

# Development and Psychometric Properties of an Instrument for Assessing Group Processes in Collaborative Problem-Solving in Higher Education

Desarrollo y propiedades psicométricas de un instrumento para la evaluación de procesos grupales en la resolución colaborativa de problemas en la educación superior

**Received:** June 6, 2023; **Reviewed:** June 26, 2024; **Accepted:** September 26, 2024

## Mauricio López-Cruz

University of Chile, Santiago, Chile.

ORCID: <https://orcid.org/0000-0002-8288-2988>

## Victoria Sánchez-Gómez<sup>1</sup>

University of Salamanca, Salamanca, Spain.

ORCID: <https://orcid.org/0000-0001-9330-895X>

## Patricio Cabello

Universidad Andres Bello, Santiago, Chile.

Universidad de Chile, Santiago, Chile.

ORCID: <https://orcid.org/0000-0001-9656-3147>

## Abstract

This article aimed to develop and analyze an instrument for assessing group processes involved in collaborative problem-solving tasks in higher education. The Scale for the Assessment of Group Processes in Collaborative Problem Solving (GROUPS) consisted of 24 self-reporting items grouped into four dimensions: Exploring and Understanding, Representing and Formulating, Planning and Executing, and Monitoring and Reflecting. The study followed an instrumental design and took place in Chile. The test was administered to 939 higher education students, all of whom answered the instrument upon completing a 3-week task on collaborative problem-solving. The task was designed by the research team in collaboration with the students' teachers. Regarding validity, the instrument structure showed appropriate fit indexes according to the theoretical model's four factors (RMSEA and SRMR under  $<.05$ ; CFI and TLI over  $.95$ ). The instrument also showed appropriate indices of reliability (ordinal alpha over  $.80$  in all factors). The different group processes observed are described. Moreover, the instrument's characteristics are discussed, as are the results obtained and their implications for developing collaborative problem-solving skills in higher education.

## Keywords

problem-solving, collaborative problem-solving, group processes, higher education, tertiary education, measurement.

## Resumen

Este estudio tuvo como objetivo desarrollar y analizar un instrumento para evaluar los procesos grupales involucrados en tareas de resolución colaborativa de problemas en educación superior. La Escala para la Evaluación de Procesos Grupales en la Resolución Colaborativa de Problemas (GROUPS) consta de 24 ítems de autoinforme agrupados en cuatro dimensiones: *Exploración y Comprensión*; *Representación y Formulación*; *Planificