greater creativity and quality of solutions stimulated by the ideas of other group members. (OECD, 2017). In this sense, achieving the development of these skills has the potential to be an effective tool to solve complex problems, especially those that we know are not possible to solve without group work, many times multidisciplinary. A concrete and recent example of this need is the COVID-19 pandemic, which has shown that in the academic, political, civil, etc. context there is an interdependence that demands complex work between the different agents and that the action of just one of them is not enough to solve the problem.

1.1.2. The Problem

The Program for International Student Assessment (PISA) 2015 conducted a large-scale assessment of the collaborative problem-solving skills of approximately 50.000 15-year-old students from 52 countries. Its results revealed that only 8% performed at the highest CPS proficiency level, while 29% performed at the lowest levels (OECD, 2018). These results may be due to the scarcity of curricula or programs to teach CPS skills (Scouler & Care, 2018). Indeed, relevant challenges are still being faced to teach these skills effectively (Fiore et al., 2017). In addition, little empirical work on CPS has been done to date, which in turn implies that there is insufficient knowledge to design interventions to develop these skills (Rosen, 2010).

On the other hand, the assessment of CPS still presents challenges. According to Sun et al. (2020), there is no consensus on a CPS framework to operationalize this construct and measure it effectively (Andrews-Todd & Forsyth, 2018). In this regard, previous studies have used various measures to assess CPS (i.e., surveys, tests, observations, and think-aloud protocols), which in turn, in terms of quality, vary widely (Oliveri et al., 2017). Consequently, guiding principles to assess CPS is still required (Andrews-Todd and Forsyth, 2018; Bause, Brich, Wesselein, and Hesse, 2018).

In short, to advance in the development of educational policies and programs for collaborative problem solving in students, which are evidence-based and help address the most urgent problems of society, it is necessary to contribute more research to determine how the cognitive, social, emotional, and motivational processes of collaborative problem solving can be evaluated, and to what extent it is possible to build environments supported by technology that favor its development empirically. (Jablansky, 2020).

1.1.3. What Is CPS

The first step to understanding what we mean by collaborative problem solving is to define what we mean by problem. An accepted definition in literature (OECD, 2012), comes from Duncker (1945) who refers to a problem when a person has a goal but does not know how to achieve it.

Therefore, as presented in Figure 1.1, a problem begins with a given state in which the person has prior or initial knowledge about the problem (Givens). In addition, there is a desired state or a goal to be reached (Goals). However, the existence of barriers hinders the path to reach that goal, such as lack of knowledge or other barriers. Consequently, using available tools or admissible actions (Operators). Overcoming barriers may involve not only cognition but also social, motivational, and effective means (Funke, 2010; OECD, 2012).

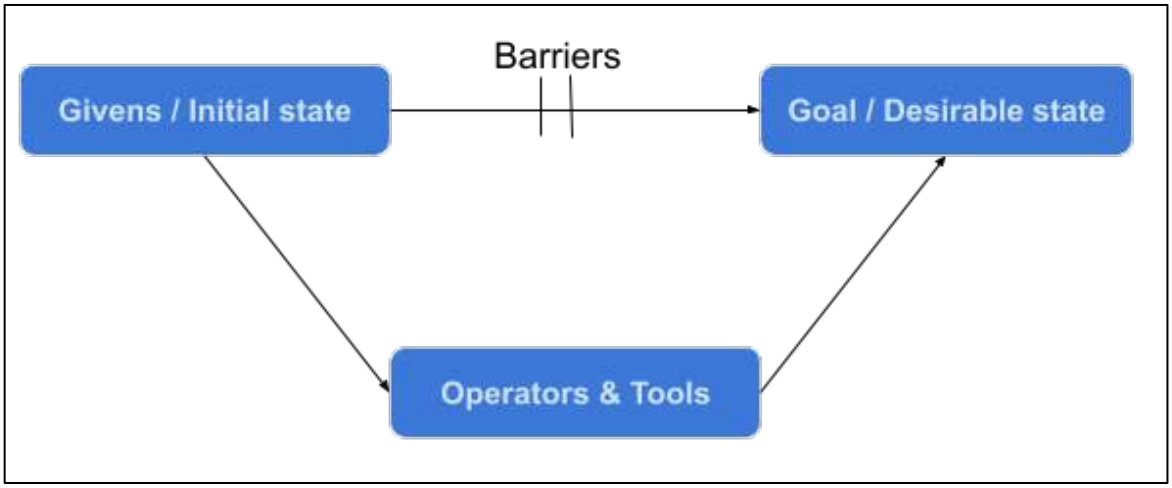


Figure 1-1: Diagram of a problem

An example of a problem is the pandemic produced by Covid-19 if we consider that what we want is to control (i.e., decrease or eradicate) its spread. In this case, the given or initial state is the information we have regarding the virus, the symptoms, the means of propagation, among others. The permitted or possible actions (operators) are to establish various strategies such as quarantines, use of masks, vaccines, etc. Then, to evaluate their effectiveness, such as PCR tests, the number of deaths, and other statistics, which would be considered as tools available to resolve the problem.

There are several aspects that can be defined to characterize a problem. The first of these corresponds to the level of knowledge that the problem demands and can be defined as knowledge-lean, which means that it does not require much prior knowledge, but only general knowledge on the solvers' part. In contrast, a problem is knowledge-rich, when it requires specific knowledge about a particular domain or field. A second aspect that characterizes a problem refers to the demand of previous experience it requires, being semantically-lean or semantically-rich. Finally, if, for example, a problem is presented with an initial state, a goal, and the operations available to solve it, clearly established (i.e., a puzzle), we will say that it is a well-defined problem (Gilhooly, 2012, p. 3; Robertson, 2016, p. 20). That is, it has a precise formulation of the problem. On the contrary, we will say that it is an ill-defined problem when there is uncertainty about how to reach the solution. The latter tend to be more frequent in everyday life and their biggest challenge is often that they require the problem solver to first clarify a definition of the problem, i.e. to convert it into a well-defined problem (Avry, 2021).

When confronted with a problem, there are a series of steps (not necessarily linear), through which the problem solver must pass (Pretz et al., 2003; see also Polya, 2004), among them are (Avry, 2021**)**:

1. Identify the problem
2. Define the problem
3. Develop a solution strategy
4. Organize the required knowledge
5. Allocate resources to solve the problem
6. Monitor progress
7. Evaluate the solution.

According to Mayer (1998), it is not enough to have only the basic and cognitive skills required to solve a problem successfully, but these must also be orchestrated and controlled (metacognitive factors); in addition, it is necessary to have the will and persistence necessary to solve the problem (motivational factors) such as individual interest, self-efficacy, etc. as these acquire an essential role during this process. (Avry, 2021)

On the other hand, unlike Cooperation, in which the members of a group divide the task into individual subparts that are distributed among each of them. Collaboration is a "coordinated and synchronous activity that is the result of a continuous attempt to build and maintain a problem´s shared conception" (Roschelle and Teasley, 1995, p. 70). In this sense, as defined by OECD (2017), in the context of its theoretical framework to measure CPS in the PISA test, because some social interactions do not involve shared goals or objectives, accommodation of the different group members' perspectives, or organized attempts to achieve expected goals or outcomes, social interaction is a necessary but not sufficient condition for collaboration.

Consequently, considering the definitions and characteristics of problem-solving and collaboration, CPS is defined as "the ability of an individual to participate effectively in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort needed to arrive at a solution and pooling their knowledge, skills, and efforts to arrive at that solution" (OECD, 2017). While there are various theoretical frameworks of CPS, which we will see below, this definition is widely accepted in literature (Scoular et al., 2017). Furthermore, there is consensus that assessing students' collaborative problem solving involves two aspects that reflect the nature of the construct: a cognitive aspect of problem-solving and a social aspect of collaboration (Andrews-Todd & Forsyth, 2020; Graesser et al., 2018).

An in-depth look at understanding the different processes underlying collaborative problem solving is described by Avry (2021) in his thesis, who presents a model defined as a dynamic process in which group members construct and update their individual representations (in memory), combining general knowledge and incoming information both from themselves (through individual processing) and from their peers (through observable outcomes) concerning the cognitive, motivational and relational aspects of the collaborative problem-solving task. In this way, Avry (2021) distinguishes the personal and interpersonal aspects of collaborative problem-solving.

1.1.3.1. Personal Aspects of Collaborative Problem Solving

From a cognitive point of view, both lower and higher-level cognitive processes are necessary to solve the problem. That is, the lower level includes processes such as memorizing, disconnecting from stimuli that are irrelevant to the task (inhibiting), and focusing attention. And analysis, synthesis, and reasoning, for higher levels. To solve the problem, problem solvers construct an internal representation of the problem (problem space), which includes the limitations or constraints, the available procedures and the possible actions to be taken to solve the problem. The latter may become too many to try with each one of them, so reduction strategies are used to narrow down the spectrum of possibilities. However, collaborative problem solving also requires monitoring and control processes to ensure the achievement of goals (metacognitive processes). These processes involve knowledge and awareness about oneself, the other members of the group, the task, and the strategies used to solve the problem, as well as skills dedicated to problem-solving regulation. In this sense, Winne and Hadwin's model (Hadwin et al., 2011) details how cognitive and metacognitive processes construct learning activities.

On the other hand, as already mentioned, the different motivational processes influence the initiation, intensity, and persistence of individuals' behavior during collaborative problem-solving. Avry (2021) differentiates three categories to classify the fundamental constructs that motivational processes involve (Usher & Morris, 2012): The first one corresponds to self-beliefs and attitudes, where self-beliefs are subdivided into four different types: first the individual's descriptive knowledge about him/herself in terms of skills, abilities, attractiveness, social acceptability, etc. (Self-concept; Bong & Skaalvik, 2003; Byrne, 1984; Mcinerney, 2012), Then, the beliefs that people have about their attributes and abilities as a person (Self-beliefs; Valentine et al., 2004). Followed by how an individual evaluates himself or herself (Self-esteem; Ruholt et al., 2015), finally, people's beliefs that they can successfully perform a task (Self-efficacy; Bong and Skaalvik, 2003; Zimmerman, 2000). As for attitudes, it refers to predispositions and habitual ways of feeling, thinking, and acting, rather enduring and general that derives from specific beliefs (Scherer, 2005). The second type, Achievement goals, and values, which are related to the personal goals and individual purpose (Achievement goals; Niemczyk, 2012; Scherer, 2005; Wolters & Taylor, 2012). As well as the incentives that promote task accomplishment (Values; Eccles & Wigfield, 2002). Finally, the third category corresponds to Attributions about success and failure, which refers to how people infer regarding the cause of their behavior. In this sense, people tend to automatically find explanations for events, and attribute causes as to why they succeeded or failed. Therefore, these attributions can influence their future motivation.

This is similar to when we talk about the cognitive and metacognitive processes required in collaborative problem-solving. We can extend the reasoning used above to understand that since individuals require motivational processes for successful collaborative problem solving, then they must also be able to monitor and regulate their motivation in a way that is conducive to achievement. Then, as Avry (2021) explains, metamotivational processes allow problem solvers to tailor the type and level of motivation that is required, through different strategies.

1.1.3.2. Interpersonal Aspects of Collaborative Problem Solving

Starting from the inclusion of other members that form a group and get involved in the resolution of a common problem, the cognitive, metacognitive, motivational, and meta-motivational processes that we distinguish in the personal aspects of collaborative problem solving, we will extend them to interpersonal aspects.

Cognitive processes become socio-cognitive processes (Suchy and Holdnack, 2013), which include three main activities (Decuyper et al., 2010): First, knowledge acquisition, which refers to the creation of shared mental models, and involves sharing, storing, and retrieving information for this purpose. Second, participation, which refers to the mutual adaptation and coordination of group members to create a shared discourse. Finally, creation occurs when group members, through co-construction and constructive conflict, co-create new knowledge (Avry, 2021). Then, analogous to the personal aspect, meta-cognitive processes, extend to socio-meta-cognitive processes, which refer, according to Decuyper et al (2010), to the capacity for group reflection, i.e. the ability to understand the situation together, establish common objectives and implement group strategies.

Following the analogy of the previous paragraph, the motivational processes of collaborative problem solving, in the interpersonal aspect, become socio-motivational processes, which include collective phenomena that influence group performance, such as beliefs, attitudes, interdependence, group cohesion, among others. Consequently, socio-meta-motivational processes help monitor and control socio-motivational processes, some challenges that require the regulation of motivation among group members are: different expectations, different work styles, different commitment to the task, different abilities to reach common ground, and external constraints (Järvelä and Järvenoja, 2011).

Finally, an additional dimension, specific to the interpersonal aspect, corresponds to the socio-relational dimension, which interacts with the socio-motivational and socio-cognitive dimensions, and refers to the type of relationships that exist between peers in the group. Examples of relational phenomena are psychological safety and conflict escalation (Burgoon & Hale, 1984). Consequently, social-meta-relational processes refer to the monitoring and control of social-relational processes to promote the groups' success in solving the task. However, to our knowledge, social-meta-relational processes appear to be understudied to date (Avry, 2021).

1.1.4. CPS's Most Relevant Theoretical Frameworks

Throughout literature, there have been several CPS models (Sun et al, 2020), most notably those proposed by Roschelle and Teasley (1995), Nelson (1999), PISA (OECD, 2017a), ATC21S (Griffin et al., 2012), a CPS ontology (Andrews-Todd & Forsyth, 2018). The following is a brief review of their characteristics:

One of the first approaches was the one proposed by Roschelle and Teasley (1995) advocating a micro-analysis of teamwork to fully grasp how CPS happens. Central to this theoretical framework is the notion of "a shared conception of the problem". It further integrates goals, descriptions of the current state of the problem, knowledge of available problem-solving actions, and associations that relate goals, characteristics of the current state of the problem, and available actions. Consequently, team members are expected to constantly monitor divergence in understanding, and cognitive convergence is essential for effective collaboration (Teasley, Fischer, Dillenbourg, Kapur, & Chi, 2008). They conclude, therefore, that successful collaborative problem solving involves cycles that alternate between periods of lower and higher intensity (Sun et al, 2020).

Another framework that addresses CPS from a pedagogical perspective is the one proposed by Nelson (1999), which provided guidelines for the implementation of collaborative problem-solving activities in authentic environments (see Lee, Huh, & Reigeluth, 2015; Merrill, 2002). In this way, the author defined a set of actions that team members can perform to demonstrate effective collaboration: for example, through the allocation of time for individual and collective work, peers should identify and share relevant knowledge and resources. In addition, to create positive interdependence, peers must acquire and apply skills in communication, leadership, conflict management, and other social skills. Next, all team members must actively participate in problem-solving and understand that there is no individual success without group success. Finally, team members must be personally accountable for their tasks within the CPS process to ensure fairness. (Sun et al, 2020).

Another theory originally designed for CPS is the teamwork model detailed by CRESST (Chung, O'Neil, and Herl, 1999; O'Neil, Chung and Brown, 1995; see Figure 5). This framework builds on the work of Salas and colleagues (Morgan, Salas, and Glickman, 1993; Salas, Dickinson, Converse, and Tannenbaum, 1992), and addresses what processes are necessary for successful collaborative problem-solving in more depth. The CRESST model, like other frameworks, divides CPS into two components, a collaborative one consisting of six skills: (a) adaptability, (b) coordination, (c) decision-making, (d) interpersonal, and (f) communication (Chung et al., 1999), and a problem-solving component, which in turn is composed of content understanding, problem-solving strategies, which can be domain-dependent or domain-independent, and self-regulation, which has two main components, motivation, and metacognition. Finally, motivation is divided into effort and self-efficacy, and metacognition is divided into self-control and planning. (See Figure 1.2, Jablansky, 2020).

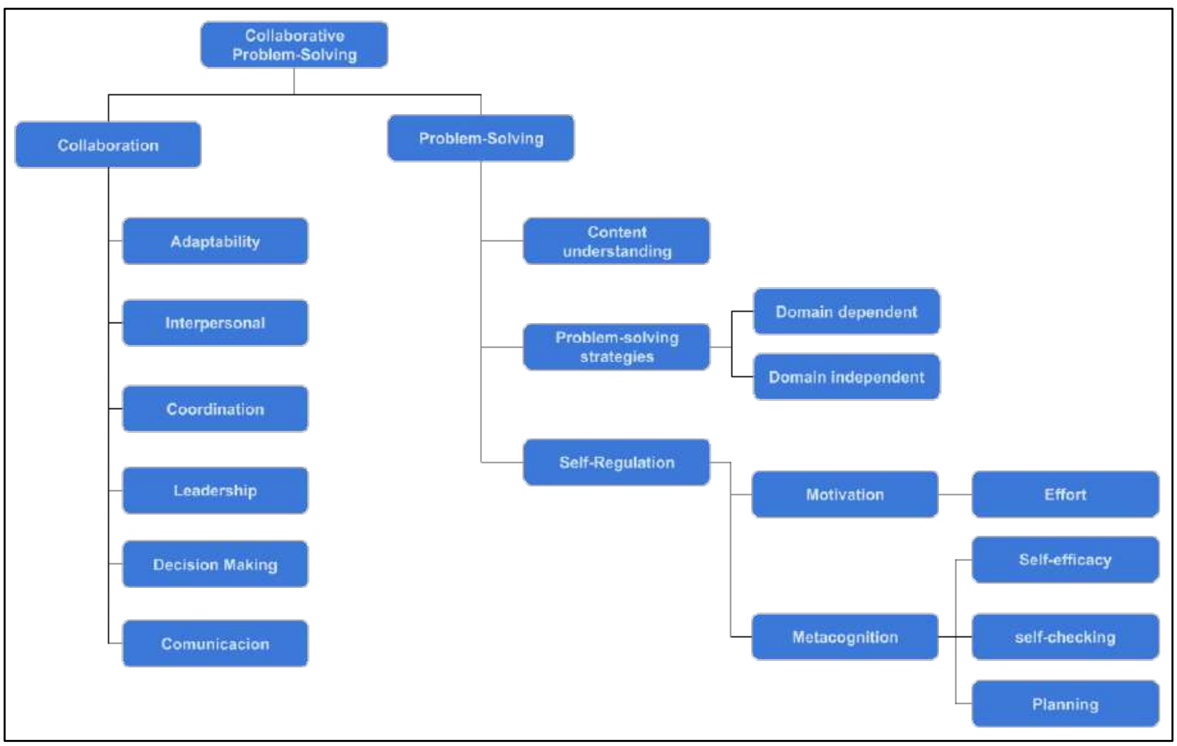


Figure 1-2. CRESST model (Chung, O'Neil & Herl, 1999; O'Neil, Chung & Brown, 1995).

A fourth CPS framework is that proposed by ATC21S which seeks to define and develop pedagogies and assess 21st-century skills in students (Care, Scoular, & Griffin, 2016). The project was undertaken as a step in the formulation of standardized CPS measures in Human-Human interactions. This framework defines CPS as a skill composed of the combination of critical thinking skills, problem-solving, decision making, and collaboration (Fiore et al., 2017; Scoular, Care, & Hesse, 2017). Similar to the models seen above, it divides CPS into a social and a cognitive dimension (See Figure 1.3. Within the social dimension, engagement refers to an individual's commitment to both teammates and the task (Hesse, Care, Buder, Sassenberg, & Griffin, 2012). Perspective-taking refers to the quality of interactions with teammates (Care et al., 2016). Finally, the category of social regulation arises from the observation that team members bring different knowledge, experiences, opinions, and strategies to a given task, and success depends in part on the group's ability to leverage these differences. On the other hand, within CPS´s cognitive dimension, task regulation refers to the competencies involved in completing the task, such as analyzing the problem, setting goals, managing resources, and gathering information (Care et al., 2016), and finally, knowledge construction reflects the ability to integrate and synthesize the contributions of other team members and further refine these representations of problems, plans, and monitoring activities (Jablansky, 2020).

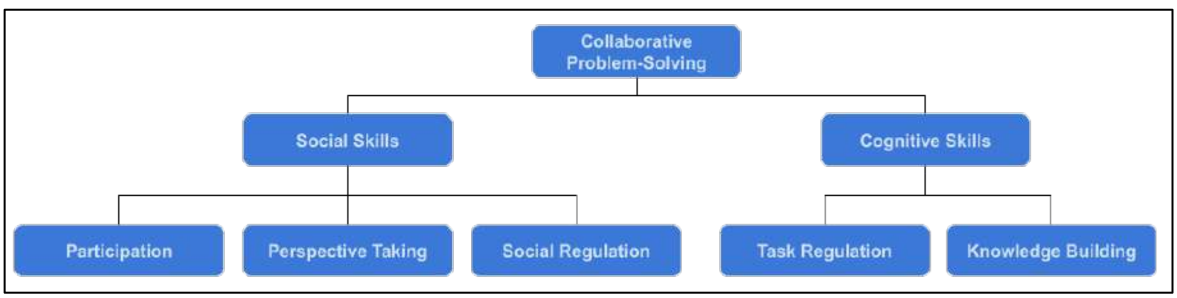


Figure 1-3: ATC21S CPS Model

A final framework to review corresponds to the one defined by the OECD for its Program for International Student Assessment (PISA), which was implemented for the first time in 2015. Specifically, it results from crossing the three collaborative (i.e., social) competencies with the four task work processes (i.e., cognitive), adapted from the 2012 PISA assessment on individual problem solving (i.e., explore and understand, represent and formulate, plan and execute, monitor and reflect). to produce 12 resulting skills on which to base the CPS assessment (See Table 1; OECD, 2017).

Table 1.1: Collaborative Problem Solving Skills Matrix (OCDE, 2017).

	1. Establish and maintain a common understanding	2. Take appropriate actions to solve the problem.	3. Establish and maintain group organization
A. Explore and understand	(A1) Discover group members' perspectives and skills	(A2) Discover the type of collaborative interaction to solve the problem according to the goals/objectives.	(A3) Understand the roles to solve the problem
B. Represent and formulate	(B1) Construct a shared representation and negotiate the meaning of the common problem	(B2) Identify and describe the tasks to be completed.	(B3) Describe roles and organize the group communicating protocol/rules of engagement
C. Plan and execute	(C1) Communicate with group members about actions to be taken	(C2) Publish plans	(C3) Follow rules of engagement by encouraging peers to perform their tasks.
D. Monitor and reflect	(D1) Monitor and repair shared understanding	(D2) Monitor action results and evaluate the problem-solving success	(D3) Monitor, give feedback and adapt the group’s organization and roles.

Establishing and Maintaining a Shared Understanding refers to identifying shared knowledge, recognizing the perspectives of other group members, establishing a shared representation of the problem, and monitoring and renegotiating shared knowledge on an ongoing basis throughout the process. Next, Taking the Actions Needed to Solve the Problem involves identifying what strategies are needed to address the problem, implementing those strategies, and finally evaluating the success or failure of the strategy. Finally, Establishing and Maintaining Team Organization involves considering the talents, resources, and assets of each team member, understanding the roles of each team member, and continually adapting strategies and reflecting on the team's progress (Fiore et al., 2017; Scoular et al., 2017; OECD, 2017).

A main difference of the PISA framework, with respect to the one proposed by ATC21S, is the composition of the working groups. While PISA uses computer-mediated communication between individuals and virtual agents, ATC21S is characterized by assessing Human-Human interactions (Scoular et al., 2017). On the other hand, a second difference is related to the purpose that both projects have. While PISA was designed as a large-scale summative assessment to inform educational systems internationally. ATC21S was created for the benefit of students and educators by identifying the steps and subskills needed to solve problems collaboratively. However, the two frameworks are considered broadly similar, as both emphasize the respective cognitive and social components of CPS (Scoular et al., 2017; Jablansky, 2020).

Despite the existence of literature reporting numerous projects aiming to assess collaborative problem-solving skills in students (Krkovic, Wüstenberg & Greiff, 2016; OECD 2017; Scoular & Care, 2019b; Stoeffler, Rosen, Bolsinova & Von Davier, 2019), more applied studies on their implementation and evaluation in real educational settings are required (Graesser, Fiore, Greiff, Andrews-Todd, Foltz & Hesse, 2018).

As mentioned in section 1.1.2, pre-post experimental designs are essential to identify which methods effectively develop collaborative problem-solving skills in students. To do so, equivalent tests are required to facilitate consistent score interpretations and reduce the practice effect (Quereshi, 2003). An assessment with two equivalent versions uses a set of questions divided into two equivalent sets ("versions"), where both sets contain questions that measure the same construct, knowledge, or skill (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014).

1.1.5. CPS Skill Development

The field of research about collaborative problem solving has yet to agree on how to implement a tool that develops this skill in a scalable way (von Davier, Hao, Liu, & Kyllonen, 2017; Funke, Fischer, & Holt, 2018). However, in Qian & Clark's (2016) a review of game-based learning and 21st-century skills that included 29 studies, one-third of the empirical work reported medium to large effect sizes as evidence of the efficacy of game-based learning in promoting 21st-century skills. Thus, although to date the knowledge and research related to game-based learning has grown considerably and has shown its effectiveness (Janakiraman, Watson, Watson & Newby, 2021), in the specific case of collaborative problem-solving skills there are still few studies that use this methodology in its development (Greipl, Moeller & Ninaus, 2020).

From a theoretical point of view, Plass et al. (2015) propose four central arguments in favor of game-based learning. (i) Motivation refers to the fact that incentive elements in the form of trophies, stars, or leaderboards, as well as certain game mechanics, stimulate situational interest and are widely used and known in entertainment games. (ii) Player Engagement: refers to the fact that games allow for a wide range of ways to engage learners, e.g., by integrating physical movements or interactions as part of the game experience (Greipl, Moeller & Ninaus, 2020). Four types of engagements (affective, cognitive, behavioral, sociocultural) are distinguished). (iii) Adaptability: this refers to the possibility of adapting to the player, e.g., to the player's capabilities or his specific situation. Finally, (iv) Graceful Failure: refers to the fact that digital games can create a safe environment in which mistakes are allowed to be made without the negative consequences of failure.

In addition, game-based learning experiences will be strongly linked to three main factors: Cognitive factors are related to the need to adapt to the player's capabilities, in case a balance between the difficulty of the task and the student's ability is not achieved, it can lead to boredom or overexertion, and finally to quitting the game (Plass, Homer, & Kinzer, 2015). Similarly, emotional factors, constitute the key to achieving meaningful learning (Schell, 2015), as they influence attention, memory, and cognition (Moreno, 2006; Izard, 2009; Plass and Kaplan, 2016). For example, motivation towards learning can be impaired by frustration(Craig et al., 2004). Finally, according to Greipl, Moeller & Ninaus (2020), social factors influence game-based learning in two ways. On the one hand, the experience changes from simply being accompanied by another peer versus being alone. And on the other hand, the comparison and/or collaboration between players, represent indirect social interactions that influence learning (Vrugte et al., 2015).

Ultimately, the basic idea of considering game-based learning for the development of collaborative problem-solving skills is that games have the potential to make otherwise boring or strenuous tasks more enjoyable to perform and, therefore, increase student engagement. In turn, this engagement leads to improved performance, through changes in students' attitudes and behaviors (Landers, 2014; Landers, Auer, Collmus, & Armstrong, 2018), motivation, and affect (Clark, Tanner-Smith, & Killingsworth, 2016; Greipl, Moeller, & Ninaus, 2020; Sailer & Homner, 2020).

1.1.6. Feedback as Part of SCP Skill Development

Feedback can be understood as information provided by an agent (e.g., a teacher, a peer, a book, etc.) about an individual's performance, understanding, or behavior (Hattie & Timperley, 2007). Therefore, it is a widely used tool in teaching both content and skills (Erhel & Jamet, 2013; Zhu, Liu & Lee, 2020), Similarly, the concept of scaffolding feedback, refers to how an adult or expert teaches someone less competent to solve a problem or complete a task (Wood, Bruner, & Ross, 1976). This concept is manifested when a tutor takes control of aspects of the task that are beyond the student's capabilities, thus allowing him or her to succeed in developing it, which would not have been possible autonomously (Plass et al 2015). Consequently, it is evident that for this feedback to be effective, the error must occur within the student's zone of close development, defined by Vygotsky (Bruner, 1985; Pea, 2004). In particular, Pea (2004) adds the existence of components that make a real scaffolding, such as the dynamic adaptation of feedback, which requires continuous student evaluation, and the gradual waning of this as skills and knowledge are acquired.

As mentioned in the previous section (Section 1.1.5), a fundamental aspect of game-based learning, which is facilitated by technology integration, is that it opens new opportunities to create adaptive and safe learning environments for the learner, as well as real-time monitoring of players' actions and interactions with a digital environment (Greipl, Moeller & Ninaus, 2020). Thus, in an optimal scenario, a digital learning environment would be able to measure learners' subjective cognitive load and adapt the difficulty of the tasks to be solved, so that it achieves a balance between tasks that is neither too easy nor too difficult. This, for example, could result in personalized feedback or scaffolding according to players' achievements and interactions (see, e.g., Chen & Law, 2016; Kao, Chiang, & Sun, 2017).

In short, to foster the development of collaborative problem-solving skills through game-based learning methodologies, we must look for mechanisms that provide feedback, but that must also respond to real educational contexts. In this sense, traditional technology (i.e., that does not use artificial intelligence techniques) has been useful to deliver predefined feedback. However, in this case, by using artificial intelligence, it is possible to identify the right timing, presentation method, and content for effective feedback (O'Donovan, den Outer, Price & Lloyd, 2019), as these techniques respond to changes in performance time and adapt to the user's behavior.

1.2. Research Questions

The following research questions guide this thesis:

- RQ1. How can we assess students' collaborative problem-solving skills?
  - RQ1.1. How can we assess collaborative problem-solving skills among elementary school students using an instrument in two equivalent forms?
- RQ2. How can we foster the development of collaborative problem-solving skills in students?
  - RQ2.1. How can we implement a collaborative game that allows students to develop their collaborative problem-solving skills?
  - RQ2.2. How can we implement a system based on artificial intelligence techniques that effectively supports collaborative game feedback and allows students to demonstrate their collaborative problem-solving skills?

1.3. Research Hypothesis

The following hypotheses were used to frame the work conducted for this thesis:

H1. It is possible to develop a valid and reliable instrument to assess collaborative problem-solving skills among students in two equivalent ways

H2. The scores obtained when using the instrument are equivalent for both ways and genders.

H3. There is no significant difference when assessing collaborative problem-solving in human-human groups versus human-agent groups using the proposed instrument.

H4. It is possible to give efficient feedback to students using collaborative games that use artificial intelligence techniques in a real educational context.

H5. The feedback provided by the system implemented in a collaborative game promotes the use of students' collaborative problem-solving skills and allows them to overcome the obstacles presented by the game to a greater extent.

H6. Using appropriate design principles and learning theories it is possible to develop a game that students can demonstrate their collaborative problem-solving skills.

H7. Students using developed games improve their attitude towards collaborative problem-solving.

1.4. Objectives

The general objective of this research is to study solutions to teach and evaluate collaborative problem-solving. Starting from this objective, this thesis will be based on the following specific objectives.

O1. Design, implement and validate an assessment instrument to evaluate Collaborative Problem Solving with two equivalent versions in elementary school students.

O2. Design, implement and validate a feedback system based on artificial intelligence techniques, which helps dynamically monitor the actions that students perform and deliver timely and effective help in a game that seeks to develop Collaborative Problem-Solving skills in real educational contexts.

O3. Design, implement and validate a game to develop Collaborative Problem Solving skills.

1.5. Methodology

Each of this thesis' objectives combines technological and educational challenges. Therefore, its design, implementation, and validation or analysis are iterative processes in which the possibility of testing the technology with real users is fundamental. In addition, a relevant and coherent aspect of this thesis and the difficulty of the objectives set forth was the collaboration of an interdisciplinary team that included researchers from engineering, education, psychology, and literature. Consequently, the methodology used was Design-Based Research (Amiel & Reeves, 2008), which is mainly characterized by combining the study of certain phenomena, and iterative design. In addition, it is flexible, providing the possibility of modifying the protocol and the design in the process. Finally, it is an open methodology, which provides the possibility to combine quantitative and qualitative methods and techniques (Savard, Bourdeau, & Paquette, 2020).

Reeves (2006) reports the five DBR steps, which are presented below with their implemented details in this thesis:

1. **Analyze practical problems with collaboration between professionals and researchers**: In this stage, a review of the respective literature was carried out, related both to the existing theoretical frameworks of collaborative problem solving, as well as to the advances in the measurement of these skills, and the initiatives to develop them in students.
2. **To develop theories, solutions, and technological innovations**: At this stage, the activities were designed and the variables involved were operationalized. In the case of the measurement instrument, the items, scores, interfaces, and additional surveys were defined. In the case of the game, the stages, mechanics, dynamics, and game elements were defined. Finally, the first versions of the three main technologies proposed by this thesis were developed (i.e., CPS assessment tool, feedback system, CPS development game).
3. **Iterative testing cycles, theory refinement, and solutions in practice:** For each technology developed, a large number of tests were carried out with real users. These tests sought to evaluate the usability and stability of the developed technologies, but also their performance in the data collection and quality for subsequent analysis.

4. **Reflecting on the production of design theories and principles and emphasizing solution implementation**: Based on the tests, both components of the software itself and its underlying theories were adjusted. In the case of the evaluation instrument, the iterations permitted substantial changes. Initially, the evaluation was performed using an open chat, whose logs were manually coded to estimate the skill level, whereas, in the last iteration, two equivalent versions were used in which students collaborated with virtual agents through predefined messages. Similarly, in the game, initially, students could chat, but the usage was very poor considering that they had to play simultaneously and typing made it difficult. Consequently, the chat was changed to recording audio conversations that the students were having out loud.
5. **Refine theories, problems, solutions, and design principles**: Finally, based on the completed tests, that is, those that met the number of users and all the conditions required to capture valid information for the quantitative and qualitative analyses used, relevant contributions to the theory were concluded, the results of which are reported in the elaboration of three papers. The methodological details of each specific objective of this thesis (Section 1.5) will be presented in the following chapters.

1.6. Results

The main findings from this thesis are the following:

- R1. We developed a new tool for assessing collaborative problem solving through an iterative and incremental design process. Using the DBR approach not only allowed us to discover important findings, but also to describe in detail how the assessment tool was developed.
- R2. The DBR process allowed us to validate the collaborative properties, the design, and the usability of the activity.
- R3. There are advantages of using virtual agents, as it prompts assessments that otherwise may not occur for every participant in a human-human group (Graesser et al., 2018). Furthermore, by using open chat for the activity, the students use a more functional style of language and opt for trying a solution and communicating what they have done.

R4. Our results showed that our instrument predominantly focuses on the problem-solving component of collaborative problem-solving.

R5. We were also able to develop two equivalent forms of the instrument, which delivered statistically equivalent results when assessing collaborative problem-solving skills.

R6. Our invariance measurement test showed that the instrument is equivalent when assessing girls and boys. This allowed us to obtain an instrument without gender bias.

R7. Our results also confirm that when using this framework there are no significant differences between human-human and human-agent groups.

R8. Our approach takes advantage of the plan´s structures to break them down into problems that can be monitored more efficiently. This allowed us to improve on the running times reported by Muise et al. (2011) and produce a list of validations that maintained their correctness. While our approach does not outperform the monitoring of a sequential plan, it allows reducing the number of calls to the planning algorithm significantly.

R9. Our results suggest that the feedback helped the students make more progress within the game in terms of the number of obstacles that they were able to overcome. Furthermore, the feedback given by the system proved to be effective as it generally prompted a conversation among the members of the group on fundamental collaborative skills, such as establishing and maintaining the organization of the group.

R10. We built a game that implements collaboration script and group awareness to develop collaborative problem-solving skills based on the OECD framework Our results reveal the existence of a relationship between collaborative scripts and group awareness as tools to support students' regulation skills during CPS activities.

R11. We also highlighted the relationship between these tools and the positive emotions reported by the students during the intervention (i.e., satisfaction).

R12. Our results show that the game has a positive impact on students’ attitudes towards collaboration.

R13. Our results also suggest that there is a relationship between the co-regulation process required by the game and the shift in emotions from frustration to satisfaction. Consequently, the proposed collaborative game develops regulation skills and positive emotions, which are key elements of CPS (OECD, 2017).

1.7. Research Limitations

This thesis establishes several findings that contribute to expanding the knowledge that exists about the measurement and development of collaborative problem-solving skills in students. However, like in many other research studies, some limitations are an important part of understanding our results, and also to consider in future research. These are detailed below:

1. The students' limited age range: the studies were conducted on elementary school students (5th and 6th grade), with ages between 10 and 13 years old. This does not generalize the results, since at different ages the students' abilities may vary. However, these are ages that have been hardly explored in literature, so it is also relevant to note that.
2. Intervention extension: The interventions carried out could not be applied over long periods during the year, as they were not directly related to the educational program (Curriculum), and therefore the number of sessions used is limited.
3. Existence of natural groups: The studies were carried out in schools, during their regular activities, so it was necessary to respect their schedules and internal organization. In this sense, neither parallel courses nor courses from different schools could be put together in a single large group to perform a random distribution.
4. Low demographic variety: The studies were conducted only with schools in Chile's metropolitan region, with the same type of administrative dependence (i.e., subsidized private schools) and similar socioeconomic levels (i.e., middle and upper-middle-class).
5. A small sample: Although the sample size used was acceptable for the methods used, increasing the number of students is recommended for better overall findings. However, student attendance is a relevant variable that should be taken into account both when increasing the number of intervention

sessions and when constructing the sample, since several cases should be discarded for this reason. In addition, when working with people in an iterative methodology such as the one used, it should be considered that each time the intervention is tested, it cannot be tested again with the same population. Therefore, the total number of participants is much higher than the number that will actually be analyzed.

1.8. Thesis Outline

This thesis focuses on two important aspects, on the one hand, the measurement of collaborative problem solving, and on the other hand the development of this skill in students. The first aspect will be related to the development of a CPS assessment instrument. The second aspect will be related to the development of a game to develop this skill, and to the development of a feedback system based on artificial intelligence techniques. Therefore, this thesis will be divided into the following chapters.

1.8.1. Chapter 2

Because pre-post experimental designs are essential to identify new methods to develop collaborative problem-solving skills, it is necessary to design equivalent assessments to facilitate consistent score interpretations and reduce the training effect. To do so, a Design-Based Research approach to design and validate an assessment tool with two equivalent versions based on a framework proposed by the OECD and applied to a collaborative activity will be presented in Chapter 2. In addition, the iterations and tests carried out to arrive at the final instrument will be presented. In the process, different hypothesis questions that have been studied in the CPS measurement framework, such as the use of virtual agents, predefined messages, and gender differences, will be addressed. Finally, the results obtained will be presented, demonstrating their validity, and especially the possibility of constructing two equivalent forms, which represents a major challenge for the execution of experimental designs.

1.8.2. Chapter 3

In this chapter, the study and development of a feedback system that uses automated planning techniques to promote communication among students will be presented. The system is designed to be used in a real-world educational environment, considering the underlying theory of when and how to give feedback. To test its effectiveness and performance, analyses are performed by using it in a game that seeks to develop CPS skills. On the backend, the system calculates the solution to the task in a partial order plan using an automated planning engine. While monitoring the plan and providing feedback to the students. The feedback will be shown to allow students to play the game better, improve their communication, and develop their collaborative problem-solving skills. It will also show a novel approach to monitoring multi-agent partial order plans that is more efficient than the approaches proposed in the literature, which facilitates its use in low-tech environments such as classrooms.

1.8.3. Chapter 4

Because there is still little research on the development of CPS skills among elementary school students. Chapter 4 analyzes how regulation skills and emotions of elementary school students are supported by a collaborative game using a collaborative script and group awareness. Its effectiveness will be presented through an intervention to 223 students, combining qualitative and quantitative analyses. The results will be explained in detail, which show an improvement in the attitudes towards collaboration in the experimental group students. Finally, the relationship between the tools used and emotions, co-regulation, and shared regulation is highlighted; providing evidence of how the game favors regulation skills and emotions, which in turn promotes the development of collaborative problem-solving skills.

1.9. Thesis Structure

The structure of this thesis is based on the research objectives described in Section 1.4. Figure 1.4 provides a model to demonstrate the connections between the objectives, hypotheses, research questions, items, and results included in this thesis.

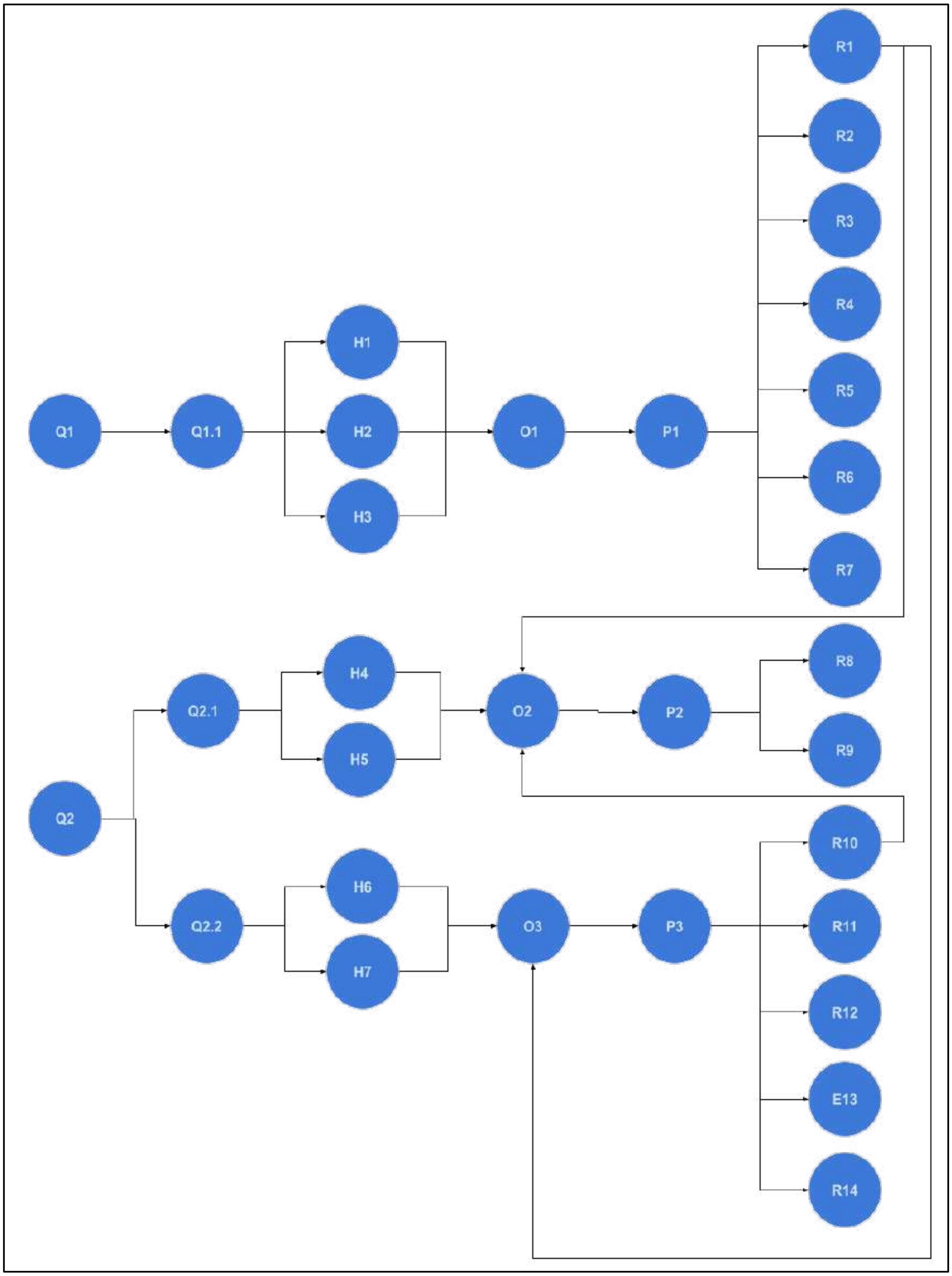


Figure 1-4: Connections between the research questions, hypotheses, objectives, papers, and results.

2 ASSESSING COLLABORATIVE PROBLEM-SOLVING SKILLS AMONG ELEMENTARY SCHOOL STUDENTS

2.1. Introduction

Recent education and curriculum reforms have focused on effectively incorporating 21st century skills into various different programs (Andrews-Todd & Forsyth, 2018; Griffin, McGaw, & Care, 2012). Growing interest in this area is reflected in the educational policy of countries such as the United States, Japan and Singapore (Csapó & Funke, 2017; Binkley et al., 2012), among others (Chang et al., 2017; Von Davier, Hao, Liu & Kyllonen, 2017). Within this context, Collaborative Problem-Solving (CPS) skills have become increasingly important as the problems faced by society become more and more complex and abstract, requiring a greater level of expertise in order to be solved (Fiore, Graesser & Greiff, 2018; Care & Griffin, 2014). Furthermore, the experience gained from such expertise is spread across different people, making their collaboration essential (Rosen, Wolf, & Stoeffler, 2019). This highlights the need for the widespread teaching and assessment of collaborative problem-solving skills (OECD, 2017; Liu, Hao, Von Davier, Kyllonen & Zapata-Rivera, 2016; Sun et al., 2020).

New assessment methods play a key role in the teaching-learning process of 21st century skills as it represents a move away from traditional methods of assessment (Halpin, von Davier, Hao & Liu, 2017). In this regard, assessments based on multiple choice or closed-ended questions are increasingly considered non-optimal as they do not pose any real cognitive challenge to the students. Instead, such assessments should focus on activities that can measure a student’s skills in understanding what he or she does when faced with a problem (Andrews-Todd & Kerr, 2019).

The first large-scale assessment of collaborative problem solving was proposed by the Assessment and Teaching of 21st Century Skills Project (ATC21s; Griffin & Care, 2014). For this assessment, two students collaborated via chat to solve a group task using a computer. A system was then used to automatically review records of the students’ actions (Hao, Liu, Kyllonen, Flor, & von Davier, 2019; Scoular & Care, 2019a). However, this automated process did not include a review of the content of the students’ messages (Adams et al., 2015).

An alternative method was proposed by the Educational Testing Service (ETS). Their method was implemented in a project called Collaborative Science Assessment Prototype (ECSAP). They made two versions of a web-based task involving a science simulation based on volcanoes, which started with students answering a question individually. They were then allowed to talk and collaborate with their peers via chat before changing their original answer if they so wished. The assessment focused mainly on two variables: the number of students who change their original answer having collaborated with others, and the change in scores before and after the collaboration. A record of the students' conversations was also kept for subsequent discourse analysis, allowing the researchers to understand how students solved problems collaboratively (Hao, Liu, Von Davier & Kyllonen, 2015; Griffin, 2017). Similar studies have looked to automatically classify messages exchanged by students by interpreting the natural language. However, scholars are still exploring how to improve these methods (Hao, Chen, Flor, Liu & von Davier, 2017; Flor, Yoon, Hao, Liu & Von Davier, 2016).

During collaborative problem solving, the different members of a group depend on one another. Several studies have therefore shown that the composition of a group can have a significant effect on student performance (Chang et al., 2017; Chen & Kuo, 2019; Stewart, Amon, Duran & D'Mello, 2020). Other studies prefer to focus on peer-to-peer collaboration, giving greater importance to authentic assessment methods and revealing a positive impact on students' problem-solving skills (Nouri, Åkerfeldt, Fors & Selander, 2017; Sun et al., 2020). Furthermore, other studies have shown improved results when students work with virtual agents (Stoeffler, Rosen, Bolsinova & Von Davier, 2019; Rosen & Tager, 2013). However, there is no consensus among researchers with regards to the best method for measuring collaborative problem-solving skills (i.e., using virtual agents or humans as peers) (Graesser, Kuo & Liao, 2017; Herborn, Stadler, Mustafić & Greiff, 2018).

In this regard, the position taken by the Programme for International Student Assessment (PISA) in its 2015 cycle was to use a computer-mediated environment in which the students interacted with virtual agents to solve different types of problems. They proposed a framework which combines the three main collaborative competences with the four stages of the individual problem-solving process (OECD, 2017). In this case, the three main collaborative competences are: establishing and maintaining shared understanding, taking appropriate action to solve the problem, and establishing and maintaining team organization (OECD, 2017). Similarly, the four stages of the problem-solving process are: exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting (OECD, 2017). However, some authors suggest that the assessment proposed by the OECD does not fully measure students' collaborative problem-solving skills. This is because the assessment focuses on the cognitive component of problem solving, rather than on the social aspects (He, von Davier, Greiff, Steinhauer & Borysewicz, 2017; Child & Shaw, 2019).

Despite the literature reporting numerous projects that aim to assess students' collaborative problem-solving skills (Krkovic, Wüstenberg & Greiff, 2016; OECD 2017; Scoular & Care, 2019b; Stoeffler, Rosen, Bolsinova & Von Davier, 2019), more applied studies on its implementation and evaluation in real-life educational settings are required (Graesser, Fiore, Greiff, Andrews-Todd, Foltz & Hesse, 2018). Therefore, pre-post experimental designs are essential for identifying which methods effectively develop these skills. To do so, equivalent tests are required to facilitate consistent score interpretations and reduce the practice effect. This leads us to our first research question, which asks "How can we assess collaborative problem-solving skills among elementary school students using an instrument with two equivalents forms?"

Finally, it is difficult to understand the complex design processes that underpin these projects. This not only makes it difficult to interpret their findings, it also makes it hard to replicate the studies and use them as a starting point for further research. In this sense, Design-Based Research emerges as a viable method as it explicitly details the advantages and disadvantages of the design decisions that are made during a lengthy, iterative and complex process such as assessing collaborative problem-solving skills. Additionally, framework generalizations are still needed (Sun et al., 2020). In this regard, one contribution would be to describe experiences involving different populations, e.g. participants of different ages. Our second research question therefore asks, "Which design principles will help develop a tool that facilitates the assessment of collaborative problem-solving skills?"

2.2. Methodology

2.2.1. Hypothesis

To answer our research questions, we must contrast different hypotheses that show the validity, reliability and equivalence of the instrument to be developed:

H1: Design-Based Research allows for the development of an instrument for assessing collaborative problem-solving skills among elementary school students with two equivalent forms.

H2: The instrument that is developed measures both the problem-solving and collaborative dimensions of collaborative problem-solving skills among students aged between 10 and 13 years old.

H3: The scores obtained on the two forms of the instrument are equivalent.

H4: The scores obtained are equivalent across genders.

H5: There is no significant difference when assessing collaborative problem solving in human-human groups versus human-agent groups using the proposed instrument.

2.2.2. Design-Based Research

The assessment of collaborative problem-solving skills is a complex task that involves both theory and technology design. The Design-Based Research (DBR) approach can support such a complex task as it allows for the intertwining of research and practice; examining the technological process while being also used as a pedagogical tool (Amiel & Reeves, 2008). This method focuses on studying learning in context, through systematic, iterative and incremental development (Design-Based Research Collective, 2003). One way of guiding this iterative process is the model proposed by Reeves (2006; see also Figure 2-1). Based on this model, we designed and validated an instrument for measuring collaborative problem-solving skills in three iterations (Table 2-1).

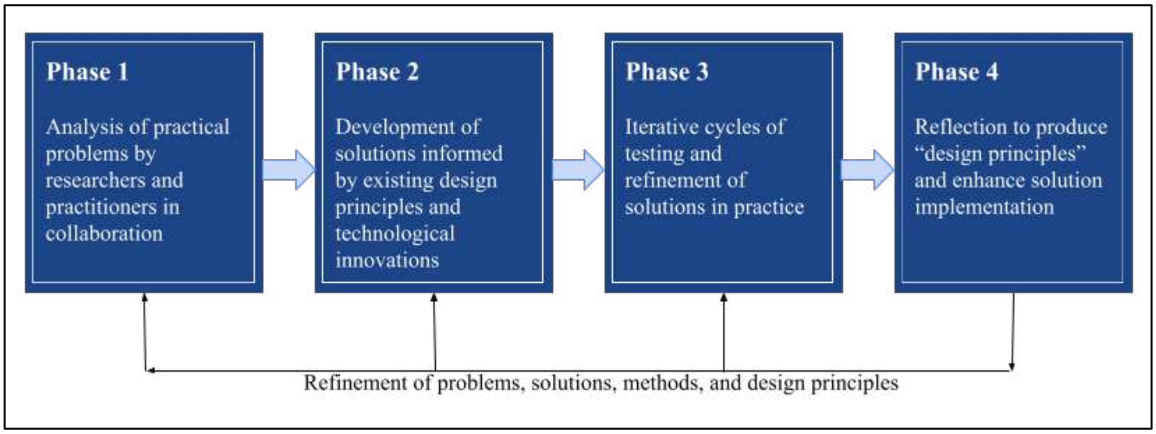


Figure 2-1: Design-Based Research model (Reeves, 2006)

Table 2-1: Summary of the Design Iterations and Validation of the Measurement Instrument

Iterations	DBR Phases	Tasks	Outcomes
First Iteration	Phase 1	Literature review and definition of the theoretical framework	Two equivalent forms of an instrument for measuring collaborative problem solving (pre and post).
	Phase 2	Operationalization of the theoretical framework	
	Phase 3	Testing and analysis of the instrument in a real-world setting	
	Phase 4	Reflection on the results and suggestions for future iterations	

Second Iteration	Phase 2	Redefinition of the assessment indicators and theoretical framework (inclusion of virtual agents)	Two versions of the test, one in which the groups comprise only humans (HH) and another in which the groups include a virtual agent (HA)
	Phase 3	Testing and analysis of the instrument in a real-world setting	
	Phase 4	Reflection on the results and suggestions for future iterations	
Third Iteration	Phase 2	Update of the interface, redefinition of the items relating to exploring problems so as to improve student engagement and the validity of the instrument	Final version of the instrument, with improvements to the interface
	Phase 3	Testing and analysis of the instrument in a real-world setting	
	Phase 4	Conclusions regarding the validity of the instrument	

Furthermore, Edelson's model (2002) divides DBR into design procedures, problem analyses, and design solutions. This model, therefore, represents a collection of decisions that designers must make in any design process, as well as providing a practical way to report on a complex DBR project.

2.2.2.1. Design procedure

Measuring collaboration is a complex task that must consider various different perspectives. Doing so therefore requires a team with different backgrounds in education and advanced levels of collaborative work (von Davier, Hao, Liu, & Kyllonen, 2017). To build a solution, we assembled a multidisciplinary team. This team included specialists in computer science, educators, designers, specialists in linguistics and literature, and psychologists.

To meet the objective of building an instrument that allows us to answer the research questions posed in the introduction, we first had to conduct a review of the literature (see Table 2-1). This allowed us to define both the construct and the characteristics of the activity that would be used to evaluate it (Phase 1, Figure 2-1). We then continued with the operationalization of the theoretical framework (Phase 2, Figure 2-1). Following this, a prototype of the instrument was built and tested in a real educational context to analyze the validity and observe how students used the solution (Phase 3, Figure 2-1). Finally, based on the results, we improved the design of the proposed instrument (Phase 4, Figure 2-1). We then repeated this cycle two more times, with new outcomes obtained in each iteration (Table 2-1).

2.2.2.1. Initial problem analysis

A literature review was conducted to find a set of guiding principles for the development of a standardized assessment of collaborative problem solving. This began by defining collaborative problem solving as *the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution* (OECD, 2017). Another possible definition is the one proposed by ATC21s. However, both initiatives are very similar in terms of their construct (Scoular, Care & Hesse, 2017).

Furthermore, ATC21s (Griffin & Care, 2014), ETS (Von Davier et al., 2017) and OECD (OECD, 2017) all measure collaborative problem-solving skills by analyzing the logs that are generated when students solve the problem via chat. In this respect, collaborative problem-solving skills can be inferred from the communication and interaction that takes place when students solve a problem as a group (Von Davier et al., 2017).

Our first task was, therefore, to develop an activity that had to be solved collaboratively. To do so, the design process considered the conditions required for an activity to be considered collaborative (Szewkis et al., 2011). This includes the existence of a common goal (Dillenbourg, 1999), positive interdependence between peers (Johnson and Johnson, 1999), coordination and communication between peers (Gutwin and Greenberg, 2004), individual accountability (Slavin, 1996), awareness of peers' work (Janssen et al., 2007) and joint rewards (Axelrod and Hamilton, 1981). Finally, the theoretical framework proposed for PISA 2015 (OECD, 2017) is used as the basis for analyzing the conversations and interactions that take place during the collaborative problem-solving process. This decision was based on the fact that the PISA framework has an enormous impact on educational policies across the globe. This framework has been used in previous studies (Rosen, 2017; Graesser, Cai, Morgan, & Wang, 2017; He et al., 2017; Herborn et al., 2018) and combines the three main collaborative competences with the four stages of the individual problem-solving process (OECD, 2017). Combining these two elements provides a matrix of twelve specific skills (OECD, 2017). In turn, each of these skills is then associated with different actions, processes and strategies in order to define each student's level.

2.2.2.3. First Iteration: Problem Analysis

To evaluate the usability of the activity, a panel of three experts in education and technology evaluated the activity's usability. An assessment tool for testing the usability of mobile applications was adapted to the specific context of collaborative problem solving (keepitusable.com, 2011). This produced a total of 25 criteria for assessing the usability of the activity (Table G.1, Appendix G).

A total of 171 students were then tested using this activity (Table 2-2, Section 2.3). The students were randomly divided into 57 groups of 3 (Zurita, Nussbaum & Salinas, 2005) and had to solve the problems that were presented by the assessment tool while communicating exclusively via chat. These groups were divided into two. One half of the groups started with Form A (Figure 2-2) before moving on to Form B (Figure 2-3). The other half started with Form B before moving on to Form A. Doing so allowed us to control for the learning effect of one form over the other. The students had to work in silence during both activities and were only allowed to communicate via the online chat, where the conversation could be recorded.

Following this, an evaluator then applied the relevant assessment criteria (Table F.1, Appendix F) to the group chats to validate the instrument.

2.2.2.4. Second Iteration: Problem Analysis

The second version of the solution was tested with a total of 325 students. This version involved two different activities (Table 2-2, Section 2.3). The aim of the first activity was to analyze differences between human-human groups and human-agent groups. This activity involved 160 students, who were randomly divided into two different kinds of groups: groups with three students (human-human) and groups with one student (human-agent). This led to a total of 40 groups of three (120 students in total) and 40 individuals, who each worked with two virtual agents. In the groups of three students, only one randomly-selected student's ability to collaborate was assessed. This is because assessing all three students would require all three of them to be given the same opportunities to be assessed, thus tripling the length of the study and requiring a larger variety of problems to be solved. Finally, both the human-human groups as well as the Human-Agent groups were asked to solve problems in Form A of the test (Figure 2-4). A score for collaborative problem-solving skills was therefore calculated for each of the students, as well as recording the time it took them to solve the problem.

The aim of the second activity was to study the construct validity and show that it is possible to develop two equivalent forms of the same tool (A and B). This activity only included Human-Agent groups and involved 165 students, plus the 40 students who worked with virtual agents in the first activity. The second activity therefore involved a total of 205 participants, all of whom completed both forms of the test (A and B, Figures 2-4 and 2-5, respectively). The students did so in the same way as the first iteration by completing one form after the other, without any additional time or activities between the two forms. The students worked in silence and were only allowed to communicate via chat using pre-defined messages.

2.2.2.5. Third Iteration: Problem Analysis

This final analysis aimed to confirm the internal structure found in the previous iteration. The activity therefore involved a total of 223 participants (Table 2-2, Section 2.3), all of whom completed both forms of the test (A and B, Figure 2-7 and 2-8, respectively). The students did so in the same way as the first and second iterations, i.e., in silence, communicating via chat using pre-defined messages and without any additional time or activities between the two forms.

2.2.3. Participants

Table 2-2: Participants in each iteration

	Age	Boys	Girls	Total
First Iteration	10 – 13	88	82	171
Second Iteration	10 – 13	166	159	325
Third Iteration	10 – 13	120	103	223

A total of 719 students participated in the different iterations of this study. The students were aged between 10 and 13 (fifth and sixth grade), an age at which children have the necessary cognitive skills for completing such tasks, often linked to abstract thinking (Dumontheil, 2014; Crook, 1998; Molnar, Greiff, & Csapo, 2013). The students all came from backgrounds of medium and medium-high socioeconomic status and participated in the study on a voluntary basis, with written consent from the parents and authorization from the school. The total sample was split across three iterations, as shown in Table 2-2.

All of the students attended schools based in Santiago, Chile. It is worth noting that, according to PISA (OECD, 2018), with scores of 452, 417 and 444 in reading, mathematics and science, respectively, Chile enjoys the best academic results in the region. However, these are still way below the global average of 487, 489 and 489 in reading, mathematics and science, respectively (OECD, 2018). In the 2015 PISA test for Collaborative Problem-Solving Chile obtained 457 points, while the global average was 500 (OECD, 2018).

2.2.4. Data collection and instruments used

The data for this study mainly came from records of the students' conversations and actions within the platform when solving problems collaboratively (Table H.1, Appendix H).

A summary of the instruments used to gather data can be found in Table 2-3.

Table 2-3: Summary of the Instruments and Analyses Performed During Each Iteration

DBR Iteration	Instrument/Data	Analysis	Hypothesis
First Iteration	Log of student conversations and actions	Reliability of instrument for measuring collaborative problem solving	H2
		Validity of instrument for measuring collaborative problem solving	H2
Second Iteration	Log of student conversations and actions	Reliability of instrument for measuring collaborative problem solving	H2
		Validity of instrument for measuring collaborative problem solving	H2
		Comparison between forms	H3
		Comparison between boys and girls	H4
		Comparison between human-human and human-agent groups	H5

Third Iteration	Log of student conversations and actions	Validity of instrument for measuring collaborative problem solving	H2

Firstly, we analyzed design principles from the real-life experience during the three iterations (H1). Secondly, we were able to contrast our second hypothesis (H2) with reliability and validity results by reviewing user logs and participant conversations from each iteration. Thirdly, based on the information collected from the user logs during the test and refine phase of iteration 2 (Intermediate design solution, Section 2.2.4), we analyzed the equivalence of the instruments (H3) by comparing the results from Form A and Form B. Then the second iteration allowed us to compare boys' and girls' results (H4). Finally, from the same iteration, we compared the results between human-human groups and human-agent groups (H5).

2.2.5. Statistical methods

SPSS 26 and R 3.5.1 were used to analyze the data. The specific methods used for each iteration of this study (Table 2-1) are described below.

2.2.5.1. First and second iteration

To analyze the validity and reliability of the proposed instrument, exploratory factor analysis (EFA) was used to test the factor structure of the instrument in the first and second iterations (Williams, Onsman & Brown, 2010). Firstly, communalities of all of the items were analyzed to identify which items should be eliminated (see Table 2-5). Following this, parallel analysis was used to determine the number of factors (see Figure 2-6). We then analyzed whether there was any relationship between the factors that were found and the framework that was adopted. This was done using the principal axis and varimax rotation methods for the two forms of the test (Table 2-6). The two forms were analyzed separately to explore their internal structures (Loehlin, 2004). Finally, McDonald's Omega (Dunn, Baguley & Brunsden, 2014) was calculated to test the reliability of both forms of the test (Table 2-6).

Following this, measurement invariance tests (Millsap, 2011) were used to test the equivalence of forms A and B of the instrument (Table 2-7), as well as to compare gender differences (Table 2-8). A series of increasingly restrictive invariance models were tested (Hirschfeld & Von Brachel, 2014), including configural (number of latent variables and the pattern of loadings are similar across the groups), weak (magnitude of the loadings is also similar across the groups), strong (intercepts are also similar across the groups), and strict (residual variances are also similar across groups) invariance. The cutoffs used to analyze invariance were the difference between chi-square and the fit indices proposed by Millsap and Cham (2012) and Cheung and Rensvold (2002), in which differences of less than .010 in CFI and .015 in RMSEA are considered indicators of equivalence. Next, multivariate analysis of variance (MANOVA; Morrison, 2005) was conducted using the average score for each factor to compare the human-human and human-agent models (Table 2-9). Finally, an analysis of variance (ANOVA; Keselman, 1998) was used to compare how long it took human-human and human-agent groups to complete the activity (see Table 2-10).

2.2.5.2. Third Iteration

Based on the results of the EFA analysis described above, a confirmatory factor analysis (CFA, Brown, 2014) with robust maximum-likelihood estimator was used to validate five models (Table 2-11). To do so, we took five fit indices and their respective cutoff values into consideration as being indicative of an acceptable fit (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1999). The first of these indices was chi-square, for which a good fit would provide an insignificant result at a 0.05 threshold. However, there are some limitations to the use of this index due to the impact of the sample size. The second index was the Comparative Fit Index (CFI), for which we used >.95 as the statistical critical value. The next index was the Tucker-Lewis Index (TLI), which considers statistical values greater than .95 as critical. The fourth index was the Root Mean Square Error of Approximation (RMSEA), which considers values <.06 as the cutoff. The final index was the Standardized Root Mean Square Residual (SRMR), which requires values of greater than .09. In addition, McDonald's Omega coefficient (Dunn, Baguley & Brunsden, 2014) was calculated to show the reliability of the instrument scores (Table 2-12).

2.2.6. Missing Data

In this study, 7% to 20% of students had missing responses on the two forms of the test, which was due to a number of reasons. The most frequent reasons were completely random. Some of the missing values were caused by a technical error (e.g. disconnection), mishandling of the device, or battery issues. Also, not all students were able to finish the tests. For some students this was due to a lack of motivation, while for others they found it difficult to finish on time. We recorded these as nonignorable missing values and used Full Information Maximum Likelihood (FIML) to treat them[1]. FIML is known to be reliable when it comes to treating missing data, making less restrictive assumptions than, for example, listwise deletion (Enders, 2010).

2.2.7. Ethical procedure

Firstly, the study was approved by the University's ethics board. Following this, the schools were invited to participate in the study and gave their written authorization to do so.

Then, before participating in the first session of each iteration, the students' parents were provided with information on the purpose of the research, the specific activities it involved and the duration of these activities, as well as the risks and benefits of participating. The voluntary nature of participating in the study was highlighted. Furthermore, the students' parents were requested to sign an informed consent form.

Finally, at the start of the first session of each iteration, the students whose parents had provided informed consent were informed about the activities in which they would be participating. Following this, the students were invited to sign an informed assent and reassured that they could stop participating in the activity at any time.

2.3. Iterative design cycles and results

We present our design considerations and the proposed solution for each of the three iterations. For each iteration, design reflection and construct validation results are also included to answer our research questions.

---

[1] We also tested the CFA models using listwise deletion. However, this resulted in either 1) convergence issues or 2) only minor differences when models did converge, when compared to FIML and missing value coding configuration.

2.3.1. First Iteration: Initial design solution

Two versions of the same activity were developed to produce two equivalent assessments that can be used as a pre- and post-test, accordingly. In this case, the problem involved choosing stars (Form A, Figure 2-2) or atoms (Form B, Figure 2-3) to build a constellation (or molecule) by following a specific pattern. Each activity in the first iteration involved 8 difficulty levels, with an increasing number of stars or atoms and an increasingly complex geometry.

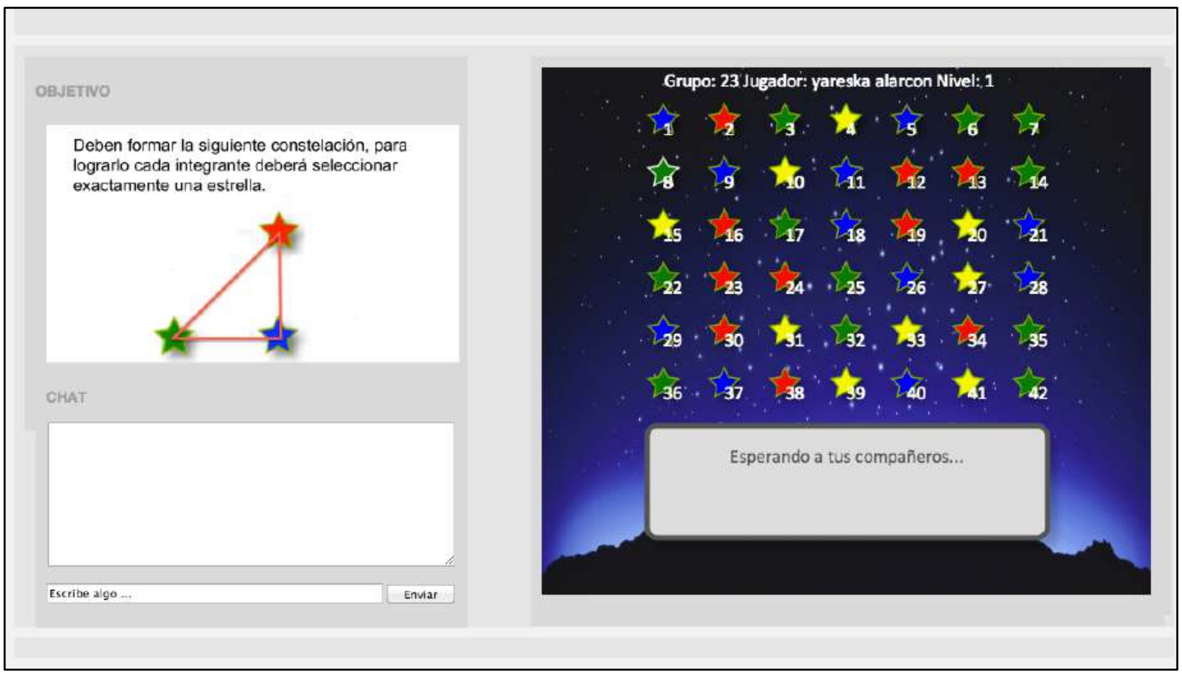


Figure 2-2: Form A

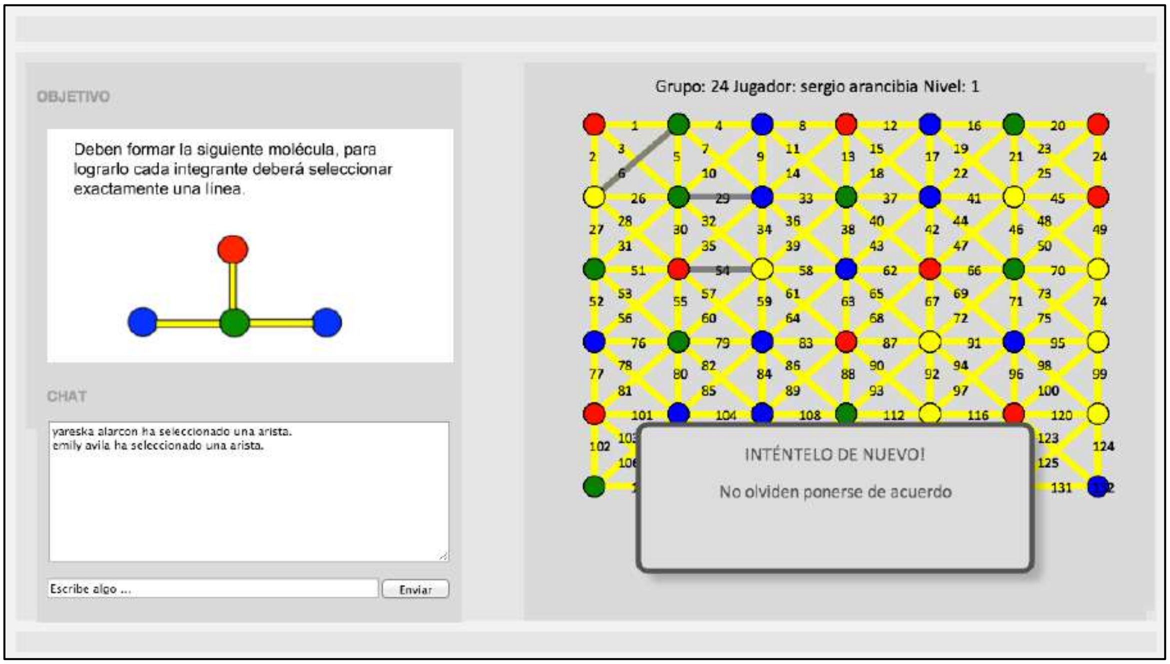


Figure 2-3: Form B

As with previous studies, peers communicated via chat (Griffin & Care, 2014; OECD, 2017; Von Davier et al., 2017). This tool was chosen as it is a simple and familiar form of communication that can facilitate data collection (OECD, 2017). Furthermore, the activity did not require any particular background knowledge so that the participants' prior knowledge would not interfere with the study. The task had to be completed in groups of three. This group size was chosen as it is recommended for collaborative activities (Zurita, Nussbaum & Salinas, 2005) and has been shown to encourage negotiation and debate among peers (Nussbaum et al., 2009; Strijbos et al., 2004). In addition, the OECD (2017) argues that triads also allow for interesting complexities within social interactions.

Furthermore, the conditions for a collaborative activity are fulfilled as follows: the three members of the group all have to create a constellation or molecule together following the pattern of colors and the distribution that is shown (common goal), with each of them choosing one or more stars or atoms (positive interdependence between peers) and communicating exclusively via chat (coordination and communication between peers). To meet their objective, each participant can only choose stars or atoms of the color that is assigned to them (individual accountability). Finally, at the end of each level the students can see the constellation or molecule that has been created (awareness of peers' work) and move on to the next level (joint reward).

Additionally, to ensure positive interdependence, each level had at least three possible solutions, where each participant had to choose at least one star or atom. This increases interdependence and ensures that the students must come to an agreement. For the group to complete a level, the students first had to settle on which solution they would build before assigning roles to each member of the group and building the required structure. To do so, the students could only communicate via chat, with the system saving a copy of the students' conversations.

In summary, a correct solution occurs when each participant selects the set of stars or atoms that allows the group to match the colors and pattern that is presented and form the required constellation or molecule.

Based on this initial solution, the results of the parallel analysis (Figure E.1, Appendix E) for Form A of the test suggests 7 factors that explained 46% of the variance (Table E.3, Appendix E). Similarly, the same analysis for Form B suggests a total of 8 factors (Figure E.2, Appendix E) that explained 42.34% of the variance (Table E.5, Appendix E). Furthermore, items 1 and 21 on Form A of the test had low eigenvalues and had to be reviewed (Table E.2, Appendix E). In the same way, items 1 and 21 on Form B of the test had low eigenvalues (Table E.4, Appendix E), while item 14 was removed from the analysis as no student managed to answer it correctly. Finally, it was not possible to detect any patterns among these factors and the corresponding items. It was therefore not possible to establish any sort of relationship between the underlying theory and the factors found from the analysis of the student conversations.

Additionally, the average Cronbach's alpha for Form A was .703 (Table E.3, Appendix E) and for Form B was .540 (Table E.5, Appendix E). The difference in these reliability indicators may be due to the fact that there was an error in the design of Form B, leading it to be more difficult. This is because the number of elements that had to be selected in Form B (Figure 2-3) of the test was greater than in Form A (Figure 2-2). Form B therefore had to be redesigned to ensure that the two forms were equivalent.

A summary of a report on the activity's usability can be found in Table E.1 (Appendix E). In this case, the experts highlighted the importance of the tutorial that is shown to the participants. However, they also suggested including a button to revisit the tutorial and an undo button as (in this version) the students were not able to reverse their decision once they had selected a star or atom. They also highlighted certain messages that were either redundant or unclear and therefore needed correcting.

Finally, the results suggested that the instrument required certain improvements. As the psychometric results were not optimal (See Appendix E), the initial assessment criteria had to be modified. On this matter, an important takeaway from this iteration was realizing how difficult it was to review the chat logs. This is because the students' communication tended to be more informative than strategic, with the students often giving updates on their actions rather than planning a solution. Because of this, coding the conversations using the criteria proposed in Appendix F became too complex and difficult to scale.

2.3.2. Second Iteration of Design-Based Research: Intermediate design solution

Based on the analysis from the initial solution, a decision was made to redefine the assessment criteria. As a result, new criteria were defined for each of the skills found in the OECD Framework (OECD, 2017). These definitions were based on the description of collaborative problem-solving skills proposed by PISA (OECD, 2017). The new criteria (Table 2-4) were reviewed on a content level by the multidisciplinary team responsible for this research. This team includes members with an extensive background in primary education, psychology and linguistics. Following this, a 24-item assessment tool was then developed (Table of Specifications in Table A.1, Appendix A). Each of these items corresponds to a unit that assesses a skill in a specific context following a particular stimulus. A stimulus is any object or event that prompts a response from the students (e.g. a change of goal or a message sent by a peer, etc.). The context, on the other hand, refers to everything that surrounds and happens during the build-up to the stimulus (e.g. the current level and the number of stars or atoms to be chosen, etc.). These items can be found throughout the activity and look to measure one of the 12 skills defined by the OECD (OECD, 2017). Each item therefore has a problem-solving component and a collaborative component. For instance, item 4 looks to measure B1 from the OECD framework (Table A.1, Appendix A), which is defined as "building a shared representation and negotiating the meaning of the problem (common ground)" (OECD, 2017). This skill includes representing and formulating (Problem-solving) and establishing and maintaining shared understanding (Collaborative competence). Consequently, to meet the criteria for B1, the student has to show that he or she established a shared vision of the problem by checking that everyone has a similar understanding, negotiating and discussing the understanding of the objective, or recapitulating and summarizing the previous discussion about the group's understanding of the problem (Table 2-4).

Table 2-4: Assessment Criteria by CPS Skill

CPS Skill	Description	Assessment criteria
A1	Discovering team members' perspectives and abilities	**Identify mutual knowledge (what each participant knows about the problem)** · Ask about the objectives of the problem · Request clarification and affirmation about the objectives · Directly request a response from the other person, or lead the discussion about the objectives **Identify group members' perspectives of the problem to be solved** · Ask about or respond to each teammate's perspective of the problem · Request clarification and affirmation on each teammate's perspective · Directly request a response from a specific teammate, or lead the discussion about the teammates' perspectives
A2	Discovering the type of collaborative interaction used to	**Identify the actions and interaction neede to solve the problem** · Ask about or respond to actions and interactions needed to solve the problem

	solve the problem, along with the goals	· Request clarification and affirmation on actions and interactions needed to solve the problem · Directly request a response from a specific teammate, or lead the discussion about actions and interactions needed to solve the problem **Identify constraints of actions when solving the problem** · Ask about or respond to how the constraints of the problem restrict their actions · Request clarification and affirmation on how the problem constraints restrict their actions · Directly request a response from a specific teammate, or lead the discussion about how the constraints of the problem restrict their actions
A3	Understanding the problem-solving roles	**Identify the necessary problem-solving roles** · Ask about or respond to the roles needed to solve the problem · Request clarification and affirmation about the roles needed to solve the problem · Directly request a response from a specific teammate, or lead the discussion about the roles needed to solve the problem **Identify team members' strengths and weaknesses** · Ask about or respond to each teammate's problem-solving skills or abilities · Request clarification and affirmation on each teammate's problem-solving skills or abilities · Directly request a response from a specific teammate, or lead the discussion about each teammate's problem-solving skills or abilities

B1	Building a shared representation and negotiating the problem's meaning (common ground)	**Establish a shared vision of the problem** · Ask to check that everyone understands the same · Negotiate and discuss the understanding of the objective of the problem · Recapitulate and summarize the previous discussion about the group's understanding of the problem
B2	Identifying and describing tasks to be completed	**Establish group goals/tasks to solve the problem** · Ask about or respond to the tasks needed to solve the problem · Negotiate and discuss the tasks needed to solve the problem · Recapitulate and summarize the previous discussion about the tasks needed to solve the problem
B3	Describing roles and team organization (communication protocol/rules of engagement)	**Establish the roles and organization needed to solve the problem** · Ask about or respond to the organization roles/tasks needed to solve the problem · Negotiate and discuss the organization roles/tasks needed to solve the problem · Recapitulate and summarize the previous discussion about the organization roles/tasks needed to solve the problem
C1	Communicating with team members about the actions to be/ being performed	**Formulate a sequence of steps or plan to solve the problem** · Ask about or respond to a plan or a possible solution from teammates · Negotiate and discuss a potential solution and integrate information from group members to develop a better solution · Request clarification and affirmation of the plan · Directly request a response from a specific teammate, or lead the discussion about a plan

		· Ask for agreement on the plan or give short, positive feedback
C2	Enacting plans	**Execute actions to attempt a solution** · Execute planned actions · Communicate actions performed · Make decisions based on the actions of others
C3	Following the rules of engagement, (e.g. prompting other team members to perform their tasks.)	**Follow rules of engagement** · Lead or conciliate as appropriate · Report the fulfillment/non-fulfilment of own commitments · Ensure adherence to rules by asking/checking with others about the progress and fulfillment of their task · Ensure group communication of important information · Maintain and encourage collaborative communication in the group · Ask questions to verify explicit information important to others · Facilitate necessary changes to repair communication breakdowns · Facilitate necessary changes to overcome obstacles · Facilitate necessary changes to optimize group performance
D1	Monitoring and repairing the shared understanding	**Monitor shared understanding of the problem** · Ask about or respond to doubts about the problem · Identify and report misunderstandings of the problem · Reaffirm or discuss what the group understands about the problem · Repair misunderstandings · Directly request a response from a specific teammate, or lead the discussion about doubts or errors regarding the group's understanding of the problem

D2	Monitoring action results and evaluating the success in solving the problem	**Monitor the group results when solving the problem** · Ask about/respond to the results of the task · Talk about the success or failure of the plan to solve the problem · Reaffirm or discuss the action performed · Give feedback about the action performed · Directly request a response from a specific teammate, or lead the discussion about doubts or errors regarding the actions performed
D3	Monitoring, providing feedback and adapting the team organization and roles	**Monitor the organization of the group** · Talk about non-adherence to the rules · Talk about team disorganization · Give feedback on the success or failure of the organization of the group in solving the problem · Directly request a response from a specific teammate, or lead the discussion about the success, failure or doubts regarding the organization of the group

Note. The CPS skills are based on the OECD Framework (OECD, 2017).

To address the difficulty of coding the students' conversations, the option of using pre-defined messages in the chat was explored. To do so, students took turns to choose a message to send to their peers. Because all of the students were choosing from the same set of pre-defined messages, only one student from each group was ultimately assessed. Doing so avoided any unnecessary redundancy or repetition. Although the other two students participated in the human-human setting, acting as 'collaborators' for the main human-human group (Rosen, 2015), they were not actually assessed.

The pre-defined messages were created based on the conversation logs from the first iteration and the minimum interaction patterns required to solve collaborative problems (Laurinen & Marttunen, 2007; He et al., 2017). By doing so, conversations for each of the assessment criteria (Table 2-4) were made up of the different messages selected by each of the students. Furthermore, two types of interaction pattern were required to cover the 12 skills defined by PISA. This included non-simultaneous communication (i.e., when participants take turns to talk) and simultaneous communication (i.e., when participants can send messages at the same time). In this way, the skills classified by PISA as *identify* and *explore* (OECD, 2017) are easily addressed through simultaneous communication. This is because they require few messages as the tasks are simple (i.e., recognizing patterns) and do not require complex communication.

As a result, the student being assessed had to communicate by selecting a message from a list of possible alternatives. Appendix B shows how the conversations for each of the skills defined in Table 2-4 were structured. For example, we can represent the turns of the three participants in the group for skill C1 (Figure B.7, Appendix B) and assign one or more messages to each member of the group to assess the skill (Table J.1, Appendix J).

Having introduced pre-defined messages for peer-to-peer communication, it was then decided to automatically score the students' responses in order to aid scalability. Each possible combination of messages (i.e., conversation) for each skill was then graded on a scale from 0 to 3. A score of 0 suggests that the skill was not present in the conversation, while a score of 1 suggests that the skill was barely present. A score of 2, on the other hand, suggests that the skill was present, while a score of 3 suggests there was outstanding use of said skill. Appendix J includes an example of a conversation in which the student scored 0 points for skill C1, and another where the student scored 3 points (Table J.2, Appendix J).

When working with a group of humans, the responses cannot be considered deterministic as they depend on each person's response (Chang et al., 2017; Rosen, 2017). In this case, it is therefore necessary to assess each of the possible combinations. When grouping a human with two virtual agents, however, it is possible to standardize the assessment as the virtual agents will follow a set script (Rosen & Tager, 2013). To compare these two cases, two versions of the assessment tool were developed. The number of members in the group, the dialogue, and the problems are the same, regardless of whether or not the group included virtual agents. This ensures that the two systems are coherent and equivalent. Appendix D shows an example of how a virtual agent is used for skill C1 from Table 8 and Appendix C describes how the activity works.

Consequently, certain aspects of the interface also had to be modified based on the usability report from the first iteration and the decision to switch to pre-defined messages (Figure 2-4, Form A and Figure 2-5, Form B).

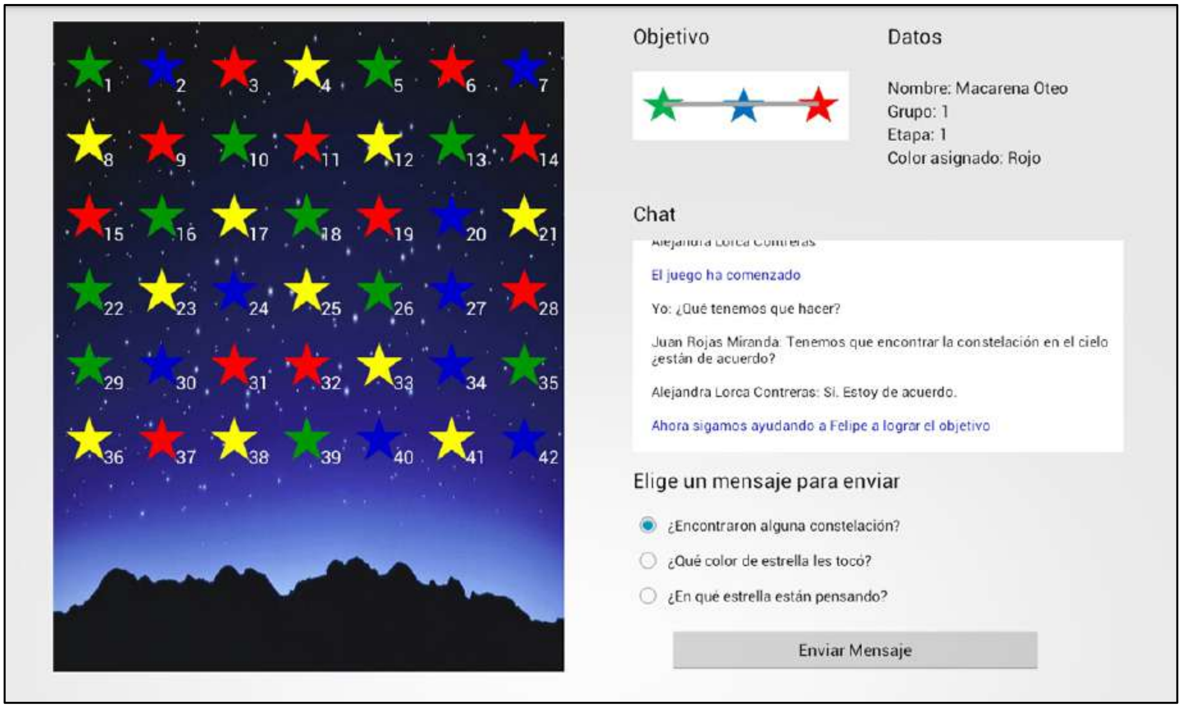


Figure 2-4: Form A

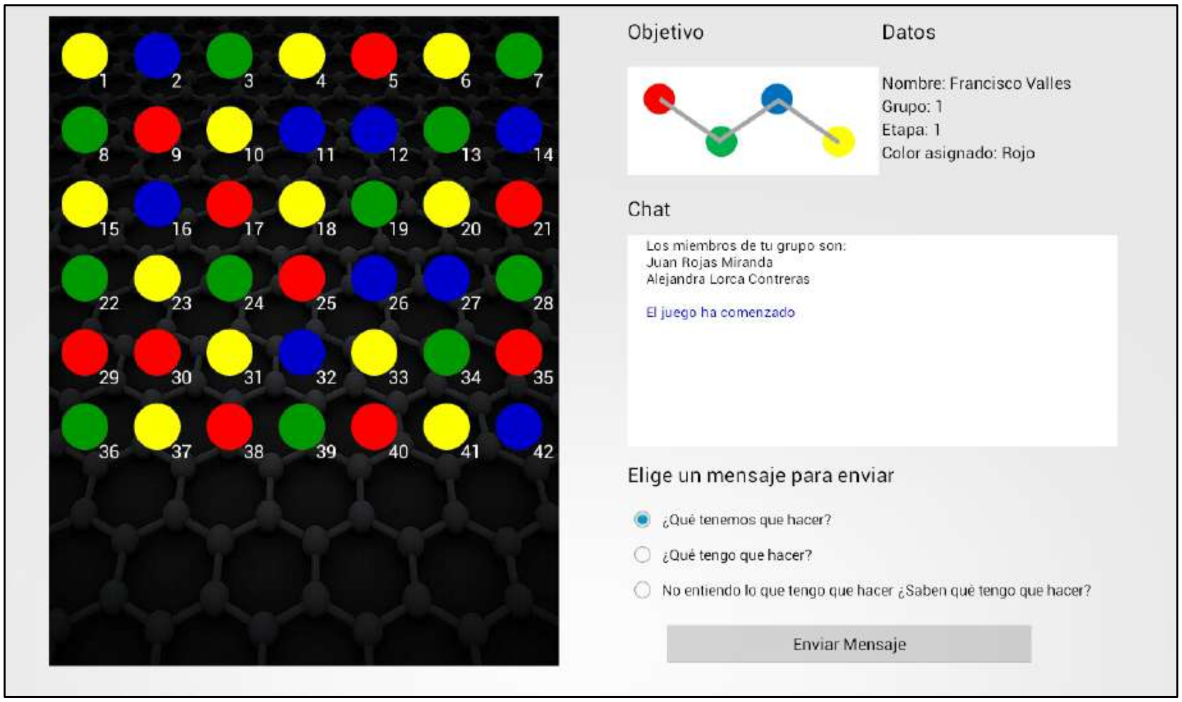


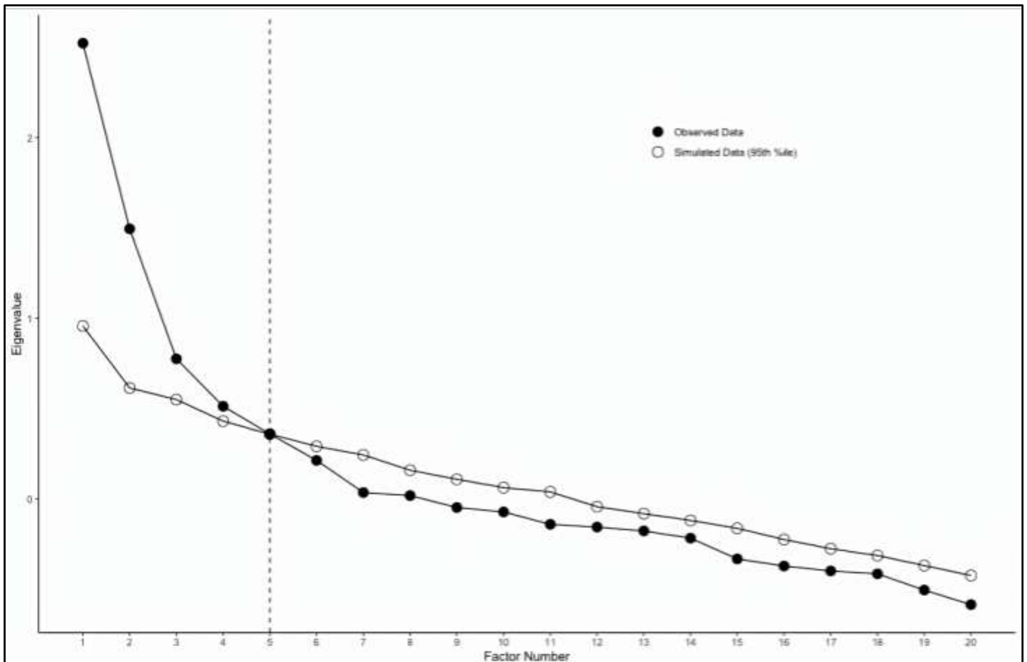
Figure 2-5: Form B

At the end of the second iteration the results were validated. The results of the Exploratory Factor Analysis were as follows:

Table 2-5: Communalities of the Items Analyzed in Iteration 2

Item	h2	u2	Item	h2	u2	Item	h2	u2	Item	h2	u2
1	.31	.69	6	.09	.91	11	.71	.29	16	.28	.72
2	.15	.85	7	.32	.68	12	.50	.50	17	.34	.66
3	.10	.90	8	.47	.53	13	.49	.51	18	.28	.72
4	.18	.82	9	.35	.65	14	.28	.72	19	.32	.68
5	.04	.96	10	.67	.33	15	.27	.73	20	.07	.93

Note. h2 (communalities) represent the proportion of the variability and u2 (uniquenesses) corresponds to the proportion of variability, which cannot be explained by a linear combination of the factors. Another way to calculate the communality is to subtract the uniquenesses from 1.



Note: The vertical dashed line indicates the number of factors that the parallel analysis suggested should be retained.

Figure 2-6: Parallel analysis scree plot

The communality was very low (<.10) for items 5, 6, and 20 in the second iteration (Table 2-5). These items were therefore removed from the analysis. Following this, the results of the parallel analysis suggested that the number of factors was 5 (Figure 2-6). Subsequent exploratory factor analysis allowed us to obtain the 5 factors presented in Table 2-1. which explain 31% of the total variance with an average reliability of .678. These factors mainly relate to the cognitive component of collaborative problem-solving and the target (goal) level, which increases in difficulty as the activity goes on. However, the factors did not perfectly match any particular theory. Instead, they hinted at the kind of models that we subsequently decided to confirm in the third iteration. This result is explained by other factors that are not well represented by the items used. This includes the social component of collaborative problem solving, such as Self-Regulation, Attitude towards collaboration, and Participation, among others.

Table 2-6: Exploratory Factor Analysis for Iteration 2

Item	Factors					CPS Skill	Level	McDonald's ω	
	F1	F2	F3	F4	F5			Dimension	Total
10	**.82**					C1	2	.839	.678
11	**.82**					C2	2		
12	**.65**					C3	2		
14		**.47**				D1	2	.636	
15		**.52**				D2	2		
16		**.52**				C1	3		
17		**.56**				C3	3		
18		**.48**				C2	3		
7			**.56**			Outcome	1		
13			**.68**			Outcome	2	.578	
19			**.47**			Outcome	3		
8				**.64**		D1	1	.583	
9				**.59**		D2	1		
1					**.54**	A2	0		
2					**.35**	A1	0	.444	
3					**.32**	A3	0		
4					**.40**	B1	0		

*Note***.** The column *CPS Skill* represents the collaborative problem-solving skills taken from the OECD framework; The column *Level* represents the sub-task levels of the activity.

The results of the measurement invariance test between Form A and Form B are presented below (Table 2-7).

Table 2-7: Fit Indices for the Assessment of Measurement Invariance of the CPS Assessment Across Forms A and B

	χ^2	*df*	CFI	RMSEA	Δ*df*	Δp-value	ΔCFI	$\Delta\chi^2$	ΔRMSEA
Configural	348.90	218	.834	.066	-	-	-	-	-
Weak	361.29	230	.826	.066	12	.09861	.008	186.015	.000
Strong	372.39	242	.835	.062	12	.76128	.009	82.999	.003
Strict	383.90	259	.838	.059	17	.72837	.003	131.162	.003
Mean	391.26	264	.835	.060	5	.13156	.004	84.828	.000

Note. χ^2 = Chi-Square; *df* = degrees of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; Δ*df* = degrees of freedom difference; ΔCFI = CFI difference; $\Delta\chi^2$ Chi-square difference; ΔRMSEA = RMSEA difference.

Configural invariance was established based on the overall model fit indices, which revealed an acceptable fit ($\chi^2(218)$ = 348.90; CFI= .834; RMSEA = .066). We then established that the magnitude of the loadings is similar across the groups (weak invariance) by comparing the fit of the weak model with the fit of the configural model, which revealed no significant differences (Δp-value = .09861). The differences for the CFI (ΔCFI = .008) and RMSEA (ΔRMSEA = .000) are less than the proposed cutoff of .010 and .015, respectively (Cheung & Rensvold, 2002). Similarly, the results showed that the intercepts (Strong) and residual variances (Strict) are similar across the groups.

Therefore, the proposed instrument for assessing Collaborative Problem-Solving Skills is equivalent in both its forms (A and B).

Next, the results of the measurement invariance test between boys and girls are presented below (Table 2-8).

Table 2-8: Fit Indices for the Assessment of Measurement Invariance of the CPS Assessment Across Gender

	χ^2	*df*	CFI	RMSEA	Δ*df*	Δvalor-p	ΔCFI	$\Delta\chi^2$	ΔRMSEA
Configural	320.14	218	.867	.058	-	-	-	-	-
Weak	329.63	230	.871	.056	12	.6967	.004	90.729	.002
Strong	336.27	242	.878	.053	12	.8824	.007	66.073	.003
Strict	352.34	259	.872	.052	17	.2173	.006	212.031	.001
Mean	353.48	264	.877	.051	5	.9492	.005	11.538	.001

Note. ;CFI = comparative fit index; RMSEA = root mean square error of approximation; Δ*df* = degrees of freedom difference; ΔCFI = CFI difference; $\Delta\chi^2$ Chi-square difference; ΔRMSEA = RMSEA difference.

Configural invariance was established based on the overall model fit indices, which revealed an acceptable fit ($\chi^2(218)$ = 320.14; CFI= .867; RMSEA = .058). We then established that the magnitude of the loadings is similar across the groups (weak invariance) by comparing the fit of the weak model with the fit of the configural model, which revealed no significant differences (Δp-value = .6967). The differences for the CFI (ΔCFI = .004) and RMSEA (ΔRMSEA = .002) are less than the proposed cutoff of .010 and .015, respectively (Cheung & Rensvold, 2002). Similarly, the results showed that the intercepts (Strong) and residual variances (Strict) are similar across the groups.

Therefore, the proposed instrument for assessing Collaborative Problem-Solving Skills is equivalent for boys and girls.

Unfortunately, because only 40 cases (120 students) were obtained for the human-human version of the instrument, the invariance measurement test between human-human groups and human-agent groups did not converge. However, it was possible to calculate the average score for each factor for each student (Table 2-9). By comparing the students' scores on each version of the assessment, we observed that there were no significant differences between the human-human group and the human-agent group in terms of their scores on Form A of the test. This is true when comparing the set of factors obtained from the confirmatory factor analysis in a MANOVA test (wilks= .98; F(6,295) = 1.28; p=.26), as well as when comparing each factor individually with an ANOVA test (Table 2-9).

Table 2-9: MANOVA and ANOVA by Factor

	Form		Gender		Group composition	
	A	B	Girls	Boys	HM	HH
Factor	*M*(*SD*)	*M*(*SD*)	*M*(*SD*)	*M*(*SD*)	*M*(*SD*)	*M*(*SD*)
F1	1.22(.99)	1.15(1,03)	1.26(1.02)	1.26(0.94)	1.43(1.04)	1.75(0.52)
F2	1.25(.70)	1.27(.65)	1.25(0.64)	1.29(0.67)	1.37(0.49)	1.40(0.61)
F3	1.43(.97)*	1.14(.92)*	1.40(0.95)	1.28(0.96)	1.50(0.92)	1.65(0.88)
F4	1.15(.95)	1.13(.95)	1.16(0.95)	1.20(0.93)	1.46(0.86)	1.20(0.99)
F5	1.78(.55)	1.64(.70)	1.71(0.65)	1.72(0.55)	1.68(0.47)*	1.93(0.49)*
All	Wilks= .97; F(5,296) = 2.08; p=.068		Wilks=.994; F(5,336)=.41; p=.843		Wilks=.881; F(5,74)=2.00; p=.088	

Note. (*) Sig. < .05; M=Mean; SD=Standard Deviation; F1, F2, F3, F4 and F5 are the factors found in exploratory factor analysis in Table 5

Moreover, there is a significant difference in the time it took the human-human group and the Human-Agent group to complete Form A of the test ($F(1,77)=68.08$; $p<.001$) (Table 2-10), with students in the Human-Agent group taking significantly less time to do so.

Table 2-10: Comparing by Duration

	Human-Human		Human-Agent		ANOVA		
Variable	*M*	*SD*	*M*	*SD*	*F*	df	*p*-value
Time	38.71	7.57	23.24	9.04	68.08	1, 78	<.001

Note. M: Mean; SD=Standard Deviation

Finally, we observe that some of the participants claimed that they did not always understand what they needed to do in the activity and that it was sometimes quite monotonous. Furthermore, although they were constantly reminded to communicate using only the chat, it was difficult for a lot of the students to refrain from making gestures or talking to their classmates. In conclusion, it was important to improve student engagement with the activity, change how the agents were named to prevent students from being distracted trying to figure out who the group members are, and improve the feedback given to the students.

2.3.3. Third Iteration: Final Solution

The final solution continued using pre-defined messages and Human-Agent groups as the main setting. Moreover, this version included a new interface design (Figure 2-7, Form A, and Figure 2-8, Form B), with a new narrative to boost student engagement (for full details of the design of the activities see Appendix I). Finally, feedback was included after each stage of the activity so that the students could see their progress, while the virtual agents were given fictitious names.

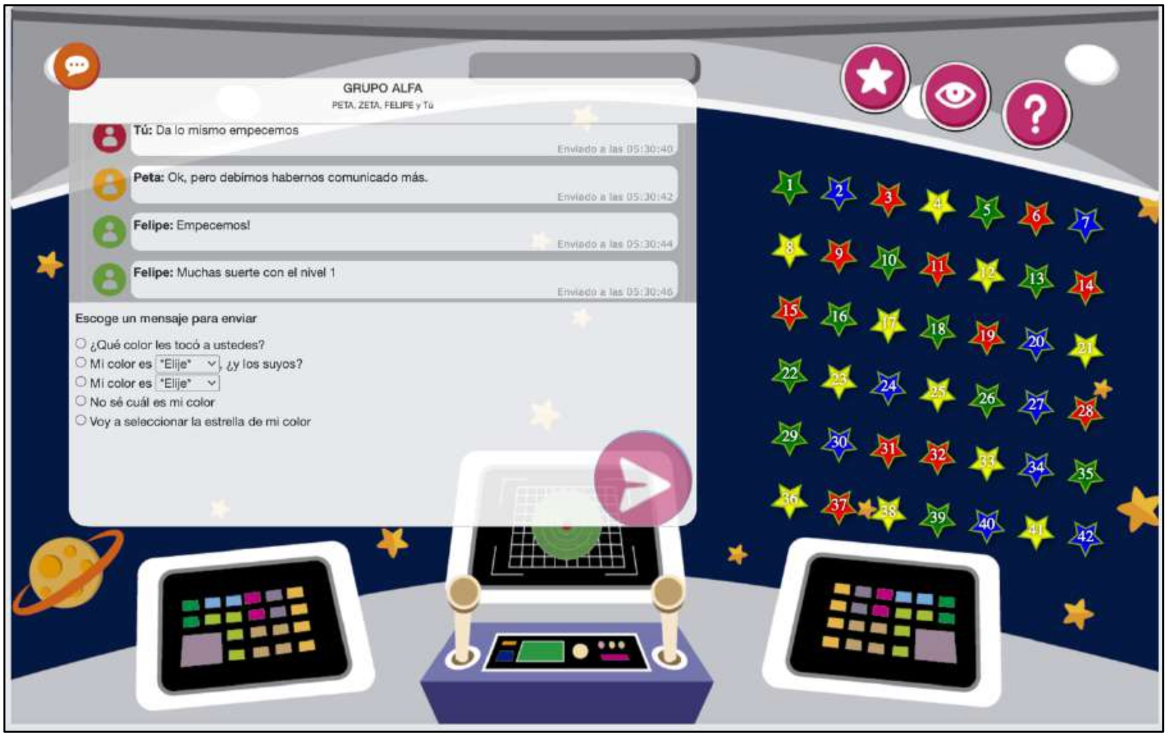


Figure 2-7: Form A

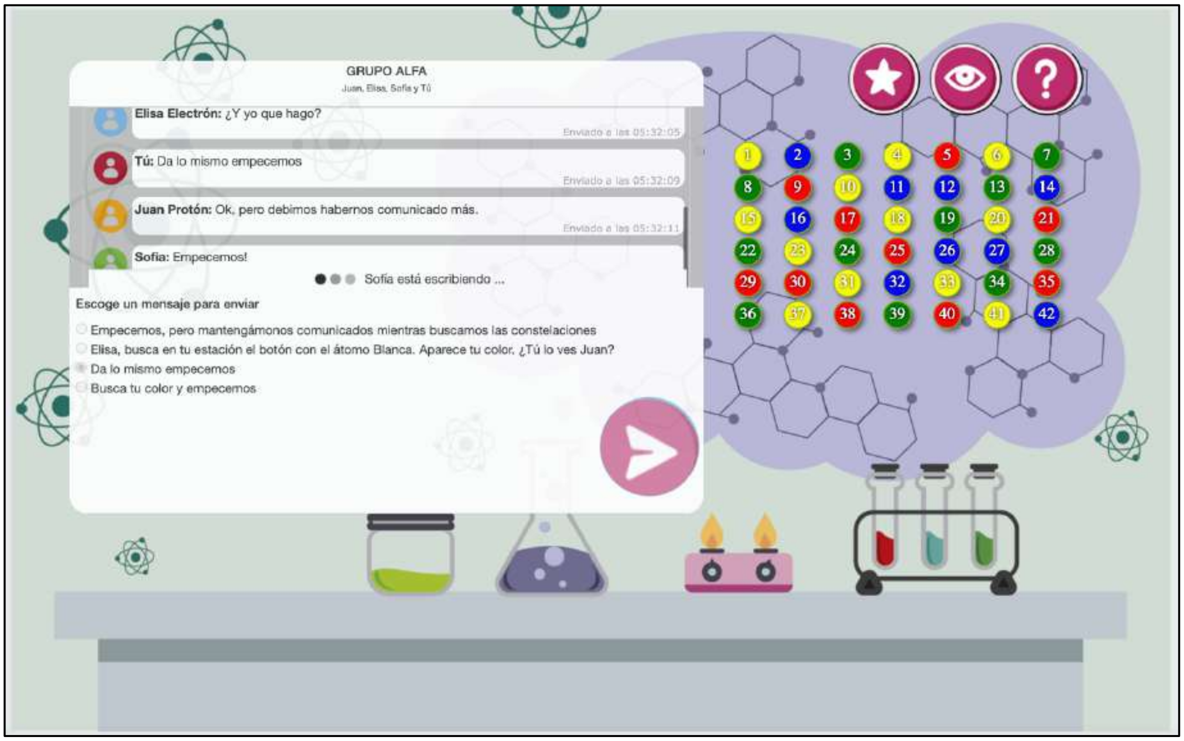


Figure 2-8: Form B

Using this version, and based on the results of the Exploratory Factor Analysis (EFA), 5 models were assessed in order to confirm the instrument's internal structure (Figures 2-9, 2-10. 2-11, 2-12 and 2-13).

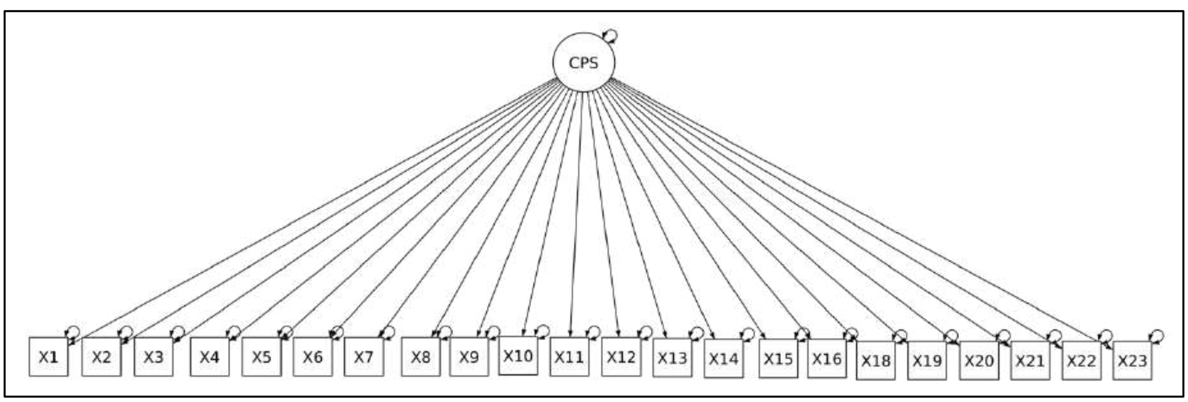


Figure 2-9: Model 1 grouped all the items of the instrument into a single factor

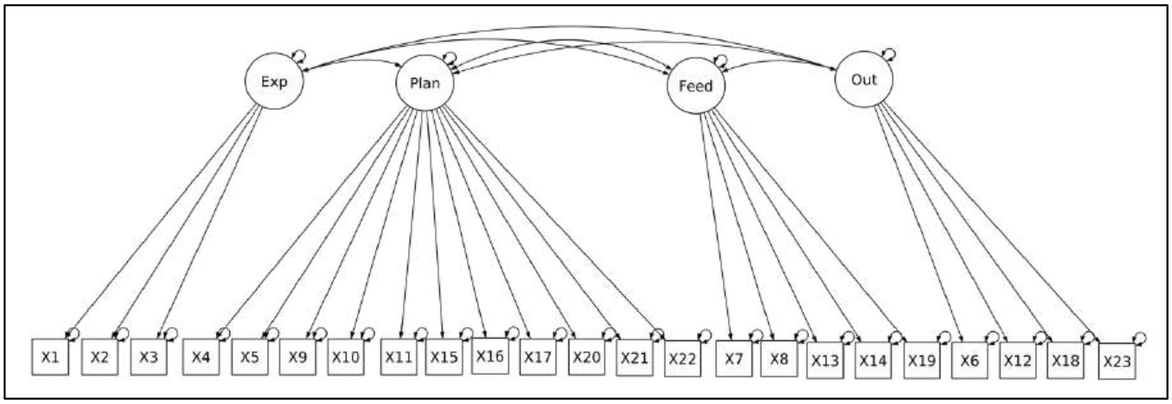


Figure 2-10: Model 2 grouped the items into 4 factors, corresponding to the cognitive processes in individual problem solving.

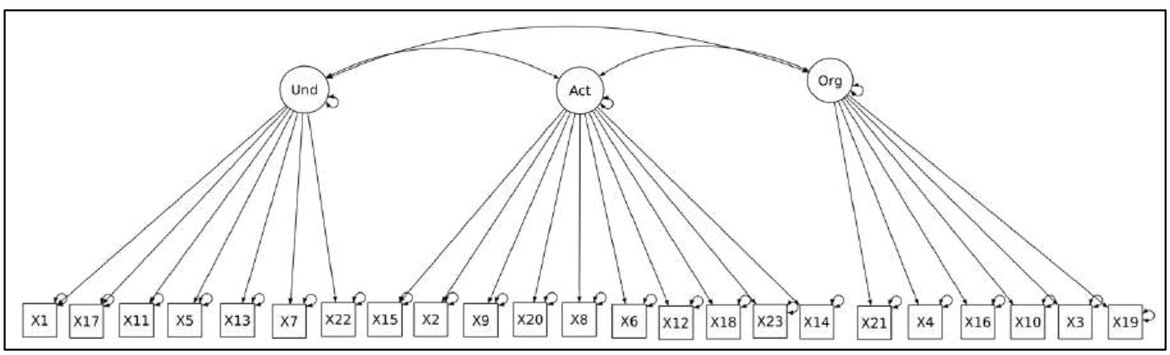


Figure 2-11: Model 3 grouped the items into 3 factors, corresponding to the 3 collaborative competences

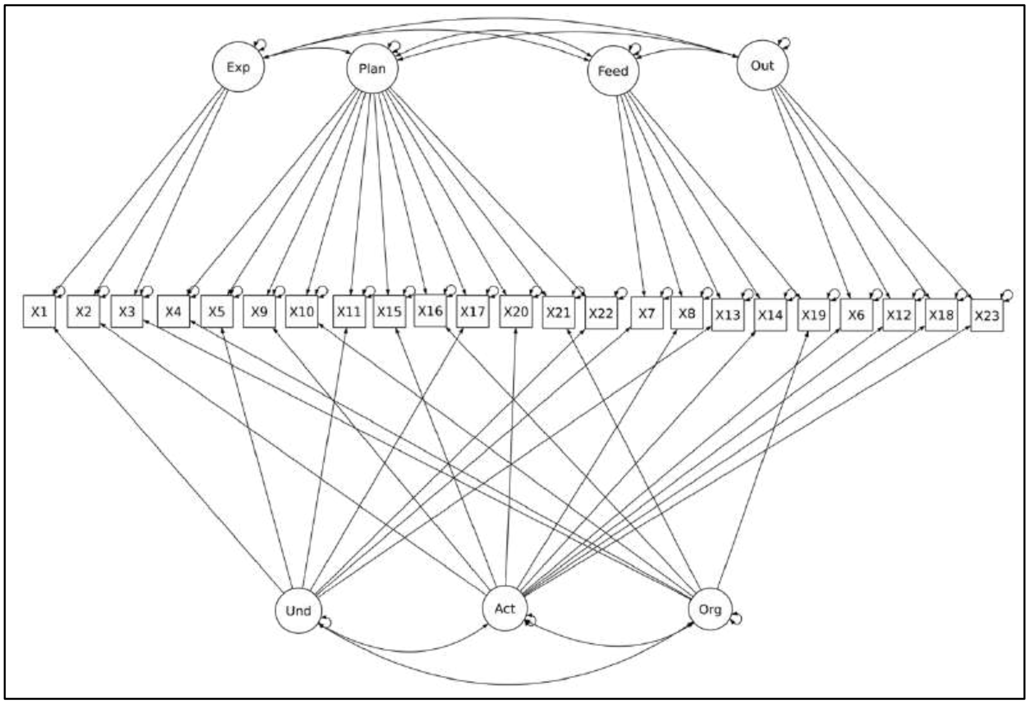


Figure 2-12: Model 4 grouped the items into 7 factors, corresponding to the 4 factors of model 2 and the 3 factors of model 3

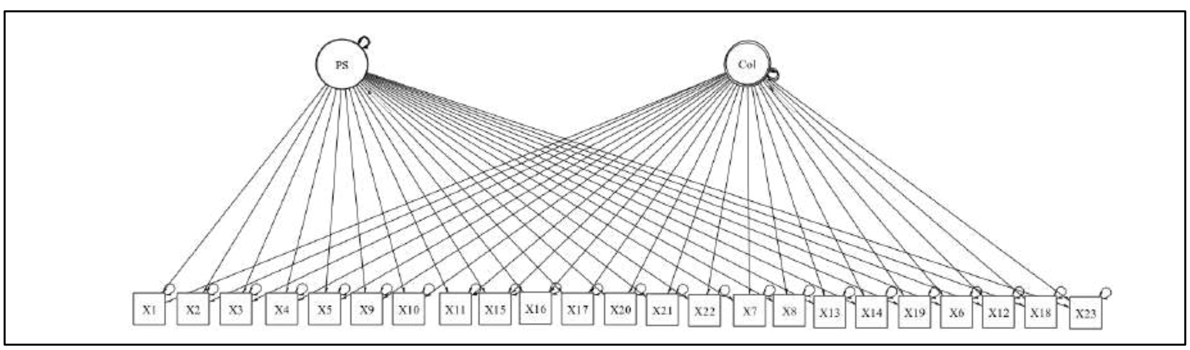


Figure 2-13: Model 5 grouped the items into 2 factors, corresponding to a problem-solving factor (PS) and a collaboration competencies factor (Col).

Model 4 was unable to converge on a solution, while Model 1 did not return acceptable fit indices (Table 2-11). Similarly, Model 3 displayed a poor fit ($\chi2$ (149) =

213.22, $p<.001$; CFI=.83; TLI=.81; RMSEA = .05, CI[.03, .06]; SRMR=.07) based on the cutoff criteria described in the methodology (section 2.2.5).

Table 2-11: Results of the Confirmatory Factor Analyses

Model	χ^2	*df*	*p*-value	CFI	TLI	RMSEA	SRMR	AIC	BIC
Model 1	405.09	230	<.001	.63	.60	.06 [.03, .06]	.09	11694.2	11842.8
Model 2	238.51	224	.24	.97	.97	.02 [.0. .036]	.06	11539.0	11707.0
Model 3	213.22	149	<.001	.83	.81	.05 [.03, .06]	.07	9195.2	9328.4
Model 4				No convergence					
Model 5	250.021	206	.02	.91	.87	.03 [.03, .06]	.06	11694.2	11842.8

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation with 90% confidence interval; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual.

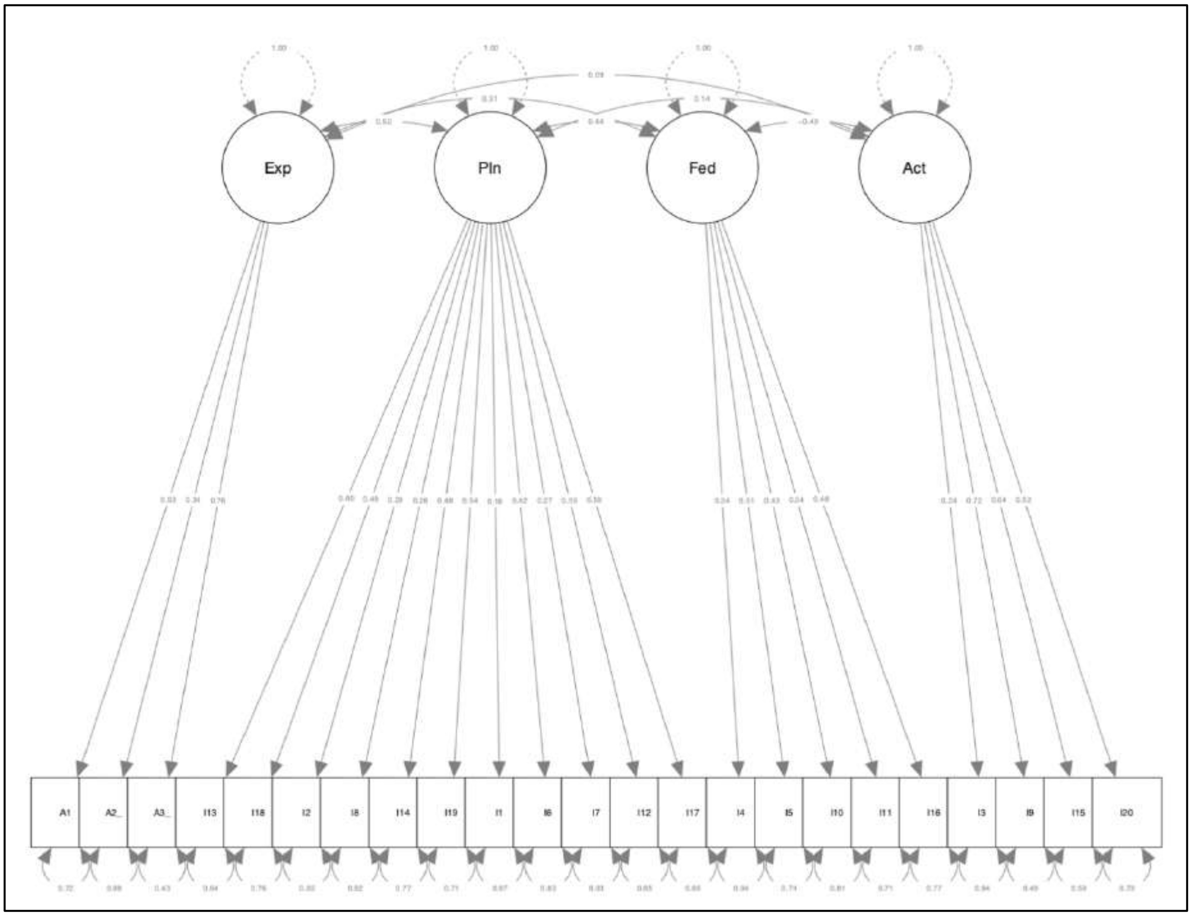


Figure 2-14: Model 2 for collaborative problem solving, where Exp = Exploring and representing the problem; Pln = Planning and executing the problem; Fed = Monitoring and reflecting; Act = Outcomes or results of solving the problem ; and i1, i2..i20 are the items that make up each dimension.

Model 2 displayed an excellent fit based on the cutoff criteria described in the methodology (section 2.2.5), ($\chi 2(224) = 238.51$, p=.24; CFI=.97; TLI=.97; RMSEA = .019, CI[.. .036]; SRMR=.062) (Table 2-11). This corresponds to the four-factor model representing the four cognitive processes involved in individual problem solving. Model 5 also showed an acceptable fit ($\chi 2$ (206) = 250.021, p=.02; CFI=.91; TLI=.87; RMSEA = .03, CI[.03, .06]; SRMR=.06).

As models 2 and 5 were non-nested, comparing them using the chi-square test was not recommended (Bentler, 1990). Nevertheless, the fit indices were better in Model 2. Additionally, the AIC and BIC obtained were lower for Model 2 (Table 2-11). We therefore considered Model 2 to be the better of the two.

Furthermore, the reliability results for Model 2 are presented in the following table (Table 2-12). Although the general reliability obtained for the scores (**ω**=.726) is at the limit of acceptance for studies in the field (Taber, 2018), the results in each dimension were low. However, this is related to the number of items that were included in each factor (Eisinga, Te Grotenhuis & Pelzer, 2013).

Table 2-12: Reliability Results for Model 2

Model	Dimensions	Items	McDonald's ω
Model 2	**Problem-solving**	**22**	**.726**
	Exploring and Representing	**3**	**.593**
	Planning and Executing	**11**	**.717**
	Monitoring and Reflecting	**5**	**.551**
	Outcomes	**4**	**.621**

Note. The column *Items* represents the number of items for each dimension.

2.4. Discussion

2.4.1. H1 DBR allows for the development of an instrument for assessing collaborative problem-solving skills among elementary school students with two equivalent forms.

We developed a new tool for assessing collaborative problem solving (Section 3) through an iterative and incremental design process. Using the DBR approach (Section 2.2) not only allowed us to discover important findings (Section 3), but also to describe in detail how the assessment tool was developed (Land & Zimmerman, 2015).

The DBR process allowed us to validate the collaborative properties, the design, and the usability of the activity. Certain problems were detected by conducting an authentic collaborative problem-solving assessment. This includes the difficulty of coding student conversations using the OECD framework, given that the framework was designed for assessing students individually and not as a group. It also highlights the advantage of using virtual agents, as it prompts assessments that otherwise may not occur for every participant in a human-human group (Graesser et al., 2018). Furthermore, by using open chat for the activity, the students use a more functional style of language and opt for trying a solution and communicating what they have done. This is instead of following a problem-solving process with clearly-defined steps (i.e., exploring, planning, etc.) that can be guided by the inclusion of a virtual agent. Finally, any assessment that requires manual scoring is obviously a challenge when it comes to scalability (Intermediate problem analysis, Section 2.2.4). This coincides with the considerations presented by the OECD regarding the underlying theory that guided the development of the PISA assessment (OECD, 2017; Nouri et al., 2017). Despite previous studies comparing human-human and human-agents groups (Stadler, Herborn, Mustafić & Greiff, 2020; Rosen & Tager, 2013), there is still little evidence on the advantages and disadvantages of assessing via the use of open chat using the OECD framework in human-human activities.

Nevertheless, some authors still use human-human assessment as they find it to be more realistic and to allow for the assessment of a large range of individual variables and social processes (Graesser et al., 2018; Care, Scoular, & Griffin, 2016). The decision to use virtual agents in this study was thus based on the end-goal of the solution. In this case, if the aim is to develop a scalable, standardized assessment tool, then facilitating the collection and correction of responses is key. On the other hand, if the objective were to teach or learn (i.e., more formative than summative assessment), then an authentic environment would be more appropriate (Sun et al., 2020).

Finally, this iterative process provided us with the possibility to progressively improve the assessment experience. This is especially important with elementary school students since the motivational factor seems to be particularly relevant. Both the interface and the dialogue with the virtual agents benefitted from the feedback obtained in each test in the context of a real classroom.

2.4.2. H2 The instrument that is developed measures both the problem-solving and collaborative dimensions of CPS among students aged between 10 and 13 years old.

The first iteration of the DBR process allowed us to study the psychometric characteristics of an initial version of the instrument. We found that the results obtained in the explanatory factor analysis were not acceptable, while also failing to show equivalence between both forms (Section 2.3.1). We therefore decided to redo the evaluation criteria in a second iteration.

The second iteration then allowed us to explore the internal structure of the assessment tool (Section 2.3.2), as well as validating the way in which collaborative problem solving was operationalized. This operationalization includes the table of specifications, the structure of the conversations, the way in which the conversations were evaluated, and how this related to the different scores given to assess collaborative problem solving. This level of detail on the internal structure of the assessment is unprecedented in the literature (Scoular & Care, 2020). Similar studies using the PISA framework have mainly referred to comparisons with other theoretical frameworks or assessments to provide validation (Herborn et al., 2018; Oliveri, Lawless & Molloy, 2017).

Finally, in the third iteration of this study, our results showed that our instrument predominantly focuses on the problem-solving component of collaborative problem-solving (Section 2.3.3). This is reflected in the fact that the items on our assessment tool are grouped together by factors, such as Exploring, Planning and Monitoring (Table 2-11), as opposed to being grouped by both problem-solving skills and collaborative competences. However, Model 5, which is a bifactor model that includes a factor for the problem-solving dimension and another factor for the collaborative dimension, obtained a relatively acceptable fit. Additional research into this model is therefore suggested. As highlighted previously, several authors have already drawn attention to the heavy cognitive focus of the OECD framework (Sun et al., 2020; Nouri et al., 2017; He et al., 2017; Scoular & Care, 2020). Our study goes beyond this by providing a better understanding of what is being assessed, revealing the tool's assessment capacity and the latent variables that it measures (Figure 2-14). Additionally, we report the results for a population of students previously not included, i.e., elementary school students aged between 10 and 13 years old.

2.4.3. H3 The scores obtained for the two forms of the instrument are equivalent

We were also able to develop two equivalent forms of the instrument (A and B, Figure 2-4 & 2-5), which delivered statistically equivalent results when assessing collaborative problem-solving skills (Table 2-7). This opens up the largely untapped potential of working with experimental designs that assess collaborative problem solving before and after an intervention, a need that has been reported in the literature (Graesser et al., 2018).

2.4.4. H4 The scores obtained are equivalent across genders

In previous research, such as PISA 2015 (OECD, 2017), significant differences were reported between boys and girls. Our invariance measurement test showed that the instrument is equivalent when assessing girls and boys (Table 2-8). This allowed us to obtain an instrument without gender bias. The characteristics (age and country of residence) of our participants may explain the difference between our results and results of previous studies. This is because the students in our sample were younger and from a specific demographic.

As a result, it is necessary to extend the sample to different ages and demographics to generalize our findings.

2.4.5. H5 There is no significant difference when assessing collaborative problem solving in human-human groups versus human-agent groups using the proposed instrument

Finally, our results also confirm that when using this framework there are no significant differences between human-human and human-agent groups (Table 2-9). Although this has been proven before (Herborn et al., 2018; Rosen & Tager, 2013; Rosen, Wolf, & Stoeffler, 2019), our results also show that students working with virtual agents take significantly less time to complete group work (Table 2-9). There are also other advantages of using the human-agent version of the activity in the classroom. This includes the ease with which the activity can be modified, as tailoring the dialogue with virtual agents is much easier that tailoring the dialogue and actions of three humans. Furthermore, working with the human-agent version does not require any connection between devices

2.5. Conclusions

The present study reports on the design and validation of an assessment tool for measuring students' collaborative problem-solving skills.

Our first research question asked "How can we assess collaborative problem-solving skills among elementary school students using an instrument with two equivalents forms?" To answer this question, we proposed an operationalization to assess collaborative problem-solving skills based on the PISA framework (OECD, 2017).

Following this, our second research question asked "Which design principles will help develop a tool that facilitates the assessment of collaborative problem-solving skills?" To answer this question, we followed an iterative DBR process to develop an assessment tool and a collaborative activity that fulfilled the conditions for collaborative activities described in the literature (Szewkis et al., 2011). In each iteration, we analyzed the design aspects of the activity and the psychometric aspects of the instrument to contrast five hypotheses.

We explored the internal consistency of the instrument over the three iterations. We considered five models to confirm the validity of the test. The results revealed that while the tool measured the cognitive dimensions of the collaborative problem-solving process, this was not the case for the social aspects of collaboration (Table 2-11). This leads us to confirm that the assessment tool, which was developed based on the OECD framework, only partially measures the social component of CPS, even though collaboration is considered a fundamental part of CPS (Andrews-Todd & Forsyth, 2018; Care, Scoular, & Griffin, 2016).

On the other hand, we described in detail how to develop two equivalent forms of our CPS assessment (Section 2.3), as well as a set of indicators for measuring the 12 skills involved in CPS (Table 2-4). Furthermore, two versions of the tool were designed: one in which the students worked among themselves (human-human); the other in which the students worked with two virtual agents (human-agent). We found that there were no significant differences between the two versions, a finding that is in line with previous studies (Herborn et al., 2018; Rosen & Tager, 2013; Rosen, 2014). However, it was also found that students take significantly longer to complete the activity in a human-human group, while there were no significant differences based on gender.

The main contributions of this study are the findings regarding the internal validation of our instrument and its ability to measure collaborative problem-solving skills among students aged between 10 to 13 years old. This shows that it is possible to measure the different stages of the problem-solving process (explore, represent, plan and monitor) with two equivalent forms of a test. This, in turn, allows us to study the impact of developing students' collaborative skills. The study also provides evidence in support of using virtual agents during assessments. Our findings reveal that while student performances in a human-human or human-agent setting are similar, the use of virtual agents can provide administrative and/or logistical advantages. Finally, one final contribution was to present evidence of the validation of an instrument for assessing CPS in a previously understudied population, such as elementary school students.

One of the limitations of the tool proposed in this study is its reliance on text-based communication. In this case, important information on the students' collaborative problem-solving skills is lost, such as details of their non-verbal communication (Chopade et al., 2018). Furthermore, students do not have complete freedom when interacting with each other as the messages used when solving the problem and communicating with one another are pre-defined. In this respect, recent research, such as the study by Chopade et al. (2018) suggests that an effective framework for assessing CPS should draw from different data sources, as well as exploring other statistical methods (Swiecki, Ruis, Farrell & Shaffer, 2020). Another limitation is that even though we developed a completely new instrument based on the OECD framework, we were not able to compare this with other instruments, e.g. the PISA assessment. This is because we did not use the same activities. Thus, both the aim of the instrument (i.e., focusing on the individual level of CPS skills) and the sample (i.e., students aged between 10 and 13 years old) were different.

Finally, future work should look to use more advanced technology involving artificial intelligence, such as image and voice recognition software, among others. Using such technology would allow us to capture a wider range of student interactions, as well as developing more challenging and realistic problems. This is particularly important as the aim of CPS is to solve problems that are individually complex (Graesser, Kuo & Liao, 2017). Further work with virtual agents is also recommended as they can facilitate standardized and scalable assessments. However, more work is needed to improve the authenticity of such an activity and therefore provide a more realistic experience (Nouri et al., 2017), thus fostering collaboration.

3 USING AUTOMATED PLANNING TO PROVIDE FEEDBACK DURING COLLABORATIVE PROBLEM-SOLVING

3.1. Introduction

In recent years, educational and curricular reform has focused on the incorporation of new skills, including critical thinking, problem solving, self-regulation, information technology, communication, and collaboration (Andrews-Todd & Forsyth, 2018; Griffin, McGaw, & Care, 2012). Within these skills, collaborative problem solving is essential (Sun, Shute, Stewart, Yonehiro, Duran & D'Mello, 2020). This is because the problems facing society are becoming increasingly complex and require greater expertise in order to be solved (Care & Griffin, 2014). However, such expertise is often distributed across a number of people, making collaboration essential if an effective solution is to be found (Andrews-Todd & Forsyth, 2020; Scoular, Care & Awwal, 2017). In this sense, previous studies have made a significant contribution to the assessment of collaborative problem-solving skills (Scoular & Care, 2020; OECD, 2017; Herborn, Stadler, Mustafić, & Greiff, 2020). However, research into methods for teaching these skills remains relatively unexplored and highly necessary (Graesser, Fiore, Greiff, Andrews-Todd, Foltz & Hesse, 2018; Rosen, Wolf & Stoeffler, 2020).

In 2015, the OECD assessed collaborative problem-solving skills among more than 50.000 fifteen-year-old students from different countries. The results showed that only 8% of them achieved the highest level, while 29% scored at the lowest end of the scale (OECD, 2017). Since the publication of this report, interest in including collaborative problem solving in educational programs across the world has increased (Stadler, Herborn, Mustafić & Greiff, 2020). However, developing these skills among students requires methodologies that are capable of considering their complex social and cognitive composition. Mechanisms are also needed to provide scaffolding to enable students to progress through their zone of proximal development and take their CPS skills to the next level (Graesser et al., 2018).

One particularly promising approach is the use of videogames for developing 21st century skills (Qian & Clark, 2016; Hewett, Zeng & Pletcher, 2020). Due to its ability to motivate and increase student awareness, teach knowledge, change behaviours, and even improve skills, the use of game-based learning has generated significant interest (Calvo-

Morata, Alonso-Fernández, Freire, Martínez-Ortiz, & Fernández-Manjón, 2020). En este sentido, como menciona Taub, Sawyer, Lester & Azevedo (2020). los juegos pueden propiciar un efecto positivo en el aprendizaje, gracias a la incorporación de componentes que fomentan el compromiso afectivo, conductual, cognitivo y social (Plass et al, 2015)
Además, las tecnologías de la información y la comunicación, permiten construir contextos complejos de aprendizaje y evaluación, que facilitan a los estudiantes desarrollar habilidades del siglo XXI (Qian y Clark 2016). ya que brindan entornos integradores, interactivos y aproximados de resolución de problemas de la vida real (Shute y Becker 2010; Chen, Cui, & Chu, 2020).

Por otra parte, feedback between peers has been shown to be uno de los factores más importantes que influyen in the teaching-learning process with these kinds of methodologies (Plass, Homer & Kinzer, 2015; Nadolny, et al., 2020; Hattie y Gan 2011), ya que, como indican Goldin, Narciss, Foltz & Bauer (2017), la retroalimentación formativa ayuda a los estudiantes que identifican cuáles son sus metas y cómo alcanzarla, a la vez que, monitorean y comprenden dónde se encuentran en un proceso de aprendizaje. Por tanto, feedback can be understood as the information provided by an agent (e.g. a teacher, a peer, a book, etc.) on the performance, comprehension or behavior of an individual (Hattie & Timperley, 2007). It is therefore a tool that is widely used in the teaching of both content and skills (Erhel & Jamet, 2013; Zhu, Liu & Lee, 2020).
Desde el punto de vista tecnológico, un sistema de retroalimentación rudimentario tiene una estructura básica en la que el alumno proporciona una respuesta a un problema planteado (input) y el sistema retorna una retroalimentación (output). En contraste, un sistema más avanzado (Inteligente), es aquel que adapta la retroalimentación que proporciona al alumno, en función de su adquisición de conocimientos y desempeño en tiempo real (Gilbert et al, 2018).

Although the aim of feedback is to move the student closer to their objective within a given activity or task, it does not always work out this way. There is evidence to suggest that feedback does not always improve learning and may even hinder it. The factors that can influence the effectiveness of feedback include content, timing, and presentation, among others (O'Donovan, den Outer, Price & Lloyd, 2019). A su vez, Gilbert et al (2020), destaca que uno de los principales desafíos de la efectividad de la retroalimentación eb tareas grupales, es que el contenido de la retroalimentación sea atendido por al menos un miembro del grupo (Nadler 1979). Por tanto, Debido a que cuando feedback has to be given on the collaborative participation of multiple agents, addressing this need in real time becomes even more complex. Our first research question therefore asks: "How can feedback be given effectively to the participants in a collaborative game when interacting simultaneously to solve a problem?"

To answer this question, the present study focuses specifically on developing a game-based learning platform with an automated system to provide such feedback in a collaborative context. Automated planning technology was used to determine the feedback that was given (Ghallab, Nau, Traverso, 2004). This involves building a system that receives a model of the problem to be solved by the players. This is described using Planning Domain Definition Language (PDDL), the standard language for describing planning problems.

Using this model, the system generates a plan containing the actions that must be completed by the players. During the game, the players carry out actions and, based on the plan that is generated, the system assesses whether or not they are making progress. When the system detects a lack of progress, it sends feedback to the players with suggestions of what they need to do in order progress. This feedback is taken from the plan and requires some form of collaboration from the students.

Furthermore, as the game goes on, it is essential to monitor adherence to the plan. Doing so can allow the system to assess the students' progress, as well as detecting whether they have reached a point in the game at which the plan becomes invalid. This happens when players reach a point at which no part of the plan can be used to help them meet their objective. In order to detect whether players have reached this point we use *plan execution monitoring* (Ghallab, Nau, Traverso, 2004).

The approach proposed by Muise et al. (2011) is an example of the state of the art for the execution monitoring of partial-order plans. Unfortunately, this general approach may require exponential running time for the whole plan. However, we observe that our plans are usually composed of 'independent chains'. We exploit this characteristic and propose a new plan-monitoring algorithm that can be substantially more efficient than Muise's. This is beneficial for our application since it allows the system to be implemented using inexpensive CPUs.

Finally, based on the objectives and conditions described previously, our second research question asks: “How can we develop an effective Feedback System based on Automated Planning when the plans involve multiple agents in a real-world educational context?”

3.2. Material and Methods

3.2.1. Overview of the Architecture

Before building a solution that can answer our research questions, we must first understand the context in which said solution will be deployed. In this sense, the context is a collaborative game for 3 players, developed specifically for tablets (Appendix K). The PDDL domain of the game is presented in Appendix B and used to implement a feedback system. Given that the technology available in most high-school classrooms is usually not high-end, this game connects the players using a laptop computer as a server.

The general architecture of the feedback system can be seen in Figure 3-1. The server computes plans, while clients for each group of students are responsible for the videogame visualization and the monitoring of the plan. The monitoring system communicates with the server, requiring new plans when needed.

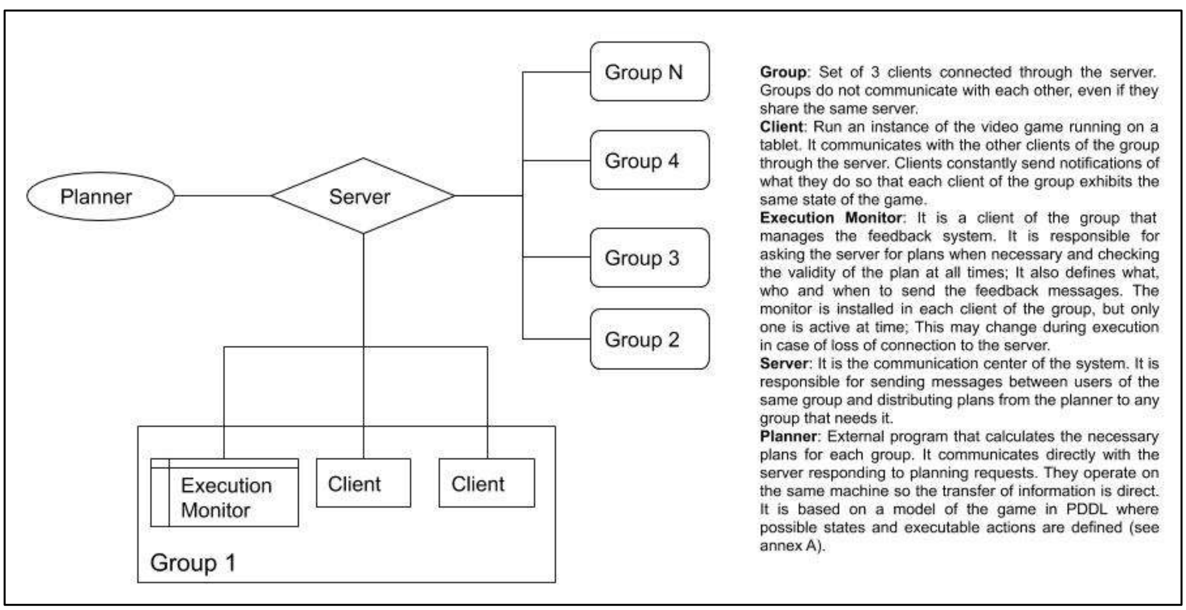


Figure 3-*1*: Diagram of the solution developed for this study

3.2.2. Planning and Plan Monitoring

A planning problem is defined as a tuple $\Pi = \langle F, A, I, G \rangle$, where F is a finite set of propositions (or facts), A is a set of actions, $I \subseteq F$ is the initial state, and $G \subseteq F$ is the goal. The state *s* is a subset of *F*. The facts contained within a state are considered *true* and those not contained within a state are considered *false*. An action $a \in A$ is defined by 3 sets. $PRE(a)$, the precondition, which contains the facts that must be true in order for action *a* to be executable; $ADD(a)$, facts that action *a* adds to the state when carried out; and $DEL(a)$, facts that action *a* deletes from the state when carried out (Kautz, McAllester & Selman, 1996). The resulting state from action *a* on *s* is defined as $\delta(s,a) = \big(s \setminus DEL(a)\big) \cup ADD(a)$. The state resulting from action *a* on *s* is defined as $\delta(s,a) = \big(s \setminus DEL(a)\big) \cup ADD(a)$. We extend the definition of δ for sequences of actions in the usual way. When a proposition p belongs to $ADD(a)$ we say that p is *added by* a. Furthermore, when p belongs to $DEL(a)$ we say that p is *deleted* by a.

An action a is defined as executable on a state s iff $PRE(a) \subseteq s$ (Muise et al., 2011). A sequence of actions $a\alpha$ is executable on a state *s* iff *a* is executable on *s* and *α* is executable on $\delta(s,a)$. Furthermore, we can define that if α is executable on s, then α achieves G from *s* iff α is executable on s and $G \subseteq \delta(s,\alpha)$. A sequential plan for Π is a sequence of actions α such that α achieves G from I. A *suffix* of a sequential plan $\alpha = [a_1, \dots, a_n]$ is the sequence $[a_i, \dots, a_n]$ where $1 \le i \le n+1$. The *prefix* of the plan is defined analogously (Muise et al., 2011).

If the environment in which the plan is executed can be affected by external events, the plans must be monitored. The aim of monitoring is to ensure that during execution the plan still allows the agent to reach the goal state to be reached from the current state (Epstein & Tripodi, 1977). We say that a sequential plan α is *valid* on a state s, iff there is a suffix of α that can achieve G from s. When the plan is not valid, the monitoring system takes measures such as fixing the plan or planning again from scratch in order to come up with a valid plan (Fritz & McIlraith, 2007).

There are several methods in the literature for carrying out this monitoring process (Epstein & Tripodi, 1977; Weld, 1994; Fritz & McIlraith, 2007). In this paper we use one of these methods, known as sequential monitoring. This is based on using the concept of *regression*, which ensures the minimal size of the resulting states. Given a problem $\Pi = \langle F, A, I, G \rangle$ and a set of facts, ψ, expressed as a set of facts, the regression of ψ is defined in terms of an action a, denoted as $R(\psi, a)$, as follows: $R[\psi, a] = (\psi \setminus ADD(a)) \cup PRE(a)$, *if* $ADD(a) \subseteq \psi$ y $DEL(a) \cap \psi = \emptyset$ (else $R[\psi, a]$ is undefined). The iterated regression on a sequence of actions α, denoted as $R^*[\psi, \alpha]$, is simply the successive application of the regression operator on each action within the sequence (assuming that this is defined for each step). For example, iff $\alpha = a_1, a_2, a_3$, then $R^*[\psi, \alpha] = R[R[R[\psi\ a_3], a_2], a_1]$ (Fritz & McIlraith, 2007). For the sake of readability, hereafter we shall use the notation R instead of R^*.

Based on the notion of regression proposed by Fritz et al., we define the *validity conditions* of a plan $\alpha = a_1 a_2 \dots a_n$, denoted as $V_\Pi(\alpha)$, as the set $\{R^*[G, a_n]$, $R^*[G, a_{n-1} a_n]$, … $, R^*[G, \alpha]\}$. We will therefore say that a plan α is *valid* on a state s iff there is a $\in V_\Pi(\alpha)$ such that $S \subseteq s$, i.e. a condition in $V_\Pi(\alpha)$ is met in s.

Another way of representing a solution to an automated planning problem is using a partial-order plan (Weld, 1994). A partial-order plan is a set of actions governed by a partial order. It can therefore be used to represent a set of sequential plans, each of which a linearization of the set, that is, is a sequence containing all actions in the set and that is such that its actions respect the partial order. Before formally defining the notion of a partial-order plan, we first introduce the notion of partial order. A partial order O over a set of actions A defines a transitive, antisymmetric and reflexive relation $\preccurlyeq_O$, where we omit the subindex when it is clear from the context. A linearization of a set of actions A with regards to a partial order O over A is a sequence of actions $\alpha = a_1 a_2 \dots a_n$ such that (1) each action of A appears only once in α and (2) if $a_i \preccurlyeq a_j \in O$ then $i \leq j$. We say that $a \prec b$ if and only if $a \preccurlyeq b$ and $a \neq b$. We denote the set of all the linearizations of A over O as $Lin(P)$. A tuple $P = \langle A, O \rangle$, where A is a set of actions and O is a partial order over A is a partial-order plan for Π if and only if every sequence in $Lin(P)$ is a plan for Π. Partial-order planning provides a compact representation of, in general, a number of linearizations which is exponential in the size of A.

The validity condition defined for sequential plans is not directly applicable to partial-order plans as there is no single sequential representation of the plan. However, it is equally possible to define a validity condition for partial-order plans. Given a planning problem Π and a partial-order plan P for Π, P is defined as *valid for a state s* iff there is a linearization of *P* that is still valid for *s*, i.e. it is executable on *s* and the resulting state of a given *suffix* contains the objective *G* (Muise et al., 2011). In more formal terms we have that $V_{\Pi}(P) = \bigcup_{\alpha \in Lin(P)} V_{\Pi}(\alpha)$.

Analyzing a large number of possible linearizations of a partial-order plan is costly. Muise et al. (2011) defined an efficient method for computing the validity conditions using the following notations: $last(\langle A, O \rangle) \stackrel{\text{def}}{=} \{a \mid a \in A \land \nexists a'\ a \prec a'\}$, which represent the set of actions for the plan where there are no restrictions on the order in which they originate; and $prefix(\langle A, O, a \rangle) \stackrel{\text{def}}{=} \langle A \setminus \{a\}, O \setminus \{a' \prec a \mid a' \in A\} \rangle$, which represents the resulting partial-order plan having deleted action *a* and all associated order constraints. This is considered indefinite when $a \notin last(\langle A, O \rangle)$ (Muise et al., 2011).

Using this notation, Muise et al. (2011) define a method where starting with a problem $\Pi = \langle F, A, I, G \rangle$ and a partial-order plan $\langle A, O \rangle$ can generate a list of conditions/actions L which is such that if $(c, a) \in L$ then for each state s that contains all propositions in c then a is an action that corresponds to a legal execution of the partial-order plan, guaranteed to reach the goal from state s.

Algorithm 1: Muise et al.'s Condition-Action List Generator (2011)

Input: POP $\langle A, O \rangle$. Planning problem $\Pi = \langle F, Act, I, G \rangle$

Output: List of ($\psi, a)$ pairs.

L = []; // L is the list of ($\psi, a)$ pairs to be returned

$\Gamma = \{\langle G, \langle A, O \rangle \rangle\}$ // Γ is a set of tuples of the form $\langle \psi, P \rangle$

for $i = 1..|A|$ **do**

    foreach $\langle \psi, P \rangle \in \Gamma$ **do**

        foreach $a \in last(P)$ **do**

L.append(($\mathcal{R}[\psi, a], a$));

/* Update to γ_{i+1} */

$\Gamma = \bigcup_{\langle\psi,P\rangle \in \Gamma} \{\langle \mathcal{R}[\psi, a], prefix(P, a)\rangle \mid a \in \text{last}(P)\}$

return L;

The method in Algorithm 1 uses Γ, a set of tuples of the form $\langle t, P\rangle$, where t represents a set of facts and P represents a partial-order. It starts with the goal state G and the plan $\langle A, O\rangle$. The algorithm iterates as long as there are actions in Π and the set Γ is updated in each iteration. This returns the actions $last(P)$, and for each action a the pair $(R[s, a], a)$ is added to the list L. At the end of each iteration the value of the set Γ is updated for the corresponding actions. If L is the list returned by running the algorithm for a planning problem Π, then Muise et al. (2011) show that a state s is valid for a plan iff any condition of L is contained within the plan, so that if the state s contains any of them then the plan is considered valid.

By way of example, we apply Muise et al.'s (2011) method to the partial-order plan $P = \langle A, O\rangle$ for the problem $\Pi = \langle F, A, I, G\rangle$, where $A = \{a_1, a_2, a_3\}$ and O is represented by the graph in Figure 3-2

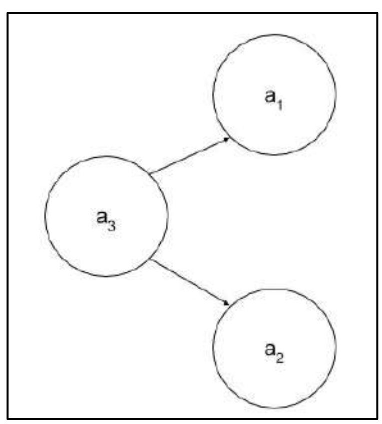


Figure 3-2: Graph of partial-order plan P

First, we must add the pair (G,P) to the list Γ in order to then start iterating on Γ. Three iterations will be performed as $|A| = 3$.

For iteration 1 we have that $\Gamma = \{(G,P)\}$ and $L = \emptyset$. $last(P)$ is computed to contain which actions could be executed last in P. Based on the graph of P (Figure 2), these are identified as the actions which no other actions depend upon, i.e. a_1 and a_2.

$\Gamma = \{(R(G,a_1),P_1),(R(G,a_2),P_2)\}$, where P_1 and P_2 correspond to $prefix(P,a_1)$ and $prefix(P,a_2)$, respectively. In iteration 2,

$\Gamma = \{(R(R(G,a_1),a_2),P'),(R(R(G,a_2),a_1),P')\}$ where

$P' = prefix(prefix(P,a_1),a_2) = prefix(prefix(P,a_2),a_1) = \langle\{a_3\},\emptyset\rangle$. The final iteration returns $\Gamma = \{(R(R(R(G,a_1),a_2),a_3),P''),(R(R(R(G,a_2),a_1),a_3),P'')\}$ where $P'' = \langle\emptyset,\emptyset\rangle$.

The worst-case running time for this algorithm is when all of the actions are independent. This is because for each step the list of elements in $last(P)$ would only decrease by 1 element per iteration for each element of Γ. This implies a worst-case exponential running time. This led to performance problems when attempting this method in real time during the game (see Section 2.3), exceeded acceptable parameters for maintaining the flow of the game and not affecting the user experience. We concluded Muise et al.'s approach was not suitable for our application.

3.2.3. Efficient Monitoring of Multi-Agent Plans

One advantage of the method presented in the previous section is that it is capable of working with partial-order plans. As explained previously, it makes sense to work with partial-order plans for our application. As there are multiple linearizations, the possibility of having to re-plan during monitoring is lower. However, as we showed above, the worst-case running time when monitoring a partial-order plan is exponential. In the worst-case, this computation requires exponential running time and memory.

In this section we describe an alternative method to the one proposed by Muise et al. (2011). Our approach can be used for monitoring the kind of multi-agent plans usually generated in multi-agent settings, where collaboration between agents is necessary but many of the actions carried out by an agent A are independent of the actions carried out by another agent B. We are particularly interested in the concept of partial-order plans with the kind of structure shown in Figure 3-3. This plan shows two chains of actions $a_1, \ldots$, executed by one agent, and $b_1, \ldots$, executed by another, which come together as actions $c_1, \ldots$, executed by any given agent.

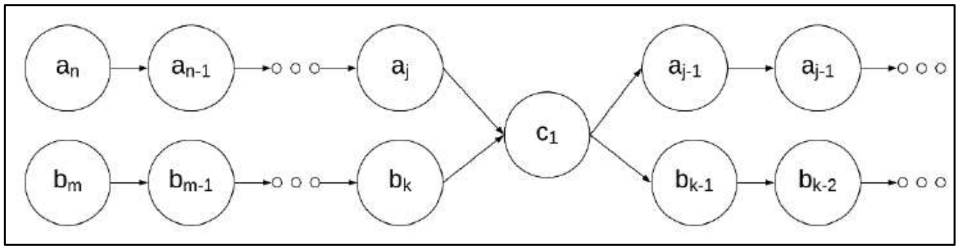


Figure 3-3: Chains of actions joined by a single action c_1

We will now show how a polynomial algorithm can be used to calculate a list similar to the list Γ described in Muise et al. (2011) when working with partial-order plans. In this case, the plans are similar to those shown in Figure 3-3 and contain independent chains of action. Doing so, provides us with an efficient way of calculating when a plan is still valid. However, before looking at the algorithm we will first look at some definitions.

Firstly, we will introduce some notation. If $P_1 = \langle A_1, O_1 \rangle$ and $P_2 = \langle A_2, O_2 \rangle$, we use $P_1 \cup P_2$ to denote $\langle A_1 \cup A_2, O_1 \cup O_2 \rangle$. If V_1, V_2 are two sets, we denote the set $\{v_1 \cup v_2 \mid v_1 \in V_1, v_2 \in V_2\}$ as $V_1 \otimes V_2$.

Our main interest for the rest of this section is to define a way of efficiently performing a regression with what we refer to as 'independent plans', which naturally occur in multiple-agent contexts. Intuitively, two plans can be considered independent when the effects of one do not delete the effects or preconditions of the other.

Definition 1: We say that $P_1 = \langle A_1, O_1 \rangle$ and $P_2 = \langle A_2, O_2 \rangle$ are independent with respect to the sets of propositions G_1 and G_2 if and only if:

1. No action in A_1 (respectively, A_2) has a delete that appears in G_2 (respectively, G_1) or in a precondition of an action in A_2 (respectively, A_1).
2. No fact added by an action in A_1 (respectively, A_2) is also deleted by an action in A_2 (respectively, A_1).

Furthermore, we say that $P_1, P_2, \dots, P_n$ are independent if they are pairwise independent. This definition extends directly to the case in which either P_1 or P_2 are sequences of actions, in which a total order is assumed between actions.

When there are two plans that are linear and independent, then there will be multiple ways of executing them. However, each execution maintains the relative order of the actions in each of the sequences. We define the concept of interleaving below. Given two sequences of actions α and β, the set $Int(\alpha, \beta)$ contains all of the interleavings of α and β. The formal definition is outlined below.

Definition 2: Let $\alpha = a_1 a_2 \dots a_n$ and $\beta = b_1 b_2 \dots b_m$ be two sequences. The set of the two interleavings of α and β are inductively defined as $Int(\alpha, \beta) = \alpha$ if $m = 0$, $Int(\alpha, \beta) = \beta$ if $n = 0$, otherwise

$$Int(\alpha, \beta) = \{a_1 \alpha' \mid \alpha' \in Int(a_2 \dots a_n, \beta)\} \cup \{b_1 \beta' \mid \beta' \in Int(\alpha, b_2 \dots b_m)\}.$$

We know that if P is a partial-order plan, each linearization of P achieves the goal. If both P_1 and P_2 are partial-order plans, each linearization of the union of these, $P_1 \cup P_2$, is an interleaving of a linearization of P_1 with a linearization of P_2. This property is formalized as follows:

Proposition1: Let $P_1 = \langle A_1, O_1 \rangle$ and $P_2 = \langle A_2, O_2 \rangle$. Then each linearization of α for $P_1 \cup P_2$ belongs to $Int(\alpha_1, \alpha_2)$, where α_1 and α_2 are linearizations of P_1 and P_2, respectively.

Proof: The proof is by contradiction. Let β be a linearization of $P_1 \cup P_2$ that does not achieve the goal. Now let β_1 and β_2 be sub-sequences of β that only contain actions stemming from A_1 and A_2, respectively. Then one of β_1 or β_2 must violate the order at O_1 (or at O_2), which would mean that β is not a linearization of $P_1 \cup P_2$.

Lemma 1: Let P_1 and P_2 be two partial-order plans for $\Pi_1 = \langle F, A, I, G_1 \rangle$ and $\Pi_2 = \langle F, A, I, G_2 \rangle$, respectively. Furthermore, let P_1 and P_2 be independent of G_1 and G_2. Then $P_1 \cup P_2$ is a plan for $\Pi = \langle F, A, I, G_1 \cup G_2 \rangle$.

Proof: Given that P_1 and P_2 are plans for Π_1 and Π_2, any linearization of both plans achieves G_1 and G_2 from I, respectively. Now let us take a linearization α of $P_1 \cup P_2$. Given Proposition 1, α must be an interleaving of the two linearizations of P_1 and P_2, let us call them α_1 and α_2. Given the definition of independence, α is executable as no precondition of an action α_1 can be invalidated by an action of α_2 and vice-versa. Finally, from the definition of independence we also get that α achieves $G_1 \cup G_2$ as no proposition added by an action of α_1 is deleted by an action of α_2 and vice-versa.

These results tell us that if we take two independent plans then the union between them is also a plan for the union of their objectives. Furthermore, any linearization of the resulting plan is an interleaving of the linearizations of the independent plans. This therefore provides an efficient way of calculating the validity conditions of an interleaving. More specifically, it tells us that when a sequence of actions comes from interleaving the sequences α_1 and α_2 then the regression of the interleaving can be calculated by only looking at α_1 and α_2. This finding is particularly interesting as when working with independent plans it means that we can focus on specific parts of the regression to calculate the regression for the whole plan.

Lemma 2: Let α_1 and α_2 be sequences of actions independent from G_1 and G_2. Let $G = G_1 \cup G_2$ and, finally, let β be a sequence of actions in $Int(\alpha_1, \alpha_2)$. Then $R(G, p) = R(G_1, \alpha_1) \cup R(G_2, \alpha_2)$.

Proof: By induction on the size of β. The base case can be verified directly because $\|\beta| = |\alpha_1| = |\alpha_2| = 0$, and $G = G_1 \cup G_2$. By definition, in that case, the regression of a set of facts for the empty sequence is the same set. Now, let us suppose that the theorem holds for any β such that $\|\beta| \leq k$, and show that this is also true for $a\beta$ where a is an action. We have two cases: that a is the first action of α_1, and that a is the first action of α_2. We will only look at the case in which $\alpha_1 = a\alpha_1'$; this therefore implies that $\beta \in Int(\alpha_1', \alpha_2)$. We therefore have that:

$$R(G, a\beta) = R(R(G, \beta), a)$$

Using the inductive hypothesis, we obtain that:

$$R(G, a\beta) = R(R(G_1, \alpha_1') \cup R(G_2, \alpha_2), a) = \big(R(G_1, \alpha_1') \cup R(G_2, \alpha_2)\big) \setminus add(a) \cup prec(a)$$

Given the assumption of independence between α_1 and α_2 from G_1 and G_2 we have that $R(G_2, \alpha_2)$ can only contain propositions in G_2 or propositions that are a precondition of an action in α_2. Therefore, it cannot contain any element in $add(a)$. We can therefore rewrite the above as:

$$R(G, a\beta) = (R(G_1, \alpha_1') \setminus add(a) \cup prec(a)) \cup R(G_2, \alpha_2)$$

Finally, using the definition of the operator R we get that $R(G, a\beta) = R(G_1, a\alpha_1') \cup R(G_2, \alpha_2)$. This concludes the proof of this lemma.

Lemma 3: Let $\alpha_1$and α_2 be sequences of actions independent of $G_1$and G_2. Furthermore, let α_1 and α_2 be plans for $\Pi_1 = \langle F, A, I, G_1 \rangle$ and $\Pi_2 = \langle F, A, I, G_2 \rangle$. Finally, let $\Pi = \langle F, A, I, G_1 \cup G_2 \rangle$ and $\beta \in Int(\alpha_1, \alpha_2)$. Then $V_\Pi(\beta) = V_{\Pi_1}(\alpha_1) \otimes V_{\Pi_2}(\alpha_2)$.

Proof:

$$V_\Pi(\beta) = \{R(G, \beta') \mid \beta' \text{ is a suffix of } \beta\}$$

$$= \{R(G, \beta') \mid \beta' \in Int(\alpha_1', \alpha_2'), \alpha_1' \text{ is a suffix of } \alpha_1, \alpha_2' \text{ is a suffix of } \alpha_2\}$$

Using Lemma 2 we get that:

$$V_\Pi(\beta) = \{R(G_1, \alpha_1') \cup R(G_2, \alpha_2') \mid \alpha_1' \text{ is a suffix of } \alpha_1, \alpha_2' \text{ is a suffix of } \alpha_2$$
$$= \{R(G_1, \alpha_1') \mid \alpha_1' \text{ is a suffix of } \alpha_1\} \otimes \{R(G_2, \alpha_2') \mid \alpha_2' \text{ is a suffix of } \alpha_2\}$$

$$= V_{\Pi_1}(\alpha_1) \otimes V_{\Pi_2}(\alpha_2)$$

Theorem 1: Let $\Pi = \langle F, A, I, G \rangle$ be a planning problem. Let $\{\Pi_i = \langle F, A, I, G_i \rangle\}_{i=1}^n$ be a sequence of planning problems such that $G = \bigcup_{i=1}^n G_i$, with $G_i \cap G_j = \emptyset$ when $i \neq j$. Let

P_i be a plan for Π_i, for every $i \in \{1, \dots, n\}$, such that $P_1, P_2, \dots, P_n$ are independent of $G_1, G_2, \dots, G_n$. Finally, let $P = \bigcup_{i=1}^{n} P_i = P$. Then $V_\Pi(P) = V_{\Pi_1}(P_1) \otimes V_{\Pi_2}(P_2) \otimes \cdots \otimes V_{\Pi_n}(P_n)$.

Proof: Using the fact that $V_\Pi(P) = \bigcup_{\alpha \in Lin(P)} R(G, \alpha)$ we can see that all of $\alpha \in Lin(P)$ is an interleaving of linearizations of $P_1, P_2, \dots, P_n$. We can then apply Lemma 3 to get the desired result.

The above theorem allows us to express the validity conditions of a plan that is based on the union of independent plans using only the validity conditions of these independent plans. Note that $V_\Pi(P)$ is exponential on the number of times that $\otimes$ is applied. However, for our task, i.e. determining whether a plan is still valid for a given state s, it is not necessary to compute $V_\Pi(P)$ as we can instead use the result of $V_{\Pi_1}(P_1), \dots, V_{\Pi_n}(P_n)$.

Specifically, we propose Algorithm 2 for monitoring the execution of $P = P_1 \cup \dots \cup P_n$ on a state s, when $P_1, \dots, P_n$ are independent.

Algorithm 2: Monitoring independent plans

Input: A partial-order plan P that is the union of independent plans $P_1, \dots, P_n$; a state s

Output: re-plan if necessary; else ok

1 Compute $V_{\Pi_i}(P_i)$, for all $i \in \{1, \dots, n\}$. (only once per plan)

2 If for every $i \in \{1, \dots, n\}$, there is a $c \in V_{\Pi_i}(P_i)$ such that $c \subseteq s$, then return ok. Else, **return** re-plan.

The correctness of this algorithm can be proven immediately based on the fact that every element in $V_\Pi(P)$ can be represented as the union of elements in $V_{\Pi_1}(P_1), \dots, V_{\Pi_n}(P_n)$. For the time complexity, observe that Line 1, each calculation of $V_{\Pi_i}(P_i)$ is proportional to $Lin(P)$, which, at the same time, has its worst-case exponential runtime on the size of P. If P is a linear plan, such as those in the experiments in educational settings found below, the computation of Line 1 is linear on the number of actions contained in the plans. This is because there is only one linearization for each plan. The condition of Line 2 requires the following running time: $O(|V_{\Pi_1}(P_1)| \cdot |V_{\Pi_2}(P_2)| \cdots |V_{\Pi_n}(P_n)|)$.

3.2.4. Extending our Method to a Broader Class of Plans

The method we have presented can be extended to a more interesting class of problems. This includes cases where we have chains of actions that are joined by a single action. For example, consider that for a given task there is a partial-order plan such as the one shown in Figure 3-3. This type of plan can often be found in the sort of contexts covered by this study. In this case, several agents can complete actions independently and, at some point, may need a synchronization action before continuing to work independently. This synchronization action may be, for example, opening a door, allowing both agents to access the same area before continuing with the plan independently.

For the discussion that follows, suppose that our partial-order plan essentially comprises p “blocks”. Each of these blocks considers the execution of two sequences of actions that are almost independent, which the reader can imagine are executed by different agents. Therefore $\alpha_1, \alpha_2, \dots, \alpha_p$ are p sequences of actions and $\beta_1, \beta_2, \dots, \beta_p$ are another p sequences of actions. Furthermore, $c_1, \dots, c_{p-1}$ are the synchronization actions. Intuitively, the actions of α_i are executed in the order given by sequence α_i. This is also the case with the actions of β_i: among themselves they are carried out in order. Additionally, the actions of α_i are executed concurrently with the actions of β_i. There is therefore no relationship of order between the actions of α_i and β_i. Finally, both the final action of α_i as well as the final action of β_i must occur *before* c_i.

Note that the linearizations of the plan described above take the form $\delta_1 c_1 \delta_2 c_2 \cdots c_{p-1} \delta_p$, where $\delta_i \in Int(\alpha_i, \beta_i)$, for all $i \in \{1, \dots, p\}$. We can therefore extend our previous method to compute the validity conditions for this plan. This can be done without considering all of the linearizations of the partial-order plan. Instead, we can take advantage of the fact that these linearizations have a certain independent structure. Before defining how to generalize our method for this type of plan, we must first accurately define the independence assumptions. Firstly, we assume that the objective G can be written as

$G = G_1 \cup G_2$. Furthermore, we also assume that α_j and β_i are independent with respect to G_1 and G_2, for every $i, j \in \{1, \dots, p\}$. Finally, we assume that the preconditions of c_i, Pr_i, for all $i \in \{1, \dots, p-1\}$, can be divided into two disjoint sets Pr_i^{α} and Pr_i^{β}, such that $Pr_i = Pr_i^{\alpha} \cup Pr_i^{\beta}$, and that α_i and β_i are also independent with respect to Pr_i^{α} and Pr_i^{β}. Intuitively, this tells us that α_i and β_i satisfy independent parts of the precondition of c_i.
Algorithm 3 can be used to monitor this kind of plan. Its correctness can be proven immediately using the above method in which we calculate the regression conditions for two plans that are executed independently by calculating the validity conditions of the independent plans. Extending this algorithm to more than two chains is simple. However, we will not present the general algorithm as it is somewhat involved.
The running time for this algorithm is $O(\sum_{i=1}^{p} |\alpha_i| \cdot |\beta_i|)$. This is substantially lower than the running time returned by Muise et al.'s (2011) algorithm, which is $\Omega(\prod_{i=1}^{p} 2^{min\{|\alpha_i|,|\beta_i|\}})$. This is because the running time for Muise et al.'s algorithm is proportional to the number of linearizations of the plan. Furthermore, it is easy to prove that $|Int(\alpha, \beta)| = 2^{min\{|\alpha|,|\beta|\}}$.

Algorithm 3: Monitoring chains with synchronization actions

Input: A partial-order plan P for objective $G_1 \cup G_2$, whose linearizations take the form $\delta_1 c_1 \delta_2 c_2 \cdots c_{p-1} \delta_p$, where $\delta_i \in Int(\alpha_i, \beta_i)$, for all $i \in \{1, \dots, p\}$, with the independence conditions described in the text.

Output: re-plan if necessary; else ok

1. **if** *algorithm is run for the first time* then
2. $G_p^{\alpha} \leftarrow G_1$; $G_p^{\beta} \leftarrow G_2$
3. **for** $i \leftarrow p \dots 1$ do
4. $R_i^{\alpha} = \{R(G_i^{\alpha}, \alpha) \mid \alpha \text{ is a suffix of } \alpha_i\}$ // validity conditions for α_i
5. $R_i^{\beta} = \{R\left(G_i^{\beta}, \beta\right) \mid \beta \text{ is a suffix of } \beta_i\}$ // validity conditions for α_i
6. $G_{i-1}^{\alpha} = R(G_i^{\alpha}, \alpha_i) \setminus add(c_{i-1}) \cup Pr_{i-1}^{\alpha}$
7. $G_{i-1}^{\beta} = R\left(G_i^{\beta}, \beta_i\right) \setminus add(c_{i-1}) \cup Pr_{i-1}^{\beta}$

---

8. **if** for some $i \in \{1, \ldots, p\}$, $G_i^{\alpha} \cup G_i^{\beta} \subseteq s$ or

there exists an $s^{\alpha} \in R_i^{\alpha}$ and an $s^{\beta} \in R_i^{\beta}$ such that $s^{\alpha} \cup s^{\beta} \subseteq s$ then

9. **return** ok

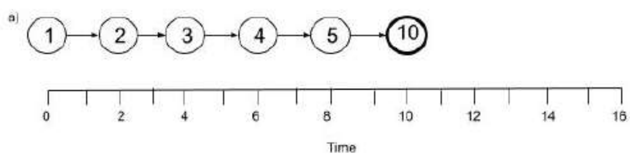
10. else

11. **return** re-plan

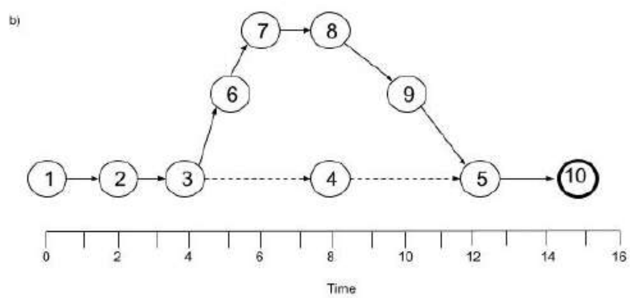
3.2.5. Feedback System

The system developed for this study must be capable of providing feedback that promotes communication among the participants of a group and allows all agents to finish the game (Appendix K). To do so, the solution is split into two stages. The first of these stages involves identifying and understanding what the agents have to do. This is done by generating and maintaining a plan of actions that must be followed by the participants in order to complete the game. This plan is obtained following the methods presented in Section 2.4. The second stage involves giving feedback to the agents and requires knowing when to give the feedback and what the content of this feedback should be.

To determine when to give the feedback, it is important to note that feedback can be given, and considered effective, at various points during a collaborative activity (Nicol, Thomson & Breslin, 2014; Boud & Molloy, 2013; Freedberg, Glass, Filoteo, Hazeltine & Maddox, 2017). Additionally, as the participants get further into an activity the level of feedback should gradually decrease (Chan, Kaur Sidhu, Narasuman, Lee & Yap, 2016). Based on these two points and using the concept of distance from objective, which we define as the number of actions required by an agent in order to meet the objective, we calculate progress as the relationship between the amount of time since the last piece of feedback was given and the total distance from the objective for all of the agents. An increase in this value suggests that the agents are not making progress or are spending too long in one place. Based on this premise, and setting an initial limit manually, we decided to provide feedback when the progress score exceeded this limit. For example, Figure 4 shows two sequences of actions to arrive at a shared objective (node 10), with the respective times. For each unit of time we can calculate the progress by multiplying the time that has elapsed since the last piece of feedback was given by the number of actions needed to meet the objective. If we set a limit of 15, then in the case of sequence a) in Figure 3-4, the agents will always be considered to be making progress. This is because the highest score they can obtain is when the time since the last feedback is 5 and the distance from the objective is 3. Furthermore, from sequence b) in Figure 3-4, we can see that after 5 units of time the distance from the objective increases. The result of multiplying this by the time elapsed since the last feedback message was sent is 25, which is greater than 15. In this case, it would be determined that the agents are not making progress.



t	tlf	d	p	result	t	tlf	d	p	result
1	1	5	5	Progressing	6	6	2	12	Progressing
2	2	4	8	Progressing	7	7	2	14	Progressing
3	3	4	12	Progressing	8	8	1	8	Progressing
4	4	3	12	Progressing	9	9	1	9	Progressing
5	5	3	15	Progressing	10	10	0	0	Progressing



t	tlf	d	p	result	t	tlf	d	p	result	t	tlf	d	p	result
1	1	5	5	Progressing	6	1	4	4	Progressing	11	6	2	12	Progressing
2	2	4	8	Progressing	7	2	4	8	Progressing	12	7	1	7	Progressing
3	3	4	12	Progressing	8	3	3	9	Progressing	13	8	1	8	Progressing
4	4	3	12	Progressing	9	4	3	12	Progressing	14	9	1	9	Progressing
5	5	5	25	No progressing	10	5	2	10	Progressing	15	10	0	0	Progressing

Figure 3-4: Representing progress in the game

To define the content of the feedback, the solutions to the different obstacles in the game were identified (Appendix K). The different points of interest within the activity were also established. It is at these points that the agents may require feedback (Table 3-1). In this sense, the content is predefined and personalized for each of the given situations.

Table 3-1: Actions for overcoming obstacles during the game

POI	Obstacle	Action	Solution
2,8,9,10	Jump	Move-jump	PJ3 must use their jumping ability in order to reach the next point

2	Breakable-wall	Break-wall	PJ2 must use their hitting ability in order to break through the wall and continue
2,14	Door	Use rune	PJ1 must use the blue rune in order to open the door so that PJ2 and PJ3 can continue
3	Linked-Rock	Break-wall	PJ2 must use their hitting ability in order to break through the wall and thus break the rock in order to continue
5	Force-Field	Use magic	PJ1 must use their magic ability to pass through the force field and continue
7	Ladder	Use machine	PJ2 must use the lever to lower the ladder and continue
2,8,9	Jump	Use rope	PJ3 must use the rope so that PJ1 and PJ2 can continue
8	Rune	Use rune	PJ1 must have the rune so as to use it on the door
		Pick rune	PJ2 or PJ3 must drop the rune so that PJ1 can pick it up
9,10	Rope	Use rope	PJ3 must have the rope in order to use it to climb up onto the platforms
		Pick rope	PJ1 or PJ2 must drop the rope so that PJ3 can pick it up
11	Gear	Use gear	PJ2 must have the gear so that they can use the machine and open the trap door
		Pick gear	PJ1 or PJ3 must drop the gear so that PJ2 can pick it up
12	Spider	Use magic	PJ1 must use their magic ability so that PJ1, PJ2 and PJ3 can enter
12	Boulder	Push boulder	PJ2 must use their hitting ability to push the boulder and continue
13	Boulder	Use gear	PJ2 must use the machine so that the boulder falls and they can continue

17	Spider	Use Magic	PJ1 must use their magic ability (12) so that PJ2 and PJ3 can enter
15	Key	Use key	PJ3 must have the key in order to open the chest
16	Key	Pick Key	PJ1 or PJ2 must drop the key so that PJ3 can pick it up

Furthermore, we must consider that if the feedback contains all of the information or the solution to the activity then it is less effective for teaching and developing skills (Wooten & Ulrich, 2017). The messages were therefore split into five levels for each action (Table 1), from a low level of detail, e.g. “You still haven’t arrived, is everything OK?” (Level 1, Table 3-2), to a high level of detail, e.g. “Player 3 must use their super jumping powers in order to cross the bridge” (Level 5, Table 3-2). These messages were sent iteratively to the participants whenever they needed it, based on their progress. The level of detail also increased gradually when the feedback was repeated for a given action. If the players moved on to another action, then the level of detail was reset.

Table 3-2: Example messages with increasing levels of detail

Level of detail	Message
1	You still haven’t arrived, is everything OK?
2	I expected to see <player> around here, do you know what happened to them?
3	Don’t forget to use your abilities
4	Did you try using your super jumping powers?
5	<player> must use their super jumping powers in order to cross the bridge

Note: <player> represents the name of one of the players

Furthermore, in order to promote communication between agents, the message is always delivered to one of the agents who is not involved in the action. This can be identified using the parameters that are set for each action.

Finally, based on the context and architecture described in Section 3.2.1, we can see that there are certain challenges that may come up when implementing the proposed solution. For example, planning for multiple groups from a single server, and the possibility of a client losing its connection to the server, among others. Measures were therefore taken to solve these issues, without changing the design. This included adding a cache to store the request for a re-plan on the server, as well as the plan that was returned by the planner. Therefore, when another group needed a plan based on the same instruction a re-calculation was not necessary, with the system instead sending the stored version of the plan. Furthermore, a reconnection system between the clients and server was developed. This system detected when a client had disconnected and switched control of the objects (monitor, variables etc.) to another client (if necessary) while the connection was re-established.

3.2.6. Implementation

This study was based on the use of a collaborative game for tablets (Appendix K). This involved 3 players on independent devices (clients) connected via a single server, run on a laptop computer. The server used the *Fast Downward* planning system to generate plans (Helmert, 2006). This system returns a sequential plan that is automatically transformed into a partial-order plan when the actions of the plan are executed by individual agents. The client monitor includes a logic for calculating the monitoring conditions for the plans and reviewing them in real time, following the steps described in Section 3.2.4.

To study our proposed solution, a test was carried out in a real-world educational setting during a single session. The sample included 69 fifth-grade students (36 male, 33 female) aged between 10 and 13 (M: 10.565, SD: .56) in Santiago, Chile. It is worth noting that, according to PISA (OECD, 2018), Chile enjoys the best academic results in the region, with scores of 452, 417 and 444 in reading, mathematics and science, respectively. However, these are still way below the global average of 487, 489 and 489 in reading, mathematics and science, respectively (OECD, 2018). In the 2015 Pisa test for Collaborative Problem Solving Chile obtained 457 points, while the average was 500 (OECD, 2018).

The session lasted for one hour, during which time the students were randomly divided into groups of 3, who then sat next to each another. Each group had 3 tablets, one per student. Half of the groups played a version of the game including the proposed feedback system, while the other half played a version without it. Furthermore, all of the tablets were connected to microphones that the students wore around their necks in order to record the conversations they had with their teammates.

The qualitative data was gathered from the conversations that the students had about problem solving during the session, recorded by the microphones (Appendix O). The quantitative data was taken from the server's logs, which recorded participants' actions during the game (Appendix D).

3.2.7. Post-game Analysis

3.2.7.1. Efficiency of monitoring

This analysis is based on a simulation that allows the performance of different methods of monitoring to be compared (i.e. sequential method, Muise et al's method and the proposed method). For the sequential method, a single sequential plan was monitored and taken as if it had been generated directly by *Fast Downward.*

The time taken by each method to calculate the validity conditions needed for the feedback system was measured in milliseconds. This was done based on a set of partial-order plans such as the one described in Figure 3-5. The set comprised 19 partial-order plans, each with between 3 and 21 actions.

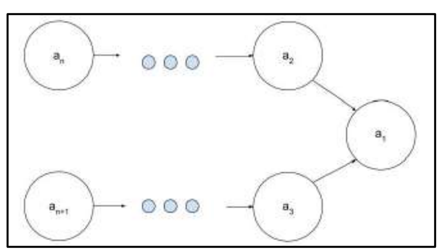


Figure 3-5: Representation of a set of test plans

This test was carried out without any agent interacting with the system and with the server connected only to one client so as to avoid other actions interfering with the results that were obtained. This includes actions such as receiving messages from other clients or sending update messages.

3.2.7.2. Efficiency of re-planning

A comparison was also made between the number of re-plans made by the game using the validity conditions for each of the monitoring methods. This was important as it is key for determining the feasibility of the proposed solution within the context described in Section 3.2.1. This comparison was made based on 5 randomly-selected records of the sequences of actions taken by the students (Appendix N) following the process described in Section 3.2.6. These sequences were reviewed following the corresponding order of each action in order to assess the state of the game after each one and determine whether the plan was valid according to the monitoring method that was used. If after a given action the state did not meet the validity conditions, the conditions were updated according to the monitoring method and testing continued until the sequence was completed. The number of times that the validity conditions were not met was calculated and considered the number of re-plans for a given monitoring method.

3.2.7.3. Effectiveness of feedback

Data from 23 groups was analyzed qualitatively. This corresponds to a total of 69 students (33 control and 36 experimental). These groups were chosen as the quality of the recordings was acceptable, allowing them to be listened to and analyzed.
Content analysis was performed using the Atlas.ti software package. Unlike open coding, the use of structured analysis allowed us to analyse the students' responses based on pre-defined concepts.

The analysis followed a recursive structure (McAlister et al., 2017), in which the researchers coded two records based on a single code for each iteration. The research concepts and methodological decisions were aligned, and a codification matrix (see Table 3-3) was designed and iterated as the codification process was developed. Each iteration of analysis was based on one of the codes. Once the concept was agreed by the investigators, a transcript was then analysed using the given code. If necessary, the concepts were realigned and the transcripts were swapped among the researchers. By doing so, each researcher completed at least four iterations per category. The unit of analysis was therefore set as each turn taken by the participants during the intervention.

The first step in the coding process was to identify moments in which the participants touched on the feedback messages delivered by the system. Accordingly, a message was coded as *feedback* whenever a participant said something that matched with the feedback given by the system developed for this study. Following this, any references to the type of feedback were then coded. The next category for analysis was collaborative problem-solving skills. This was based on the OCED framework, which combines the three main collaborative competences with the four stages of the individual problem-solving process (OECD, 2017). Furthermore, any conversations or messages between the members of a group following the feedback they received were considered a *consequence* of said feedback. The effectiveness of the feedback is therefore understood as when both a *consequence* and an associated collaborative problem-solving skill are present following an item of feedback.

Any instances in which the participants communicated after receiving a feedback message were therefore coded based on these 12 skills. Subsequently, categories for classifying obstacles present in the game were defined and coded using the audio files. In other words, we can establish the progress made by each group within the game, as well as the number of messages referring to the obstacles that they managed to identify. We are also able to see whether or not they managed to make it to the final obstacle (i.e. the boulder).

Table 3-3: Matrix for coding and analysis

Dimension	**Code**	**Criteria**
Feedback	Movement	A message suggesting that one or more of the participants go or come to a specific place.
	Ability	A message advising the use of one of the character’s special abilities in order to overcome an obstacle.

	Action	A message suggesting an action (Jump, Shoot, etc.) in order to overcome an obstacle.
	Status	A message inviting participants to reflect on their current state by asking a question (e.g. Is everything OK?).
Obstacles	Platform	The participants must discover that at the top of the tree there is a button that, when pressed, activates a platform that will allow them to access a path that is otherwise unreachable by jumping (e.g. “There’s a yellow button in the tree”).
	Individual Path	A part of the game where each player must follow a path based on their colour (e.g. “No, I’ve already been down and been everywhere, I’ve done everything”).
	Items	To overcome obstacles later in the game, each of the participants must find an item of their colour (e.g. “Here’s another item!”).
	Door	An obstacle that allows players to access the final part of the game and that must be opened using one of the items found previously (e.g. “I’m at the door”).
	Spider	An obstacle that can only be overcome by simultaneous and coordinated participation from the 3

		members of the group (e.g. "Oh!! Look, there's a Boss!").
	Machine	A machine that they must activate in order to get past the spider and can only be activated using one of the items they find (e.g. "Over there there's a tool; I don't know what it is!")
	Boulder	The final obstacle in the game, which must be pushed in order to get past it (e.g. "Yeah, you have to activate the boulder!").
CPS Skills	(A1) Discovering team members' perspectives and abilities	Identify mutual knowledge (i.e. what each participant knows about the problem)
		Identify group members' perspectives of the problem to be solved
	(A2) Discovering the type of collaborative interaction used to solve the problem, along with the goals	Identify the actions and interaction needed to solve the problem
		Identify constraints of actions when solving the problem.
	(A3) Understanding the problem-solving roles	Identify the necessary problem-solving roles
		Identify team members' strengths and weaknesses
	(B1) Building a shared representation and negotiating the problem's meaning (common ground)	Monitor shared understanding of the problem

	(B2) Identifying and describing tasks to be completed	Establish group goals/tasks to solve the problem
	(B3) Describing roles and team organization (communication protocol/rules of engagement)	Establish the roles and organization needed to solve the problem
	(C1) Communicating with team members about the actions to be/being performed	Formulate a sequence of steps or plan to solve the problem
	(C2) Enacting plans	Execute actions to attempt a solution
	(C3) Following the rules of engagement, (e.g., prompting other team members to perform their tasks.)	Follow rules of engagement
	(D1) Monitoring and repairing the shared understanding	Monitor shared understanding of the problem
	(D2) Monitoring action results and evaluating the success in solving the problem	Monitor the group results when solving the problem
	(D3) Monitoring, providing feedback and adapting the team organization and roles	Monitor the organization of the group

3.2.8. Ethical Considerations

The study was approved by the University's ethics board. The school was then invited to participate in the study and provided their written authorization. Before participating in the session, the students' parents received information on the purpose of the research and the specific activities that were involved, as well as the risks and benefits of participating. The voluntary nature of participating in the study was made clear. The students' parents were then asked to sign an informed consent form. Finally, at the beginning of the session, the students whose parents had given their consent were told about the activities they would be participating in and were invited to sign an informed consent form, while they were also advised that they could stop participating in the activity whenever they wanted.

3.3. Results

3.3.1. Performance

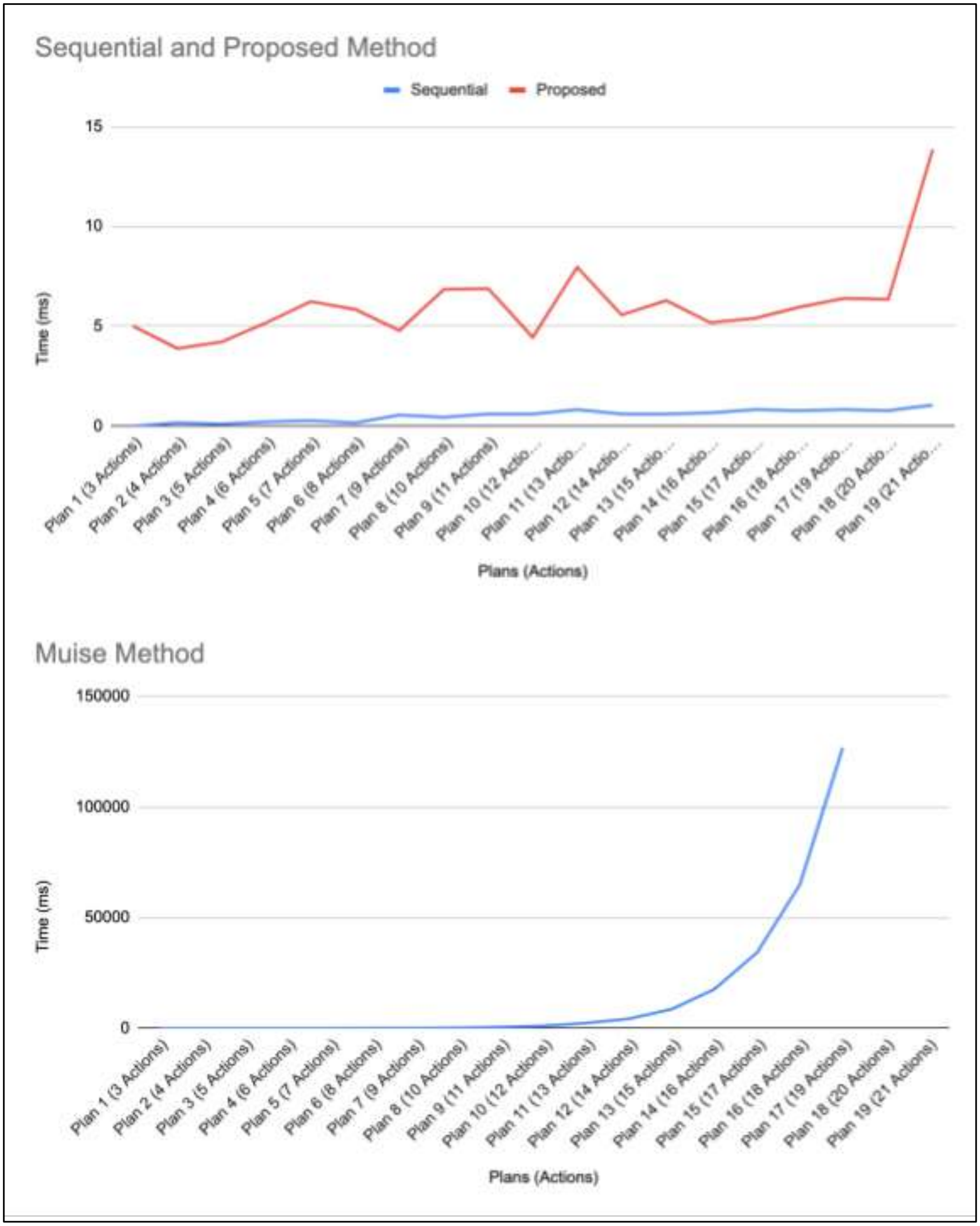
The performance of our automated planning and monitoring depends on two factors: first, the efficiency of monitoring and, second, the efficiency of planning. Indeed, plans are needed for feedback generation and for assessing progress. We want to minimize the number of replans since each replan requires significant computation. Second, we want to minimize the time required to monitor. Below we include an analysis of both aspects.

To evaluate monitoring efficiency, a simulation was carried out with 19 partial-order plans (See Section 3.2.7.1) according to the method depicted in Figure 3-5. Nine tests were conducted for each of these plans in order to measure the time it took to calculate the validity conditions. This led to a total of 171 measurements for each of the three monitoring methods that were used in this study (Appendix P). The average performance for each method based on the number of actions in each plan is shown in Table 3-4 and displayed as a graph in Figure 3-6.

Table 3-4: Summary of the performance per plan for each method

	Performance (ms)		
Plans (Actions)	Sequential	Muise et al	Our approach
Plan 1 (3 Actions)	.00	6.01	5.01
Plan 2 (4 Actions)	.17	11.34	3.89
Plan 3 (5 Actions)	.11	15.15	4.22
Plan 4 (6 Actions)	.22	18.47	5.18
Plan 5 (7 Actions)	.28	26.60	6.24

Plan 6 (8 Actions)	.17	28.88	5.85
Plan 7 (9 Actions)	.56	50.87	4.79
Plan 8 (10 Actions)	.45	84.77	6.85
Plan 9 (11 Actions)	.61	158.17	6.88
Plan 10 (12 Actions)	.61	287.58	4.45
Plan 11 (13 Actions)	.83	574.48	7.96
Plan 12 (14 Actions)	.61	1136.09	5.57
Plan 13 (15 Actions)	.61	2416.91	6.29
Plan 14 (16 Actions)	.67	4341.74	5.18
Plan 15 (17 Actions)	.83	8661.33	5.40
Plan 16 (18 Actions)	.78	17767.30	5.96
Plan 17 (19 Actions)	.83	34350.28	6.40
Plan 18 (20 Actions)	.78	64931.09	6.35
Plan 19 (21 Actions)	1.06	126648.40	13.86



Sequential and Proposed Method
Sequential
Proposed
15
10
5
0
Time (ms)
Plan 1 (3 Actions)
Plan 2 (4 Actions)
Plan 3 (5 Actions)
Plan 4 (6 Actions)
Plan 5 (7 Actions)
Plan 6 (8 Actions)
Plan 7 (9 Actions)
Plan 8 (10 Actions)
Plan 9 (11 Actions)
Plan 10 (12 Actio...
Plan 11 (13 Actio...
Plan 12 (14 Actio...
Plan 13 (15 Actio...
Plan 14 (16 Actio...
Plan 15 (17 Actio...
Plan 16 (18 Actio...
Plan 17 (19 Actio...
Plan 18 (20 Actio...
Plan 19 (21 Actio...
Plans (Actions)
Muise Method
150000
100000
50000
0
Time (ms)
Plan 1 (3 Actions)
Plan 2 (4 Actions)
Plan 3 (5 Actions)
Plan 4 (6 Actions)
Plan 5 (7 Actions)
Plan 6 (8 Actions)
Plan 7 (9 Actions)
Plan 8 (10 Actions)
Plan 9 (11 Actions)
Plan 10 (12 Actions)
Plan 11 (13 Actions)
Plan 12 (14 Actions)
Plan 13 (15 Actions)
Plan 14 (16 Actions)
Plan 15 (17 Actions)
Plan 16 (18 Actions)
Plan 17 (19 Actions)
Plan 18 (20 Actions)
Plan 19 (21 Actions)
Plans (Actions)

Figure 3-6: Graph of performance for the sequential method, Muise's method and the proposed method

To measure the replanning performance (see Section 3.2.7.2), we calculate the percentage the number of actions requiring a re-plan vs. the total number of actions in the sequence that was analyzed. The average distance between re-plans (in terms of number of actions) is also calculated for each method of monitoring. These two measures give us an idea of how frequently is planning required. These scores are given as averages so as to facilitate analysis.

Table 3-5: Summary of actions in an optimal and real plan

	Optimal Plan		Real Plan		
	Best Group	Total	Minimum	Maximum	Mean
Actions	46	42	50	90	70

When running Muise's method in a real-world context the system did not manage to complete the process. The above comparison was therefore only made between the sequential method and our method. The optimal plan for completing the stage of the game that was analyzed (Appendix K) comprised 46 actions (Table 3-5). The group that made most progress in the game managed to complete 42 of these actions. However, the number of actions made by the groups varied between 50 and 90 actions.

Table 3-6: Percentage of re-plans per monitoring method

	Re-plans			Distance (Actions)	
Method	Mean	Minimum	Maximum	Mean	Mode
Muise et al.	-	-	-	-	-

Sequential	.74	.58	.89	2.3	1
Our approach	.46	.38	.89	3.4	1

The average percentage of re-plans for the sequential method was 74% (Table 3-6), with a range between 58% and 89%. The average distance between re-plans was 2.3 actions, with a mode of 1. This shows that our approach, which relies on the use of partial-order plans rather than sequential plans, requires fewer replans.

For our approach, the average percentage of re-plans was 46% (Table 3-6), with a range between 38% and 89%. The average distance between re-plans was 3.4 actions, with a mode of 1.

3.3.2. Effectiveness of feedback

Following the method described in Section 3.2.7.3, a total of 102 quotes from the students were considered as feedback. 80% of these produced a consequence, while the remaining 20% did not (Figure Q.2, Appendix Q).

Most of the feedback given by the system (57%) was classified as Movement, i.e. advising one of the characters to go somewhere (Figure Q.3, Appendix Q). This was followed by Ability, with 27%, and Action and Status, with 10% and 6%, respectively. This is in line with the objective of the game, which consisted of reaching the end of the game by overcoming obstacles. A larger amount of feedback on the players' Movement was therefore to be expected.

Furthermore, we can see that the students' messages identified as Feedback (Figure Q.4, Appendix Q) were mainly related to skills B3 and C3 for collaborative problem solving (Table M.1, Appendix M). These correspond to the skills of organizing the group, establishing ground rules, and ensuring they are followed. In this sense, messages coded as B3 (e.g. "Everyone has to get here" or "The red player has to go to the tower by the lake") helped organize the group, while those coded as C3 (e.g. "You still haven't arrived, is everything OK?") encourage the students to stay on track.

Similarly, from Figure Q.4 (Appendix Q) we can see that the consequences of the feedback are related to a wider range of collaborative problem-solving skills (Table M.1, Appendix

M). Indeed, in this case there are messages associated with each skill (Table Q.1, Appendix Q). For example, some of the messages are associated with a shared understanding, such as those related to understanding the problem (e.g. "No I don't know what I have to do") or clarifying misunderstandings (e.g. "Yeah, the red player has to pass, because over there are some yellow and green things"). In other cases, there were conversations about the actions needed to solve the problem, such as when students proposed a sequence of actions (e.g. "First you have to jump over there, then jump over there") or shared what they had done (e.g. "I already shot him"). Finally, most of the participants communicated about how to organize the group (A3, B3, C3 and D3). This includes messages about understanding the roles needed to solve a problem (e.g. "What's my ability?"), setting ground rules (e.g. "P2 you have to get there"), following rules (e.g. "Look, I already shot that yellow thing"), and monitoring and adapting the organization and roles within the group (e.g. "P2 look, it looks like one has to go over there and not come over here, the other, the red player, has to go over there. And what about me? Where do I go?").

Finally, having identified the obstacles overcome by each group, we were able to compare how much progress was made by the participants in the experimental group vs. those in the control group (Figure Q.5, Appendix Q). We can see that 6 groups in the control group and 10 groups in the experimental group managed to make it to the first obstacle, coded as Platform. Meanwhile, none of the control groups and only 2 of the experimental groups managed to make it to the final obstacle, coded as Boulder. This difference between the control and experimental groups is evident for each of the obstacles. Furthermore, the total number of groups at each obstacle also decreases progressively.

3.4. Discussion

Analyzing the results from the previous section, we are able to compare the performance of the 3 monitoring methods that were tested. The sequential method was the fastest, with the running time failing to exceed 2.5 milliseconds for any of the tests (Table 3-2). These running times were expected as this method calculates based on a single linearization of the plan, regardless of its form (Figure 3-5). Muise et al.'s (2011) method returned an exponential running time (Figure 3-6) that increased with each additional action, taking almost 2 minutes to run the final plan (Table 3-2). According to Muise et al. (2011), this behavior is expected for cases where there is a parallel domain (i.e. actions that are independent of one another), as is partially the case with the plan that was tested (Figure 3-5). The results for the proposed method (Table 3-2) were considerably better than for Muise et al.'s (2011) method for the plan with the greatest number of actions (21 actions), but worse than the sequential method for every plan (Fritz & McIlraith, 2007).

The measurements from the first experiment show that the values for each method and plan move within a well-defined range, except for certain tests that returned values that were clearly out of this range. This includes the case of Test 1 for a 5-action plan using Muise et al.'s (2011) method (Table 3-2). These values can be explained by understanding that the running time depends not only on the test but also on the condition of the hardware.

Furthermore, we can see that almost 75% of actions with the sequential method require a re-plan. However, the time needed to obtain the validity conditions once the new plan has been generated is negligible. With our proposed method, only 45% of actions require a re-plan and, although the time for obtaining the validity conditions is not as good as with the sequential method, it is still unnoticeable during a normal execution.

Therefore, the main consequence of our study relates to the suitability of the proposed algorithm within an educational setting. In this case, re-planning means asking for an execution from an external planner, which can take between 3 to 5 seconds to generate a new plan. It is therefore essential to minimize the number of re-plan requests that are made by the chosen monitoring method. The proposed method therefore meets the requirements regarding running time and re-planning requests in a real-world classroom setting (Grover, Chakraborti & Kambhampati, 2018).

It is important to note that these results are valid for plans that take the form described in Section 3.2.4. This means plans that generate independent paths between the actions and have few shared actions joining these paths (in this case one). In the game presented here, this takes place between the individual actions of the 3 agents and the shared actions that must be completed at various stages in the game.

The second main consequence of this study relates to the effectiveness of the feedback. In this sense, we were able to code the participants' communication based on: type of feedback, whether or not it produced a consequence, associated collaborative problem-solving skills (both for the feedback as well as the consequence), and the obstacles that were overcome within the game. This allowed us to show that the feedback not only improved the students' performance, it also encouraged them to use their collaborative problem-solving skills. This finding is particularly relevant as there is currently little research into the development of collaborative problem solving (Graesser et al., 2018; Graesser, Greiff, Stadler & Shubeck, 2020). Furthermore, our study provides evidence on how to give effective feedback during collaborative problem solving, both of which are key skills in modern society (Rosen, Wolf & Stoeffler, 2020).

Finally, using Automated Planning to generate feedback has been seen in other areas, such as science and mathematics, with mixed results (Fyfe & Rittle-Johnson, 2016). The implementation described here could be applied in other contexts if it is suitably adapted. For example, by correcting the messages and adjusting the parameters, it could be used to design games and activities that foster other 21st century skills, where real-time feedback plays a key role. This includes skills such as creative thinking and creativity, among others.

3.5. Conclusions

This study sought to answer the question "How can feedback be given effectively to the participants in a collaborative game when interacting simultaneously to solve a problem?" To do so, we studied the viability of using Automated Planning systems to provide feedback in collaborative games. As these games operate in real time, the response time and running time must be low to allow a good user experience. Furthermore, the solution must be able to work with the technology in real-world educational settings. The study therefore also looked to answer the question "How can we develop an effective Feedback System based on Automated Planning when the plans involve multiple agents in a real-world educational context?"

Some monitoring approaches for partial-order plans, such as the one proposed by Muise et al. (2011), compute all of the linearizations and their possible validity conditions. However, our results show that the running time for such approaches make them impractical for use in a real-world educational setting. In this sense, an important contribution was developing a way to compute validity conditions for monitoring partial-order plans within a reasonable amount of time for the given context. Our approach takes advantage of the structure of the plans in order to break them down into problems that easier to solve. This allowed us to improve on the running times reported by Muise et al. (2011) and produce a list of validations that maintained their correctness. While our approach does not outperform the monitoring of a sequential plan, it allows reducing the number of calls to the planning algorithm significantly. Por lo tanto, los resultados obtenido constituyen una contribución relevante a la aplicación de Automated Planning en contextos educativos reales, en los cuales se requieren tiempos de respuesta que permitan una buena experiencia de usuario y exigencias de computo factibles con el hardware disponible, lo cuál se logra al reducir the number of calls to the planning.

Another achievement of this study is the feedback system, which promotes communication among group members and helps them develop collaborative problem-solving skills within the context of a collaborative game (Appendix K). Indeed, the results suggest that the feedback helped the students make more progress within the game in terms of the number of obstacles that they were able to overcome (Section 3.3.3). Furthermore, the feedback given by the system proved to be effective as it generally prompted a conversation among the members of the group on fundamental collaborative skills, such as establishing and maintaining the organization of the group. The results of our study therefore provide evidence regarding the effect of feedback on collaborative problem solving and suggest a first step towards creating technology-mediated collaborative environments to promote the development of collaborative problem-solving skills among primary school students.

Nevertheless, there are certain limitations of this study that should be highlighted. One such limitation is that the monitoring method proposed here can only be applied to specific types of plan (see Figure 3-5). Another limitation is that the system that was developed is strongly linked to the domain of the problem that was used by the planner to generate the plans (Appendix L). In other words, any change to the domain would need to be reflected in the logic of the monitoring system. Furthermore, the inclusion of an external planner represents an additional limitation as running times cannot be further optimized without altering the source code. A final limitation relates to the size of the sample and how unrepresentative it is. In this sense, the sample only includes students from a single country and is too small to generalize the results.

Future work should look for algorithms that can efficiently monitor more complex plans than the one described here and that also work in real-world contexts such as a classroom.

Furthermore, developing collaborative problem-solving skills not only involves helping students to solve the problem at hand, it also involves developing their collaborative and communication skills. Mechanisms that include feedback on the level of communication among the participants should therefore be explored. This would include providing feedback on their interactions, agreements, comprehension of the problem and level of organization, among others.

4 INTEGRATING COLLABORATIVE SCRIPT AND GROUP AWARENESS TO SUPPORT GROUP REGULATION AND EMOTIONS TOWARDS COLLABORATIVE PROBLEM SOLVING

4.1. Introduction

In 2015, the OECD carried out a large-scale, global assessment of Collaborative Problem Solving (CPS) through its PISA testing program (OECD, 2017). The results revealed that only 8% of students from 52 different countries were able to achieve the highest level of CPS proficiency (OECD, 2017). Furthermore, CPS skills have become increasingly important for solving the complex problems faced by today's society (Graesser, 2018b; Griffin & Care, 2014). As the expertise needed to solve these complex problems is spread among many people, collaboration has become essential (Rosen, Wolf & Stoeffler, 2020). This therefore highlights the need for the widespread teaching and assessment of CPS skills (OECD, 2017; Sun et al., 2020). Consequently, research on understanding and measuring CPS skills has increased considerably (Sun et al., 2020). However, there is still little research on developing CPS skills among elementary-school students (Rojas, 2021; Graesser, 2018a; Graesser, 2018b; Rosen, Wolf & Stoeffler, 2020).

CPS skills are defined as the capacity to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution, and pooling their knowledge, skills and effort to arrive at that solution (OECD, 2017). Another possible definition is "Collaborative problem solving can be defined as a joint activity where dyads or small groups execute a number of steps in order to transform a current state into a desired goal state" proposed by the Assessment and Teaching of Twenty-first Century Skills (ATC21S, Hesse, Care, Buder, Sassenberg, & Griffin, 2015) However, the two initiatives are very similar in terms of their construct (Scoular, Care & Hesse, 2017). In this sense, both agree that CPS involves the integration of cognitive and social skills (Sun et al., 2020). The cognitive dimension refers to problem-solving skills and is related to how students manage the task (i.e., exploring, representing, planning and monitoring the problem). The social dimension refers to the collaborative aspect of CPS and is related to group organization and coordination (i.e., establishing and maintaining a shared understanding, taking appropriate action to solve the problem, establishing and maintaining team organization) (OECD, 2017; Sun et al., 2020).

From a socio-cognitive perspective, in order to solve problems, students need to develop a set of skills that will help them direct their learning process. This is known as self-regulated learning (SRL) and involves setting objectives, monitoring, regulation and control over sources of both personal (i.e., cognitive, motivational and emotional) and behavioral or socio-environmental influence (Zimmerman, 2013). In a classroom setting, regulation is usually supported by the teacher. However, regulation can also be supported by peers or, increasingly, by technology. In this sense, technology offers huge potential by providing rich contexts in which to support the functional coordination and construction of knowledge (Dillenbourg, Jarvela & Fischer, 2009; Miller & Hadwin, 2015). SRL can also be transferred to a context of collaborative learning, a concept known as socially shared regulation of learning (SSRL) (Järvelä, Jarvenoja, Malmberg, & Hadwin, 2013). Based on the model for a collaborative task proposed by Hadwin, Järvelä & Miller (2018), SSRL refers to the group's metacognitive control of the task, specifically when groups regulate together as a collective i.e., co-construct knowledge, align monitoring perceptions, and monitor and evaluate progress (Järvelä et al., 2013). Within this same context, it is also possible to define the concept of Co-Regulated Learning (CoRL). This refers to the regulatory mechanism that is stimulated when switching between self-regulation and shared regulation (Lim & Lim, 2020). CoRL frequently occurs during social interactions that are supported or prompted by and with others, such as scaffolded activities (Järvelä & Hadwin, 2013). Therefore, in the context of CPS, both the OECD framework (OECD, 2017) and ATC21S (Hesse et al., 2015) agree that students who are faced with solving a collaborative problem must not only deal with the cognitive aspects of the problem-solving process; they must also deal with the motivational, emotional, and social aspects of the process (Lyons et al., 2020). For example, previous studies have reported that emotions can be related to performance on CPS tasks (Camacho-Morles, Slemp, Oades, Morrish, & Scoular, 2019). Indeed, Camacho-Morles et al. (2019) demonstrated that emotions such as enjoyment, boredom, and anger can influence student performance on CPS tasks as they affect their motivation to invest effort in the activity. Consequently, Pekrun (2017) suggests that positive emotions are related to the use of advanced SRL strategies, such as planning, monitoring, elaboration, and critical thinking.

Negative emotions, on the other hand, are related to the use of shallow SRL strategies, such as rote memorization and recall (Pekrun, 2017; Buono, Zdravkovic, Lazic & Woodruff, 2020). As a result, a relationship can be drawn between emotions and self-regulated learning strategies during the learning process.

Given the above, and considering that many students do not have the necessary regulation skills and are not able to communicate effectively in order to solve collaborative tasks, finding tools to support these regulation skills is essential (Järvenoja, Järvelä & Malmberg, 2020). In this sense, although previous studies have developed technological tools to support SRL, few of them have focused on developing tools to support students' regulation skills in a collaborative setting (Järvelä et al., 2015).

Within the field of Computer Supported Cooperative Learning (CSCL), there are two tools that have been used extensively and appear to support regulation (i.e., Self-regulation, Shared-regulation & Co-regulation; Yilmaz & Yilmaz, 2020; Hadwin, Bakhtiar & Miller, 2018). One such tool is known as a collaborative script, which defines the rules and instructions needed in order to guide and help students behave during CSCL so that everyone benefits from the process (Hadwin, Bakhtiar & Miller, 2018). Fischer et al. (2013) distinguish between internal and external collaboration within a collaborative script. The former refers to flexible cognitive structures that are based on the knowledge of specific collaborative practices. The latter helps learners overcome dysfunctional internal collaboration scripts by providing the external information needed to engage in productive collaborative learning processes. The authors therefore argue that a failure to engage in high-level collaborative processes indicates a lack of internal collaborative scripts. Collaborative scripts can be understood as a means of providing scaffolding or guidance for students in terms of the learning activities, the roles they require, and the order in which specific tasks must be completed (Popov et al., 2019). Moreover, recent studies have shown that collaborative scripts have a positive effect on learning and collaborative skills (i.e., CPS skills). However, there is also a suggestion that collaborative scripts may have a negative effect on motivation (Dillenbourg, 2002). In this sense, the meta-analysis conducted by Radkowitsch, Vogel & Fischer (2020) did not find any evidence to suggest that collaborative scripts negatively affect student motivation. Instead, they recommend increasing the level of freedom afforded by scripts or adapting them to the learners' needs (Rau et al., 2017) in order to limit the possible negative effects of CSCL scripts on learner autonomy.

The second tool for supporting regulation skills is known as group awareness, which is defined as a student's knowledge of their peers' current status and level of commitment. This can help groups coordinate and complete the relevant parts of a shared task (Yilmaz & Yilmaz, 2020). En este sentido, previous studies (Li, Li, Zhang & Li, 2021) have shown that students who participate collaboratively in CSCL environments lack the necessary perception of the other members of the group if no specific attention is paid to this aspect, making it difficult to construct knowledge in such environments (Janssen & Bodemer, 2013). Consequently, research has revealed that group awareness may promote equal participation (Pifarré, Cobos & Argelagós, 2014), support better coordination and regulation of group activities (Janssen, Erkens & Kirschner, 2011), facilitate the knowledge needed to complete collaborative tasks (Dehler, Bodemer, Buder & Hesse, 2011), and promote knowledge construction (Li, Li, Zhang & Li, 2021). Moreover, the literature also suggests that there are three types of group awareness considered crucial for effective collaborative learning: behavioural, cognitive, and social awareness (Bodemer & Dehler, 2011). As Ma, Liu, Liang & Fan (2020) suggest, information on teammates' activities, such as what they are doing or what they will do later, is an element of behavioural awareness (Janssen et al., 2011). Cognitive awareness includes information on the teammates' knowledge, beliefs or goals (Chavez & Romero, 2012). Finally, social awareness refers to information on the participants' perception of how well the group is working together (Bødker & Christiansen, 2006).

Our study looks to address the lack of research into understanding and observing the effects of combining a collaborative script and group awareness in order to support SRL (Schnaubert et al., 2020). Furthermore, the current vision of the learning process involves the interaction of other processes (i.e., cognitive, motivational, emotional, and social; Zimmerman & Schunk, 2011). Developing and implementing innovative technologies in order to aid successful learning and regulation is therefore essential (Noroozi, Järvelä & Kirschner, 2019). Consequently, the use of game-based learning has generated significant interest due to its ability to motivate and increase student awareness, teach knowledge, change behaviors, and even improve skills (Calvo-Morata, Alonso-Fernández, Freire, Martínez-Ortiz, & Fernández-Manjón, 2020).

One way to enhance the quality of interactions during gameplay is to script the collaboration. This imposes certain constraints that can help increase the level of interaction (Van der Meij, Veldkamp & Leemkuil, 2020). Furthermore, when playing a collaborative game, awareness of a peer's current state (i.e., their location, abilities, status, etc.) is essential for achieving both individual and group goals (Teruel, Condori-Fernandez, Navarro, González & Lago, 2018). The script must therefore give players the freedom to communicate in a way that is not overly restricted (Van der Meij et al., 2020). In this sense, over-scripting, where the communication becomes too heavily constrained by the rules that are imposed (Clark, Tanner-Smith & Killingsworth, 2016; Dillenbourg & Jermann, 2007), should be avoided.

In summary, despite recent progress in the measurement of CPS skills, further research is required in order to develop these skills among students. A first step would therefore be to find methodologies that enhance not just the cognitive aspect of these skills, but also allow students to effectively regulate their emotions, motivation and behavior. This includes both their own emotions, motivation and behavior, as well as those of their fellow group members or the group as a whole. Doing so is a fundamental requirement of collaboration. In this sense, further studies are needed to show how the design elements of a collaborative environment influence the student experience with CPS (Dindar, Järvelä, & Järvenoja, 2020). However, there are very few studies that look at the motivational potential of games involving a collaborative script and group awareness for building CPS skills among elementary school students. Our research therefore looks at how elementary school students' regulation skills and emotions are supported by a collaborative game involving a collaborative script and group awareness tools. Consequently, we conducted a quasi-experimental design to answer the following research question:

RQ: "How does a game-based collaborative script that scaffolds group awareness affect students' regulation skills and emotions during collaborative problem solving?"

4.2. Methodology

4.2.1. Participants

Table 4-1: Participant per school, gender and group

			Students					
			Intervention			Focus Group		
Group	**School**		**Boys**	**Girls**	**Total**	**Boys**	**Girls**	**Total**
Control		1	19	13	32	4	4	8
		2	34	38	72	4	4	8
	Total Control		53	51	104	8	8	16
Experimental		1	35	26	61	4	4	8
		2	31	27	58	4	4	8
	Total Experimental		66	53	119	8	8	16
	Total		119	104	223	16	16	32

Note: School column refers to the school to which the students belong; Group refers to the group, either experimental or control, to which the students were assigned; Intervention refers to the number of students assigned to each group during the intervention; Finally, Focus group refers to the number of students from each group that participated in the focus group.

The sample (See Table 4-1) included 223 students (119 boys, 104 girls) aged between 10 and 13 (M:10.65, SD:.56). The students were all in sixth grade at two different schools. Both schools belong to the same school group, where they share the same methodologies, resources and assessments. The students were of medium and medium-high socioeconomic status (SES) and participated voluntarily in the study with consent from their parents and authorization from the schools. In the 2015 Pisa test for CPS Chile obtained 457 points, while the average was 500 (OECD, 2018).

4.2.2. Collaborative Game

We designed a game that uses a collaborative script to scaffold group awareness, based on the game design theories described in section R.1 (Appendix R). This includes design principles for achieving a state of flow (Kiili, De Freitas, Arnab, & Lainema, 2012), and the theoretical framework for designing educational videogames (Plass, Homer & Kinzer, 2015).

Furthermore, the game was developed based on the CPS skills matrix proposed by OECD (2017) and the conditions required for a collaborative activity (Szewkis et al., 2011). This includes the existence of a common goal (Dillenbourg, 1999), positive interdependence between peers (Johnson & Johnson, 1999), coordination and communication between peers (Gutwin and Greenberg, 2004), individual accountability (Slavin, 1996), awareness of peers' work (Janssen et al., 2007) and joint rewards (Axelrod & Hamilton, 1981). However, one of the main challenges when designing games for learning is balancing the playful experience with the expected learning objectives (Quinn, 2005; Qian & Clark, 2016). In the case of CPS, one difficulty highlighted by previous studies is keeping each of the twelve skills from the CPS matrix (OECD, 2017) separate and scripting them independently. Interdependence among dimensions is an inherent challenge when developing CPS dynamics (Funke, Fischer & Holt, 2018; von Davier, Hao, Liu & Kyllonen, 2017). In our case, considering the game had six levels, the twelve skills proposed by the OECD were grouped based on these six levels (Table 4-3). This was done according to the complexity of the problem and the level of interdependence it required (Table T.1, Appendix T). Furthermore, the skills in one level tended to depend on the skills in the previous level.

Our game consists of 46 tasks (Appendix S) and requires 3 peers to solve the tasks collaboratively (Section R.2, Appendix R) while working simultaneously on separate devices. As a result, the game requires effective coordination of individual and group processes in terms of both collaboration and problem solving. Furthermore, the game also includes a set of basic actions and abilities that can be used by the players, as well as the general rules defined by the collaborative script. Group awareness is enabled by the different displays described in Table 4-2. It is important to note that the collaborative script and group awareness are closely related as each depends on the other, while combining the two is essential for successfully completing the game (Schnaubert et al., 2020).

The level of interdependence and communication within each team is also highlighted by including a feedback system (prompts) represented by a non-playing character. This system provides the players with distributed information, i.e., Player A receives key information based on Player B's progress, meaning that they have to communicate with each other in order to share said information and make progress within the game.

Table 4-2: General game characteristics

Collaborative Script			**Group Awareness**	
Basic Actions	Roles	Game Rules	Display	Prompt
- Move (Left, Right) - Jump - Activate & shoot - Use item - Activate & deactivate special ability - Recover health and/or energy	- 3 different roles, one for each participant - 1 special ability for each role	Rules to advance between levels - Reach the end of the stage together - Gradual sequence of tasks that increase in difficulty. Health, energy, and experience rules - Energy, health and experience are shared by the group	Avatar is differentiated based on the role Graphic elements that indicate the need for actions and ability of another players. Graphic elements that indicate the need for simultaneous actions with another player. Graphic elements that indicate the status of shared resources (Health, Energy, Experience Points)	Prompt invites reflection on team strengths and weaknesses. Prompt suggests that you ask for abilities from other teammates Prompt invites you to communicate with teammates to coordinate actions Prompt recommends you to communicate and negotiate solutions or overcome obstacles within the game

- Inventory items	- There are locations to recover energy - There are locations to recover health Common action rules - Move forward and backward freely within the level Ability roles - The shared energy decreases when players use their special ability - Prompt appears after a certain amount of time without making progress	Signalling mechanisms Display to other teammates when player is using special ability Shared map with individual camera for each user, shows in real time what participants are doing if they are in the same location

The tasks presented to the students maintain a similar level of challenge in terms of the collaborative dynamics that are required (Dindar et al., 2020). This was built into our collaborative script in order to regulate the learning process for each dimension of CPS based on the students' zone of proximal development. The CPS skills are therefore presented to the students in 6 stages of increasing levels of difficulty (Table 4-3).

Table 4-3: Game Characteristics by Stage

Stage	**Objective**	**Collaborative Script (Level advancement rules)**	**CPS Skills Required (Predominant)**	**Difficulty**
1	Students have to familiarize themselves with the environment and basic actions within the game. No complex cognitive tasks are required. They have to solve simple tasks and understand that individual performance is key to solving the group task given the existence of a common goal.	- Perform basic actions to reach the end of the stage. - Overcome obstacles that require specific role-based abilities. - Perform role-specific actions.	(A2) Discovering the type of collaborative interaction to solve the problem, along with goals. (A3) Understanding roles to solve problem	1
2	Students have to solve a problem with a common goal. The problem consists of individual tasks that require the formulation of hypotheses about the other players by analysing the information they provide in order to solve the problem. They have to recognize the impact of the decisions made by the other team members and use this information to solve the task.	- Perform basic actions to reach the end of the stage. - Overcome obstacles that require specific role-based abilities. - Perform role-specific actions. - Overcome obstacles that require coordinated action.	Skills from previous level (B1) Building a shared representation and negotiating the meaning of the problem (common ground) (B2) Identifying and describing tasks to be completed	2

			(C2) Enacting plans	
3	Students have to solve a problem with a common goal. The problem consists of individual tasks that require information from the other team members. Students must recognize that other team members may have important information, based on their own perspective, that is needed to solve the problem and the task.	- Perform basic actions to reach the end of the stage. - Overcome obstacles that require specific role-based abilities. - Activate actions specific to your role - Overcome obstacles that require coordinated action. - Use shared abilities with other teammates in a coordinated manner. - Share your special abilities in coordination with other teammates.	Skills from previous level (A1) Discovering perspectives and abilities of team members (B3) Describe roles and team organization (communication protocol/rules of engagement)	3
4	Students have to solve a problem with a common goal. The problem can be solved using multiple strategies; students must reach a consensus regarding the most appropriate strategy. Individual tasks are also presented, while maintaining information exchange and interdependence. The	- Perform basic actions to reach the end of the stage. - Overcome obstacles that require specific role-based abilities. - Perform role-specific actions. - Overcome obstacles that require coordinated action. - Make decisions on how to move forward together	Skills from previous level (C1) Communicating with team members about the actions to be/ being performed (C3) Following rules of engagement, (e.g., prompting other	4

	cognitive level of the task increases. Students have to analyse possible strategies for solving a problem, evaluating each one in order to reach a consensus on the most appropriate strategy.	when following different paths.	team members to perform their tasks.) (D2) Monitoring results of actions and evaluating success in solving the problem.	
5	Students have to solve a problem with a common goal. The problem can be solved using multiple strategies; students must reach a consensus regarding the most appropriate strategy. Individual tasks are also presented and assigned to each participant by the group, based on each player's strengths and weaknesses. The cognitive level of the task increases and requires greater synchronization.	- Perform basic actions to reach the end of the stage. - Overcome obstacles that require specific role-based abilities. - Perform role-specific actions - Overcome obstacles that require coordinated action. - Make decisions on how to move forward together when following different paths - Overcome more complex obstacles by combining the above rules.	Skills from previous level (D3) Monitoring, providing feedback and adapting the team organization and roles	5
6	Students have to solve a problem with an ill-defined common goal. This involves creative and CPS , role assignments, monitoring,	- Perform basic actions to reach the end of the stage. - Overcome obstacles that require specific role-based abilities.	Skills from previous level (D1) Monitoring and repairing the	7

and coordination. The cognitive load of this task is higher as it requires creative skills, as well as collaborative work and coordination in order to come up with and implement a solution.	- Perform role-specific actions. - Overcome obstacles that require coordinated action. - Make decisions on how to move forward together when following different paths - Overcome more complex obstacles by combining the above rules. - 6 different solutions to complete the level. - 25 different locations between which the participants can move independently or together.	shared understanding

Note: The CPS Skill Required column refers to the predominant CPS skills required by the game stage, based on the OECD framework (OECD, 2017); The Difficulty column refers to the estimated difficulty for the game stage based on what is set by Funke et al. (2018).

To set the difficulty level for each stage of the game (Table T.1, Appendix T), we considered three main aspects from the proposed by PISA for modelling collaborative tasks (OECD, 2017): 1) the level of interdependence among the players; 2) the information available to each player when starting the tasks; 3) whether information on the other players' actions is explicit or implicit. We also considered the criteria proposed by Funke et al. (2018): 1) whether or not an expected outcome is given; 2) whether the methods for solving the problem are familiar or unknown to the user; and 3) whether the data needed to solve the problem are given or incomplete.

4.2.3. Data collection and instruments used

A scale measuring attitudes towards collaboration was used to analyze the effect of the game on the students (Cronbach's $\alpha = .912$), adapted from Hwang, Yang & Wang (2013) and Ødegård (2006). This scale included 14 items (Table V.1, Appendix V), which the students evaluated during the first session (before the intervention had started) and the final session (see Section 4.2.4). A 1 on the scale showed complete disagreement with the sentence, while a 5 showed complete agreement.

Given their relationship with student self-efficacy (Travis & Bunde, 2020) and as a means of measuring academic performance, the students' grades from the corresponding school year (GPA) were also obtained. These were provided directly by the schools having received consent from the students' parents or guardians.

Finally, the students participated in a group interview, during which they answered a series of questions designed by the research team. The aim of these questions was to understand the students' experiences and opinions of group activities that they do in and outside of school, as well as with the game (in the case of the experimental group) (Appendix U). These interviews were recorded and subsequently transcribed.

4.2.4. Procedure

An intervention was carried out across a total of 5 sessions (Figure 4-1), held on consecutive days, and lasting 60 minutes each. During the first session, the participants signed an informed consent form and completed a pre-test on attitudes towards collaboration. The classes at each school were randomly assigned to either a control or experimental group. Then, during the second session, the experimental group was divided into random sub-groups of 3 students each. The sub-groups played the game together on individual tablets. Team members were seated together so that they could talk to each other. The focus of the study was to understand the dynamics of collaborative learning that stem from the interaction prompted by the collaborative script. The experimental group therefore worked collaboratively, scaffolded by the game. The students in the control group attended regular classes with the corresponding subject teacher, without playing the game or receiving any training or instructions relating to collaboration. During the fifth session, all of the participants completed the post-test on attitudes towards collaboration. Finally, a week after the intervention, a 30-minute focus group was held at each school with 8 students from the experimental group, and another with 8 students from the control group (32 students in total). The participants for each focus group were chosen at random and asked a series of questions developed by the research team so as to structure the conversation (see Appendix U).

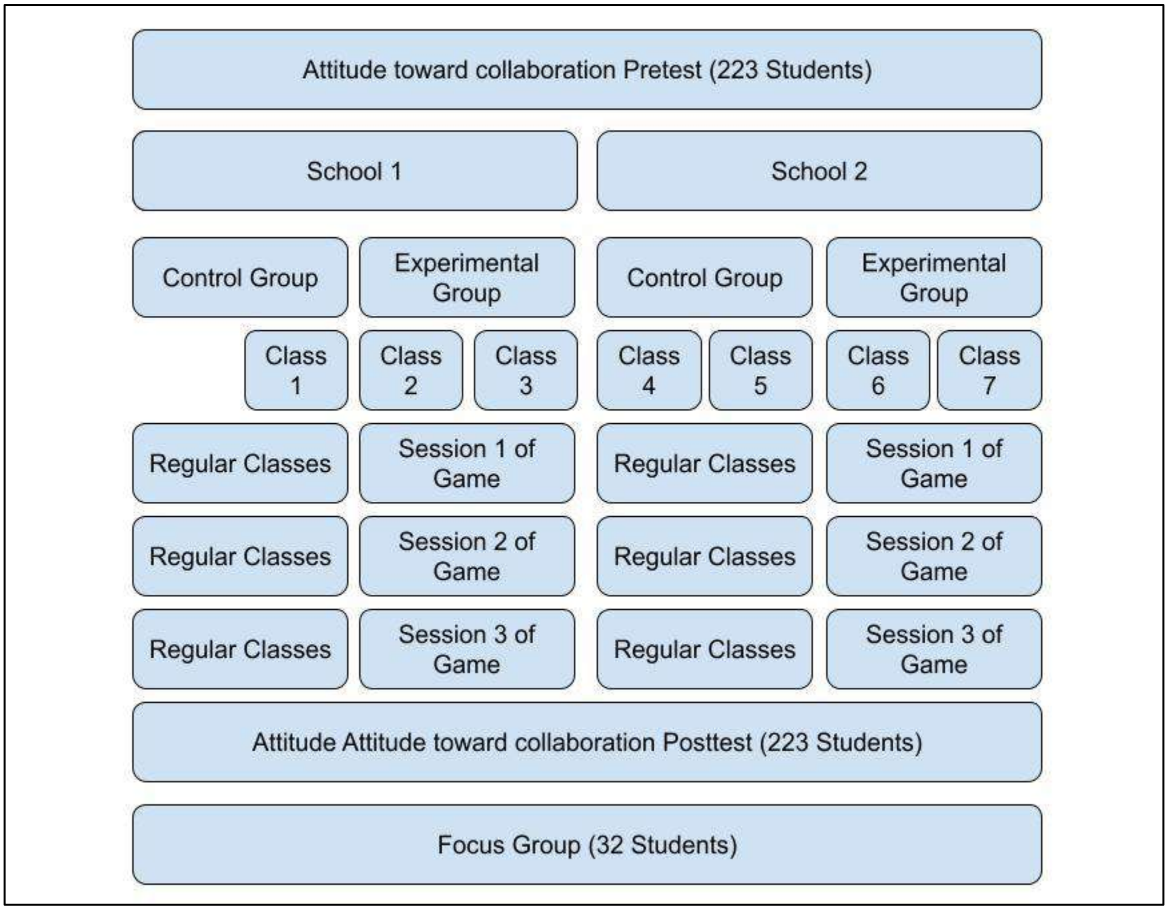


Figure 4-1: Procedure of the study.

4.2.5. Ethical considerations

The present study was approved by University's ethics committee. Following this, the schools were then invited to participate in the study and gave their written authorization to do so.

Before participating in the first session, the students' parents were provided with information on the purpose of the research, the specific activities it involved and the duration of these activities, as well as the risks and benefits of participating. The voluntary nature of participation in the study was highlighted. Furthermore, the students' parents were asked to sign an informed consent form. At the start of the first session, the students whose parents had provided informed consent were told about the activities in which they would be participating. Following this, the students were then invited to sign an informed consent form and reassured that they could stop participating in the activity whenever they liked.

Finally, two additional informed consent forms were sent to the participants of the focus groups and signed by the students and their parents.

4.2.6. Qyantitative Method

Statistical analysis was performed using R (version 4.1.1) (R Core Team, 2021). Descriptive statistics were reported per group as a mean and standard deviation for quantitative variables and as a number and percentage (%) for categorical variables. The significance of differences in quantitative variables meeting the normality and homogeneity of variance assumptions was evaluated using t-tests. Categorical variables were evaluated using Pearson's chi-squared test. A p-value less than 0.05 was considered statistically significant.

Linear regression was proposed for analyzing the association between the score for final attitude and the variables intervention, initial attitude score, GPA and gender. Mathematically, this model is written as follows:

$$Y_i = \boldsymbol{X}_i^{\top}\boldsymbol{\beta} + \epsilon_i, \qquad (1)$$

where Y_i represents the final attitude score of the ith student and $\boldsymbol{X}_i = (X_{1i}, X_{2i}, \dots, X_{pi})$ is her/his variable vector with coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$. Finally, ϵ_i denotes the error term and follows a Normal(.σ^2) distribution.

4.2.7. Qualitative Method

The group interviews were recorded and transcribed for subsequent structured content analysis using the NVivo software package (NVivo(X), QSR). Unlike open coding, the use of structured analysis allowed us to analyse the students' responses based on pre-defined concepts.

The analysis followed a recursive structure (McAlister et al., 2017), in which the researchers coded two transcripts based on a single code for each iteration. The research concepts and methodological decisions were aligned and a codification matrix (see Table 4-4) was elaborated and iterated as the codification process was developed. Every iteration of analysis was based on one of the codes. Once the concept was agreed between investigators, a transcript was analysed using the given code. If necessary, the concepts were again realigned and the transcripts were swapped among the researchers. By doing so, each researcher completed at least four iterations per category. The format of the interviews did not allow for extensive nor overly-complex input from the participants. The unit of analysis was therefore set as each turn taken by the participants during the interview. The interrater reliability between coders was obtained through Cohen's Kappa coefficient, with a 98.9% of agreement, while Cohen's k coefficient was 0.6, indicating a substantial agreement.

Table 4-4: Matrix for coding and analysis

Code Name		Description	Criteria	Excludes
Regulation	Self-Regulation	An individual regulates his/her own thinking, behaviour, motivation, and emotions in the joint task	Conversation turn includes a reflection about the users thinking, behavior, motivations or emotions in relation to a collaborative task	Reflections about thinking, behavior, motivation or emotions not related to a collaborative task
	Co-regulation	Individual(s) temporarily guide, prompt, nudge and support each other's self-regulation	Conversation turn includes a narration or reflection on how group members momentarily coordinate individual	Narration or reflection on how group members momentarily coordinate individual

	in the joint task	processes regarding a collaborative task	processes outside a defined collaborative task
Shared-Regulation	The group collectively regulates their thinking, behavior, motivation, and emotions in the joint task	Conversation turn includes a narration or reflection on how the whole group regulated their thinking, behavior, motivations or emotions towards the task at hand	Narrative or reflection on how the whole group regulated their thinking, behavior, motivations or emotions outside a defined collaborative task
Group Awareness	A group member deliberately prompts or asks information about cognitive, behavioral or social processes in other group members during a collaborative task	Conversation turn narrates or reflects upon a process where there is information exchange about members cognitive, behavioral or social processes during the task	Process where there is information exchange about members cognitive, behavioral or social processes with no task at hand.
Collaborative Script	Activities related to a set of rules that prescribe a sequence of collaborative actions,	Conversation turn includes the description of a task with a set of rules and instructions that structure a form of collaborative task	Description of tasks with no rules for structured collaboration. Description of structured

				collaboration with no defined task
Emotion	Anger	A sustained and strong negative emotional state of a group member during a collaborative process.	Conversation turn refers to a sustained or extreme negative emotional state of a group member during a collaborative process.	Either short or low intensity emotional negative states related to tasks.
	Boredom	An inactive emotional state derived from a low valuation of a task or process and a low arousal from the challenge derived from it.	Conversation turn refers to an inactive emotional state that may derive from a low valuation of a task or process.	Mentions on inactive emotional states that do not derive from a specific task or do not convey a low valuation of a process
	Excitement	An active and positive emotional state aroused from the challenge or intrinsic characteristics of the task	Conversation turn refers to a positive emotional state of joy and expectancy.	Short term satisfaction derived from specific tasks or processes.
	Frustration	Negative emotion aroused upon encountering	Conversation turn refers to a negative emotional state or process evoked by the impossibility of	Sustained negative emotional states evoked by external or long-termed

	an obstacle in the achievement of a task, goal, or expectation, or in satisfying one's needs.	performing necessary tasks	processes within an activity.
Happiness	General state of well-being	Conversation turn refers to a prolonged and general state of well-being related to a collaborative activity.	Short positive emotions aroused by the interaction with a collaborative task.
Indifference	General and prolonged state derived from an inactive emotional state and low valuation of the task	Conversation turn refers to a prolonged inactive and negative emotional state.	Short termed inactive and negative emotional state derived from a specific part of the task.
Interest	General and prolonged state of positive valuation of a task, process or activity that leads to intrinsic motivation	Conversation turn refers to the positive valuation of a task, process or activity.	Positive valuation of elements that do not refer either to specific tasks, processes or activities.
Sadness	General and prolonged state in a negative and inactive emotion.	The conversation turn refers to a prolonged and general state dominated by a negative and inactive emotion.	Short term negative emotions aroused by specific task or internal

	Satisfaction	Positive emotion aroused upon surpassing an obstacle and gaining control of a task, goal, Expectation or process.	Conversation turn either refers to a positive emotion aroused upon surpassing an obstacle and gaining control of a task, goal, Expectation or process.	Prolonged states of general well-being or positive emotions derived from extrinsic elements of an activity.
	Shame	Negative and inactive emotion derived from a global negative criticism of oneself. (Wilson, Shame and collaborative learning in second language classes)	Conversation turn elaborates on a negative emotional state derived from a global negative perception of oneself abilities towards a task or process.	Negative and inactive emotions derived from task elements.
	Worry	Active and negative emotional state derived from a low level of confidence in achieving a highly valued task.	Conversational turn expresses a negative and active emotional state derived from a high expectancy of the task outcome and a low confidence in the abilities and extrinsic conditions to achieve a specific task or process.	Active and negative emotional states not derived from expectancy of achievement or low confidence about a highly valued task.

The first step in the coding process was to identify moments in which the idea of group awareness was touched on by the participants. Bodemer & Dehler (2011) propose a model of group awareness in which this factor can be found in three different instances. Firstly, group awareness can be found in the behavior of the group members, particularly when providing their reflections on the tasks and activities that are being carried out by the other members of the group. Secondly, group awareness can also be found in the expression of the beliefs, ideas and knowledge held by the other group members. Finally, group awareness is considered a form of social awareness, i.e., each member of the group's awareness of the group dynamics. Based on this information, the students can then constantly monitor and modify the organization of the group. We chose this model as it allowed us to integrate the sort of group awareness that occurs naturally as a by-product of the interaction with the awareness that comes from the scripts embedded in the game. Furthermore, the model allowed us to code the students' conversations based on structured patterns of turn-taking. Doing so allowed us to explicitly show the elements that raised awareness, as well as the process the students followed when said elements were introduced by the collaborative script. This is because Bodemer & Dehler's (2011) framework identifies these two instances.

Following this, any references to any kind of collaborative script were then coded by the researchers. As mentioned in the introduction (Section 4.1), a collaborative script is a system that defines the rules and instructions that are needed to assist and guide students during a collaborative task (Hadwin et al., 2018). Analysis was based on the speech units, which focused on the gameplay and audiovisual resources used to guide the collaboration. A collaborative script focuses on the moments in which the system attempts to scaffold the interaction between the users, rather than the content-related dynamics of the activity (Fischer et al., 2013). In this sense, within the students' discourse we were able to identify descriptions of elements of the game that were meant to make them aware of their group and regulate their joint efforts.

The next category for analysis was regulation (Table 4-8), based on the model proposed by Hadwin, Järvelä & Miller (2018). Regulation in this model is socially situated, while human agency takes on a central role. In this sense, each agent is understood to have their own purpose, intentions and objectives, which will not always necessarily align with the group's objective. Given this, there are three distinct modes of regulation in highly-interactive and collaborative contexts: SRL, CoRL and SSRL. Any instances in which the participants referred to group regulation were therefore coded using these three categories. While self-regulation can be found in both individual and group conditions, co- and shared-regulation can only be elicited in group conditions. In our case, the shared context for both groups was the classroom. The aim was to understand the social situation in which the students in our study showed their regulation skills. This could then be contrasted with the social situation that was also mediated by the collaborative script. However, coding and classifying the discourse is not always straightforward, especially when it comes to distinguishing between self-regulation and co-regulation in group contexts. This has previously been highlighted as a cause of confusion in the literature (Hadwin et al., 2018, p.93). In our case, the distinction was mainly made by identifying the transitional nature of co-regulation. Whenever a student highlighted regulatory processes that ultimately led to positive or negative consequences that affected only them personally, this was coded as self-regulation. This was the case of one participant, who showed self-regulation of certain emotional processes:

"if you're on a team of more than four, and they kill you, you don't have to get angry, because it was just a joke. Instead, if you get angry they're going to think you're a bad person…" (Student 3, Control, School 2)

In this case, although the participant goes through a regulation process within a group setting, the effects only impact them individually, without affecting the group's collective goals. A student from the experimental group at the same school, on the other hand, understood the process as a whole and commented on the support that they needed from their classmates.

"In the game I was always falling over or couldn't go on, and my classmates always had to go back because each had their own special ability and had to help the others." (Student 2, Experimental, School 2).

In this case, the participant shows their awareness of the fact that help and support was distributed across the team members. By doing so, they understand the transitional nature of the regulation, which enables or restricts their ability to complete the tasks at hand, while also allowing for the presence of other kinds of regulation (Hadwin et al., 2018, p.94).

This model (Hadwin et al., 2018, p.94) also understands that regulation is a multifaceted process. Consequently, coding is not only based on the presence of moments of meta-cognition. Instead, there are also other cognitive, motivational, behavioral and emotional factors at play. As Isohätälä, Näykki & Järvelä (2020) have shown, it is possible to distinguish cognitively-oriented and socioemotionally-oriented tasks from task-focused interaction in collaborative learning. In this sense, it is not only essential to acknowledge the importance of regulating emotions, but also acknowledge how collaborative learning processes are activated (Malmberg, Järvelä & Järvenoja, 2017). Given this, we decided to search for emotions by operationalizing a series of categories for classifying them (Table 4-10). This was based on the work by Reis et al. (2018), who analyzed more than 58 studies on affective states and socio-emotional factors in computer-supported collaborative learning. Their findings established 29 types of emotion displayed by students in computer-supported collaborative learning environments. The heuristic and empirical nature of their research resulted in a useful, although redundant categorization of emotions. For the present study, these 29 emotions were reduced to 11 based on conceptual overlap and repetition, as is the case of the codes "sad" and "sadness". From these 11 emotions a description was made based on a dimensional approach that describes emotional experiences through a small number of affective dimensions, being two the most relevant when it comes to variation in affects: Valence (negative or positive affective state) and Activation (activating, deactivating) (Loderer, Pekrun, Plass, 2020). Furthermore, Chi-square test and Fisher's Exact test were performed to show the association among the variables studied. Next, to show the difference between each pair of columns (i.e., control and experimental group), the column proportions (for each row) were compared using a z test with a conservative Bonferroni method.

Finally, the results of the previous procedure are grouped into two major findings (section 4.3). The first refers to the relationship between regulation and the collaborative script (Section 4.3.2). While the second refers to the emotional factors present in the collaboration (Section 4.3.3).

4.3. Results

4.3.1. Attitude toward collaboration

The descriptive statistics for the continuous variables (i.e., initial attitude, final attitude and GPA) are included in Table 4-5. A series of t-tests were conducted on a sample in order to compare the mean of each variable between the experimental and control groups. In this sense, we can see that there are significant differences in attitudes towards collaboration following the intervention (p-value = .001), but not prior to the intervention (p-value = .148). Finally, there are also significant differences in the GPA scores between the two groups (p-value = .017).

Tab**le 4-5:** Descriptive statistics

	Control Group				Experimental Group				
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max	p-value
GPA	5.818	.651	4.4	7.0	5.624	.558	4.3	6.9	.017
Initial Attitude	4.165	.871	1.143	5.0	4.484	.578	1.571	5.0	.148
Final attitude	4.183	.907	1.0	5.0	4.523	.537	1.429	5.0	.001

Similarly, for the categorical variables group (control or experimental) and gender (girls and boys), we include a crosstab (Table 4-6), revealing no statistically significant differences in the proportion of boys and girls who participated in the intervention (□2(1) = .451; p=.502).

Table 4-6: Crosstab Group vs Gender

Group	Gender		
	Girls	Boys	Total
Control	51	53	104
Experimental	53	66	119
Total	104	119	223

The association between the score for final attitude and belonging or not to the experimental group (the variable intervention), the students' gender, their GPA, and their initial attitude towards collaboration was analyzed using the linear regression model (1). Table 7 shows a summary of the regression parameters for this model.

Table 4-7: Multiple regression summary (Adjusted $R^2 = .274$).

Parameter	**Interpretation**	**Estimate**	**SD**	**95% CI**	**P-value**
β_0	intercept	1.497	.468	(.576, 2.419)	.002
β_1	intervention	.242	.090	(.064, .420)	.008
β_2	initial attitude	.423	.062	(.301, .546)	< .001
β_3	GPA	.167	.077	(.016, .319)	.031
β_4	gender	-.102	.089	(-.278, .074)	.253

This means that, on average, students in the experimental group scored .242 more points on the final attitude test than students in the control group. On the other hand, for every one-unit increase in the score for initial attitude, the predicted value of the final attitude score increases by .423. Similarly, a one-unit increase in GPA leads to an average increase of .167 units in the score for final attitude. Gender was not significant.

4.3.2. The collaborative script allows students to reflect on their regulation processes

As is shown in the Table 8, there is a relationship between the number of reflections on the regulation process made by participants from the experimental group and participants from the control group ($\chi2(2) = 9.391$, $p = .009$). Specifically, we can observe significant differences in the proportion of codes categorized as Self-Regulation between the control group (34.1%) and the experimental group (12.7%). The same occurs with codes categorized as Shared-Regulation, which are more frequent in the experimental group (45.6%) than the control group (24.4%). Although the number of codes categorized as Co-Regulation for the experimental group (33) was higher than for the control group (17), when comparing these as a proportion of the total number of codes for each group, there are no significant differences between the two (41.5% vs. 41.8%, respectively).

Table 4-8: Regulation type per group

	Group		
Regulation	Control	Experimental	Total
Co-Regulation	17_a (41.50%)	33_a (41,80%)	50 (41,70%)
Self-Regulation	14_a (34,10%)	10_b (12,70%)	24 (2.00%)
Shared-Regulation	10_a (24,40%)	36_b (45,60%)	46 (38,30%)
Total	41 (100%)	79 (100%)	120 (100%)

Note: Each cell shows the number of codes identified in the students' responses during the focus groups (Experimental and Control Groups) and categorized as Co-Regulation, Self-

Regulation and Shared-Regulation. The final row Total refers to the total number codes identified in the students' responses for each group, while the column Total refers to the total number of codes identified in the students' responses for each type of regulation in the experimental and control groups respectively. The column proportions test assigns a subscript letter to the categories of the column variable. Thus, different subscript letters (a and b) denote a subset of Group categories whose column proportions differ significantly from each other at the .05 level.

This may be explained in part by the experience with the collaborative script, which provides students with a framework for group awareness and regulation. In order to talk about their experience of playing the game, the students from the experimental group necessarily had to refer to the regulation process. When a participant from the control group was asked about their experience of working in a group, their responses tended to refer more to their general feelings rather than specific group dynamics.

> "Because it's more fun and because you're with your friends, you've got time to have a laugh; to do the work. You get it done quicker. We get together at someone's house and get the work done, then we play or go to the park. We do fun stuff. It's probably better and quicker, and more fun than doing it on your own."
> (Student 6, Control Group, School 1)

Here, the student does not refer to any specific procedures, agreements or actions. Instead, they talk about the pleasure of working in a group ("it's more fun", "you've got time to have a laugh", "we do fun stuff"). The only reference they make to the actual group work refers exclusively to its increased efficiency ("you get it done quicker") and location ("we get together at someone's house and get the work done"). While the student displays a positive attitude towards group work, and collaboration in general, they do not manage to describe the actual processes that this involves. The sorts of regulation process that they adopt therefore remains to be seen.

In comparison, the following response from a student in the experimental group provides a detailed reflection of the strategies that were adopted to regulate the group dynamic. Indeed, the reflection specifically references the roles, tasks and challenges that were set out by the collaborative script.

> "There were different coloured gem stones, so if the red player had the red gem they had to use their powers to open another level or path. And every player had to do the same thing with their gem stone, until we got to a point where we had to split up and each go our own way, before we finally realized that we were in the same place but in different corridors and that if we didn't help each other we wouldn't be able to keep going"
> (Student 7, Experimental Group, School 2)

In this case, the participant refers to the need to collaborate based on the demands, restrictions and requirements of the game. The participant describes the process through which the collaborative script revealed the interdependence among the players and the difficulty of maintaining group awareness. Based on the different colours, the student understands that each character has a different role to play. Furthermore, the task seems to be an individual task until the interdependence among the players is revealed. As such, the student suggests that "if we didn't help each other we wouldn't be able to keep going". In this response, which refers to elements from the collaborative script, we not only find a description of the game but also of the scaffolding process for fostering group awareness.

This hints at another of the benefits of the collaborative script: the level of reflection on the process of self-regulation and shared regulation. The collaborative script forces the students to coordinate their actions so as to meet a shared goal. To do so, the script requires them to balance the other players' decisions with their own individual rules. These only start to make sense once they are considered in association with the group rules. However, there are two interesting observations from the students' comments when they refer explicitly to problems involving the regulation process. The first is that the participants could perceive the scaffolded structure of the collaborative script, clearly distinguishing between the different levels of difficulty presented by the tasks as well as the different regulation strategies needed to complete them. This is highlighted by one of the participants, who refers to the format of the game, in which characters in different locations had to come to an agreement in order to advance.

> "It was also difficult because they were in different rooms, and sometimes someone would do something wrong, they'd connect when they shouldn't and the whole system would crash, and then it would go back to normal, but it's hard to move on when everyone's in a different place"
> (Student 5, Experimental Group, School 2)

The second point of interest is the way in which the groups perceived the problem and went about finding a solution. In the collaborative script, the solution did not come from trying different individual solutions. Instead, the team members needed to evaluate and reflect on each other's status in order to come up with a solution as a group. This is reflected in the students' comments about the game, where, by describing their experience of the problem, they showed their understanding of the need for collective regulation. For example, see the following comment by a student from the experimental group:

> "[...] you need help from your teammates, who are in different places, you need their help in order to get out and meet up, so everyone, because for example there were like walls, and everyone in their own room had to do something to help the other players, so then you started talking to each other and would say "green, open my portal", and then the red player could go through, and the yellow player can move up and then they all met up and could help the green player, and so on."
> (Student 1, Experimental Group, School 2)

This student describes the strategy used to solve the problem, based on organizing a team spread across different locations. This group in particular recycled some of the vocabulary from the game and used it in their own interactions ("green, open my portal"), so as to set off a series of actions required by the collaborative script in order to make progress in the game.

As mentioned previously, reflection on the group work within the control groups tended to focus more on the tasks outcomes in previous experiences and not so much on the process. Satisfaction in the control groups therefore tended to be associated with co-regulation, but mostly expressed in individual terms and without any suggestion as to how these two factors interact with one another.

> "I prefer playing games when you're with others. I prefer ones where you play as a group because I'm normally never any good at these things, like, I always end up falling behind. So, I like it when they wait for me and help me."
> (Student 1, Control Group, School 1)

Thanks to the increased group awareness within the experimental groups, the participants tended to focus more on the process and provide greater reflection on the group work. In simple terms, there were a large number of comments that started with the word "because", which is generally used to introduce reasons or causes (see Table 4-9). There was a significant difference ($t(175) = 20.71.\ p < .001$) of the presence of this word in experimental groups (51 codes) compared to control groups (125 codes). In other words, the collaborative script not only helps the students understand the process of regulation, but also to understand the underlying emotions that were revealed during this process.

Table 4-9: Reflections on the group work (Experimental Group)

School	**Student**	**Comments**
1	5	"Because, like Student 6 was saying, it's more fun and sometimes much easier because everyone does their bit and makes the work more fun, it's easier because everyone does their bit and I think the end result is better. Because when I do it on my own it doesn't work out so well, but when I do it with Student 6 and my other classmates, we always have a laugh

		and play around. They always tell us off, but the work always turns out well."
1	5	"I try, because, for example, we've done work before with other classmates and they've started playing around, and started arguments, and it's not all as fun as it seems, but I think I'm OK because I work with the others if I need to…"
1	7	"Because when you play on your own, for example, you die and they can't help you. But if you play in a group they can go and help you. Everyone had their own special power so if someone got stuck or didn't know where to go then we'd say "Come! Come!" and then everyone would use their special powers and that made it easier."
2	7	"Well, because we all contribute something, and maybe we can do a great job and still have fun"
2	3	"Because I always coordinate with the people I work with, and it's not just work, there's also fun and games"

4.3.3. The function of emotions

As explained by Jarvenoja et al. (2020), emotions can have a considerable and complex effect on the regulation process within a group. Experienced collaborative learners therefore not only actively design strategies for monitoring and managing the cognitive load of the group work, they also do the same for emotions. In this sense, the regulation process required managing the emotions involved in the process of playing the game and learning, as articulated by the collaborative script.

Table 4-10 shows the effect that emotions have on the regulation process within the collaborative script; an association between emotions and the intervention was observed, $\chi2(9) = 16.914$, $p = .05$ (Fisher's Exact Test p-value = .04). Also, we can observe significant differences in sadness and worry, which are greater in the control group (Table 4-10). This may suggest that the exposure to the collaborative script solved emotional issues related to external factors (e.g., friendship among peers) related to collaborative tasks. For example, in the control group of school 2, student 4 declares:

> "Because I forget things and sometimes they want to get together at houses and I cant, they won't let me, and my dads are separated and my dad sometimes is really mean to me because he does not like it".
> (Student 4, Control Group, School 2)

On the other hand, a student from the experimental group of school 2 relates that he is always left out for similar external issues in most games and group activities, but in this game, he could play as well ("but now I played"). In this sense the collaborative script, for being played during school time, seemed to be able to bypass the emotional load students could bring from permanent problems that surrounded their collaborative work experiences.

Table 4-10: Emotions described by each group

Emotion	Group		Total
	Control N (%)	Experimental N (%)	N (%)
Anger	1_a (2.20%)	4_a (7.40%)	5 (5.10%)
Boredom	2_a (4.40%)	5_a (9.30%)	7 (7.10%)
Excitement	1_a (2.20%)	4_a (7.40%)	5 (5.10%)
Frustration	10_a (22.20%)	8_a (14.80%)	18 (18.20%)
Happiness	6_a (13.30%)	9_a (16.70%)	15 (15.20%)
Interest	6_a (13.30%)	5_a (9.30%)	11 (11.10%)
Sadness	4_a (8.90%)	0_b (0 %)	4 (4.00%)
Satisfaction	10_a (22.20%)	19_a (35.20%)	29 (29.30%)
Shame	1_a (2.20%)	0_a (0%)	1 (1.00%)
Worry	4_a (8.90%)	0_b (0%)	4 (4.00%)

Indifference	0_a (0%)	0_a (0%)	0 (0%)
Total	45 (100%)	54 (100%)	99 (100%)

Note: Each cell shows the number of times each emotion was coded in the students' responses in the different focus groups (Experimental 1, Control 1, Experimental 2 and Control 2) based on the emotions of Anger, Boredom, Excitement, Frustration, Happiness, Indifference, Interest, Sadness, Satisfaction, Shame and Worry. The column 'Control' shows the total for both control groups, while the column 'Experimental' shows the total for both experimental groups. The column proportions test assigns a subscript letter to the categories of the column variable. Thus, different subscript letters (a and b) denote a subset of Group categories whose column proportions differ significantly from each other at the, 05 level.

In the case of the game-based collaborative script that was developed for this study, the emotions most frequently described by the students in both experimental groups when it came to group awareness were satisfaction (19), happiness (9) and frustration (8) (Table 4-6). While happiness tends to refer to overall perceptions of collaboration in general, satisfaction and frustration refer directly to more specific tasks. In this sense, there is an interesting dynamic between these two emotions when developing CPS skills. Indeed, balancing these two emotions plays an important role in achieving the desired state of flow (Costikyan, 2013; Juul, 2013). The aim of the game was to foster the students' CPS skills. The design therefore looked to gradually increase the complexity of the tasks (Table 4-3) based on 6 levels (Funke et al., 2018), where the information available to each player and regulating each role were considered key (section 4.2.2). The results suggest that the collaborative activity was a satisfactory experience, while the level of difficulty was considered stimulating. Further research is required in order to understand whether the optimum balance was reached, as few groups made it to the final level. This was mainly due to the increasing level of difficulty (Table 4-3; Table T.1, Appendix T) and the low level of CPS skills (OECD, 2018) and may have led to an increased feeling of frustration for some groups. Furthermore, there were also some issues of system stability (i.e., connectivity issues, batteries losing their charge, etc.).

In terms of regulation, the most active participant from the focus groups described the difficulty of working in a group in purely emotional terms. Anger, for example, was mostly mentioned to refer to the emotional outcome a collaborative work might generate. Though no significant results were found involving anger and boredom, it was clear in the qualitative analysis that negative emotions played an important role for maintaining collaboration inside the team.

> "You had to come to an agreement a lot in order to move on to the next level and working with others is difficult because sometimes they get angry and break off from the rest of the group."
> (Student 2, Experimental Group, School 1)

By describing the process of following the collaborative script, the student reveals an understanding of the emotional difficulty of finding a suitable process for regulating group work. In this sense, the system also revealed instances of students being able to make emotional adjustments in order to successfully complete the group task.

> "Before, we didn't want to work with Student 9, so I was like, you three and me don't want to work with Student 9, and we didn't want to work with them, but then we realized that Student 9 wasn't so bad after all, because for example if we couldn't get to the next stage they'd tell us off a bit, but, for example, Student 9 would help us and tell us how to get to the next stage and everything"
> (Student 2, Experimental Group, School 1)

It is worth noting that it is during co-regulation that some of the negative emotions turn into positives (Table 4-11). Although it was not possible to statistically determine the relationship between the variables regulation and emotions, a stronger relationship can be observed in the experimental group that is suggested to be explored in a larger sample. In this sense, it is understandable that in the experimental group satisfaction is the emotion most frequently associated with co-regulation (10), followed by happiness (7). While in the control group, satisfaction (6) is the most frequently associated with co-regulation, followed by frustration which is associated with co-regulation (4) and self-regulation (5) (Table 4-11).

Table 4-11: Relationship between regulation type and emotion for each group

Group	Emotion	Regulation type			Total
		Co-regulation	**Self-regulation**	**Shared regulation**	
Experimental	Anger	1	2	0	3
	Boredom	0	0	1	1
	Excitement	2	0	2	4
	Frustration	4	1	1	6
	Happiness	7	1	3	11
	Indifference	0	0	0	0
	Interest	2	0	1	3
	Sadness	0	0	0	0
	Satisfaction	10	2	4	16
	Shame	0	0	0	0
	Worry	0	0	0	0
	Total	26	8	12	46
Control	Anger	1	1	1	3
	Boredom	0	0	0	0
	Excitement	1	0	0	1
	Frustration	4	5	2	11
	Happiness	1	0	0	1
	Indifference	0	0	0	0
	Interest	1	1	1	3
	Sadness	2	2	0	4
	Satisfaction	6	1	2	9
	Shame	0	1	0	1
	Worry	2	1	0	3

Total	18	12	6	36

Note: Each cell shows the number of times each type of regulation (Co-regulation, Self-regulation and Shared regulation) was associated with the emotions Anger, Boredom, Excitement, Frustration, Happiness, Indifference, Interest, Sadness, Satisfaction, Shame and Worry.

However, from the comments made by the control group, the emotion most frequently associated with the process in their previous collaborative experiences was frustration (10 occurrences, Table 4-10), without actually addressing the steps needed to overcome this feeling. One student from Control Group, School 1 suggested the following:

> "One thing that really annoys me about group work is that you make a decision and then the rest of the group contradicts you and so no one agrees with you and they all agree with the others. So, then it creates more conflict, we take longer, and, as Student 5 says, you can't choose the people you like being with or, for example, sometimes we're all in a group and we see a boy on their own, who doesn't want to be with anybody. So, then I ask if he wants to join us. For example, today we were with Student 2 and Student 3, and we went to ask Student 9 if he wanted to be with us, but he didn't want to be with anyone, and I asked in the group if he could do the drawings, but no one wanted to work with him. Another thing that annoys me: sometimes they don't want to work with the boys because they don't like them, because of this or the other."
> (Student 6, Control Group, School 1)

This comment shows how both internal and external factors can lead to negative emotions: having to reach an agreement on the one hand and personal grievances on the other. It is noticeable how the use of the collaborative script provides a space for interacting, in which external factors become less important, as we saw from Student 2's comments "we didn't want to work with them, but then we realized that Student 9 wasn't so bad after all" (Experimental Group, School 1). In this sense, the game-based design of the collaborative script produces a magic circle (Huizinga, 2014) in which, for example, Student 8 from

Experimental Group (School 2) suggests having enjoyed the game because no one usually plays with him ("but now I played"). In the same way, the negative emotions stemming from internal factors, such as the difficulty of coming to an agreement, find a space where they are rationed out and balanced through the scaffolding included in the design. Increasing the level of complexity (Table 4-2) when generating group awareness made this possible and allowed CPS skills to be integrated into the experience proposed. This was highlighted by the following students from Experimental Group, School 2:

> "Obviously the main thing we learned was how to work in a team, that's what the game wanted us to l earn, and we also learned how to listen to each other and take turns and that you have to have different roles, like for example, a leader, so that you can organize things better."
> (Student 5, Experimental Group, School 2)

> "To work as a team, to come to an agreement, that we have to be united and talk to each other to come to an agreement."
> (Student 8, Experimental Group, School 2)

> "To learn how to work as a team in real life and in games, as well as to listen to each other so we can finish the game and stuff like that."
> (Student 2, Experimental Group, School 2)

> "It was like an example of everyday life, because a lot of the time you don't do things on your own. You're not alone in the world and you can't do everything by yourself."
> (Student 3, Experimental Group, School 2)

4.4. Discussion

We designed a game for developing CPS skills. Our results suggest there is a positive relationship between the intervention and students' attitudes towards collaboration (Table 4-7). Consequently, when analyzing the students' responses from the focus group, the students from the experimental group referred more frequently to the emotion satisfaction (Table 4-

10). This relates to the motivational potential that has been shown by games (Calvo-Morata et al., 2020). Furthermore, the improvement in attitudes towards collaboration among students in the experimental group did not reveal any differences in terms of gender. However, the improvement is explained significantly by their attitude towards collaboration prior to the intervention, as well as their GPA. In this sense, the game design (Appendix R) and inclusion of a collaborative script to scaffold group awareness (Section 4.2.2) were both fundamental in allowing collaboration to take place. The students' socio-emotional interactions were positive, which benefited their individual experiences during the collaborative activities and boosted their self-efficacy and beliefs in their own abilities. These latter findings should be investigated in greater detail (Parker, 2021; Mänty, Järvenoja & Törmänen, 2020).

Previous research has highlighted between game and collaborative scripts (Van der Meij et al., 2020). However, studies have not yet looked at how combining collaborative scripts and group awareness in a game can foster the development of CPS skills. Our results therefore provide evidence of this relationship and suggest that further research should look at how this combination can foster the development of other skills, such as critical thinking, creativity and communication. As mentioned in Section 4.2.2 and detailed in Section R.1 (Appendix R), the game design principals played a fundamental role. More specifically, this included the use of the framework proposed by Plass et al. (2015), balancing difficulty level based on positive interdependence between peers, and the conditions required for an activity to be considered collaborative (Szewkis et al., 2011).

Students who followed the collaborative script also reflected more on the regulation process (Section 4.3.2). The script gave rise of moments of reflection by regulating the need for group awareness in order to complete the task. To a certain extent, this heightened awareness of the regulation process among students who followed the collaborative script was due to the fact that the game provided an environment in which they could explicitly discuss these processes (Van der Meij et al., 2020). The visual and interactive elements of the game therefore equipped the students with the language they need to discuss how to regulate collaboration in groups. In this sense, they were able to describe the roles, tasks and challenges that they faced in the same terms that were provided by the collaborative script. Our results provide evidence to support the existence of a relationship between regulation skills and the effective combination of collaborative scripts and group awareness, further highlighting the need for this kind of study, as suggested in recent publications (Schnaubert et al., 2020).

One of the main effects that the game had on the groups was that it provided the participants with the tools they need to discuss processes such as self-regulation, co-regulation and shared regulation (Table 4-8). In this sense, the participants acknowledged the scaffolded structure of the collaborative tasks in terms of the regulation processes they required. The tasks also provided the students with the opportunity to reflect on the regulation process, which itself proved to be an additional tool for solving collaborative tasks in a digital environment. Our results highlight the collaborative nature of the game as the tasks required more collaborative than individual work. This was thanks to the collaborative script, which increased the level of group awareness required by the game. This finding is also consistent with the recent literature (Järvenoja, Järvelä & Malmberg, 2020).

Another key aspect for ensuring this level of reflection on the regulation process was how the game balanced the difficulty level. The participants were able to distinguish between different levels of difficulty based on the regulation process required by each task (Section 4.3.2). The solution to these tasks could not be found individually through trial and error. Instead, they required the students to reflect on and evaluate their teammate's progress in order to come up with collaborative solutions that required a series of coordinated actions. Our results therefore suggest that finding a suitable balance in difficulty level can foster students' CPS skills (Graesser, 2018b). Furthermore, our research also addresses the lack of studies proposing more gradualness when it comes to the level of difficulty and interdependence that is required (Table 4-3). Until now, no studies have built on the levels proposed by the OECD framework and widely used in the PISA test (OECD, 2017). By achieving this, we are making the teaching or training of CPS more accessible for schools (Scoular & Care, 2018). Future research should therefore look at how the tasks can be better adapted to the students' zone of proximal development.

Furthermore, when following the collaborative script, the emotions most frequently highlighted by the students were satisfaction and frustration (Table 4-10). To this end, the students' comments during the focus group show how a state of flow was successfully achieved, as the students transitioned from an initial sense of frustration towards a feeling of satisfaction upon completing the collaborative task set out by the script (Tyack, Wyeth, & Johnson, 2020). The students also showed a willingness to view problems of group regulation as problems that required a change in emotions. Instances of co-regulation (Table 4-11) were key to this process, as they allowed a lot of the negative emotions to be re-articulated by the groups and converted into positive emotions, thus replicating the state of flow promoted by the game's design. Along this line of reasoning, recent studies have highlighted the effect that emotions can have on students' regulation skills (Dindar et al., 2020). Järvenoja et al. (2020) provided evidence that the regulation of motivation and emotions is an inherent part of the regulation of learning during collaboration. We therefore add to this by providing evidence of how incorporating a collaborative script, more gradualness in terms of CPS Skills (OECD, 2017), and a growing increase in interdependence can effectively balance the challenges of said regulation at different moments during the game.

Finally, as mentioned in the introduction (Section 4.1), there is a need not only to measure CPS skills, but also to find mechanisms that can help develop them (Rojas, 2021). In fact, several factors have been reported as being important for developing CPS skills, i.e., attitudes towards collaboration (OECD, 2017), regulation skills (Dindar et al., 2020), and achievement emotions (Camacho-Morles et al., 2019). In this sense, our results show how combining these factors in the game that was designed helps develop the social dimension of CPS. Furthermore, the collaborative script clearly changed the focus of the discussion on collaboration, as the students in the experimental groups showed a preference for providing arguments (i.e., reasoning and strategies) to justify the regulation processes that were adopted in order to complete the collaborative tasks (Table 4-9). Our research therefore opens the door to future studies.

4.5. Conclusion

The present study provided details of the regulation processes and emotions that emerge when combining a collaborative script with group awareness during CPS activities.

Our research question asked: "How does a game-based collaborative script that scaffolds group awareness affect group regulation and emotions during collaborative problem solving?" To answer this question, we developed a game that looked to develop

CPS skills among elementary-school students by using a collaborative script and tools for boosting group awareness.

Our results reveal the existence of a relationship between collaborative scripts and group awareness as tools to support students' regulation skills during CPS activities. In addition to this, we also highlighted the relationship between these tools and the positive emotions reported by the students during the intervention (i.e., satisfaction). The game that we developed followed the design principles detailed in the literature review and included a script to guide collaboration, as well as tools for facilitating group awareness. In this sense, our results show that the game has a positive impact on students' attitudes towards collaboration. Furthermore, our study also builds on the OECD's CPS framework (OECD, 2017) by providing a level of gradualness as part of the game's design (Table 4-3). Finally, the results also suggest that there is a relationship between the co-regulation process required by the game and the shift in emotions from frustration to satisfaction. Consequently, the proposed collaborative game develops regulation skills and positive emotions, which are key elements of CPS (OECD, 2017).

The limitations of our study are related to the duration of the intervention. More prolonged and profound effects could be studied if students were further exposed to experiences such as the one described here (Popov et al., 2019). Furthermore, previous studies (Järvenoja et al., 2020; Noroozi, Järvelä & Kirschner, 2019) have recommended measuring regulation at various stages. This is because there are certain temporal and cyclical characteristics that may not be observed when only measuring before or after the intervention. In addition to this, the limited size and demographics of the sample means that our results cannot be generalized. A final limitation is related to the use of a single measurement tool (i.e., group interviews). Combining this with other methodological tools would only serve to strengthen the impact of our findings.

Finally, future work should continue to look at combining the motivational potential of games with the implementation of collaborative scripts and group awareness, while maintaining the balance in the difficulty level of the CPS tasks proposed here. We also recommend using mixed methods research to effectively show the impact of the proposed methodology on a series of factors related with the social dimension of CPS. We firmly believe that this represents an exciting opportunity for further developing these skills.

5 CONCLUSIONS AND FUTURE WORK

5.1. Conclusions

The main objective of this thesis is to study how to assess and develop collaborative problem-solving skills in students through technology. To achieve this, we developed and validated a computer-supported assessment, whose main feature is that it was designed with two equivalent versions based on the framework to measure OECD collaborative problem solving (OECD, 2017). Then, we designed a feedback system, based on automated planning techniques, that in addition to providing effective feedback in a collaborative game, would achieve a good performance in a real educational context. Finally, we designed and implemented a game that seeks to develop collaborative problem-solving skills in students and integrates group awareness and collaborative scripting tools.

Among this thesis' contributions, the use of Design-based Research as a methodology to develop and validate a collaborative problem-solving instrument stands out, since it facilitates the instrument's comprehension and replication. Likewise, we were able to contribute evidence that validated a new evaluation instrument, with two equivalent versions, constituting an advance towards the execution of experimental designs that seek to develop CPS skills; the validation showed adequate results and contributes to the knowledge regarding the use of virtual agents, as well as the framework proposed by the OECD. Then, regarding the development of CPS skills, a relevant contribution is related to the proposal of an automated planning algorithm that improves the performance and feasibility of previous algorithms in the area and provides evidence of its feasibility and benefits in a collaborative environment subject to the technological limitations that the context of a real classroom may have. Finally, the contributions of the game built in this thesis are relevant, as they provide evidence of the OECD framework use, not only for assessment but also for the teaching or development of CPS skills. Likewise, the work done in this thesis, shows how through group awareness and collaborative script tools, students benefit from an experience that impacts their social regulation skills, both in motivational, affective, and social aspects; which is the first step to achieve a profound learning experience in which, as they advance in the game, they progressively develop their CPS skills.

5.2. Future Research

This thesis, as well as most of the literature related to CPS, has complex challenges since it involves both theoretical and technological advances. In this sense, it is essential to create sophisticated collaborative environments, which benefit from different sources of information that show the process of problem-solving that a group performs. As presented in this work, elements such as the use of virtual agents, the use of artificial intelligence techniques to deliver feedback or adapt the activities to the students, the motivational and potential challenge of game-based learning, among other elements, are fundamental. Therefore, the next step, having considered these elements, is the design of longer interventions that gather more information from students, at different levels of schooling, and that focus on the motivational, emotional, social, and cognitive aspects of collaborative problem-solving.

REFERENCES

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic Coding Procedures for Collaborative Problem Solving. In Assessment and Teaching of 21st Century Skills (pp. 115–132). Springer Netherlands.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). The standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. Journal of educational technology & society, 11(4), 29-40.

Andrews-Todd, J., & Forsyth, C. M. (2018). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. Computers in Human Behavior.

Andrews-Todd, J., & Forsyth, C. M. (2020). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. Computers in human behavior, 104, 105759.

Andrews-Todd, J., & Kerr, D. (2019). Application of Ontologies for Assessing Collaborative Problem Solving Skills. International Journal of Testing, 19(2), 172–187.

Avry, S. (2021). Beyond the Dichotomy between the Socio-cognitive and Socio-emotional Spaces: the Pervasive Role of Emotions in Collaborative Problem-Solving (Doctoral dissertation, University of Geneva).

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. Science, 211(4489), 1390.

Bauer, M. (2000). Classical content analysis: a review. In Bauer, M. W., & Gaskell, G. Qualitative researching with text, image and sound (pp. 132-151). London: SAGE Publications Ltd doi: 10.4135/9781849209731

Bause, I. M., Brich, I. R., Wesslein, A. K., & Hesse, F. W. (2018). Using technological functions on a multi-touch table and their affordances to counteract biases and foster collaborative problem solving. International journal of computer-supported collaborative learning, 13(1), 7-33.

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological bulletin, 107(2), 238.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century Skills. In Assessment and Teaching of 21st Century Skills.

Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. Computers in Human Behavior, 27(3), 1043-1045.

Bødker, S., & Christiansen, E. (2006). Computer support for social awareness in flexible work. Computer Supported Cooperative Work (CSCW), 15(1), 1-28.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really?. Educational psychology review, 15(1), 1-40.

Boud, D., & Molloy, E. (2013). Feedback in higher and professional education: understanding it and doing it well. Routledge.

Brown, T. A. (2014). Confirmatory factor analysis for applied research. Guilford Publications.

Bruner, J. (1985). Narrative and paradigmatic modes of thought. Teachers College Record, 86(6), 97-115.

Buono, S., Zdravkovic, A., Lazic, M., & Woodruff, E. (2020). The effect of emotions on self-regulated-learning (SRL) and story comprehension in emerging readers. In Frontiers in Education (Vol. 5, p. 218). Frontiers.

Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. Review of educational research, 54(3), 427-456.

Calvo-Morata, A., Alonso-Fernández, C., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2020). Serious games to prevent and detect bullying and cyberbullying: A systematic serious games and literature review. Computers & Education, 157, 103958.

Camacho-Morles, J., Slemp, G. R., Oades, L. G., Morrish, L., & Scoular, C. (2019). The role of achievement emotions in the collaborative problem-solving performance of adolescents. Learning and Individual Differences, 7. 169-181.

Care, E., & Griffin, P. (2014). An Approach to Assessment of Collaborative Problem Solving. Research & Practice in Technology Enhanced Learning, 9(3).

Care, E., Scoular, C., & Griffin, P. (2016). Assessment of collaborative problem solving in education environments. Applied Measurement in Education, 29(4), 250–264.

Chan, Y. F., Kaur Sidhu, G., Narasuman, S., Lee, L. F., & Yap, B. W. (2016). The relationship between collaboration in learning, quantity and timing of feedback, and self-efficacy of students in higher education/Chan Yuen Fook…[et al.]. Social and Management Research Journal, 13(1), 60-75.

Chang, C.-J., Chang, M.-I., Chiu, B.-C., Liu, C.-C., Chiang, S.-I. F., Wen, C.-T., & Liu, C.-C. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. Computers & Education, 114, 222–235.

Chavez, J., & Romero, M. (2012). Group awareness, learning, and participation in Computer Supported Collaborative Learning (CSCL). Procedia-Social and Behavioral Sciences, 46, 3068-3073.

Chen, C. H., & Law, V. (2016). Scaffolding individual and collaborative game-based learning in learning performance and intrinsic motivation. Computers in Human Behavior, 55, 1201-1212.

Chen, C. M., & Kuo, C. H. (2019). An optimized group formation scheme to promote collaborative problem-based learning. Computers and Education, 133(64), 94–115.

Chen, F., Cui, Y., & Chu, M. W. (2020). Utilizing Game Analytics to Inform and Validate Digital Game-based Assessment with Evidence-centered Game Design: A Case Study. International Journal of Artificial Intelligence in Education, 30(3), 481-503.

Chen, Z. H., Lu, H. De, & Chou, C. Y. (2019). Using game-based negotiation mechanism to enhance students' goal setting and regulation. Computers and Education, 129, 71–81. https://doi.org/10.1016/j.compedu.2018.10.011

Chung, G. K. W. K., O'Neil Jr, H. F., & Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. Computers in Human Behavior, 15(3-4), 463-493.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural equation modeling, 9(2), 233-255.

Child, S. F. J., & Shaw, S. (2019). Towards an operational framework for establishing and assessing collaborative interactions. Research Papers in Education, 34(3), 276–297.

Chopade, P., Khan, S. M., Stoeffler, K., Edward, D., Rosen, Y., & Davier, A. (2018). Framework for effective teamwork assessment in collaborative learning and problem solving. In in 19th International Conference on Artificial Intelligence in Education (AIED), LAS, The Festival of Learning, ARL Workshop "Assessment and Intervention during Team Tutoring" AIED-2018, 2153, 48–59.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. Review of educational research, 86(1), 79-122.

Collective, D.-B. R. (2003). Design-based research: An emerging paradigm for educational inquiry. Educational Researcher, 32(1), 5–8.

Costikyan, G. (2013). Uncertainty in games. Mit Press.

Crook, C. (1998). Children as computer users: The case of collaborative learning. Computers & Education, 30(3-4), 237-247.

Csapó, B., & Funke, J. (2017). The nature of problem solving. In Using research to inspire 21st century learning.

Csikszentmihalyi, M. (1991). Flow: The psychology of optimal experience In: New York: HarperPerennial.

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. Computers in Human Behavior, 27(3), 1068-1078.

Dillenbourg P. (1999) What do yuo mean by collaborative leraning?. In P. Dillenbourg (Ed) Collaborative-learning: Cognitive and Computational Approaches. (pp.1-19). Oxford: Elsevier

Dillenbourg, P., & Jermann, P. (2007). Designing integrative scripts. In Scripting computer-supported collaborative learning (pp. 275-301). Springer, Boston, MA.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In Technology-enhanced learning (pp. 3-19). Springer, Dordrecht.

Dindar, M., Järvelä, S., & Järvenoja, H. (2020). Interplay of metacognitive experiences and performance in collaborative problem solving. Computers & Education, 154, 103922.

Domagk, S., Schwartz, R. N., & Plass, J. L. (2010). Interactivity in multimedia learning: An integrated model. Computers in Human Behavior, 26(5), 1024-1033.

Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: The role of rostrolateral prefrontal cortex. Developmental cognitive neuroscience, 1. 57-76.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British journal of psychology, 105(3), 399-412.

Edelson, D. C. (2002). Design research: What we learn when we engage in design. The Journal of the Learning Sciences, 11(1), 105–121.

Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?. International journal of public health, 58(4), 637-642.

Enders, C. K. (2010). Applied missing data analysis. Guilford press.

Epstein, I., & Tripodi, T. (1977). Research techniques for program planning, monitoring, and evaluation. New York: Columbia University Press.

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. Computers & Education, 67, 156-167.

Fiore, S. M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., ... & von Davier, A. (2017). Collaborative problem solving: Considerations for the national assessment of educational progress.

Fiore, S. M., Graesser, A., & Greiff, S. (2018). Collaborative problem-solving education for the twenty-first-century workforce. Nature human behaviour, 2(6), 367.

Fischer, F., Kollar, I., Stegmann, K., Wecker, C., Zottmann, J., & Weinberger, A. (2013). Collaboration scripts in computer-supported collaborative learning. 10.4324/9780203837290.ch23.

Flor, M., Yoon, S.-Y., Hao, J., Liu, L., & Von Davier, A. A. (2016). Automated classification of collaborative problem solving interactions in simulated science tasks (pp. 31–41).

Freedberg, M., Glass, B., Filoteo, J. V., Hazeltine, E., & Maddox, W. T. (2017). Comparing the effects of positive and negative feedback in information-integration category learning. Memory & cognition, 45(1), 12-25.

Fritz, C., & McIlraith, S. A. (2007, September). Monitoring Plan Optimality During Execution. In ICAPS (pp. 144-151).

Funke, J. (2010). Complex problem solving: A case for complex cognition?. Cognitive processing, 11(2), 133-142.

Funke, J., Fischer, A., & Holt, D. V. (2018). Competencies for complexity: problem solving in the twenty-first century. In Assessment and teaching of 21st century skills (pp. 41-53). Springer, Cham.

Fyfe, E. R., & Rittle-Johnson, B. (2016). The benefits of computer-generated feedback for mathematics problem solving. Journal of experimental child psychology, 147, 140-151.Ghallab, M., Nau, D., & Traverso, P. (2004). Automated Planning: theory and practice. Elsevier.

Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., ... & Winer, E. (2018). Creating a team tutor using GIFT. International Journal of Artificial Intelligence in Education, 28(2), 286-313.

Goldin, I., Narciss, S., Foltz, P., & Bauer, M. (2017). New directions in formative feedback in interactive learning environments. International Journal of Artificial Intelligence in Education, 27(3), 385-392.

Graesser, A. C., Cai, Z., Morgan, B., & Wang, L. (2017). Assessment with computer agents that engage in conversational dialogues and trialogues with learners. Computers in Human Behavior, 76, 607–616.

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018a). Advancing the Science of Collaborative Problem Solving. Psychological Science in the Public Interest, 19(2), 59–92.

Graesser, A. C., Foltz, P. W., Rosen, Y., Shaffer, D. W., Forsyth, C., & Germany, M. L. (2018b). Challenges of assessing collaborative problem solving. In Assessment and Teaching of 21st Century Skills (pp. 75–91). Springer.

Graesser, A. C., Greiff, S., Stadler, M., & Shubeck, K. T. (2020). Collaboration in the 21st century: The theory, assessment, and teaching of collaborative problem solving.

Graesser, A., Kuo, B.-C., & Liao, C.-H. (2017). Complex Problem Solving in Assessments of Collaborative Problem Solving. Journal of Intelligence, 5(2), 10.

Greipl, S., Moeller, K., & Ninaus, M. (2020). Potential and limits of game-based learning. International Journal of Technology Enhanced Learning, 12(4), 363-389.

Griffin, P. (2017). Assessing and Teaching 21st Century Skills: Collaborative Problem Solving as a Case Study. In D. A., Z. M., & K. P (Eds.), Innovative Assessment of Collaboration. Methodology of Educational Measurement and Assessment. Springer.

Griffin, P., & Care, E. (2014). Assessment and teaching of 21st century skills: Methods and approach. Springer.

Griffin, P., Care, E., & McGaw, B. (2012). The changing role of education and schools. In Assessment and teaching of 21st century skills (pp. 1-15). Springer Netherlands.

Griffin, P., McGaw, B., & Care, E. (2012). Assessment and teaching of 21st century skills. Assessment and Teaching of 21st Century Skills, 9789400723245, 1–345.

Grover, S., Chakraborti, T., & Kambhampati, S. (2018). What can Automated Planning do for Intelligent Tutoring Systems. ICAPS SPARK.

Gutwin, C., & Greenberg, S. (2004). The importance of awareness for team cognition in distributed collaboration. In E. Salas & S. M. Fiore (Eds.), Team cognition: Understanding the factors that drive process and performance (pp. 177–201). American Psychological Association.

Hadwin, A., Bakhtiar, A., & Miller, M. (2018). Challenges in online collaboration: Effects of scripting shared task perceptions. International Journal of Computer-Supported Collaborative Learning, 13(3), 301-329.

Hadwin, A., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. Teachers College Record, 113(2), 240-264.

Hadwin, A., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In D.H. Schunk, J.A. Greene (Eds.), Handbook of self-regulation of learning and performance (2nd ed.), Routledge, New York (2018), pp. 83-106.

Halpin, P. F., Davier, A. A., Hao, J., & Liu, L. (2017). Measuring student engagement during collaboration. Journal of Educational Measurement, 54(1), 70–84.

Ham, Y., & Hwang, J. (2021). Mathematical Literacy and Collaborative Problem-Solving: Comparison between Korean and US Students in PISA2015. Journal of Educational Research in Mathematics, 31(3), 299-320.

Han, J., Zhao, W., Jiang, Q., Oubibi, M., & Hu, X. (2019). Intelligent Tutoring System Trends 2006-2018: A Literature Review. In 2019 Eighth International Conference on Educational Innovation through Technology (EITT) (pp. 153-159). IEEE.

Hao, J., Chen, L., Flor, M., Liu, L., & von Davier, A. A. (2017). CPS-Rater: Automated Sequential Annotation for Conversations in Collaborative Problem-Solving Activities. Research Report. ETS RR-17-58. In ETS Research Report Series. ERIC.

Hao, J., Liu, L., Kyllonen, P., Flor, M., & von Davier, A. A. (2019). Psychometric Considerations and a General Scoring Strategy for Assessments of Collaborative Problem Solving. In ETS Research Report Series (Vol. 2019, Issue 1). Wiley Online Library.

Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. (2015). Assessing Collaborative Problem Solving with Simulation Based Tasks Methods. Proceedings of the 11th International Conference on Computer Supported Collaborative Learning.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), 81-112.

Hawlitschek, A., & Joeckel, S. (2017). Increasing the effectiveness of digital educational games: The effects of a learning instruction on students' learning, motivation and cognitive load. Computers in Human Behavior, 72, 79-86.

He, Q., von Davier, M., Greiff, S., Steinhauer, E. W., & Borysewicz, P. B. (2017). Collaborative problem solving measures in the Programme for International Student Assessment (PISA). In A. A. von Davier, M. Zhu, & P. C. Kyllonen, (Eds.), Innovative assessment of collaboration (pp. 95-111). Dordrecht, Netherlands: Springer.

Helmert, M. (2006). The fast downward planning system. Journal of Artificial Intelligence Research, 26, 191-246.

Herborn, K., Stadler, M., Mustafić, M., & Greiff, S. (2020). The assessment of collaborative problem solving in PISA 2015: Can computer agents replace humans?. Computers in Human Behavior, 104, 105624.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A Framework for teachable collaborative problem solving skills. In P. Griffin, & E. Care (Eds.).

Hewett, K. J., Zeng, G., & Pletcher, B. C. (2020). The acquisition of 21st-century skills through video games: Minecraft design process models and their web of class roles. Simulation & Gaming, 51(3), 336-364.

Hirschfeld, G., & Von Brachel, R. (2014). Improving Multiple-Group confirmatory factor analysis in R–A tutorial in measurement invariance with continuous and ordinal indicators. Practical Assessment, Research, and Evaluation, 19(1), 7.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Evaluating model fit: a synthesis of the structural equation modelling literature. 7th European Conference on Research Methodology for Business and Management Studies, 195–200.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6, 1–55.

Huizinga, J. (2014). Homo ludens ils 86. Routledge.

Hwang, G. J., Yang, L. H., & Wang, S. Y. (2013), A concept map-embedded educational computer game for improving students' learning performance in natural science courses, Computers & Education, 69,121-130.

Isohätälä, J., Näykki, P., & Järvelä, S. (2020). Cognitive and socio-emotional interaction in collaborative learning: Exploring fluctuations in students' participation. Scandinavian Journal of Educational Research, 64(6), 831-851.

Jablansky, S. (2020). Relational Reasoning and Socially Shared Regulation of Learning in Collaborative Problem Solving (Doctoral dissertation, University of Maryland, College Park).

Janakiraman, S., Watson, S. L., Watson, W. R., & Newby, T. (2021). Effectiveness of digital games in producing environmentally friendly attitudes and behaviors: A mixed-methods study. Computers & Education, 16. 104043.

Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. Educational Psychologist, 48(1), 40-55.

Janssen, J., Erkens, G., & Kirschner, P. A. (2011). Group awareness tools: It's what you do with it that matters. Computers in human behavior, 27(3), 1046-105

Janssen, J., Erkens, G., Kanselaar, G., & J, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? Computers & Education, 49(4), 1037–1065.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. Educational psychologist, 48(1), 25-39.

Järvelä, S., & Järvenoja, H. (2011). Socially constructed self-regulated learning and motivation regulation in collaborative learning groups. Teachers College Record, 113(2), 350-374.

Järvelä, S., Järvenoja, H., Malmberg, J., & Hadwin, A. F. (2013). Exploring socially shared regulation in the context of collaboration. Journal of Cognitive Education and Psychology, 12(3), 267-286.

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., ... & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. Educational Technology Research and Development, 63(1), 125-142.

Järvenoja, H., Järvelä, S., & Malmberg, J. (2020). Supporting groups' emotion and motivation regulation during collaborative learning. Learning and Instruction, 7. 101090. https://doi.org/10.1016/j.learninstruc.2017.11.004

Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. Theory into practice, 38(2), 67-73.

Juul, J. (2013). The art of failure: An essay on the pain of playing video games. MIT press.

Kautz, H., McAllester, D., & Selman, B. (1996). Encoding plans in propositional logic. KR, 96, 374-384.

Kao, G. Y. M., Chiang, C. H., & Sun, C. T. (2017). Customizing scaffolds for game-based learning in physics: Impacts on knowledge acquisition and game design creativity. Computers & Education, 113, 294-312.

Keepitusablecom. (2011). Mobile App Usability Checklist. http://www.keepitusable.com/keepitusablemobile-app-usability-checklist.pdf

Keselman, H. J. (1998). Testing treatment effects in repeated measures designs: An update for psychophysiological researchers. Psychophysiology, 35(4), 470-478.

Kiili, K., De Freitas, S., Arnab, S., & Lainema, T. (2012). The design principles for flow experience in educational games. Procedia Computer Science, 15, 78-91.

Krkovic, K., Wüstenberg, S., & Greiff, S. (2016). Assessing collaborative behavior in students. An experiment-based approach. European Journal of Psychological Assessment, 32, 52-60.

Land, S. M., & Zimmerman, H. T. (2015). Socio-technical dimensions of an outdoor mobile learning environment: a three-phase design-based research investigation. Educational Technology Research and Development, 63(2), 229–255.

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. Simulation & gaming, 45(6), 752-768.

Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science, its history and future: Definitions and a research agenda. Simulation & Gaming, 49(3), 315-337.

Laurinen, L. H., & Marttunen, M. J. (2007). Written arguments and collaborative speech acts in practising the argumentative power of language through chat debates. Computers and Composition, 24(3), 230–246.

Li, Y., Li, X., Zhang, Y., & Li, X. (2021). The effects of a group awareness tool on knowledge construction in computer-supported collaborative learning. British Journal of Educational Technology.

Lim, J. Y., & Lim, K. Y. (2020). Co-regulation in collaborative learning: Grounded in achievement goal theory. International Journal of Educational Research, 103, 101621.

Liu, L., Hao, J., Davier, A. A., Kyllonen, P., & Zapata-Rivera, J. D. (2016). A tough nut to crack: Measuring collaborative problem solving. In Handbook of research on technology tools for real-world skill development (pp. 344–359). IGI Global.

Loderer, K., Pekrun, R., & Plass, J. L. (2020). 5 Emotional Foundations of Game-Based Learning. Handbook of Game-Based Learning, 111.

Loehlin, J. C. (2004). Latent variable models: An introduction to factor, path, and structural equation analysis. Psychology Press.

Lyons, K. M., Lobczowski, N. G., Greene, J. A., Whitley, J., & McLaughlin, J. E. (2021). Using a design-based research approach to develop and study a web-based tool to support collaborative learning. Computers & Education, 161, 104064.

Ma, X., Liu, J., Liang, J., & Fan, C. (2020). An empirical study on the effect of group awareness in CSCL environments. Interactive Learning Environments, 1-16.

Maine, F. (2017). Collaborative and dialogic meaning-making: How children engage and immerse in the storyworld of a mobile game. In The Case of the iPad (pp. 211-225). Springer, Singapore.

Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. Contemporary Educational Psychology, 49, 160-174.

Mänty, K., Järvenoja, H., & Törmänen, T. (2020). Socio-emotional interaction in collaborative learning: Combining individual emotional experiences and group-level emotion regulation. International Journal of Educational Research, 102, 101589.

Mayer, R. E. (2014). Computer games for learning: An evidence-based approach. MIT Press.

McAlister, A., Lee, D., Ehlert, K. M., Kajfez, R. L., Faber, C. J., & Kennedy, M. S. (2017). Qualitative coding: An approach to assess inter-rater reliability. In American Society for Engineering Education (pp. 5-9).

MCINERNEY, D. M. (2012). Parents, teachers, and peers: Who has most influence on student motivation and achievement?.

Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. Computers in Human Behavior, 52, 573-588.

Millsap, R. E. (2011). Statistical approaches to measurement invariance. Routledge.

Millsap, R. E., & Cham, H. (2012). Investigating factorial invariance in longitudinal research. In B. Laursen, T. D. Little, & N. A. Card (Eds.), Handbook of developmental research methods (pp. 109–126). Guilford Press.

Molnar, G., Greiff, S., & Csapo, B. (2013). Inductive reasoning, domain specific and complex problem solving: relations and development. Thinking Skills and Creativity, 9, 35-45.

Morgan Jr, B. B., Salas, E., & Glickman, A. S. (1993). An analysis of team evolution and maturation. The Journal of General Psychology, 120(3), 277-291.

Morrison, D. F. (2005). Multivariate analysis of variance. Encyclopedia of Biostatistics, 5.

Muise, C., McIlraith, S. A., & Beck, J. C. (2011). Monitoring the execution of partial-order plans via regression. In IJCAI Proceedings-International Joint Conference on Artificial Intelligence (Vol. 22, No. 3, p. 1975).

Nadolny, L., Valai, A., Cherrez, N. J., Elrick, D., Lovett, A., & Nowatzke, M. (2020). Examining the characteristics of game-based learning: A content analysis and design framework. Computers & Education, 103936.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. Assessment & Evaluation in Higher Education, 39(1), 102-122.

Noroozi, O., Järvelä, S., & Kirschner, P. A. (2019). Multidisciplinary innovations and technologies for facilitation of self-regulated learning. Computers in Human Behavior, 10. 295–297. https://doi.org/10.1016/J.CHB.2019.07.020

Nouri, J., Åkerfeldt, A., Fors, U., & Selander, S. (2017). Assessing collaborative problem solving skills in technology-enhanced learning environments - the PISA framework and modes of communication. International Journal of Emerging Technologies in Learning, 12(4), 163–174.

Nussbaum, M., Alvarez, C., McFarlane, A., Gomez, F., Claro, S., & Radovic, D. (2009). Technology as small group face-to-face Collaborative Scaffolding. Computers & Education, 52(1), 147-153.

OECD (2010). PISA 2012 field trial problem solving framework. Accessed at: http://www.oecd.org/dataoecd/8/42/46962005.pdf.

OECD (2012). Better skills, better jobs, better lives. Highlights of the OECD skills strategy. Accessed at:http://www.oecd-ilibrary.org/docserver/download/fulltext/9112165e5.pdf.

OECD. (2017). PISA 2015 Collaborative Problem Solving Framework (pp. 1–93).

OECD. (2018). PISA 2015: Results in focus. PISA, 67, 16.

O'Donovan, B. M., den Outer, B., Price, M., & Lloyd, A. (2019). What makes good feedback good?. Studies in Higher Education, 1-12.

Ødegård, A. (2006). Exploring perceptions of interprofessional collaboration in child mental health care. International Journal of Integrated Care. http://doi.org/10.5334/ijic.165

Oliveri, M. E., Lawless, R., & Molloy, H. (2017). A literature review on collaborative problem solving for college and workforce readiness. In ETS Research Report Series (Vol. 2017, Issue 1). Wiley Online Library.

O'Neil Jr, H. F., Chung, G. K., & Brown, R. S. (1995). Measurement of teamwork processes using computer simulation (CSE Report 399). Los Angeles: University of California, Los Angeles. National Center for Research on Evaluation, Standards, and Student

Parker, P. C., Perry, R. P., Hamm, J. M., Chipperfield, J. G., Pekrun, R., Dryden, R. P., ... & Tze, V. M. (2021). A motivation perspective on achievement appraisals, emotions, and performance in an online learning environment. International Journal of Educational Research, 108, 101772.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. The journal of the learning sciences, 13(3), 423-451.

Pekrun, R. (2017). Achievement emotions. Emotions in Late Modernity, 142.

Pifarré, M., Cobos, R., & Argelagós, E. (2014). Incidence of group awareness information on students' collaborative learning processes. Journal of Computer Assisted Learning, 30(4), 300-317.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. Educational Psychologist, 50(4), 258-283.

Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In Emotions, technology, design, and learning (pp. 131-161). Academic Press.

Pretz, J. E., Naples, A. J., & Sternberg, R. J. (2003). Recognizing, defining, and representing problems. The psychology of problem solving, 30(3).

Polya, G. (2004). How to solve it: A new aspect of mathematical method (No. 246). Princeton university press.

Popov, V., Biemans, H. J., Fortuin, K. P., van Vliet, A. J., Erkens, G., Mulder, M., ... & Li, Y. (2019). Effects of an interculturally enriched collaboration script on student attitudes, behavior, and learning performance in a CSCL environment. Learning, Culture and Social Interaction, 21, 100-123.

Qian, M., & Clark, K. R. (2016). Game-based Learning and 21st century skills: A review of recent research. Computers in Human Behavior, 63, 50-58.

Quereshi, M. Y. (2003). Absence of parallel version for the traditional individual intelligence tests. Current Psychology, 22(2), 149-154.

Quinn, C. N. (2005). Engaging learning: Designing e-learning simulation games. John Wiley & Sons.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, version 4.0.2. URL https://www.R-project.org/

Radkowitsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. International Journal of Computer-Supported Collaborative Learning, 15(1), 5-47.

Rau, M. A., Bowman, H. E., & Moore, J. W. (2017). An adaptive collaboration script for learning with multiple visual representations in chemistry. Computers & Education, 109, 38-55.

Reeves, T. (2006). Design research from a technology perspective. In Educational design research (pp. 64–78). Routledge.

Reis, R. C. D., Isotani, S., Rodriguez, C. L., Lyra, K. T., Jaques, P. A., & Bittencourt, I. I. (2018). Affective states in computer-supported collaborative learning: Studying the past to drive the future. Computers and Education, 120(February), 29–50. https://doi.org/10.1016/j.compedu.2018.01.015

Resnick, M. (2017). Lifelong kindergarten. RESNICK, M. Lifelong Kindergarten-Cultivatin

Rocha, J. B., Mascarenhas, S., & Prada, R. (2008). Game Mechanics for Cooperative Games. ZON Digital Games, 72–80. https://doi.org/10.1523/JNEUROSCI.1920-10.2010

Rojas, M., Nussbaum, M., Chiuminatto, P., Guerrero, O., Greiff, S., Krieger, F., & Van Der Westhuizen, L. (2021). Assessing collaborative problem-solving skills among elementary school students. Computers & Education, 175, 104313.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In Computer supported collaborative learning (pp. 69-97). Springer, Berlin, Heidelberg.

Rosen, Y. (2014). Comparability of Conflict Opportunities in Human-to-Human and Human-to-Agent Online Collaborative Problem Solving. Technology, Knowledge and Learning, 19(1–2), 147–164.

Rosen, Y. (2015). Computer-based assessment of collaborative problem solving: Exploring the feasibility of human-to-agent approach. International Journal of Artificial Intelligence in Education, 25(3), 380-406.

Rosen, Y. (2017). Assessing Students in Human-to-Agent Settings to Inform Collaborative Problem-Solving Learning. Journal of Educational Measurement, 54(1), 36–53.

Rosen, Y., & Tager, M. (2013). Computer-based Assessment of Collaborative Problem-Solving Skills: Human-to-Agent versus Human-to-Human Approach. Pearson Research Report.

Rosen, Y., Wolf, I., & Stoeffler, K. (2019). Fostering collaborative problem solving skills in science: The Animalia project.

Rosen, Y., Wolf, I., & Stoeffler, K. (2020). Fostering collaborative problem solving skills in science: The Animalia project. Computers in Human Behavior, 104, 105922.

Ruholt, R. E., Gore, J., & Dukes, K. (2015). Is parental support or parental involvement more important for adolescents?. Undergraduate Journal of Psychology, 28(1), 1-to.

Ruipérez-Valiente, J. A., & Kim, Y. J. (2020). Effects of solo vs. collaborative play in a digital learning game on geometry: Results from a K12 experiment. Computers & Education, 159, 104008.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training.

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis.

Savard, I., Bourdeau, J., & Paquette, G. (2020). Considering cultural variables in the instructional design process: A knowledge-based advisor system. Computers & Education, 145, 103722.

Scalise K., Mustafic M., Greiff S. (2016) Dispositions for Collaborative Problem Solving. In: Kuger S., Klieme E., Jude N., Kaplan D. (eds) Assessing Contexts of Learning. Methodology of Educational Measurement and Assessment. Springer, Cham.

Schnaubert, L., Vogel, F., Bodemer, D., Fischer, F., Radkowitsch, A., Schmidmaier, R., Fischer, M. R., Tsovaltzi, D., Puhl, T., & Azevedo, R. (2020). Combining Scripts, Group Awareness Tools and Self-Regulated Learning – Theoretical Implications and Practical Implementations. In Gresalfi, M. and Horn, I. S. (Eds.), The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 202. Volume 1 (pp. 350-357). Nashville, Tennessee: International Society of the Learning Sciences.

Scoular, C., & Care, E. (2018). Teaching twenty-first century skills: implications at system levels in Australia. In Assessment and Teaching of 21st Century Skills (pp. 145-162). Springer, Cham.

Scoular, C., & Care, E. (2019a). A Generalized Scoring Process to Measure Collaborative Problem Solving in Online Environments. Educational Assessment, 24(3), 213–234.

Scoular, C., & Care, E. (2020). Monitoring patterns of social and cognitive student behaviors in online collaborative problem solving assessments. Computers in Human Behavior, 104, 105874.

Scoular, C., Care, E., & Awwal, N. (2017). An Approach to Scoring Collaboration in Online Game Environments. Electronic Journal of e-Learning, 15(4).

Scoular, C., Care, E., & Hesse, F. W. (2017). Designs for operationalizing collaborative problem solving for automated assessment. Journal of Educational Measurement, 54(1), 12-35.

Seif El-Nasr, M., Aghabeigi, B., Milam, D., Erfani, M., Lameman, B., Maygoli, H., & Mah, S. (201. April). Understanding and evaluating cooperative games. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 253-262).

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. Contemporary educational psychology, 21(1), 43-69.

Stadler, M., Herborn, K., Mustafić, M., & Greiff, S. (2020). The assessment of collaborative problem solving in PISA 2015: An investigation of the validity of the PISA 2015 CPS tasks. Computers & Education, 157, 103964.

Stewart, A. E., Amon, M. J., Duran, N. D., & D'Mello, S. K. (n.d.). Beyond Team Makeup: Diversity in Teams Predicts Valued Outcomes in Computer-Mediated Collaborations. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13.

Stoeffler, K., Rosen, Y., Bolsinova, M., & Davier, A. A. (2019). Gamified performance assessment of collaborative problem solving skills. Computers in Human Behavior, February), 106036.

Strijbos, J. W., Martens, R. L., & Jochems, W. M. G. (2004). Designing for interaction: Six steps to designing computer-supported group-based learning. Computers &Education, 42, 403–424.

Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D'Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. Computers and Education, 143(August 2019), 103672.

Swiecki, Z., Ruis, A. R., Farrell, C., & Shaffer, D. W. (2020). Assessing individual contributions to collaborative problem solving: a network analysis approach. Computers in Human Behavior, 104, 105876.

Szewkis, E., Nussbaum, M., Rosen, T., Abalos, J., Denardin, F., Caballero, D., ... & Alcoholado, C. (2011). Collaboration within large groups in the classroom. International Journal of Computer-Supported Collaborative Learning, 6(4), 561-575.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. Research in Science Education, 48(6), 1273-1296.

Taub, M., Sawyer, R., Lester, J., & Azevedo, R. (2020). The impact of contextualized emotions on self-regulated learning and scientific reasoning during learning with a game-based learning environment. International Journal of Artificial Intelligence in Education, 30(1), 97-120.

Teasley, S., Fischer, F., Dillenbourg, P., Kapur, M., Chi, M., Weinberger, A., & Stegmann, K. (2008). Cognitive convergence in collaborative learning.

Teruel, M. A., Condori-Fernandez, N., Navarro, E., González, P., & Lago, P. (2018). Assessing the impact of the awareness level on a co-operative game. Information and Software Technology, 98, 89-116.

Travis, J., & Bunde, J. (2020). Self-regulation in college: The influence of self-efficacy, need satisfaction, and stress on GPA, persistence, and satisfaction. Current Psychology, 1-11.

Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C. C. (2017). Some guidance on conducting and reporting qualitative studies.

Tyack, A., Wyeth, P., & Johnson, D. (202. April). Restorative play: Videogames improve player wellbeing after a need-frustrating event. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-15).

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. Educational psychologist, 39(2), 111-133.

Van der Meij, H., Veldkamp, S., & Leemkuil, H. (2020). Effects of scripting on dialogues, motivation and learning outcomes in serious games. British journal of educational technology, 51(2), 459-472.

Von Davier, A. A., Hao, J., Liu, L., & Kyllonen, P. (2017). Interdisciplinary research agenda in support of assessment of collaborative problem solving: lessons learned from developing a collaborative science assessment prototype. Computers in Human Behavior, 76, 631–640.

ter Vrugte, J., de Jong, T., Vandercruysse, S., Wouters, P., van Oostendorp, H., & Elen, J. (2015). How competition and heterogeneous collaboration interact in prevocational game-based mathematics education. Computers & education, 89, 42-52.

Weld, D. S. (1994). An introduction to least commitment planning. AI magazine, 15(4), 27.

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. Australasian journal of paramedicine, 8(3).

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of child psychology and psychiatry, 17(2), 89-100.

Wooten, J. O., & Ulrich, K. T. (2017). Idea generation and the role of feedback: Evidence from

Yilmaz, R., & Yilmaz, F.G.K. (2020). Examination of the effectiveness of the task and group awareness support system used for computer-supported collaborative learning. Educational Technology Research and Development, 1-26. 68:1355–1380.

Zhu, M., Liu, O. L., & Lee, H. S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. Computers & Education, 143, 103668.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. Contemporary educational psychology, 25(1), 82-91.

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. Educational psychologist, 48(3), 135-147.

Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. Handbook of self-regulation of learning and performance, 15-26.

Zurita, G., Nussbaum, M., & Salinas, R. (2005). Dynamic grouping in collaborative learning supported by wireless handhelds. Journal of Educational Technology & Society, 8(3).

APPENDIX

APPENDIX A: SPECIFICATION MATRIX OF THE ASSESSMENT

Table A.1: Specification Matrix of the CPS Assessment

Stage	Item	CPS Skill	Description
0	1	A2 Discovering the type of collaborative interaction to solve the problem, along with goals	TT must identify that talking through the chat they can make agreements to solve the problem.
	2	A1 Discovering perspectives and abilities of team members	TT must discover what his/her teammates see on their tablets and conclude that they all see the same thing.
	3	A3 Understanding roles to solve problem	TT must identify that each one was assigned a color and that corresponds to the star color they should look for.
1	4	B1 Building a shared representation and negotiating the meaning of the problem (common ground)	TT must conclude with his/her teammates that there are several solutions and that they must agree.
	5	B3 Describe roles and team organisation (communication protocol/rules of engagement)	TT must understand what roles / colors his teammates have and communicate his / her own, in order to organize and compromise the rules of the game.

	6	C1 Communicating with team members about the actions to be/ being performed	TT and his/her teammates must devise a plan and communicate to execute it and achieve the task.
	7	C2 Enacting plans	TT and his teammates execute the plan, each one has a responsibility regarding the plan they discussed.
	8	D1 Monitoring and repairing the shared understanding	TT monitors or reinforces what they understand about the problem and reiterates the importance of following the plan.
	9	D2 Monitoring results of actions and evaluating success in solving the problem	TT asks or communicates their doubts about what they should do, evaluates the achievement of the task and reinforces the idea of communicating to achieve the solution.
2	10	B2 Identifying and describing tasks to be completed	TT asks or communicates changes in the objectives of the problems, particularly when it moves to a next stage.
	12	C1 Communicating with team members about the actions to be/ being performed	TT and his/her teammates must devise a plan and communicate to execute it and achieve the task.
	13	C2 Enacting plans	TT and his/her teammates execute the plan, each one has a

			responsibility regarding the plan they discussed.
	14	D1 Monitoring and repairing the shared understanding	TT monitors or reinforces what they understand about the problem and reiterates the importance of following the plan.
	15	D2 Monitoring results of actions and evaluating success in solving the problem	TT asks or communicates their doubts about what they should do, evaluates the achievement of the task and reinforces the idea of communicating to achieve the solution.
3	16	B2 Identifying and describing tasks to be completed	TT asks or communicates changes in the objectives of the problems, particularly when it moves to a next stage.
	17	C3 Following rules of engagement, (e.g., prompting other team members to perform their tasks.)	Given that at this stage they are not explicitly assigned a color, TT must be organized with their peers to define their roles.
	18	C1 Communicating with team members about the actions to be/ being performed	TT and his/her teammates must devise a plan and communicate to execute it and achieve the task.
	19	C2 Enacting plans	TT and his/her teammates execute the plan, each one has a responsibility regarding the plan they discussed.

	20	D3 Monitoring, providing feedback and adapting the team organisation and roles	TT evaluates whether his/her teammates followed the rules of engagement, on what color to choose or communicates their doubts.
4	21	B2 Identifying and describing tasks to be completed	TT asks or communicates changes in the objectives of the problems, particularly when it moves to a next stage.
	22	C3 Following rules of engagement, (e.g., prompting other team members to perform their tasks.)	Given that at this stage they are not explicitly assigned a color, TT must be organized with their peers to define their roles.
	23	C1 Communicating with team members about the actions to be/ being performed	TT and his/her teammates colleagues must devise a plan and communicate to execute it and achieve the task.
	24	C2 Enacting plans	TT and his/her teammates execute the plan, each one has a responsibility regarding the plan they discussed.

Note*: The first column corresponds to the stage or level of the activity; The second column is the item; The third column indicates the skill CPS defined by the OECD (OECD, 2017); And finally, the fourth column is a brief item description, where TT corresponds to the student evaluated (Test Taker).*

APPENDIX B: CONVERSATION STRUCTURES BY CPS SKILL

The following diagrams correspond to structures of a generic conversations around each of the collaborative problem solving skills

Dialog structure A1

CPS Skill: Discovering perspectives and abilities of team members

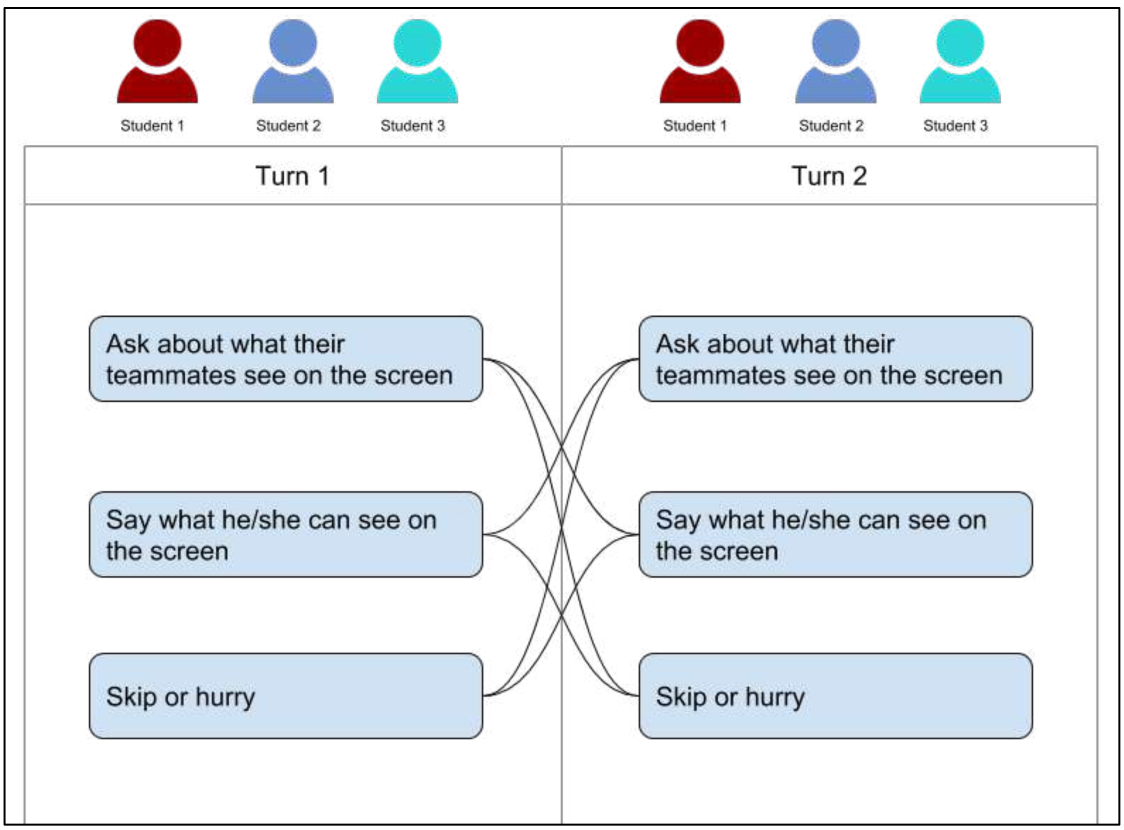


Figure B.1: Conversation structure to assess A1 skill

Dialog structure A2

CPS Skill: Discovering the type of collaborative interaction to solve the problem, along with goals

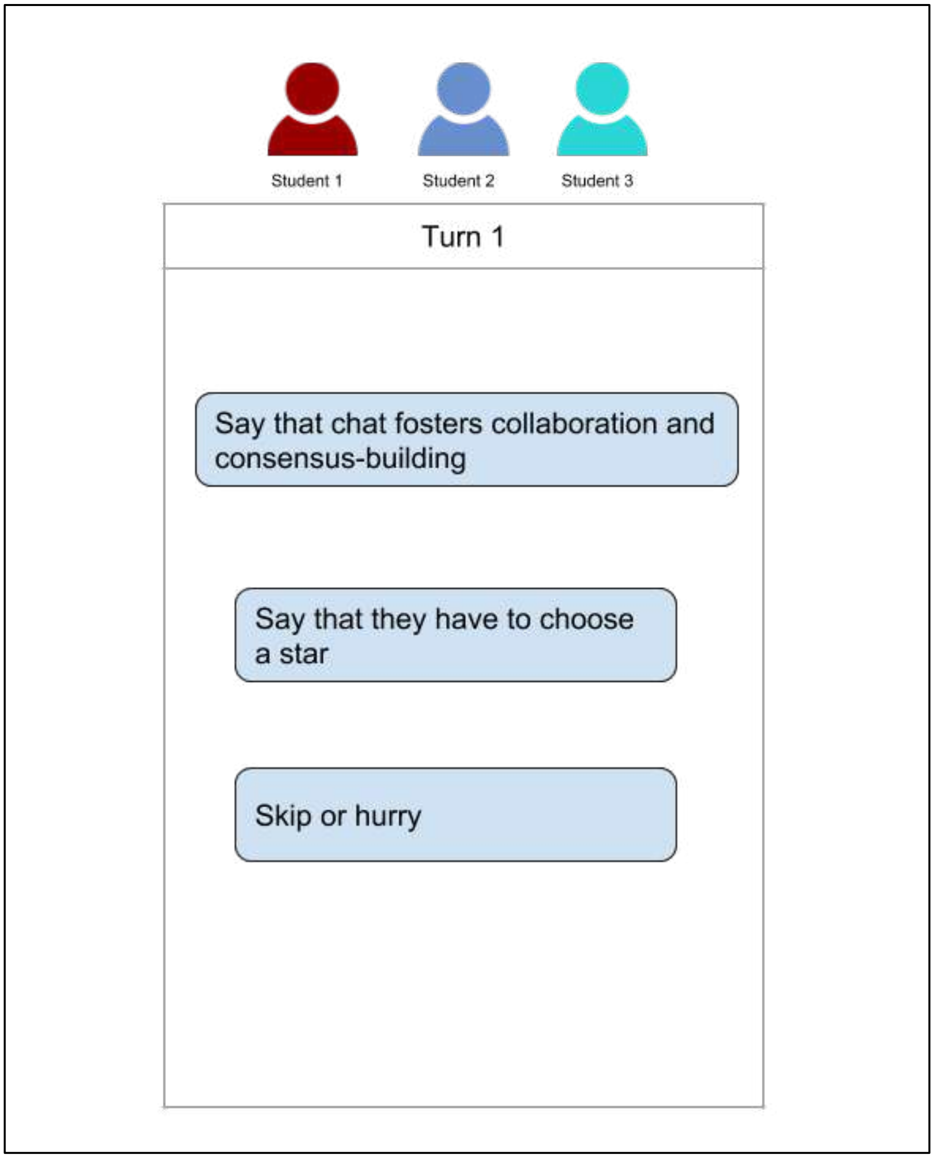


Figure B.2: Conversation structure to assess A2 skill

Dialog structure A3

CPS Skill: Understanding roles to solve problem

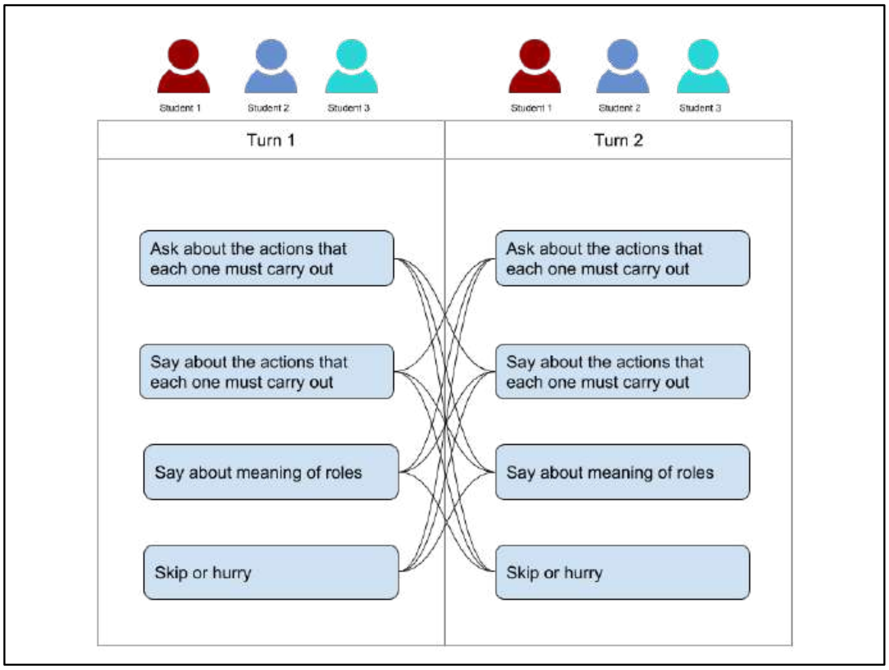


Figure B.3: Conversation structure to assess A3 skill

Dialog structure B1

CPS Skill: Building a shared representation and negotiating the meaning of the problem (common ground)

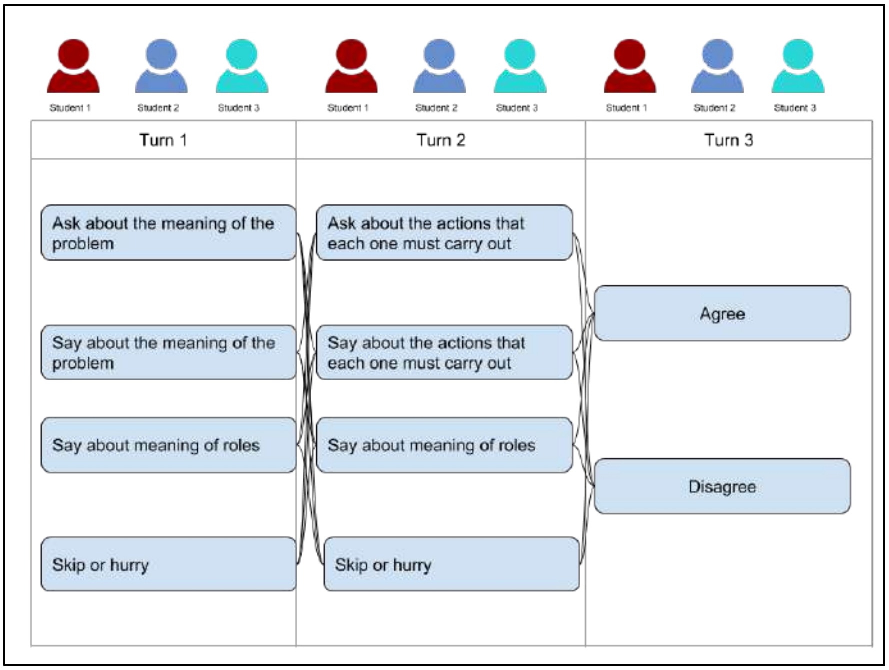


Figure B.4: Conversation structure to assess B1 skill

Dialog structure B2

CPS Skill: Identifying and describing tasks to be completed

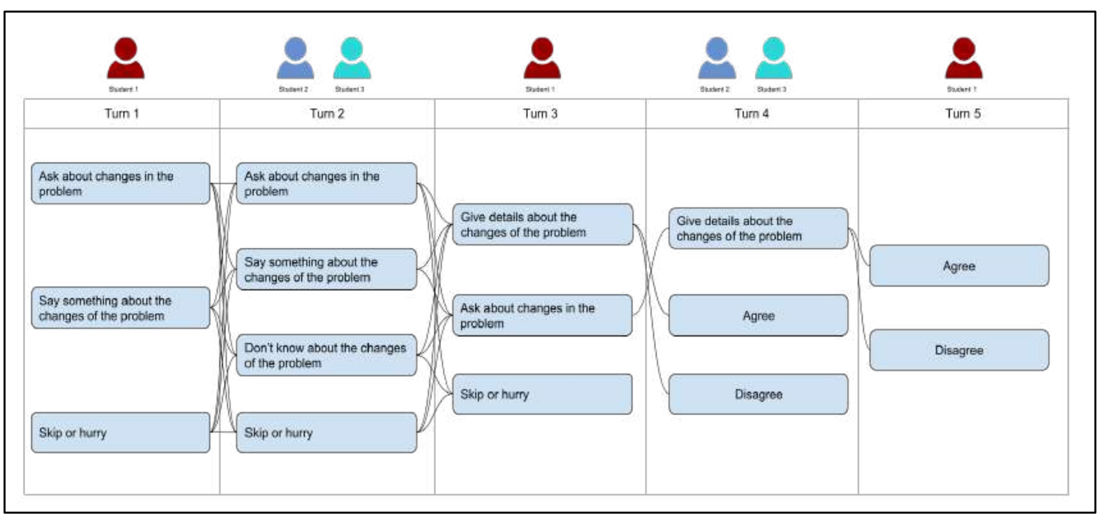


Figure B.5: Conversation structure to assess B2 skill

Dialog structure B3

CPS Skill: Describe roles and team organization (communication protocol/rules of engagement)

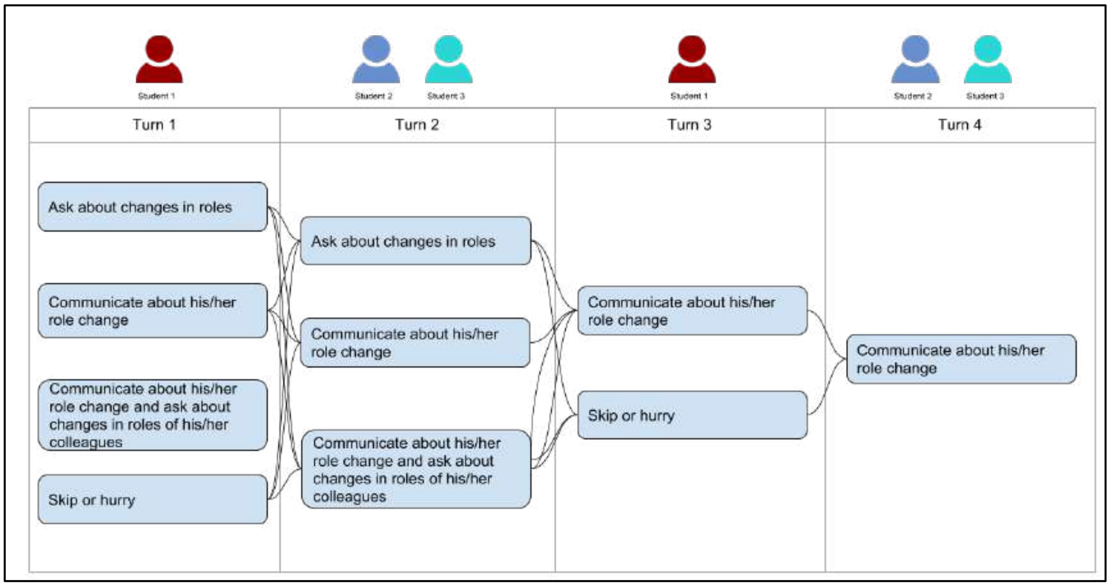


Figure B.6: Conversation structure to assess B3 skill

Dialog structure C1

CPS Skill: Communicating with team members about the actions to be/ being performed

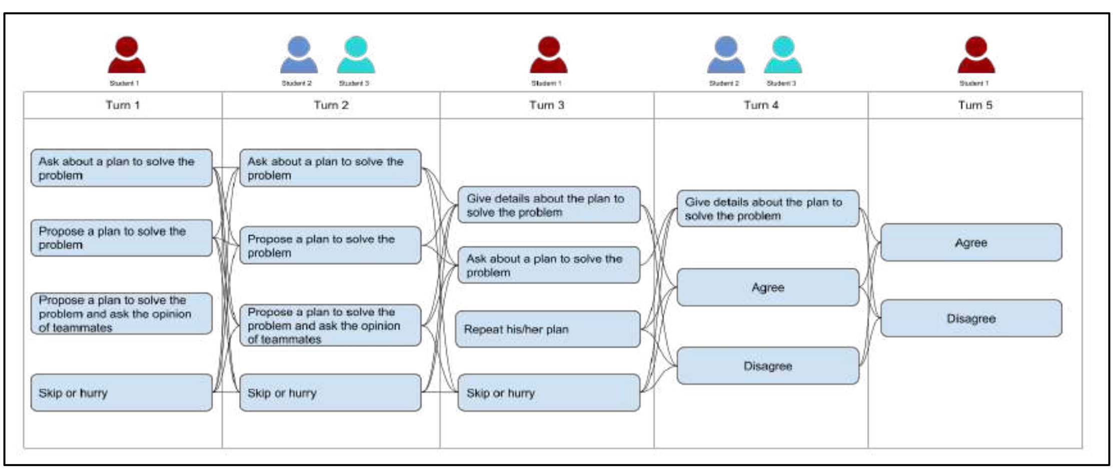


Figure B.7: Conversation structure to assess C1 skill

Dialog structure C2

CPS Skill: Enacting plans

This skill was related to actions that test takers took to solve the problem. As a result, no dialog structure is created.

Dialog structure C3

CPS Skill: Following rules of engagement, (e.g., prompting other team members to perform their tasks.)

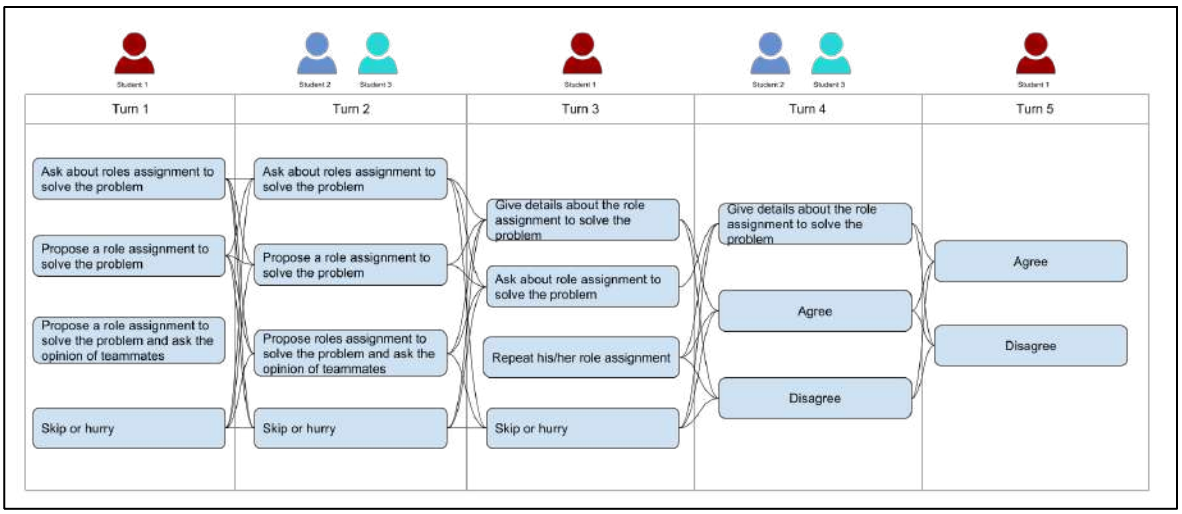


Figure B.8: Conversation structure to assess C3 skill

Dialog structure D1

CPS Skill: Monitoring and repairing the shared understanding

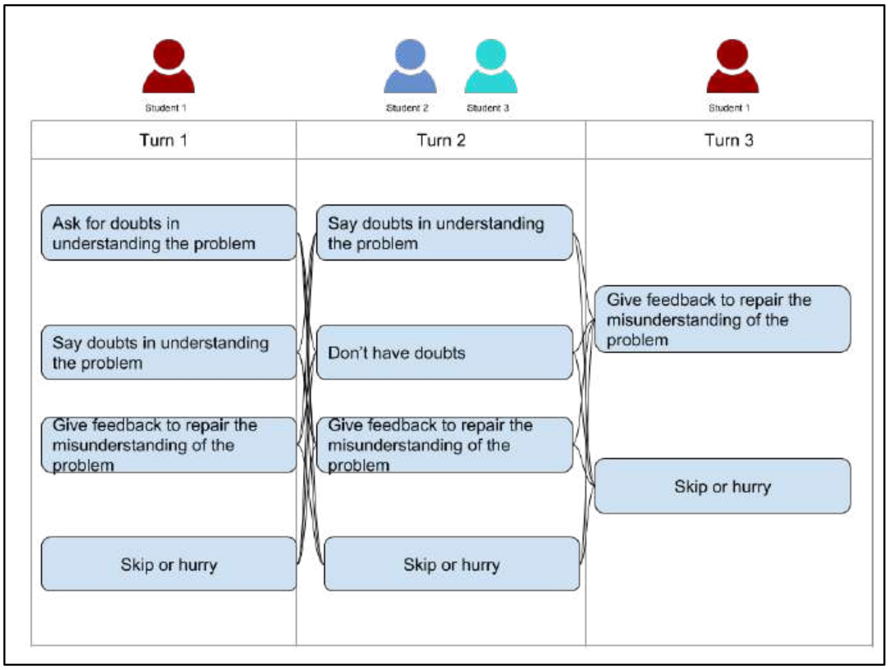


Figure B.9: Conversation structure to assess D1 skill

Dialog structure D2

CPS Skill: Monitoring results of actions and evaluating success in solving the problem

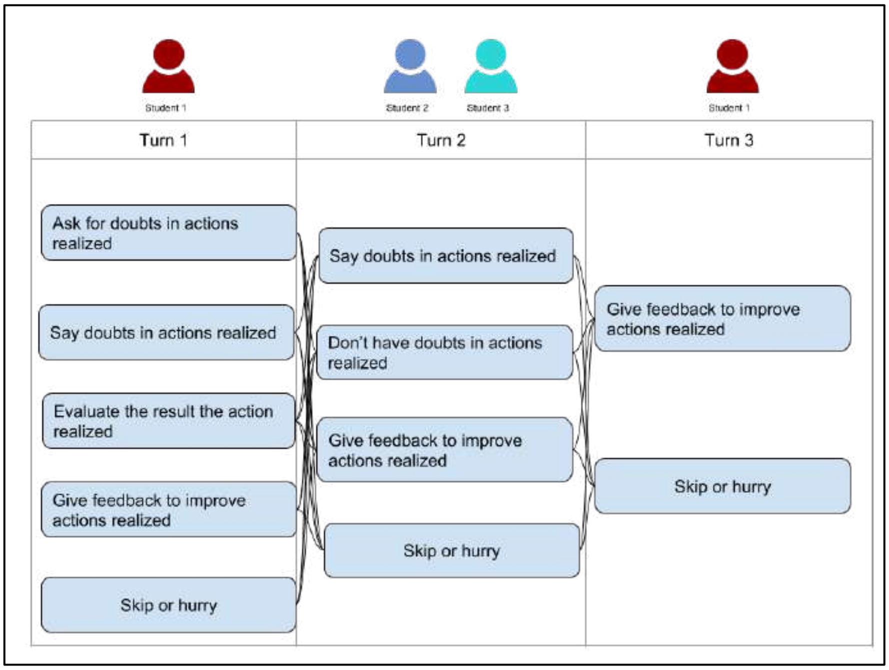


Figure B.10: Conversation structure to assess D2 skill

Dialog structure D3

CPS Skill: Monitoring, providing feedback and adapting the team organisation and roles

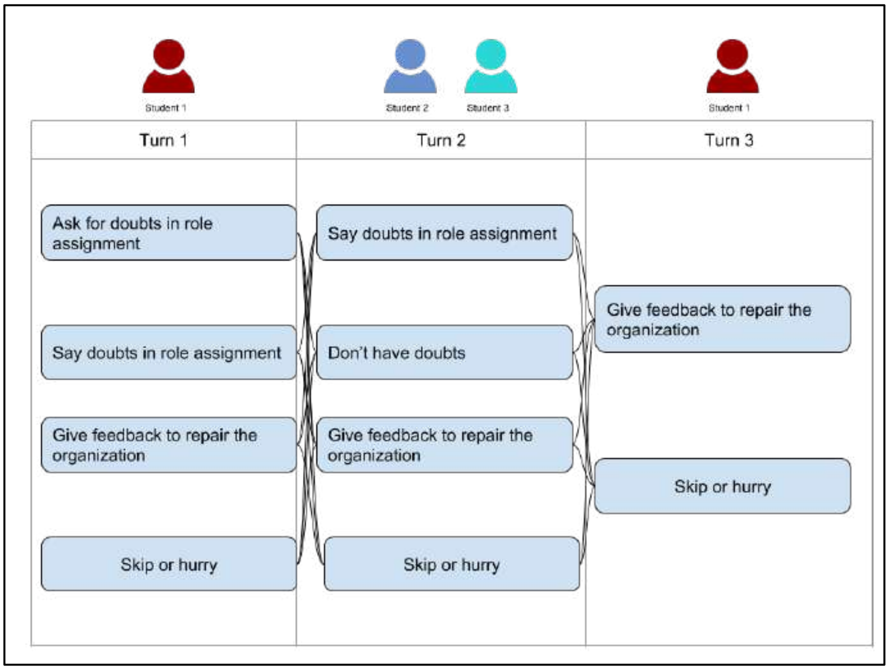


Figure B.11: Conversation structure to assess D3 skill

APPENDIX C: ACTIVITY EXPLANATION

Figure C.1 shows the interface that is presented to the students, in both the Human-Human version and the Human-Agent version. In the particular case of this example, the development of level 1 of 4 is presented. Students should look for the constellation showed to them, indicated as (4) in Figure C.1, among the set of stars (left half of the image C.1). The student must agree with his/her teammates about the solution they will select. In this case, one of the possible solutions are those indicated by the stars 1, 2 and 3, showed as (1) in Figure C.1. (Other solutions are stars 26, 27 and 28, (2) in Figure C.1, and 29, 30 and 31, (3) in Figure C.1). Because there are multiple solutions, three in this case, there is a high interdependence between peers. Also, the student who is taking part in the image (Figure C.1), has the role of selecting the blue star of the solution, (5) in Figure C.1, and must communicate to his/her teammates which blue star he/she will select. In (6) in Figure C.1, the messages sent by the three participants of the group are showed, while in (7) in Figure C.1, a message is displayed by the system. In (8) in Figure C.1, the messages that the student could send in this turn are displayed. Finally, in (9) in Figure C.1 is the button to send the selected message to his/her teammates.

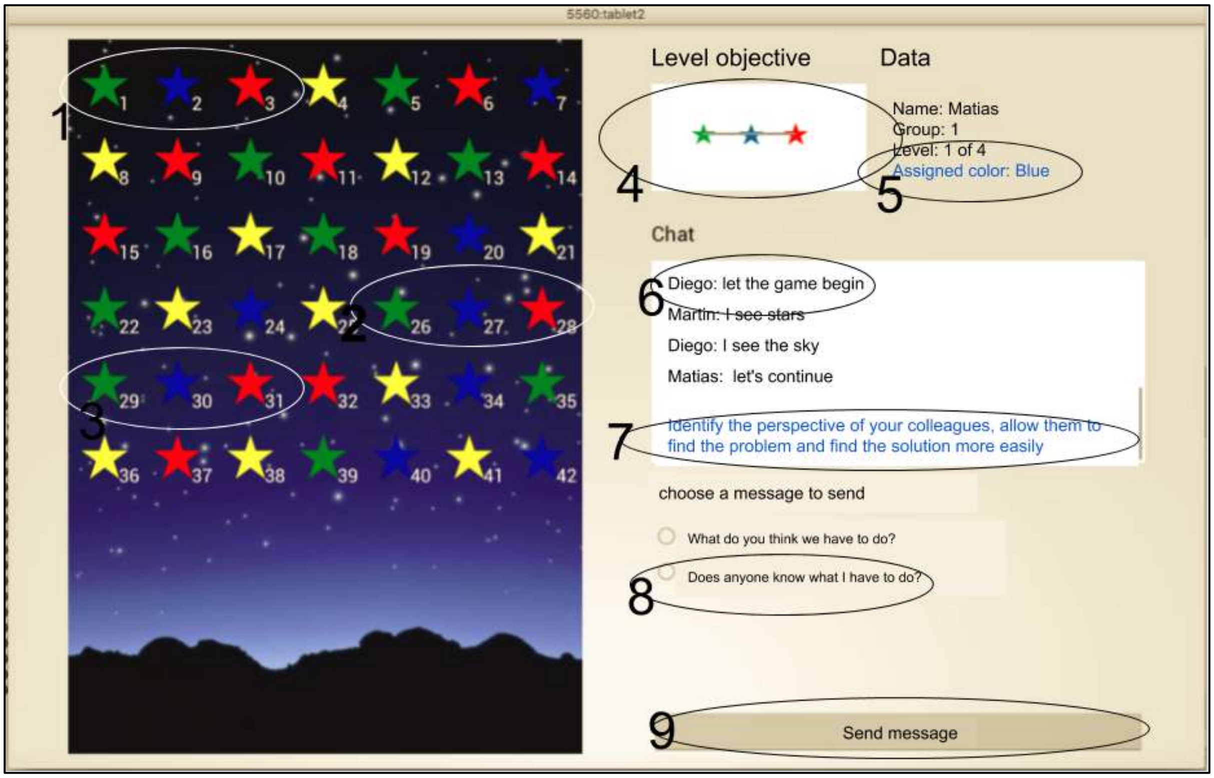


Figure C.1 User interface

APPENDIX D: VIRTUAL AGENTS EXAMPLE

As is described in Section 3.2, the conversations are assigned with a score between 0 to 3 points, where 0 indicates that there is no presence of the skill and 3 that the skill is achieved outstandingly. In this case, the message "select your stars" is not a message that asks or delivers relevant information to make a plan to solve the problem, rather it can be considered a mandatory order, even though the student does not know what solution they will implement, nor what star they will choose. Therefore, the agents will respond with a certain coherence and with intentions of recovering by giving another opportunity to the student to collaborate (Table D.1).

Table D.1: Virtual Agent Messages Example C1 Skill

Turn	Student 1	Agent 1	Agent 2
1	*Select your stars*		
2		*I do not know which stars we should choose*	
3			Does anyone else know what to choose?
4	*If we select the stars ##, ## and ##, we form the same objective constellation, because it has the same colors and order.*		
5		*I'm going to choose the stars ## and ##*	
6			*I'm going to choose the stars ## and ##*
7	*Yes, I agree.*	*Yes, I agree XX..*	

Note*: The symbols ## are replaced by a selector in the interface, which allows the user to select a number between 1 to 46, corresponding to the index of the star or atom of which he is speaking.*

XX is replaced by the interface by the name of student 1

APPENDIX E: ITERATION 1 RESULTS

E.1 Usability

Table E.1: Summary of Usability Evaluation

	Item	Expert 1	Expert 2	Expert 3	Comments
1	It is easy to learn how to use the application.	Yes	Yes	Yes	2: Tutorial must have an option to return 3: The instructions read at the beginning are useful
2	It is easy to use the application.	Yes	Yes	Yes	2: It is important that when you have to select the atom/star, the message is below the circles and people are paying attention to the chat, it is better to put the message in a way that looks better. 3: I had to stop to think how to start at the beginning.
3	Steps to use the application are easy to follow.	Yes	Yes	Yes	

	Item	Expert 1	Expert 2	Expert 3	Comments
4	Labels and texts on buttons are clear and concise.	Yes	No	Yes	1: It may be strange that there is "Send" and "Send response", that may confuse. 2: They should only change the word "constellation" to "stars"
5	Conserve overall consistency and behavior with the mobile platform.	Yes	No	Yes	2: The same message of point 2
6	Minimalist design - Excessive features are removed.	Yes	Yes	Yes	2: black background can be analyzed if it can be a little more concise with the rest of the interface 3: here was nothing left over, although at the beginning the chat seemed to me that I could write
7	Content is concise and clear.	Yes	Yes	Yes	
8	Provide feedback to the user about the state of the system.	Yes	No	Yes	3: It is very good to know what my teammates are doing

	Item	Expert 1	Expert 2	Expert 3	Comments
9	Number of buttons/links are reasonable.	Yes	Yes	Yes	
10	Elements of the interface provide visual feedback when are pressed.	Yes	Yes	Yes	
11	Ensures that not any message/feedback is being covered by the user's hand or finger.	Yes	No	Yes	
12	Colors used provide a good contrast.	Yes	Yes	Yes	2: Same as point 6 3: I found it hard to see that my atom was selected, I think that white does not contrast
13	Colors used provide good readability.	Yes	Yes	Yes	
14	Icons are clear to understand - there is no ambiguity.	Yes	Yes	Yes	
15	Font size and spacing ensure good readability.	Yes	No	No	2: The font size of the chat may be a little larger and could support color blindness

	Item	Expert 1	Expert 2	Expert 3	Comments
16	Speak the language of the users (not technical)	Yes	Yes	Yes	
17	It helps users recognize, diagnose and recover errors.	Yes	No	Yes	
18	Error messages are free of technical language	Yes	Yes	Yes	
19	Error messages clearly explain how to solve the problem.	Yes	Yes	Yes	
20	Help messages are clear and unambiguous.	Yes	Yes	Yes	
21	Instructions are easily visible or easily recoverable when appropriate.	Yes	No	No	2: Same as point 2
22	By making mistakes, mistakes can be easily corrected.	No	Yes	Yes	1: Once you make a mistake there is no going back, but I think it is part of the design of the activity so it is good that it is so.
23	By making a mistake, external help is required	No	No	Yes	2: It could be allowed to send a message after a time if the student does not

	Item	Expert 1	Expert 2	Expert 3	Comments
	to continue using the application.				answer, to help him/her if he/she does not know what to do
24	Images (stars and/or atoms) allow to identify their number and associated color.	Yes	Yes	Yes	
25	The tutorial is clear and allows you to understand the dynamics of the game.	Yes	Yes	Yes	2: Tutorial must have an option to return, n case it is not clear

E.2 Validity and reliability

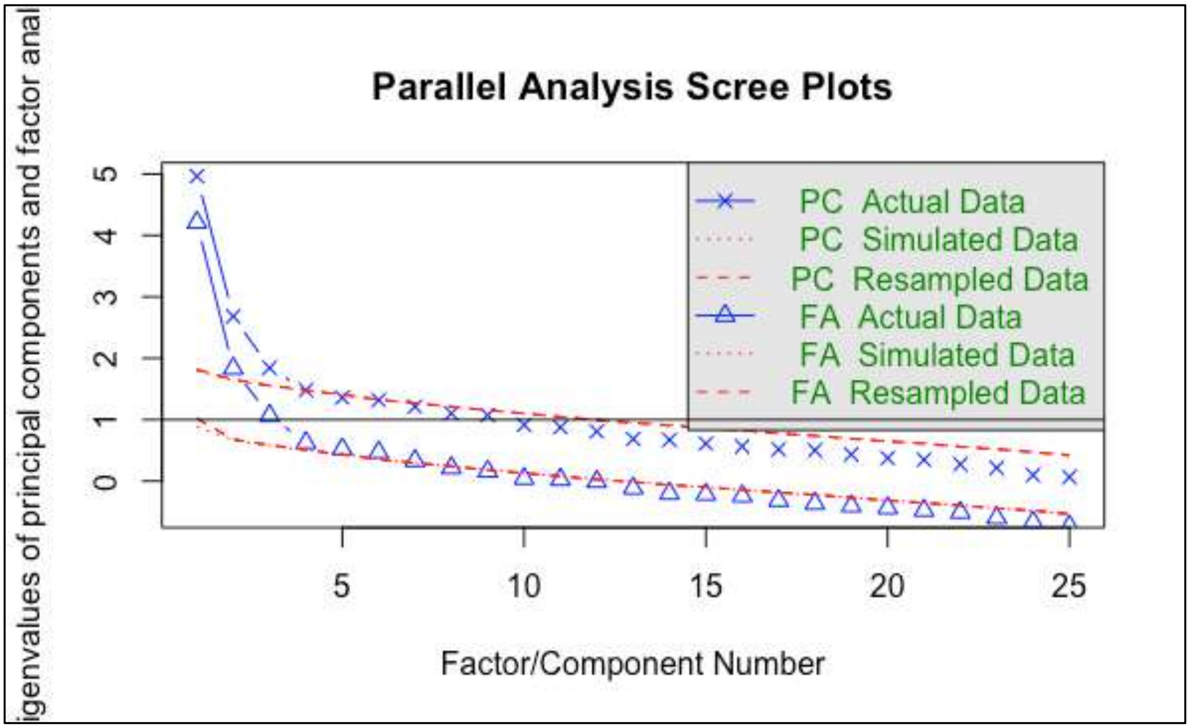


Figure E.1: Parallel analysis scree plot form A

Table E.2: Communalities Form A

Item	h2	u2	Item	h2	u2	Item	h2	u2	Item	h2	u2
1	**.071**	.929	8	.224	.776	15	.26	.74	22	.34	.66
2	.413	.587	9	.268	.732	16	.657	.343	23	.957	.043
3	.467	.533	10	.22	.78	17	.695	.305	24	.334	.666
4	.548	.452	11	.851	.149	18	.402	.598	25	.848	.152
5	.631	.369	12	.894	.106	19	.222	.778			
6	.107	.893	13	.404	.596	20	.782	.218			
7	.321	.679	14	.958	.042	21	**.046**	.954			

Table E.3: Exploratory Factor Analysis Form A

Item	1	2	3	4	5	6	7	Assessment Criteria	Cronbach's Alpha	Variance %
7	.31							Mention actions/plans to solve the problem.	0.718	.4600
16	.72							Mention actions to be taken together to solve the problem.		
17	.75							Mention actions to be carried out by the other members of his/her group (control).		
19		.45						Evaluate the success of the strategy to solve the problem.	0.749	
20		.85						He/She proposes changes in the actions to reach the solution.		
24		.48						Evaluate the success of the partners' actions to solve the problem.		
25		.87						He/She proposes changes in the actions of his/her partners if the group can not reach a solution.		
12			.90					Identify errors in the understanding of the problem by the group.	0.951	
14			.93					Mention ideas to solve errors in understanding the problem.		

6	.12			Ask about others' ideas about the purpose of the problem.	0.554
8	.24			Ask about actions/plans of other group members to solve the problem.	
10	.37			Ask about what is necessary to solve the problem.	
18	.41			Mention the strategy (actions) to solve the problem.	
22	.42			Mention roles to be taken by him/her and his/her teammates.	
23	.91			Mention actions to be carried out by the other members of his/her group (control) according to the assigned role.	
3		.64		Mention what he/she understand of the problem to be solved.	0.671
5		.78		Mention his/her ideas about the purpose of the problem	
9		.42		Mention what is necessary to solve the problem.	
11			.98	Identify gaps (lack of information) in the understanding of the problem by his/her group.	0.713

13	.55		Mention ideas to resolve gaps (lack of information) in understanding the problem.	
1		.26	Mention his/her ideas/opinions when necessary.	0.567
2		.33	Ask about the ideas/opinions of other group members.	
4		.30	Ask about what others understand about the problem to be solved.	
15		.23	Ask about joint actions to be taken to solve the problem.	
21		.20	Ask his/her teammates about the roles they must take to solve the problem.	

Note*: Rotated factors matrix, method of analysis of principal factors with varimax rotation.*

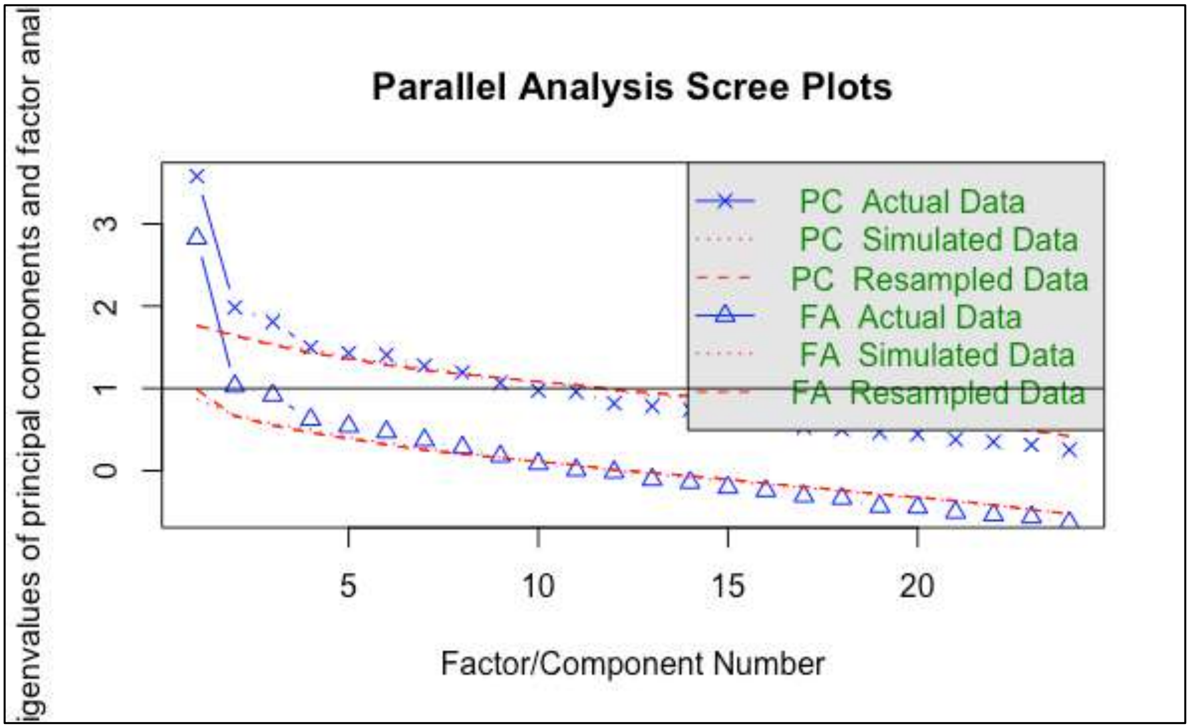


Figure E.2: Parallel analysis scree plot form B

Table E.4: Communalities Form A

Item	h2	u2	Item	h2	u2	Item	h2	u2	Item	h2	u2
1	**.062**	.938	8	.309	.691	16	.454	.546	23	.934	.066
2	.444	.556	9	.167	.833	17	.541	.459	24	.512	.488
3	.245	.755	10	.287	.713	18	.252	.748	25	.648	.352
4	.435	.565	11	.504	.496	19	.435	.565			
5	.706	.294	12	.144	.856	20	.557	.443			
6	.855	.145	13	.296	.704	21	**.048**	.952			
7	.466	.534	15	.323	.677	22	.539	.461			

Table E.5: Exploratory Factor Analysis Form B

Item	1	2	3	4	5	6	7	8	Assessment Criteria	Cronbach's Alpha	Variance %
2	.60								Ask about the ideas/opinions of other group members.	0.617	.4234
4	.64								Ask about what others understand about the problem to be solved.		
8	.45								Ask about actions/plans of other group members to solve the problem.		
15	.50								Ask about joint actions to be taken to solve the problem.		
22		.71							Mention roles to be taken by him/her and his/her teammates.	0.823	
23		.95							Mention actions to be carried out by the other members of his/her group (control) according to the assigned role.		

12	.34				Identify errors in the understanding of the problem by the group.	0.561
25	.72				He/She proposes changes in the actions of his/her partners if the group can not reach a solution.	
20	.72				He/She proposes changes in the actions to reach the solution.	
6		.92			Ask about others' ideas about the purpose of the problem.	0.53
10		.48			Ask about what is necessary to solve the problem.	
19			.64		Evaluate the success of the strategy to solve the problem.	0.641
24			.70		Evaluate the success of the partners' actions to solve the problem.	
1				.22	Mention his/her ideas/opinions when necessary.	0.646

7	.46			Mention actions/plans to solve the problem.	
16	.52			Mention actions to be taken together to solve the problem.	
17	.57			Mention actions to be carried out by the other members of his/her group (control).	
18	.26			Mention the strategy (actions) to solve the problem.	
21	.19			Ask his/her teammates about the roles they must take to solve the problem.	
3		.36		Mention what he/she understand of the problem to be solved.	0.412
5		.83		Mention his/her ideas about the purpose of the problem	
9		.35		Mention what is necessary to solve the problem.	
11			.70	Identify gaps (lack of information) in the	0.563

		understanding of the problem by his/her group.
13	.53	Mention ideas to resolve gaps (lack of information) in understanding the problem.

***Note**: Rotated factors matrix, method of analysis of principal factors with varimax rotation.*

APPENDIX F: ASSESSMENT CRITERIA ITERATION 1

Table F.1 : Assessment criteria by CPS Skill

CPS Skill		Assessment criteria
A1	Discovering perspectives and abilities of team members	Mention his/her ideas/opinions when necessary.
		Ask about the ideas/opinions of other group members.
B1	Building a shared representation and negotiating the meaning of the problem (common ground)	Mention what he/she understand of the problem to be solved.
		Ask about what others understand about the problem to be solved.
		Mention his/her ideas about the purpose of the problem
		Ask about others' ideas about the purpose of the problem.
C1	Communicating with team members about the actions to be/ being performed	Mention actions/plans to solve the problem.
		Ask about actions/plans of other group members to solve the problem.
		Mention what is necessary to solve the problem.
		Ask about what is necessary to solve the problem.

D1	Monitoring and repairing the shared understanding	Identify gaps (lack of information) in the understanding of the problem by his/her group. Identify errors in the understanding of the problem by the group. Mention ideas to resolve gaps (lack of information) in understanding the problem. Mention ideas to solve errors in understanding the problem.
A2	Discovering the type of collaborative interaction to solve the problem, along with goals	Ask about joint actions to be taken to solve the problem. Mention actions to be taken together to solve the problem.
B2	Identifying and describing tasks to be completed	Mention actions to be carried out by the other members of his/her group (control).
C2	Enacting plans	Mention the strategy (actions) to solve the problem.
D2	Monitoring results of actions and evaluating success in solving the problem	Evaluate the success of the strategy to solve the problem. He/She proposes changes in the actions to reach the solution.
A3	Understanding roles to solve problem	Ask his/her teammates about the roles they must take to solve the problem.
B3	Describe roles and team organization (communication protocol/rules of engagement)	Mention roles to be taken by him/her and his/her teammates.

C3	Following rules of engagement, (e.g., prompting other team members to perform their tasks.)	Mention actions to be carried out by the other members of his/her group (control) according to the assigned role.
D3	Monitoring, providing feedback and adapting the team organization and roles	Evaluate the success of the partners' actions to solve the problem. He/She proposes changes in the actions of his/her partners if the group can not reach a solution.

APPENDIX G: USABILITY

Table G.1: Usability checklist

	Item	Yes/No	Comments
1	It is easy to learn how to use the application.		
2	It is easy to use the application.		
3	Steps to use the application are easy to follow.		
4	Labels and texts on buttons are clear and concise.		
5	Conserve overall consistency and behavior with the mobile platform.		
6	Minimalist design - Excessive features are removed.		
7	Content is concise and clear.		
8	Provide feedback to the user about the state of the system.		
9	Number of buttons/links are reasonable.		
10	Elements of the interface provide visual feedback when are pressed.		
11	Ensures that not any message/feedback is being covered by the user's hand or finger.		
12	Colors used provide a good contrast.		
13	Colors used provide good readability.		
14	Icons are clear to understand - there is no ambiguity.		
15	Font size and spacing ensure good readability.		

	Item	Yes/No	Comments
16	Speak the language of the users (not technical)		
17	It helps users recognize, diagnose and recover errors.		
18	Error messages are free of technical language		
19	Error messages clearly explain how to solve the problem.		
20	Help messages are clear and unambiguous.		
21	Instructions are easily visible or easily recoverable when appropriate.		
22	By making mistakes, mistakes can be easily corrected.		
24	Images (stars and/or atoms) allow to identify their number and associated color.		
25	The tutorial is clear and allows you to understand the dynamics of the game.		

APPENDIX H: LOG FILE EXAMPLE

Table H.1: Example of log file

Time	Group	Test Taker	Meesage
13:22:30	1	Student 2	Hello, we can talk on the Chat and thus agree.
13:22:33	1	Student 3	Hello, we can talk on the Chat and thus agree.
13:23:01	0	Student 5	Let's start
13:23:04	1	Student 6	Let's start
13:23:04	0	Student 4	Hello, we can talk on the Chat and thus agree.
13:23:19	1	Student 3	What do they see on their tablets?
13:23:29	1	Student 2	I see the figure that we must find and you?
13:23:32	1	Student 6	I see atoms
13:23:41	0	Student 1	Let's start
13:23:43	1	Student 3	It seems that we all see the same
13:23:50	1	Student 2	How do you think we have to do it?
13:23:52	1	Student 6	How do you think we have to do it?

13:23:54	0	Student 4	Let the game begin.
13:24:07	0	Student 1	Let the game begin.
13:24:08	1	Student 3	There are some colors but I don't know what they are for, let's continue.
13:24:13	1	Student 2	I think everyone can only select atoms of the color that was assigned to them.
13:24:15	0	Student 5	I see atoms and you?
13:24:25	1	Student 6	Ya, let's continue?
13:24:26	1	Student 3	Ya, let's continue?
13:24:30	0	Student 4	It seems that we all see the same
13:24:41	0	Student 1	It seems that we all see the same
13:24:48	1	Student 2	I already know which atom to choose
13:24:49	1	Student 6	I think we must find in the sky the molecule of the figure And you?
13:24:52	0	Student 4	We were all assigned a color that we can select.
13:24:53	0	Student 5	It seems that we all see the same
13:25:03	1	Student 3	I think there is more than one possible molecule
13:25:10	1	Student 2	It's all clear, are we still?

13:25:10	0	Student 1	There are some colors but I don't know what they are for, let's continue.
13:25:18	1	Student 6	You have to agree on which atoms to choose
13:25:18	0	Student 4	Ya, let's continue?
13:25:27	0	Student 1	Ya, let's continue?

APPENDIX I: ACTIVITY DETAILS

I.1 Objective

The objective of the activity is that students should form constellations (Form A) or molecules (Form B), which meet the exact pattern of stars or atoms that are presented to them as targets of each level. For this, students should work in teams in which each participant will have to select, among a set of stars or atoms, the quantity and color of elements indicated. In addition, they should consider that there are multiple solutions and that communicating is the way to take agreement.

I.2 Narrative

In Form A of the instrument, test taker is introduced to an astronaut, who invites he or she to help him get to his destination (Figure I.1).

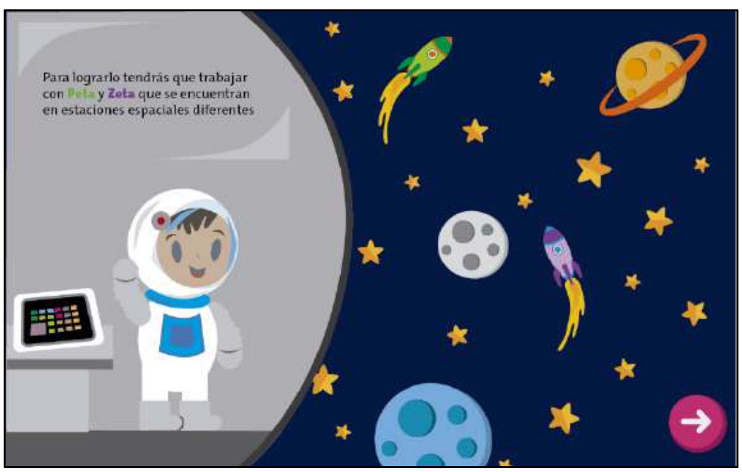


Figure I.1: Activity introduction Forma A

Then, as the first level begins, the ship's co-crew, Alpha and Beta, discuss with test taker the objectives of finding constellations, and talk about the buttons on the interface (Figure I.2).

Figure I.2: Activity main view Forma A

At the end of each level, results are presented to the test taker with the achievement or not of the actions they carried out (Figure I.3).

Figure I.3: End of a level Forma A

On the other hand, in Form B, test taker is introduced to a scientist, who invites he or she to help her in the laboratory (Figure I.4).

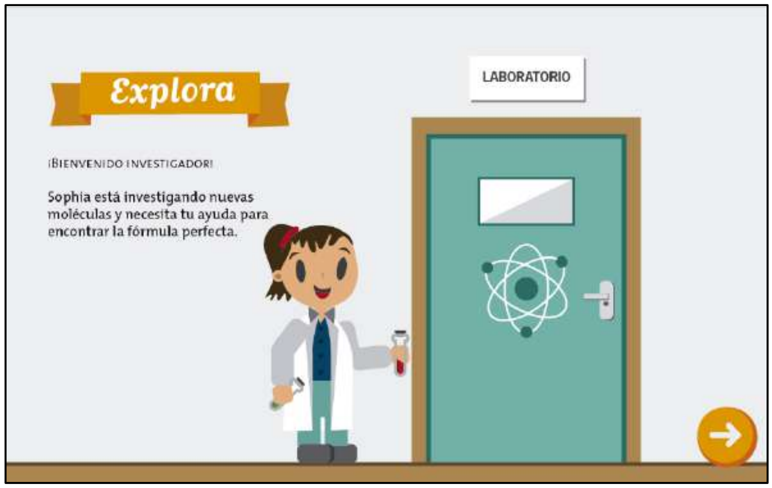


Figure I.4: Activity introduction Forma B

Then, at the beginning of the first level, the laboratory colleagues, Neutron and Proton, talk with test taker about the objectives of finding molecules, and talk about the buttons on the interface (Figure I.5).

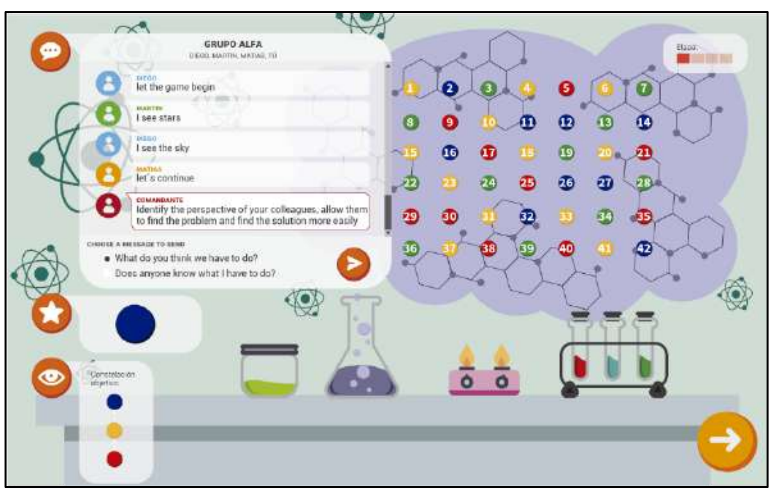


Figure I.5: Activity main view Forma B

At the end of each level, results are presented to the test taker with the achievement or not of the actions they carried out (Figure I.6).

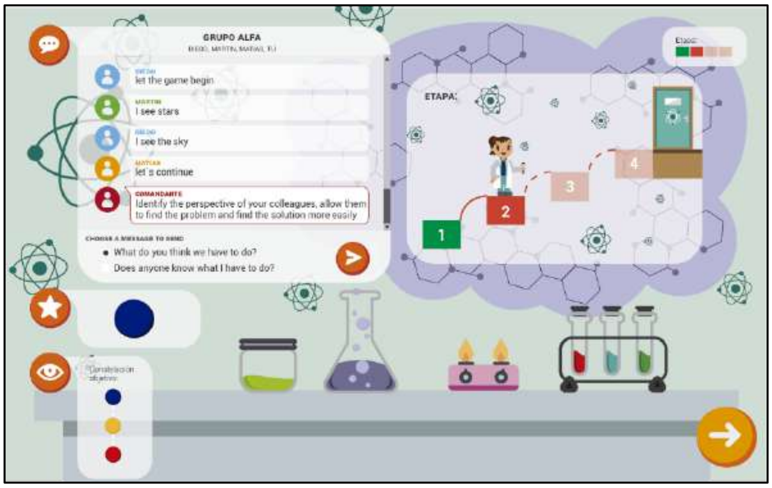


Figure I.6: End of a level Forma B

I.3 Levels target

Form A

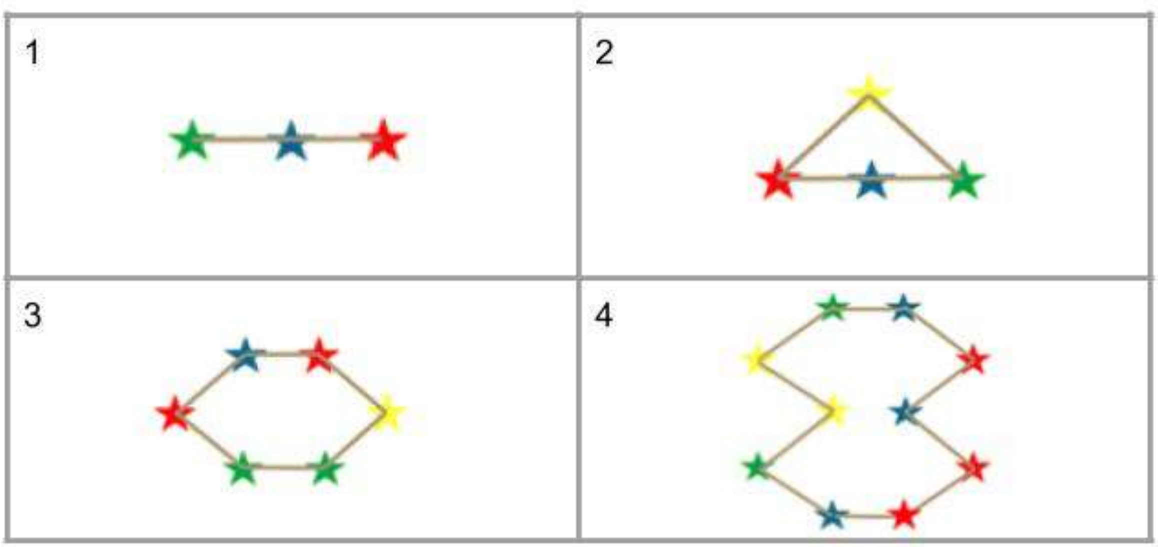


Figure I.7: Constelation target in each level of Forma A

Form B

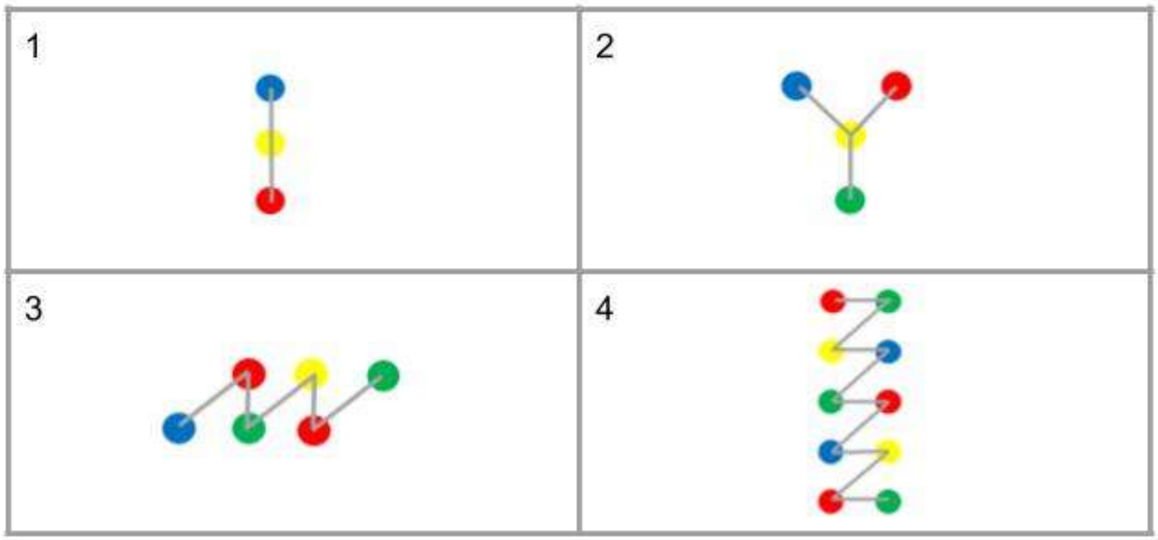


Figure I.8: Molecule target in each level of Form B

APPENDIX J: DIALOG AND SCORING EXAMPLE

Table J.1: Example of messages to related with C1 skill

Turn	Student 1	Student 2	Student 3
1	*What stars do we choose?*		
1	*I'm going to choose the stars ## and ##*		
1	*I'm going to choose the stars ## and ##. What do you think?*		
1	*XX select the ## and your ZZ the ##*		
1	*Select your stars*		
2		I think I have a plan.	
2		I do not know which stars we should choose.	
2		Does anyone else know what to choose?	
2		Yes I agree.	
2		I disagree.	
3			*I think I have a plan.*

3			*I think I have a different plan from XX.*
3			*I do not know which stars we should choose.*
3			*I agree with XX.*
3			*I disagree with XX.*
4	*If we select the stars ##, ## and ##, we form the same objective constellation, because it has the same colors and order.*		
4	*I'm going to choose the stars ## and ##*		
4	*I'm going to choose the stars ## and ##. What do you think?*		
4	*XX select the ## and your ZZ the ##*		
4	*Do what I said!*		
5		*If we select the stars ##, ## and ##, we form the same objective constellation, because it has the same colors and order.*	

5		*I'm going to choose the stars ## and ##*	
5		*I'm going to choose the stars ## and ##. What do you think?*	
5		*XX select the ## and your ZZ the ##*	
5		*I [agree, desagree] with XX.*	
6			*If we select the stars ##, ## and ##, we form the same objective constellation, because it has the same colors and order.*
6			*I'm going to choose the stars ## and ##*
6			*I'm going to choose the stars ## and ##. What do you think?*
6			*XX select the ## and your ZZ the ##*
6			*I [agree, desagree] with XX.*
7	*Yes I agree XX.*	*Yes I agree XX.*	*Yes I agree XX.*
7	*I disagree XX.*	*I disagree XX.*	*I disagree XX.*

***Note**: The symbols ## are replaced by a selector in the interface, which allows the user to select a number between 1 to 46, corresponding to the index of the star or atom of which he is speaking.*

XX, YY, ZZ, are replaced by the interface by the names of student 1, student 2 and student 3 respectively or by the names of the agents in the human-agent version.

Table J.2: Scoring example skill C1, 3 points

Turn	Student 1	Student 2	Student 3
1	*What stars do we choose?*		
2		*I'm going to choose the stars ## and ##. What do you think?*	
3			I'm going to choose the ##
4	*If we select the stars ##, ## and ##, we form the same objective constellation, because it has the same colors and order.*		
5		*I'm going to choose the stars ## and ##*	
6			*I'm going to choose the stars ## and ##*
7	*Yes, I agree XX.*	*Yes, I agree XX.*	

Table J.3: Scoring example skill C1, 0 points

Turn	Student 1	Student 2	Student 3
1	*Select your stars*		

2		*I'm going to choose the stars ## and ##. What do you think?*	
3			I'm going to choose the ##
4	*Do what I said!*		
5		*If we select the stars ##, ## and ##, we form the same objective constellation, because it has the same colors and order.*	
6			*I'm going to choose the stars ## and ##*
7	*I do not agree XX.*	*Yes, I agree XX..*	

***Note**: The symbols ## are replaced by a selector in the interface, which allows the user to select a number between 1 to 46, corresponding to the index of the star or atom of which he is speaking.*

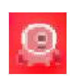
XX, YY, ZZ, are replaced by the interface by the names of student 1, student 2 and student 3 respectively or by the names of the agents in the human-agent version.

APPENDIX K: OVERVIEW OF THE VIDEOGAME

The U-Forest game consists of six stages graduated with difficulty maintaining a scaffolding process in dynamics and mechanics that exercise the use of the skills associated with collaborative problem-solving.

The game was designed based on skills matrix proposed for PISA 2015 (OECD, 2017) (Appendix L). There are three roles that players can assume when they enter the game (Table K.1). For neutrality, we summarise them into three names: 'Green', 'Red', and 'Yellow'. Each of these avatars has a unique ability to interact with the environment.

Table K.1: Game roles

avatar	Display	Ability
green		Protect players within the energy field.
yellow		Increased speed and double jump.
red		Destroy and move heavy objects.

There are six basic actions in the game: move horizontally, jump, shoot, activate/deactivate special ability, recover health and energy, and use basic items. These six actions are common to all three users.

This particular study only involved the second level of the game. This was because it was considered comprehensive enough to explore the effectiveness of the proposed solution and answer the research questions. Figure K.1 gives a general overview of level 2 of the game.

Figure K.1: Screenshot of the collaborative videogame

This level is split into two main sections. The first section ensures that the three players stick to their roles and can carry out the necessary actions in order to make progress as a team. They must first complete two shared tasks before coming to a fork. At this point, each player follows a separate path where they must complete character-specific challenges.

During the second section, the system tests whether the students are able to solve a task that requires greater levels of interdependence and synchronization among all three members of the group. In order to reach the end of the level, the players must protect themselves and avoid a large spider, before pushing a large boulder that is blocking their path. This can only be pushed into a pit if the yellow player activates a machine that is at the top of the map, protected by the spider. To do so, the rest of the group must distract the spider from below with protection from the green player. This increases the level of complexity of the collaboration among the members of the group.

The interface is managed by the teacher or operator (Figure K.2) and consists of two sections grouped on the right side of the screen. The first one is to activate the server (start the game), enter the number of players, and the stage at which each game room will start.

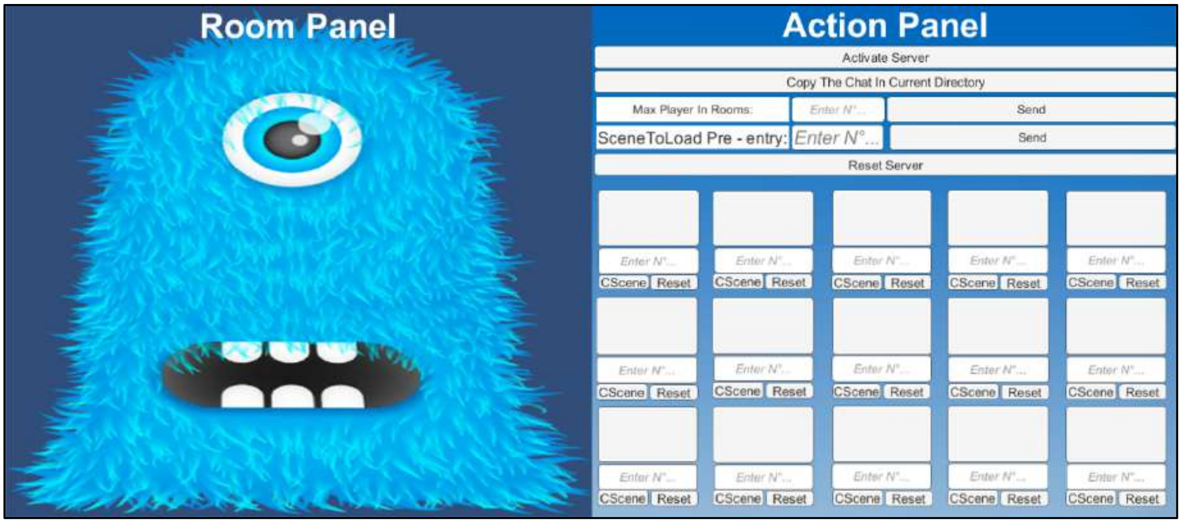


Figure K.2. Interface server control panel

In the second section, there are fifteen boxes. In these, the operator can monitor that the players enter the game correctly. Each of these units can be manipulated by the teacher. The system can: 1) restart the level; 2) Send players to their last checkpoint; and 3) change the players' stage. This section is helpful in case of problems with the game or if teachers want to move a team from one stage to another for a particular reason.

APPENDIX L: PDDL DOMAIN

```
(define (domain collabGame)
(:requirements :typing)
(:types player poi obstacle item switch - object
        mage warrior inventor - player
        rollable door jump barrier enemy - obstacle
        gear rune - item
        step lever machine doble triple - switch
)
(:predicates (player-at ?p - player ?x - poi)
             (player-distinct ?p ?r - player)
             (enemy-at ?e - enemy ?x - poi)
             (enemy-edge ?e - enemy ?x ?y - poi)
             (item-at ?x - item ?y - poi)
             (switch-at ?x - switch ?y - poi)
             (route-to ?x ?y - poi)
             (route-block ?x ?y - poi ?z - obstacle)
             (luring ?m - mage)
             (blocked ?x - obstacle)
             (open ?x - obstacle)
             (player-inventory ?p - player ?y - item)
             (linked-switch ?x - switch ?y - obstacle)
             (door-rune ?x - door ?y - rune)
             (door-route ?x ?y - poi ?z - door)
             (machine-gear ?x - machine ?y - gear)
             (machine-loaded ?x - machine)
             (switch-on ?s - switch)
             (item-assign ?i - item ?p - player)
             (switch-assign ?s - switch ?p - player)
             (rollable-locked ?r - rollable)
```

```
            (rollable-open ?r - rollable)
)
(:action move
   :parameters (?p - player ?x ?y - poi)
   :precondition (and (player-at ?p ?x)
                      (route-to ?x ?y)
                      (not (exists (?z - obstacle)
                           (and (route-block ?x ?y ?z)
                                (blocked ?z)))))
   :effect (and (not (player-at ?p ?x))
                (player-at ?p ?y)
                (forall (?e - enemy)
                      (when (and (enemy-edge ?e ?x ?y)
                                 (luring ?p)
                                 (open ?e))
                            (and (blocked ?e)
                                 (not (open ?e))
                                 (not (luring ?p))))))
)
(:action move-jump
   :parameters (?p - inventor ?x ?y - poi ?z - jump)
   :precondition (and (player-at ?p ?x)
                      (route-to ?x ?y)
                      (route-block ?x ?y ?z)
                      (blocked ?z))
   :effect (and (not (player-at ?p ?x))
                (player-at ?p ?y))
)
(:action move-through
   :parameters (?p - mage ?x ?y - poi ?z - barrier)
   :precondition (and (player-at ?p ?x)
```

```
                (route-to ?x ?y)
                (route-block ?x ?y ?z)
                (blocked ?z))
   :effect (and (not (player-at ?p ?x))
               (player-at ?p ?y))
)
(:action move-distract
   :parameters (?m - mage ?x ?y - poi ?z - enemy)
   :precondition (and (player-at ?m ?x)
                      (enemy-at ?z ?y)
                      (route-to ?x ?y)
                      (route-block ?x ?y ?z)
                      (blocked ?z))
   :effect (and (not (blocked ?z))
                (open ?z)
                (not (player-at ?m ?x))
                (luring ?m)
                (player-at ?m ?y))
)
(:action lever-on
   :parameters (?p - player ?s - lever ?x - poi)
   :precondition (and (player-at ?p ?x)
                      (switch-at ?s ?x)
                      (switch-assign ?s ?p)
                      (not (switch-on ?s)))
   :effect (and (switch-on ?s)
             (forall (?y - obstacle)
                  (when (and (linked-switch ?s ?y)
                             (blocked ?y))
                    (and (not (blocked ?y))
                             (open ?y)))))
```

```
)
(:action lever-off
    :parameters (?p - player ?s - lever ?x - poi)
    :precondition (and (player-at ?p ?x)
                       (switch-at ?s ?x)
                       (switch-assign ?s ?p)
                       (switch-on ?s))
    :effect (and (not (switch-on ?s))
                 (forall (?y - obstacle)
                     (when (and (linked-switch ?s ?y)
                                (open ?y))
                   (and (blocked ?y)
                        (not (open ?y))))))
)
(:action machine-on
    :parameters (?p - player ?s - machine ?x - poi)
    :precondition (and  (player-at ?p ?x)
                        (switch-at ?s ?x)
                        (switch-assign ?s ?p)
                        (not (switch-on ?s))
                        (machine-loaded ?s))
    :effect (and (switch-on ?s)
                 (forall (?y - rollable)
                     (when (and (linked-switch ?s ?y)
                                (rollable-locked ?y))
                           (and (not (rollable-locked ?y))
                                (rollable-open ?y)))))
)
(:action item-pick
    :parameters (?p - player ?i - item ?x - poi)
    :precondition (and (player-at ?p ?x)
```

```
                  (item-at ?i ?x))
  :effect (and (not (item-at ?i ?x))
              (player-inventory ?p ?i))
)
(:action item-drop
  :parameters (?p - player ?i - item ?x - poi)
  :precondition (and (player-at ?p ?x)
                     (player-inventory ?p ?i))
  :effect (and (not (player-inventory ?p ?i))
              (item-at ?i ?x))
)
(:action rune-use
  :parameters (?p - player ?r - rune ?x - poi ?z - door)
  :precondition (and (player-at ?p ?x)
                     (player-inventory ?p ?r)
                     (item-assign ?r ?p)
                     (blocked ?z)
                     (door-rune ?z ?r))
  :effect (forall (?y - poi)
              (when (and (route-to ?x ?y)
                         (route-block ?x ?y ?z)
                         (door-route ?x ?y ?z))
                    (and (not (blocked ?z))
                         (open ?z)
                         (not (player-inventory ?p ?r)))))
)
(:action gear-use
  :parameters (?p - player ?r - gear ?x - poi ?z - machine)
  :precondition (and (player-at ?p ?x)
                     (switch-at ?z ?x)
                     (player-inventory ?p ?r)
```

```
                    (item-assign ?r ?p)
                    (not (machine-loaded ?z))
                    (machine-gear ?z ?r))
  :effect (and (not (player-inventory ?p ?r))
               (machine-loaded ?z))
)
(:action step-on
  :parameters (?p - player ?s - step ?x - poi)
  :precondition (and (player-at ?p ?x)
                    (switch-at ?s ?x)
                    (switch-assign ?s ?p)
                    (not (switch-on ?s)))
  :effect (and (switch-on ?s)
               forall (?y - obstacle)
                    (when (and (linked-switch ?s ?y)
                               (blocked ?y))
                          (and (not (blocked ?y))
                               (open ?y)))))
)
(:action triple-switch
  :parameters (?p ?r ?s - player ?t - triple ?x - poi)
  :precondition (and (player-at ?p ?x)
                    (player-at ?r ?x)
                    (player-at ?s ?x)
                    (switch-at ?t ?x)
                    (switch-assign ?t ?p)
                    (switch-assign ?t ?r)
                    (switch-assign ?t ?s)
                    (player-distinct ?p ?r)
                    (player-distinct ?r ?s)
                    (player-distinct ?p ?s))
```

```
    :effect (forall (?z - obstacle)
                       (when (and (linked-switch ?t ?z)
                                  (blocked ?z))
                             (and (not (blocked ?z))
                                  (open ?z))))
)
(:action doble-switch
    :parameters (?p ?r - player ?t - doble ?x - poi)
    :precondition (and (player-at ?p ?x)
                       (player-at ?r ?x)
                       (switch-at ?t ?x)
                       (switch-assign ?t ?p)
                       (switch-assign ?t ?r)
                       (player-distinct ?p ?r))
    :effect (forall (?z - obstacle)
                       (when (and (linked-switch ?t ?z)
                                  (blocked ?z))
                             (and (not (blocked ?z))
                                  (open ?z))))
)
(:action push-boulder
    :parameters (?p - warrior ?r - rollable ?x ?y - poi)
    :precondition (and (player-at ?p ?x)
                       (route-to ?x ?y)
                       (route-block ?x ?y ?r)
                       (blocked ?r)
                       (rollable-open ?r))
    :effect (and (not (blocked ?r))
                 (open ?r))
)
)
```

APPENDIX M: COLLABORATIVE PROBLEM-SOLVING SKILL MATRIX PISA

Table M.1: *CPS Framework PISA 2015 (OECD, 2013)*

	1. Establishing and maintaining shared understanding	2. Taking appropriate action to solve the problem	3. Establishing and maintaining team organisation
A. Exploring and Understanding	(A1) Discovering perspectives and abilities of team members	(A2) Discovering the type of collaborative interaction to solve the problem, along with goals	(A3) Understanding roles to solve problem
B. Representing and Formulating	(B1) Building a shared representation and negotiating the meaning of the problem (common ground)	(B2) Identifying and describing tasks to be completed	(B3) Describe roles and team organisation (communication protocol/rules of engagement)
C. Planning and Executing	(C1) Communicating with team members about the actions to be/ being performed	(C2) Enacting plans	(C3) Following rules of engagement, (e.g., prompting other team members to perform their tasks.)
D. Monitoring and Reflecting	(D1) Monitoring and repairing the	(D2) Monitoring results of actions	(D3) Monitoring, providing feedback

	shared understanding	and evaluating success in solving the problem	and adapting the team organisation and roles

APPENDIX N: ACTION LOG EXAMPLE

12:08:03 Player 0 is going right from (-20.47,-43.9401)

12:08:04 Player 2 is going right from (-19.736,-43.94174)

12:08:11 Player 0 is going right from (-18.34056,-43.9401)

12:08:12 Player 2 is going right from (-18.36893,-43.94174)

12:08:12 Player 0 is going right from (-16.051,-43.9401)

12:08:13 Player 1 is going right from (-20.13,-43.94174)

12:08:13 Player 1 is going left from (-19.24194,-43.94174)

12:08:14 Player 2 is going right from (-15.09915,-43.94174)

12:08:14 Player 1 jumped from (-20.31115,-43.94174)

12:08:14 Player 1 is going right from (-20.31115,-42.88129)

12:08:15 Player 2 is going right from (-13.36335,-43.94174)

12:08:15 Player 0 is going right from (-16.04792,-43.9401)

12:08:15 Player 0 is going right from (-16.04177,-43.9401)

12:08:15 Player 1 attacked

12:08:15 Player 0 is going right from (-16.03561,-43.9401)

12:08:15 Player 0 is going right from (-16.01631,-43.9401)

12:08:15 Player 0 is going right from (-15.95362,-43.9401)

12:08:16 Player 2 attacked

12:08:16 Player 1 used his power

12:08:16 Player 0 is going right from (-15.03481,-43.9401)

12:08:17 Player 2 is going left from (-10.42481,-43.94174)

12:08:17 Player 2 attacked

12:08:17 Player 1 is going right from (-19.52016,-43.94174)

12:08:18 Player 2 attacked

12:08:18 Player 0 used his power

12:08:18 Player 2 is going right from (-11.63062,-43.94174)

12:08:18 Player 2 attacked

12:08:19 Player 2 attacked

12:08:19 Player 0 is going right from (-7.62048,-43.9401)

12:08:19 Player 0 jumped from (-7.563948,-43.9401)

12:08:19 Player 0 is going right from (-7.521948,-43.64953)

12:08:19 Player 2 attacked

12:08:20 Player 2 attacked

12:08:20 Player 0 jumped from (-5.473409,-43.94174)

12:08:20 Player 0 is going right from (-5.333408,-43.65117)

12:08:20 Player 2 attacked

12:08:20 Player 2 is going left from (-6.033129,-43.94174)

12:08:20 Player 2 attacked

12:08:21 Player 0 jumped from (-3.313319,-43.94174)

12:08:21 Player 0 is going right from (-3.313319,-43.65117)

12:08:21 Player 2 attacked

12:08:21 Player 2 is going right from (-7.641231,-43.94174)

12:08:21 Player 2 jumped from (-7.236178,-43.94174)

12:08:21 Player 2 is going right from (-7.138179,-43.65117)

12:08:22 Player 0 is going right from (-1.772227,-42.99001)

12:08:22 Player 0 stopped using his power

12:08:22 Player 2 jumped from (-4.648123,-43.94174)

12:08:22 Player 2 is going right from (-4.578123,-43.79155)

12:08:23 Player 0 used his power

12:08:24 Player 0 jumped from (-0.9758069,-42.99001)

12:08:24 Player 1 jumped from (-18.42769,-43.94174)

12:08:25 Player 2 jumped from (2.963484,-42.99001)

12:08:25 Player 2 is going right from (3.103483,-42.69944)

12:08:25 Player 1 is going right from (-18.42769,-43.12999)

12:08:25 Player 2 jumped from (5.349709,-42.79501)

12:08:25 Player 0 stopped using his power

12:08:25 Player 2 is going right from (5.489709,-42.50444)

12:08:26 Player 2 is going left from (6.998141,-42.13402)

12:08:26 Player 1 stopped using his power

12:08:26 Player 2 is going right from (6.32847,-42.13401)

12:08:26 Player 0 is going right from (-0.9758069,-42.99001)

12:08:27 Player 2 jumped from (6.603915,-42.13401)

12:08:27 Player 2 is going right from (6.645915,-41.98382)

12:08:27 Player 1 is going right from (-16.79326,-43.94174)

12:08:27 Player 0 attacked

12:08:27 Player 2 is going left from (8.234917,-41.36348)

12:08:28 Player 1 is going right from (-16.27929,-43.94174)

12:08:28 Player 2 is going right from (8.087702,-41.36348)

12:08:28 Player 0 used his power

12:08:29 Player 0 jumped from (4.147079,-42.99)

12:08:29 Player 0 is going right from (4.147079,-42.83981)

12:08:30 Player 0 jumped from (6.08812,-43.94177)

12:08:30 Player 2 is going right from (9.747133,-42.09348)

12:08:30 Player 0 is going right from (6.08812,-42.71897)

12:08:31 Player 2 is going right from (11.15234,-42.79501)

12:08:31 Player 0 jumped from (8.185307,-43.94174)

12:08:31 Player 2 jumped from (11.86539,-42.79501)

12:08:31 Player 2 is going right from (12.00539,-42.50444)

12:08:31 Player 0 is going right from (8.185307,-42.84785)

12:08:32 Player 1 jumped from (-4.454295,-43.94174)

12:08:32 Player 1 is going right from (-4.314294,-43.65117)

12:08:32 Player 2 is going left from (13.82538,-42.83472)

12:08:32 Player 0 is going right from (8.692822,-43.94174)

12:08:32 Player 2 is going right from (13.71455,-42.83472)

12:08:32 Player 0 jumped from (9.066798,-43.94174)

12:08:32 Player 0 is going right from (9.164798,-43.65117)

12:08:33 Player 0 jumped from (11.4389,-43.94174)

12:08:33 Player 0 is going right from (11.5789,-43.65117)

12:08:33 Player 1 jumped from (1.330727,-42.99001)

12:08:34 Player 2 is going right from (15.98289,-42.83472)

12:08:34 Player 1 is going right from (1.330727,-42.99001)

12:08:34 Player 2 is going right from (16.13095,-42.83472)

12:08:34 Player 1 is going left from (3.080727,-42.61736)

12:08:34 Player 0 jumped from (15.30538,-42.83472)

12:08:34 Player 2 is going right from (16.32802,-42.83472)

12:08:34 Player 0 is going right from (15.44538,-42.54416)

12:08:34 Player 1 is going right from (3.027303,-42.99001)

12:08:35 Player 2 jumped from (16.46293,-42.83472)

12:08:35 Player 2 is going right from (16.49793,-42.68454)

12:08:35 Player 1 jumped from (3.243671,-42.99001)

12:08:35 Player 1 is going right from (3.313671,-42.69944)

12:08:35 Player 2 jumped from (19.32985,-43.94174)

12:08:35 Player 1 jumped from (5.541125,-42.79501)

12:08:36 Player 2 is going right from (19.39985,-43.79155)

12:08:36 Player 1 is going right from (5.668152,-42.50444)

12:08:36 Player 0 jumped from (19.83428,-43.94174)

12:08:36 Player 0 is going right from (19.90428,-43.79155)

12:08:36 Player 1 is going left from (6.08812,-41.57221)

12:08:36 Player 1 is going right from (5.909584,-42.13401)

12:08:37 Player 1 jumped from (6.181104,-42.13401)

12:08:37 Player 1 is going right from (6.223104,-41.98382)

12:08:37 Player 0 attacked

12:08:38 Player 0 jumped from (26.18911,-43.94174)

12:08:38 Player 0 is going right from (26.39601,-43.34489)

12:08:38 Player 2 is going left from (25.96076,-43.94174)

12:08:39 Player 1 is going right from (8.370479,-41.36348)

12:08:39 Player 1 jumped from (8.37663,-41.36348)

12:08:39 Player 1 is going right from (8.37663,-41.36348)

12:08:39 Player 0 is going right from (26.58475,-43.30501)

12:08:40 Player 2 is going left from (25.54223,-43.94174)

12:08:40 Player 2 attacked

12:08:40 Player 1 is going right from (12.11561,-42.85387)

12:08:40 Player 0 is going left from (26.58475,-43.30501)

12:08:40 Player 0 is going left from (26.52514,-43.30501)

12:08:40 Player 1 is going left from (12.35502,-43.94174)

12:08:41 Player 2 is going right from (24.24265,-43.94174)

12:08:41 Player 1 is going right from (12.30549,-43.94174)

12:08:41 Player 1 jumped from (12.30857,-43.94174)

12:08:41 Player 0 is going left from (25.35225,-43.94174)

12:08:41 Player 1 is going right from (12.33657,-43.65117)

12:08:41 Player 0 jumped from (25.33603,-43.94174)

12:08:41 Player 2 attacked

12:08:41 Player 0 is going left from (25.32203,-43.79155)

12:08:41 Player 2 attacked

12:08:42 Player 0 is going left from (24.51668,-43.09188)

12:08:42 Player 0 jumped from (24.5136,-43.09188)

12:08:42 Player 0 is going left from (24.5066,-42.94169)

12:08:42 Player 2 attacked

12:08:42 Player 2 is going left from (24.35264,-43.94174)

12:08:42 Player 0 is going right from (24.2266,-41.88939)

12:08:42 Player 2 attacked

12:08:43 Player 1 jumped from (16.67025,-42.83472)

12:08:43 Player 1 is going right from (16.81025,-42.54416)

12:08:43 Player 0 is going right from (24.64208,-42.24188)

12:08:43 Player 0 jumped from (24.70477,-42.24188)

12:08:43 Player 0 is going right from (24.72577,-42.09169)

12:08:43 Player 2 is going right from (24.19758,-43.94174)

12:08:43 Player 2 is going left from (24.30311,-43.94174)

12:08:44 Player 0 is going right from (26.53386,-43.23995)

12:08:44 Player 1 jumped from (21.226,-43.94174)

12:08:44 Player 2 jumped from (23.88767,-43.94174)

12:08:44 Player 1 is going right from (21.436,-43.5206)

12:08:44 Player 2 is going right from (23.88767,-43.79155)

12:08:44 Player 0 is going right from (26.58476,-43.30501)

12:08:44 Player 0 is going right from (26.58476,-43.30501)

12:08:45 Player 2 jumped from (24.97202,-43.09188)

12:08:45 Player 2 is going left from (24.97202,-42.67074)

12:08:45 Player 1 is going left from (23.25599,-43.94174)

12:08:45 Player 2 is going right from (24.67102,-42.24188)

APPENDIX O: EXAMPLE OF A RECORDED CONVERSATION

R: I’m the red one

G: I’m the green one

Y: I’m the yellow one

Y: Has it already started?

G: It hasn’t started yet

R: But it looks like the others have already started

R: Sir, has it already started?

EXTERNAL: Not until I say so

R: Hey, we don’t start until the teacher says so

R: I’m the red one

G: Yeah, look

G: Look at your screen

ANOTHER: THOSE WHO ARE READY CAN START

G: Can we start?

G: Right, let’s get more comfortable.

R: But wait, I can’t move

G: I can’t move

G: I can’t move

G: Weeeeee

R: How did you jump?

R: Where are you guys?

R: G, G, G how did you jump?

G: No, I fell

G: Wait

G: I don’t even care

G: I sound so stupid

Y: I can’t get up

Y: How did you jump?

G: Weeee, weeeee. No, I fell again

G: Weeee

G: Uh, there's a big one. I ate it. You didn't eat it.

R: What do I do here?

Y: Is this like a race?

R: Yeah, I don't know what I'm doing

G: Ah, I'm the green one, right? I got stuck.

R: I can't get up

G: Me neither, I got stuck

G: How did you get down? I look like a mummy, I can't move hahahaha. There we go. Weeee!

R: Hahaha

G: Weeee

Y: But I don't know how to jump

G: I can't. Wait

R: OK, look. I know what to do

G: I can't, I get stuck

Y: Me too, I don't know what to do

R: Help us

R: Hey, have you already been over there?

OTHERS: we haven't been able to get there

R: Ah OK

G: How do you shoot?

G: Oh, I'm shooting! Weeeee!

G: Oh, I fell down again

Y: Let's go!!

G: No, I fell

G: I'm back

G: Hey, who's the yellow one?

Y: Me

R: Look, I got to this bit

G: Ooh

R: How did you shoot?

G: With that one. With that one, look. Weeee

APPENDIX P: PERFORMANCE OF EACH METHOD

Table P.1: *Sequential method*

	Time for calculating the plan for each test								
Plan (Actions)	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9
Plan 1 (3 Actions)	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Plan 2 (4 Actions)	.0000	.0000	.0000	.5024	.5011	.5005	.0000	.0000	.0000
Plan 3 (5 Actions)	.0000	.5011	.0000	.0000	.0000	.0000	.5005	.0000	.0000
Plan 4 (6 Actions)	.0000	.5011	.0000	.5005	.0000	.5011	.0000	.0000	.5018
Plan 5 (7 Actions)	.0000	.0000	1,0036	.4986	.0000	.5017	.0000	.0000	.5011
Plan 6 (8 Actions)	.0000	.0000	.0000	.0000	.0000	.5030	1,0016	.0000	.0000
Plan 7 (9 Actions)	.5024	.5011	.0000	2,0052	.5018	.0000	.4998	.5024	.5044
Plan 8 (10 Actions)	.5018	.0000	.5005	.5012	1,0016	.5011	.0000	1,0009	.0000
Plan 9 (11 Actions)	1,0023	.5005	.4985	.5005	.4980	.5012	.4998	1,0016	.5011
Plan 10 (12 Actions)	1,5008	.5011	.4973	.5030	.4998	.5005	.5005	.5011	.5024

Plan 11 (13 Actions)	1,5022	1,5034	.0000	.5005	.4992	.5037	1,0004	1,5021	.5018
Plan 12 (14 Actions)	.5005	1,0010	.4979	.4986	.5012	1,0042	.4986	.5005	.5024
Plan 13 (15 Actions)	1,5041	.5012	.5012	.5030	.5018	.5005	.4979	.5018	.5011
Plan 14 (16 Actions)	1,0029	.5018	.5012	.5024	.4980	1,5015	.4985	.5012	.4992
Plan 15 (17 Actions)	.9978	.5024	1,0029	.5011	2,0032	.5005	.4998	1,0010	.5005
Plan 16 (18 Actions)	.5011	.5005	.4986	.5005	1,0048	1,0023	.5011	.5004	2,0019
Plan 17 (19 Actions)	.4998	.4999	1,0010	1,0004	1,0029	.4999	.5024	1,0023	1,5021
Plan 18 (20 Actions)	.4998	2,0051	.5011	.5005	.5005	1,0004	.5011	.5005	1,0010
Plan 19 (21 Actions)	.5011	.5011	1,0017	.5050	1,0042	2,0026	1,0010	1,0022	2,0013

Table F.2

Muise's method

Number of actions	Time for calculating the plan for each test						
	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7
Plan (Actions)	6,01	6,51	5,51	6,01	6,51	5,50	5,51
Plan 1 (3 Actions)	26,46	6,51	6,51	8,01	7,01	6,52	7,01
Plan 2 (4 Actions)	42,99	8,01	8,01	7,75	7,51	8,51	8,51
Plan 3 (5 Actions)	28,55	3.05	1.02	9,51	3.55	9,51	9,51
Plan 4 (6 Actions)	13,02	42,57	32,05	14,02	14,52	38,56	38,56
Plan 5 (7 Actions)	57,59	39,56	21,03	2.53	2.03	4.57	2.03
Plan 6 (8 Actions)	34,06	34,06	54,09	34,56	57,09	36,56	55,12
Plan 7 (9 Actions)	138,29	8.13	77,66	97,71	78,67	77,63	59,10
Plan 8 (10 Actions)	259,15	139,73	152,25	134,72	151,25	133,21	151,24
Plan 9 (11 Actions)	249,44	251,16	335,20	25.91	30.82	372,97	252,47
Plan 10 (12 Actions)	517,22	513,38	63.68	484,60	766,38	519,25	515,14
Plan 11 (13 Actions)	976,20	1279,60	943,08	1327,49	1277,82	1354,03	981,06
Plan 12 (14 Actions)	2218,07	2415,58	2532,65	2373,27	231.23	2654,35	2515,22

Plan 13 (15 Actions)	4691,71	3771,58	5069,29	4516,12	4825,84	3833,76	3802,53
Plan 14 (16 Actions)	882.46	9759,23	9118,82	8307,42	7855,91	8075,68	8283,52
Plan 15 (17 Actions)	16737,74	18676,62	17581,18	16733,75	19328,31	17307,07	18964,62
Plan 16 (18 Actions)	34246,84	3419.17	35841,27	31973,38	35505,94	3450.73	33058,53
Plan 17 (19 Actions)	64285,18	64597,30	63898,86	61231,13	64093,09	70519,44	66392,15
Plan 18 (20 Actions)	131784,62	129194,59	112238,17	119603,72	128186,17	141038,88	119505,88

Table F.3

Proposed method

Number of actions	Time for calculating the plan for each test								
	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9
Plan (Actions)	3,00	3,01	2,50	9,02	9,52	3,00	9,02	3,00	3,00
Plan 1 (3 Actions)	3,01	3,00	3,50	3,00	3,00	3,00	9,01	3,50	4,01
Plan 2 (4 Actions)	3,00	3,51	3,50	4,04	3,00	1.52	3,00	3,51	3,86
Plan 3 (5 Actions)	4,51	3,51	4,01	9,52	4,01	4,01	4,00	3,51	9,52
Plan 4 (6 Actions)	4,51	5,01	5,01	4,54	5,01	4,51	5,51	16,53	5,51
Plan 5 (7 Actions)	1.02	1.52	3,50	4,51	3,51	3,01	1.52	3,51	3,51

Plan 6 (8 Actions)	4,01	3,51	4,01	4,51	4,01	3,50	11,52	4,01	4,01
Plan 7 (9 Actions)	18,03	5,01	5,51	5,01	5,01	6,02	6,01	5,51	5,51
Plan 8 (10 Actions)	5,51	18,03	5,51	5,51	5,51	5,01	5,34	6,01	5,51
Plan 9 (11 Actions)	4,51	4,01	4,01	4,51	4,00	4,51	5,01	4,01	5,51
Plan 10 (12 Actions)	4,51	21,54	6,01	6,01	7,01	6,01	6,01	9,01	5,51
Plan 11 (13 Actions)	11,52	4,01	4,01	3,50	11,02	4,01	4,01	4,01	4,00
Plan 12 (14 Actions)	4,01	3,50	4,51	12,52	4,51	6,01	12,52	4,51	4,51
Plan 13 (15 Actions)	13,02	4,01	4,01	4,51	4,51	4,01	4,01	4,01	4,51
Plan 14 (16 Actions)	4,51	4,51	4,01	12,52	4,01	5,01	5,01	4,01	5,01
Plan 15 (17 Actions)	4,01	4,51	4,01	4,00	4,01	13,02	4,01	4,01	12,02
Plan 16 (18 Actions)	4,01	4,51	5,01	12,02	4,51	4,01	4,51	5,51	13,52
Plan 17 (19 Actions)	4,51	4,01	4,01	4,54	14,02	5,01	13,02	4,04	4,01
Plan 18 (20 Actions)	23,54	7,01	7,01	6,51	19,03	6,51	24,04	24,04	7,01

APPENDIX Q: QUALITATIVE ANALYSIS

Q.1. Concept map of the categories

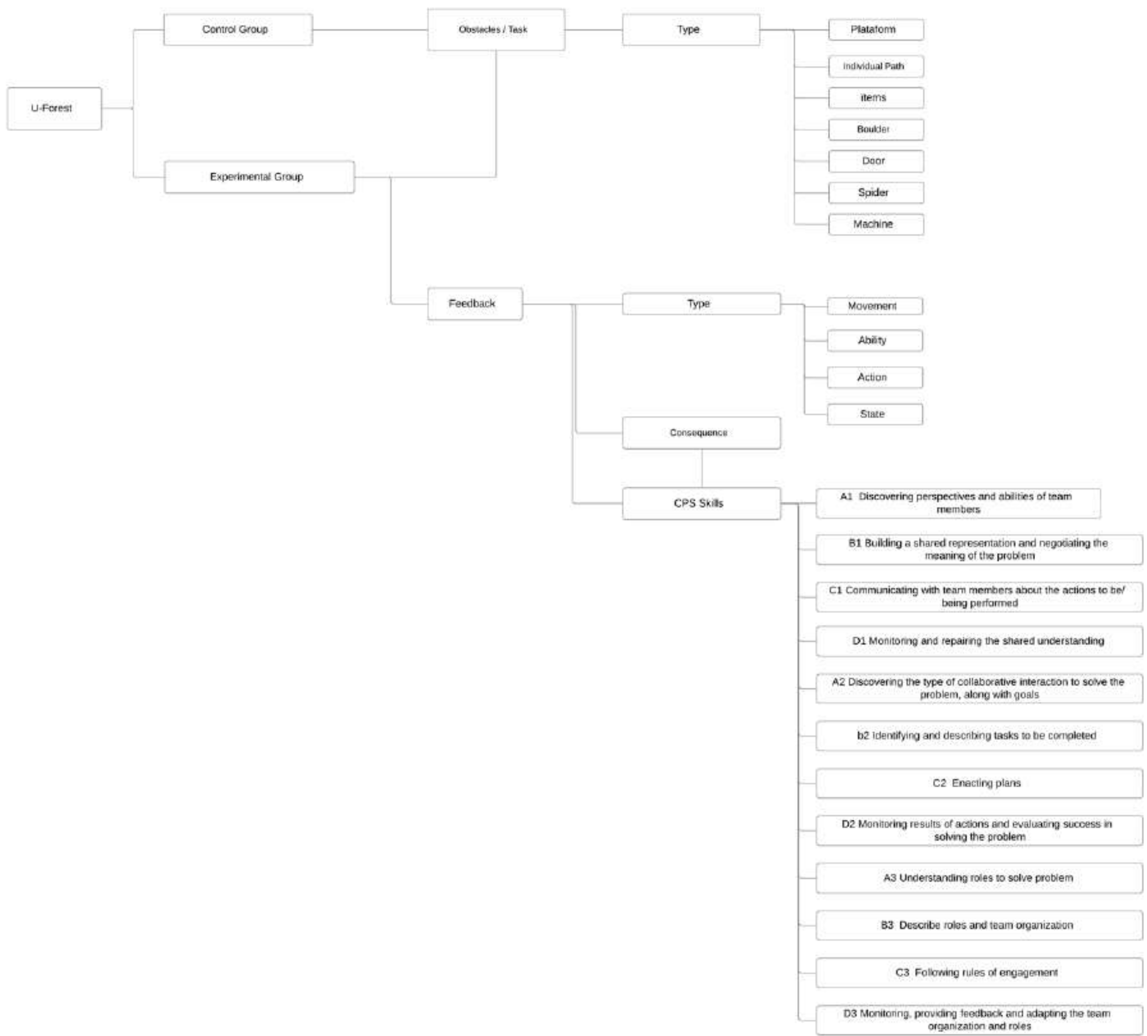


Figure Q.1: Concept map of the categories

Q.2. Effectiveness of the feedback

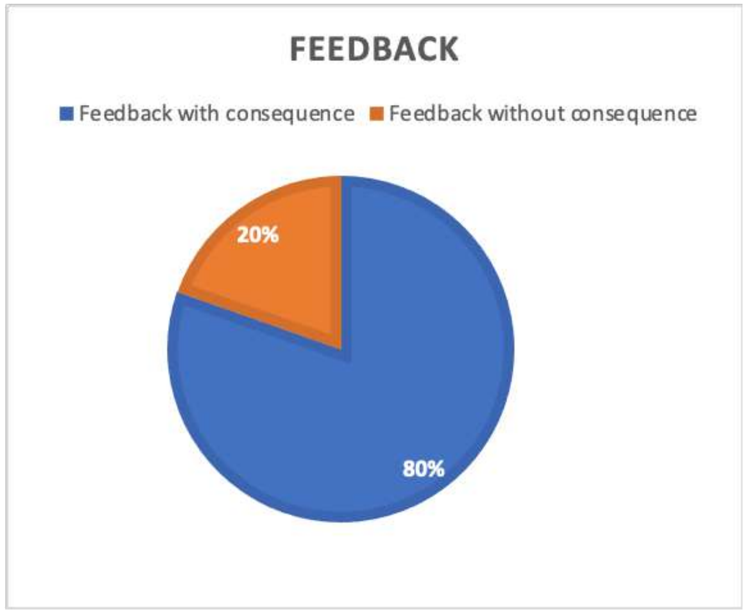


Figure Q.2: Graph showing the amount of feedback with a consequence.

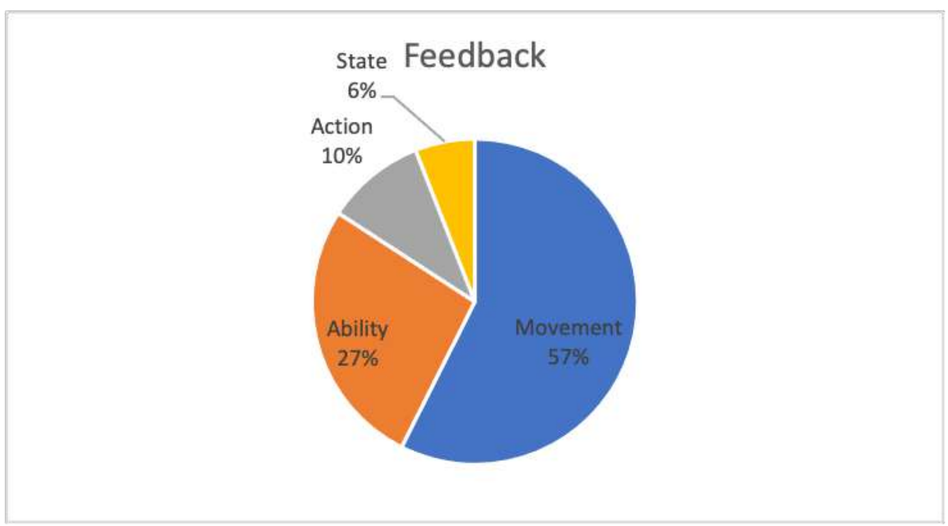


Figure Q.3: Graph of messages identified as feedback, by type

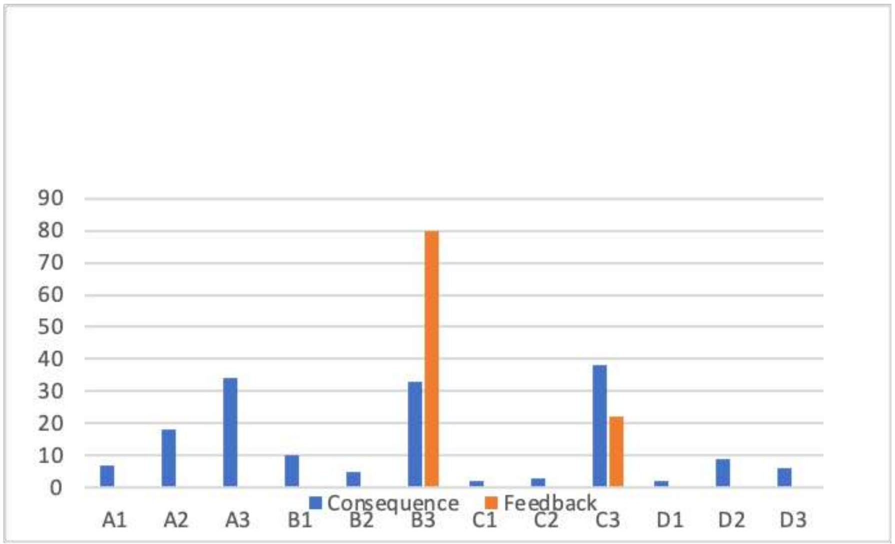


Figure Q.4: Graph of messages identified as feedback and consequences, by CPS skill

Table Q.1: Example student feedback and consequences, organized by skill.

	Feedback		Consequence	
Type	**CPS**	**Example**	**CPS**	**Example**
Ability	B3	"The yellow player has to use their special ability to get to path going over the fallen branch"	A1	"I don't know what I have to do" "What fallen branch?"
Movement	B3	"Look, it says you have to go up"	A2	"P1: It's giving you hints P2: Ah, OK OK…"
Ability	B3	"The yellow player has to use their special ability to get to	A3	"What's my ability?"

		path going over the fallen branch”		
Moveme nt	C3	“Did you go where they told you to, to the tree?”	B1	“Which tree? This tree?”
Ability	B3	“The yellow player has to use their special ability to get to path going over the fallen branch”	B2	“Yeah, you have to jump” “First you have to jump there and then you have to jump there”
Moveme nt	C3	“You still haven’t arrived. Is everything OK?”	B3	“P2 you have to get to here” “On my way!”
Ability	B3	“It says the yellow player has to use their special ability to get to path going over the fallen branch”	C1	“Come up here and try to shoot”
Ability	B3	“It says the yellow player has to use their special ability to get to path going over the fallen branch”	C2	“I already shot him”
Action	B3	“The yellow player has to use their special ability to get to path going over the fallen branch”	C3	“Look, I already shot that yellow thing”
Moveme nt	B3	“The red player has to go to the tower by the lake”	D1	“Yeah, the red player has to pass, because over there there are some yellow and green things”
Moveme nt	B3	“Yellow player, come here. You have to press this”	D2	“Done. I pressed a yellow button and the platform went up”
Moveme nt	B3	“Everyone has to come here”	D3	“P2 look, it looks like one has to go over there and not come over here, the other, the red player, has

			to go over there. And what about me? Where do I go?
Movement	B3	"It says you have to go to the tower"	No consequence
Ability	C3	"The yellow player has to use their special ability"	No consequence
Action	B3	"The blue and yellow players have to activate the hidden switch…"	No consequence

Q.3. Comparing progress between the experimental and control groups

Table Q.2: Example messages identified as obstacles in student conversations

Obstacle/Task	Control Group	Experimental Group
Platform	"I can't get up there. How do you do that super jump?" "How did you get up? I'm trapped, I can't get up… Hey, P2, there's a yellow button here, come on, come down here!" "But P2, go and press the button so we can go up"	"So P2, you have to press the yellow button!" "There's a yellow button in the tree" "Done. I pressed a yellow button and the platform went up"

Individual path	"The yellow player has to go up there and the red player has to go down there" "I'm the only one here. Where are you?" "I'm higher up, I'm coming down…"	"No, I've already been down and been everywhere, I've done everything" "Hey yeah, it looks like there are three different paths" "We have to go up and down!"
Items	"How do you use the item?" "I found an item but I don't know how it works"	"Did you get the item?" "There's another item here!" "Oh! You also found an item!"
Puerta	"There was a door and I want to get to the other side" "I found something, like a key to the door" "There's a door, and I had to put this thing and… it opened!"	"Done. I activated it!" "Where do I open the door?" "I'm at the door" "But look, there's a door!"
Machine		"No, you need to activate this, the machine, that machine!" "Over there there's a tool; I don't know what it is!"
Spider	"I'm by the spider. Now what do you guys have to do?"	"Oh!! Look, there's a Boss!"

	"I'm with the spider, come here"	"I need your help, because the spider here is massive"
Boulder		"Yeah, you have to activate the boulder!"

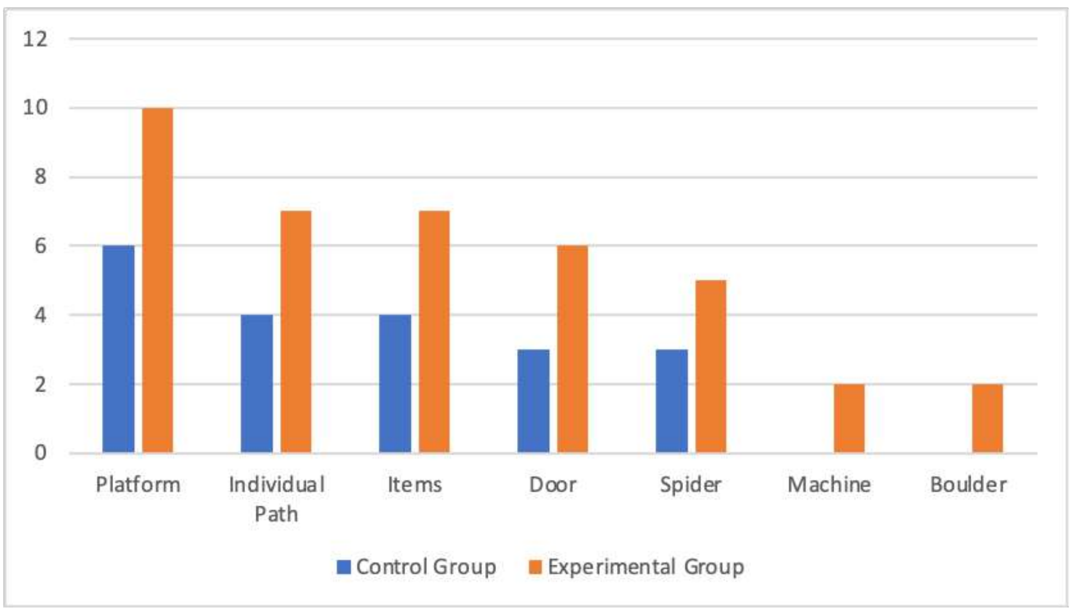


Figure Q.5: Graph showing the number of control and experimental groups at each obstacle

APPENDIX R: DESIGN CONSIDERATIONS

R.1 Design Principles

There are many articles covering tools and strategies to support the process of designing games for learning. Kiili, De Freitas, Arnab, & Lainema (2012), for example, describes a series of design principles to achieve the state of flow (Csikszentmihalyi, 1991) during the intervention: clear goals, immediate feedback, sense of control, and playability. Add to that, an increasing level of challenge can lead the user to focus on a rewarding experience in which he or she loses a degree of self-awareness and feels a temporary distortion. This type of experience, to the extent that the user maintains a significant cognitive load through the challenge, produces the mental effort (Mayer, 2014) that leads to learning.

As the central design axis, the theoretical framework proposed by Plass et al. (2015) was chosen. The framework proposes a flexible structure that can be applied to any game-based learning design process. It is articulated in a concept map whose dimensions can support the necessary scaffolding for game-based learning. The base level in this scheme is the INTERACT model (Domagk, Schwartz & Plass, 2010), where four main dimensions are articulated: 1) affective engagement; 2) behavioural engagement; 3) cognitive engagement, 4) and socio-cultural engagement. Achieving engagement in these four dimensions is one of the main objectives of the design so that the game must guarantee a complete experience aligned with the required learning (Plass et al., 2015). These four provided dimensions serve as reference points to find weaknesses in the system design and the interaction it proposes to students. As a way to regulate these dimensions in the user's relationship with the system, Plass et al. (2015) proposes that the following fundamentals be articulated from the design: 1) cognitive foundations; 2) motivational foundations; 3) affective foundations and; 4) socio-cultural foundations. As a tangible example of this, the authors propose seven dimensions that the game design must include to achieve a ludic learning experience:

R.1.1 Content and Skill

In this case, the game is not designed to teach a specific curricular content but instead to develop CPS as a skill. The use of roles for each player was prioritized for the design of the collaborative game (Nussbaum et al., 2009; Seif El-Nasr et al., 2010). The rules governing the system were then articulated based on these roles (See section R.2, Appendix R). These roles are assigned to each player according to the order in which they connect to the system. They are complementary roles that can form different types of relationship with the game environment depending on the team's interdependent interactions. These roles allow users to formulate different hypotheses and strategies for solving problems according to the characteristics of each character and the tasks within the game.

To ensure that our task was collaborative, we met the conditions established by previous studies (Szewkis et al., 2011): a common team goal was maintained (Dillenbourg, 1999; Seif El-Nasr et al., 2010); the puzzles were based on positive peer interdependence (Johnson & Johnson, 1999) and coordination and communication between them (Gutwin and Greenberg, 2004); the responsibilities of each player within the team were defined (Slavin, 1996); evidence of the rest of the team's actions was included in order to ensure there was group awareness (Janssen et al., 2007); the reward system, although minimal, was always collective (Axelrod & Hamilton, 1981); and all restrictions, such as health and energy, were shared by the team (Seif El-Nasr et al., 2010).

R.1.2 Incentive system

In this sense, when designing the game's incentive system, a distinction was made between the proposed Digital Game-Based Learning application and a gamification system. The latter focuses on using incentive systems to encourage and guide users towards a specific behavior with extrinsic motivation (Resnick, 2017; Hawlitschek & Joeckel, 2017). In this case, rewards are external to the game and given when the user demonstrates the required learning dynamics. Collaboration should therefore be driven by a desire to play, explore and solve the game with peers, rather than a reward that does not contribute towards this goal. When designing this system, we opted for mainly internal incentives so that the difficulty of the task would help motivate users to develop more effective collaborative strategies in order to solve it. In this sense, the game only includes an experience point system, in which players accumulate points as they advance and solve puzzles within the game. However, these points cannot be used for anything inside the game and do not lead to any prizes for the students.

R.1.3 Learning mechanics

Appendix S, together with Appendix W, give a detailed explanation of the evolution of tasks during the game. The design includes the use of complex scaffolding and soft scaffolding (Chen & Law, 2016), including text messages that seek to guide the students in order to recognize the forms of interaction that are available to them. On the other hand, the logic of graceful failure was also incorporated (Juul, 2013), in which any mistake made by a player can easily be amended or, at least, would not be serious enough to make the player give up too soon. In this sense, the health and energy system described in Appendix A is has a long duration and is easy to recharge. The system therefore avoids frustration becoming a barrier to learning, while also not being overly simple and negatively affecting the cognitive learning process (Kiili et al., 2012).

R.1.4 Assessment mechanics

The assessment mechanics within the game are based on the increasingly interdependent interactions that the team must have in order to progress. The team's success or failure at a task is recorded in a log. To ensure the tasks are clear and so as to avoid any unnecessary confusion, the game includes a feedback system based on a character who gives information in the form of a dialogue. We therefore use a cross-delivery feedback system between students in order to foster interaction based on the skills required by each task. In this sense, the Green player receives information that is relevant for the Red and/or Yellow player, the Red player receives information that is relevant for the Green and/or Yellow player, and the Yellow player receives information that is relevant for the Green and/or Red player.

R.1.5 Aesthetic design

The game interface (See Figure R.1) is detailed in Appendix R. It was designed to emulate standard controls on modern consoles and to allow clear interaction for students.

The game was designed so that the three players can interact with each other via oral and kinesthetic speech freely throughout the experience. In early prototypes, we could see that users were not able to express their ideas or organize the team effectively through text-based chat alone. Using the keyboard on the tablets was not a fluid process; it could not keep up with the rhythm of interaction offered by the system, leading to general frustration at the somewhat uncomfortable and unusable mechanism.

R.1.6 Narrative design

The narrative looked for the players to co-create their own story. In this sense, the system does not include a single, particular narrative. Instead, it shows or reveals a series of images, to which the group can then attach a meaning or emotion, based on whatever they co-construct (Maine, 2017). The feedback provided by the game and the little dialogue that is included all reference aspects of collaborative problem solving, such as the difficulties that can be faced when solving collaborative problems, as well as its value in everyday life.

R.1.7 Musical Score

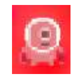
Finally, the music consisted of 6 open source songs from the chiptune album called “Music for an Unmade Forest World” by the artist Visager (2017), taken from the Free Music Archive (2018). Sound effects for the players’ actions were also included.

R.2 Roles

There are three roles that players can assume when they enter the game (Table R.1). For neutrality, we summarise them into three names: 'Green', 'Red', and 'Yellow'. Each of these avatars has a unique ability to interact with the environment.

The primary consideration to define the roles was through a special ability system. Each player pressed the B button to start consuming energy. This action triggered a change where the player acquired a new ability. Different types of aura and sound were activated to represent this to each player. Through this new ability, we figured out what Rocha & Mascarenhas (2008) suggested.

Table R.1: Game roles

avatar	Ability	Display
green	Protect players within the energy field.	
red	Destroy and move heavy objects.	
yellow	Increased speed and double jump.	

R.3 Basic Actions and Interface

There are six basic actions in the game: move horizontally, jump, shoot, activate/deactivate special ability, recover health and energy, and use basic items. These six actions are common to all three users and are represented through the game interface in Figure R.1.

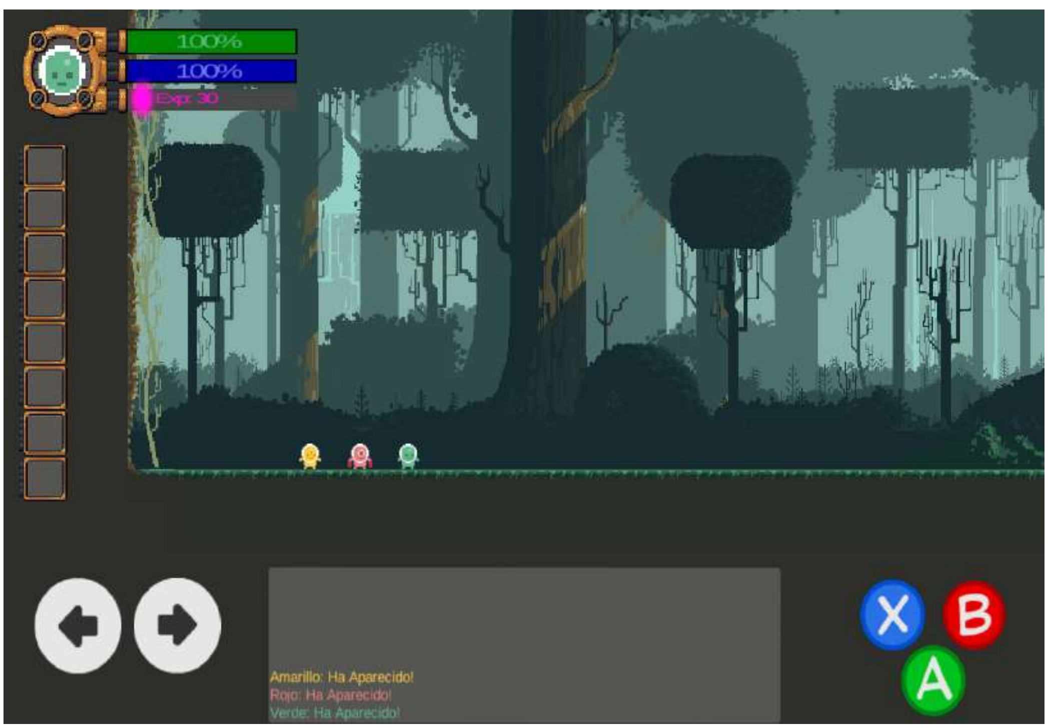


Figure R.1. Game Interface

Figure R.2. Avatar, Health, Energy and Experience Indicator

At the top left of the screen (Figure R.2), to avoid confusion and recognise one's role, the system shows the player which avatar corresponds to him/her. Besides, health and energy indicators can be seen, as well as experience level. The team shares these statistics during the game, i.e., everyone is dependent on the same energy and health supply. To renew energy and recover their health, players must approach fountains that are scattered throughout the game.

Figure R.3. Inventory of Items Collected

At the lower left (Figure R.3) is the found items inventory. Here are the items showed that the players could collect to use later within the game. The items do not change the form of interaction, but they are helpful to open doors (among others) within the game, for example.

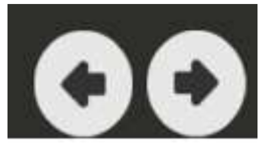


Figure R.4. Motion buttons

At the lower-left corner (Figure R.4), there are the horizontal scrolling arrows. At the bottom centre is a log where the players receive relevant information about the game; for now, it only summarises the connection and disconnection of their teammates from the game.

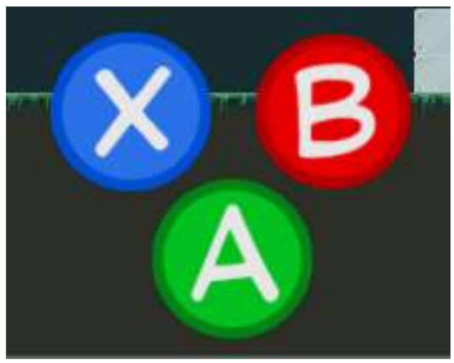


Figure R.5. Basic interaction buttons

Finally, the game's basic interaction buttons are at the bottom right (Figure R.5). The B button is to shoot the opponent the player faces in a horizontal direction, the A button is to jump, and the X button activates the special ability of each player's role. The tutorial ensures that players fully understand these dynamics.

R.4 Server Interface

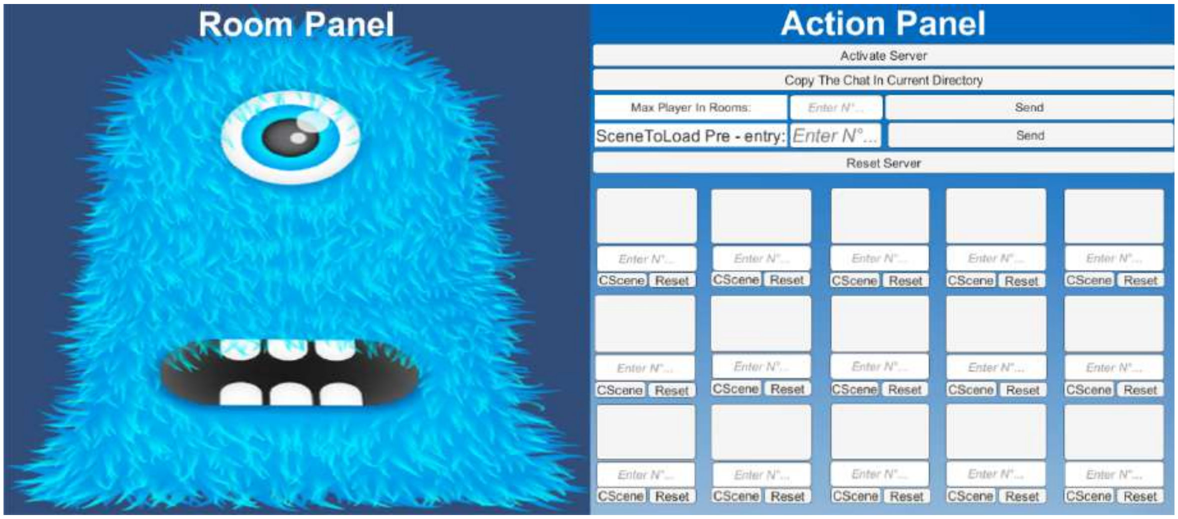


Figure R.6. Interface Server Control Panel

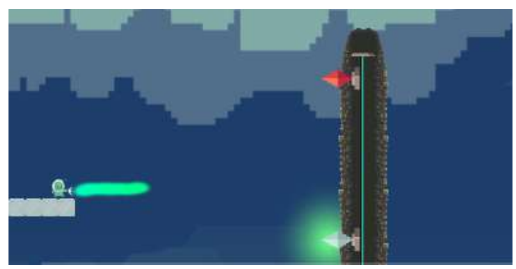
The interface is managed by the teacher or operator (Figure R.6) and consists of two sections grouped on the right side of the screen. The first one is to activate the server (start the game), enter the number of players, and the stage at which each game room will start.

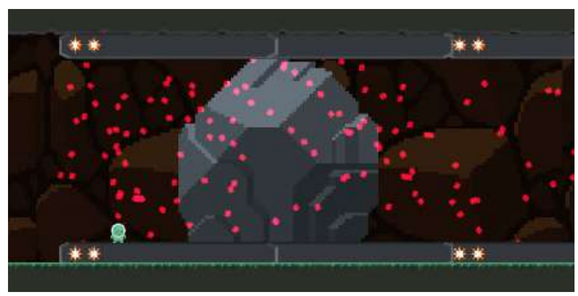
In the second section, there are fifteen boxes. In these, the operator can monitor that the players enter the game correctly. Each of these units can be manipulated by the teacher. The system can: 1) restart the level; 2) Send players to their last checkpoint; and 3) change the players' stage. This section is helpful in case of problems with the game or if teachers want to move a team from one stage to another for a particular reason.

R.5 Collaborative Interactions.

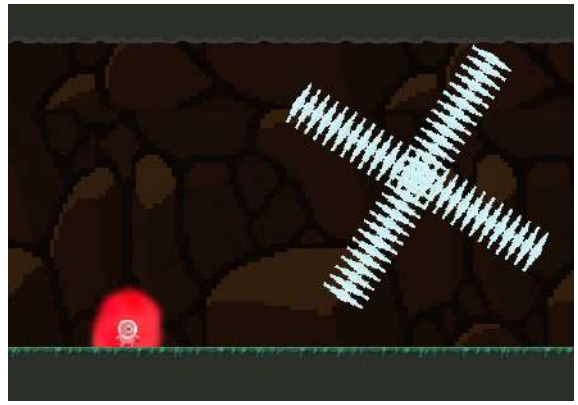
Interdependence plays a fundamental role in understanding the task outlined in the PISA framework (OECD, 2017). Therefore, the game consists of a set of basic rules and interactions to generate puzzles that include the skills of each role. As a result of this principle, the following seven systems have been generated (Table R.2).

Table R.2: Basic collaborative interactions.

Floor Switch	Button on which the player must be positioned to activate. They can be individual or group, synchronous or diachronic, and assignable to each role or a generic purple switch.	
Switch Shot	Switches activated by the respective player's shoot. It has permanent and temporary activation modes in need of special ability activated or without. They can be for a specific player or generic.	
altar	In the game, there are platforms called "Altars". These players can transport their skill to another demarcated area of the	

	map. This altar allows players to share their skills in remote areas. The green player activates a protected area; the red player an area where whoever fires will do so with projectiles borrowed from his role; the yellow player transmits a change in gravity.	
Shared Skill Zone	The recently mentioned altars activate areas like this. These areas can be activated from a distance via altars or switches and can also be activated simply by contact with the right player. The colour of the stars on the edges defines the role that activates the area.	

Health and Energy recovery.	There are water fountains scattered throughout the game. When players approach these fountains, they begin to recover health and energy. For each player approaches, the regeneration frequency increases.	
Teleporter Player	Fucsia coloured portals that, after contact with a player, transport it to a default area.	
Teleporter Shot	By shooting at these portals, players will be able to teleport their shots to other locations. In this way, players can intervene in another player's zone under their signal.	

Destructible obstacles	Some obstacles can only be destroyed or displaced with the special 'Red' ability. Within these items, some disappear forever, and some are temporarily disabled only to encourage and strengthen players' ability to coordinate.	

APPENDIX S: STAGES OF THE GAME

The *U-Forest* game consists of six stages graduated with difficulty maintaining a scaffolding process in dynamics and mechanics that exercise the use of the skills associated with CPS.

S.1 Stage 1: Tutorial

The first level seeks to lay the foundation for interaction with the system, to begin with collaborative work.

Figure S.1. Home Stage Tutorial

This level is divided into two zones that the system displays as a single excellent task. The first zone is to explore the fundamental interactions common to all roles (Figure S.1). The second zone (Figure S.2) is to understand the particular mechanics for each role. A colour filter prevents other players from passing through. In this area, each player must use their special skill twice to advance to the end of the stage.

Figure S.2. Second Section Stage Tutorial

S.2 Stage 2: Recognizing Roles

The second level is articulated in two large sections. Task 1 through 5 (Figure S.3) ensures that all three players follow the commitments associated with their role and allow the necessary changes for the progress of the rest of the team. In principle, two common tasks have been afforded. After this, each player follows a particular path (Figures S.4, S.5 & S.6) to perform tasks specific to their role within the game.

Figure S.3. Initial section Stage 2. Tasks 1-5

Figure S.4. Green player section. Tasks 6-7

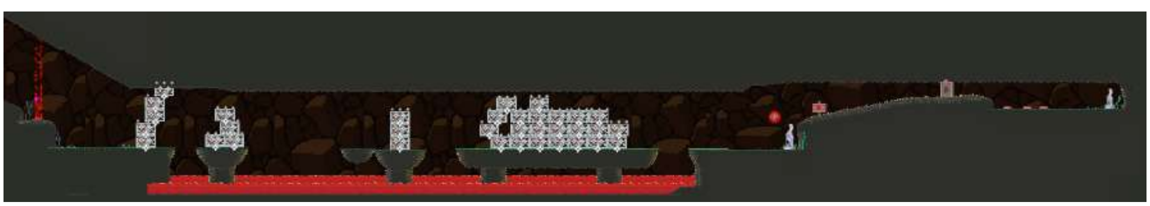


Figure S.5. Red player section. Tasks 6-7

Figure S.6. Yellow player section. Tasks 6-7

During the second section, between tasks 8 and 9 (Figure S.7), the system proves that students can solve a task that requires greater interdependence and sync in coordinating the three-team members. This task considers an increase in complexity concerning the variables that make up the users' collaborative dynamics. Organising and representing the problem clearly to generate a common action plan that is monitored becomes necessary in achieving the solution.

Figure S.7. Section 2 Stage 2. Tasks 8-9

S.3 Stage 3: Exchange perspectives and information relevant to the problem.

The third stage aims to enable players to exchange relevant information about the problem, exchange perspectives about the task's status to be resolved, and verify the essential roles and perspectives during the interaction. For this, the third level design proposes a zone where players recognise a new form of interaction with their skill and with their teammates: the use of magic zones and firing teleporters (see Appendix R).

The level is articulated in 3 large sections. The first section introduces new interaction mechanics both individually, collaboratively and remotely. The second section adds interdependence and sync factors to the mechanics presented during section one. The last section reviews a basic collaborative dynamic. To ensure that team members have complementary perspectives that they use in rebuilding the problem, players are in different play areas during the level. These areas require the effective and constant exchange of relevant information to overcome the task from the perspective of each member of the group.

B.4 Stage 4: Analysing possible problem-solving strategies

The fourth level seeks to ensure that the team can generate a common representation of the problem and weigh possible strategies. For this, it is proposed a level with three sections (Figure B.17), of which it is necessary to overcome at least two to finish the level.

Figure S.8. Home Stage 4

Each path has an equivalent level of difficulty with non-identical activities. At the end of these, players can find one of the four gears available to overcome the level; reaching two of these gears can now advance to the next stage.

S.5 Stage 5: Global task of multiple strategic solutions solved in consideration of the strengths and weaknesses of team members

The fifth level reinforces the skills worked on the previous level. However, this increases the possible paths and strategies to solve the stage. Players face six paths of upward difficulty. Students know the character of incremental difficulty and can choose the most exciting and appropriate tasks according to their abilities as a team. In this way, users must activate four switches that become affordable as the paths proposed by the level end.

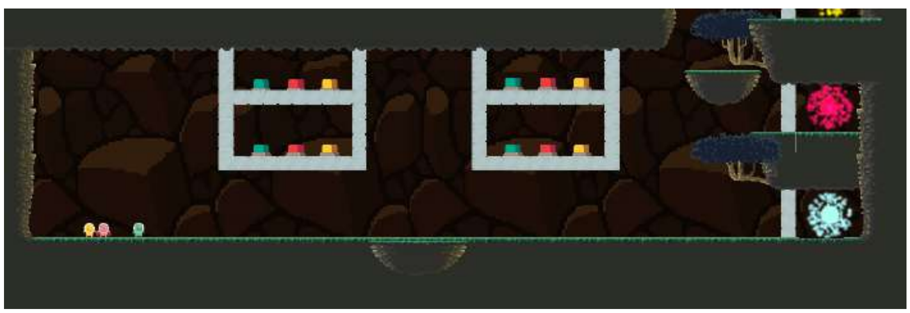


Figure S.9. Start stage 5

The central area of the stage consists of a large room with six corridors distributed upwards (Figure S.9). The NPC alerts players to the difficulty level of the area they are accessing. At the same time, it warns of the need to remain coordinated during the development of tasks.

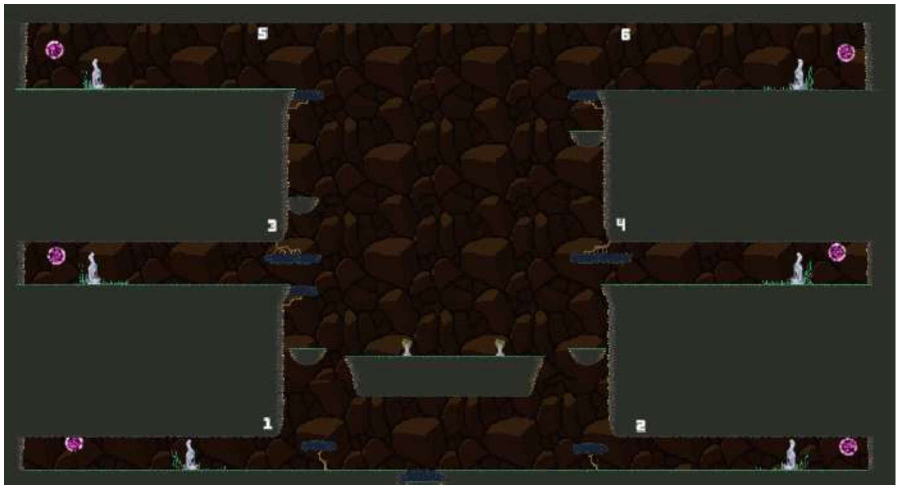


Figure S.10. Stage 5 Central Zone

S.6 Stage 6: Global ill-defined task

The sixth level seeks players to be able to create their ways to solve the problem. For this, the design proposes a level whose main mechanics and dynamics are formed through a maze in which players must meet and find solutions that allow them to reach the game's endpoint. Once the exit is found, the team can decide whether to finish the game or return to continue exploring possibilities within the zone.

Figure S.11. Panoramic Level 6. (Trimming)

Level 6 consists of 25 small sections organised as a diamond (Figure S.11). There are two maps within each of these sections and a group of between one or five teleporters that allow them to move through it (Figure S.12). The maps presented to the team allow them to visualise where they are and where they will go if they enter through that portal, using a rock with small lights representing the entire map. The blue light indicates the current section, and the purple section indicates the destinations of the portals. The only exception to this is the initial section, where players are divided by portals with particular destinations to each role and are represented by lights of each colour. In other zones, purple is a general rule because portals are accessible to all.

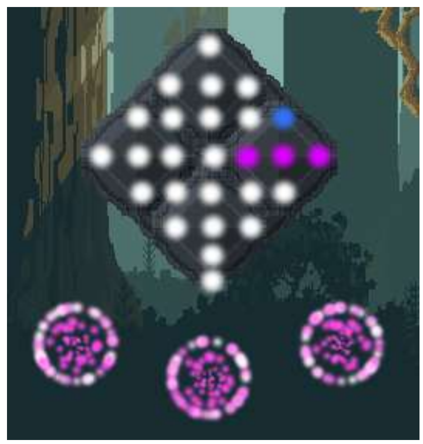


Figure S.12. Map location level 6.

There are five zones and a sixth switch-based mode that allow them to reach the end of the game. Each of these sections contains a different puzzle aligned with the complexity of level six tasks. Once the players reach one of the endings, they can decide to return to the puzzle where they can find a new challenge and ending for their own intrinsic motivation.

APPENDIX T: GAME TASKS

This matrix (Table T.1) presents each task (T), the points of interest involved in the resolution of each (POI), the difficulty of the task, and the associated skill according to the PISA skill matrix (OECD, 2017).

The matrix for stage 1 was not considered because that is not related to a specific skill. However, the success of the players on stage 1 was stored in the log as a single task to analyse the indicators of time and effort, it took each player to learn the basic mechanics.

Table T.1: Task Matrix Stage 2 - 6

Stage	*Task (POI)*	*difficulty*	*Skill required*		
			Player 1	*Player2*	*Player3*
2	*T1 (POI 2)*	*1*	*A2*	*A2*	*A2*
	T2 (POI 3.4)	*2*	*C3*	*C3*	*A3*
	T3 (POI 5)	*2*	*A3*	*A3*	*A3*
	T4(POI 6,8,9,11,12)	*3*	*A - B3*	*A - B3*	*A - B3*
	T5 (POI 7,1.13)	*2*	*C3*	*C3*	*C3*
	T6 (POI 14)	*3*	*A3*	*A3*	*C3*
	T7 (POI 15,16,17)	*2*	*A - B3*	*A - B3*	*A - B3*
	T8 (POI 18.19)	*4*	*B – C - D*	*B - C - D*	*B-C - D*
	T9 (POI 20.21)	*1*	*A3*	*A3*	*A3*
3	*T1 (POI 0)*	*1*	*A3*	*A3*	*A3*
	T2 (POI 1-2)	*3*	*A1 A2 C*	*C-D*	*to*
	T3 (POI 3)	*3*	*C - D*	*to*	*A1 A2 C*
	T4 (POI 4)	*3*	*A1 A2 C*	*C - D*	*to*
	T5 (POI 5)	*3*	*C - D*	*to*	*A1 A2 C*

	T6 (POI 6)	*3*	*C - D*	*A1 A2 C*	*to*
	T7 (POI 7-9)	*4*	*A1 A2 C*	*to*	*C - D*
	T8 (POI 10-14)	*3*	*A1 A2 C - D*	*A1 A2 C - D*	*A1 A2 C - D*
	T9 (POI 15)	*3*	*A1 A2 C - D*	*C - D*	*to*
	T10 (POI 16)	*3*	*to*	*C - D*	*A1 A2 C - D*
	T11 (POI 17)	*3*	*to*	*A1 A2 C - D*	*C - D*
	T12 (POI 18)	*3*	*to*	*C - D*	*A1 A2 C - D*
	T13 (POI 19)	*3*	*C - D*	*A1 A2 C - D*	*to*
	T14 (POI 20-21)	*2*	*C1 C3 D3*	*A3*	*A3*
4	*T1 (POI 2-6)*	*4*	*A3 B3 C - D*	*A3 B3 C - D*	*A3 B3 C - D*
	T2 (POI 7-8)	*5*	*B - C - D*	*B - C - D*	*B - C - D*
	T3 (POI 9-10)	*2*	*A3*	*A3*	*A3*
	T4 (POI 19)	*6*	*A - B – C - D*	*A - B – C - D*	*A - B – C - D*
	T5 (POI 20-24)	*5*	*B – C - D*	*B – C - D*	*B – C - D*
	T6 (POI 25-28)	*5*	*A - B – C - D*	*A - B – C - D*	*A - B – C - D*
5	*T2 (POI 1-2)*	*3*	*A3*	*A3*	*A3*
	T2 (POI 8-12)	*4*	*A - B – C – D*	*A - B – C – D*	*A - B – C – D*
	T3 (POI 13-16)	*5*	*A - B – C – D*	*A - B – C – D*	*A - B – C – D*
	T4 (POI 3-7)	*6*	*A - B – C – D*	*A - B – C – D*	*A - B – C – D*
	T5 (POI 17-31)	*7*	*A - B – C – D*	*A - B – C – D*	*A - B – C – D*
	T6 (POI 32-45)	*8*	*A - B – C – D*	*A - B – C – D*	*A - B – C – D*
	T7 (POI 46)	*3*	*A3*	*A3*	*A3*
6	*T2 (POI 1-2)*	*3*	*A - B – C – D*	*A - B – C – D*	*A - B – C – D*
	T2 (POI 3, 21)	*5*	*A - B – C – D*	*A - B – C – D*	*A - B – C – D*
	T3 (POI 4-6)	*8*	*A - B – C – D*	*A - B – C – D*	*A - B – C – D*
	T4 (POI 7)	*7*	*A - B – C – D*	*A - B – C – D*	*A - B – C – D*

T5 (POI 9-11)	*8 A - B – C – D*	*A - B – C – D*	*A - B – C – D*
T6 (POI 12-16)	*8 A - B – C – D*	*A - B – C – D*	*A - B – C – D*
T7 (POI 17-19)	*8 A - B – C – D*	*A - B – C – D*	*A - B – C – D*
T8 (POI 22)	*6 A - B – C – D*	*A - B – C – D*	*A - B – C – D*

Note: T represent the task, POI refers to the point of interest, A, B, C & D refer to the complete row of problem-solving in the OECD framework.

APPENDIX U: FOCUS GROUP INTERVIEW GUIDELINES

These guidelines are presented to structure the focal study carried out with the students of both schools. The section written in italics corresponds exclusively to the experimental group, while the rest of the text constitutes what was also common to the control group.

1. **Initial Greeting**
    a. Today I am going to spend some time here talking to you because it is important for me to meet children your age and know how you think... that helps us design things that are entertaining and interesting to you.
    b. My name is Andrea; I live here in Santiago with my two cats, one is male and is called Willi-Willy, and the other is female and is called Lila... I have three brothers smaller than me… and I really like to draw and talk to people. Can you tell me their names and their favorite pastime?
    c. (Let them show up and stick a sticker with their name on their uniform with their hobby).
    d. Now let's start our conversation.
2. **interview**
    a. Tell me a little about your favorite activities.
    b. And at school, what tasks do you enjoy doing?
        i. Do you like group assignments?
        ii. What is so positive about them? What do you like about group tasks?
        iii. What don't you like about group assignments?
    c. You should know that not all people are good at everything, there are some things where we are better than in another truth
        i. Do you think they're good or bad at working as a group?
            1. Those who say they're good at working as a group, why do they think so?
            2. Those who say they're bad, why do they think they're bad?

   ii. Do you play on the computer or consoles?
   iii. What are your favorite games?
   iv. Why are they your favorites?

d. I've known some games can only be played as a group, is that true?
   i. Do you like these games?

e. Is it true that you guys were testing a game during the week in class?
   i. Could you tell me what the game was about?
   ii. How was it played?
   iii. What did you have to do to pass the stages in this game?
   iv. Did this game have anything unique or different compared to what they always play?
      1. What was different?
      2. How did it look like the other games they play?
   v. Raise your hand who didn't like the game... and now raise your hand who did like the game...
      1. Those in the group who didn't like the game, can you tell me why you didn't like it?
      2. Now the other way around, who liked the game, can you tell me why you liked it?

f. I was told that this game has to be played as a group, is that true?
   i. And why can't you play alone?
   ii. What do you think of this game can only be played as a group?
   iii. Is that good for you or not?
   iv. Was the game easy or difficult for you? What element of the game was easy/difficult?
   v. If I told you that now other children will play this game for the first time, what recommendations would they give them?
   vi. Do you feel like you learned anything by playing this game? What did they learn?
   vii. What do you prefer, playing console games in a group or individually? Why?

g. Finally, let's do an activity where we're going to make a collage. Let's send a message to the inventor of this game. The important thing is that the message they send must be a message agreed by everyone.
h. Finally, we're going to do an activity where we're going to make a collage. There is a person who wants to make a game for them to play as a group. Let's send a message to the developer of this game. What do you recommend? The important thing is that the message you send must be a message agreed by everyone.

APPENDIX V: ATTITUDE TOWARD COLLABORATION SCALE

Adapted from Hwang, Yang & Wang (2013) and Ødegård (2006).

Table V.1: Attitude toward collaboration scale

Regarding collaborative work:	1	2	3	4	5
1. I believe that learning to work in group is interesting and valuable.					
2. I would like to learn more to work in groups during classes.					
3. It is worth learning about teamwork.					
4. It is important for me to be able to work well as a team.					
5. It is important to be able to work as a team with the people around me and the rest of society.					
6. I will seek to practice and improve my skills of working in a group.					
7. It is important for everyone to learn to work in a team.					
8. I consider it valuable to work in group					
9. It is important to make a personal commitment when working in a group					
10. I always have a clear role and goals when I work in a group					
11. I believe that our roles should always be clearly defined					
12. I often see that the best working groups have a clear and defined leader					

13. It is important that the group leader organize the work in a way that helps the group achieve its goals

14. Participants in a groupwork often feel frustrated with each other

15. Whenever I work in a group I receive relevant comments about my contributions to the team

16. In the groups that I participate my classmates are willing to listen to me if I have problems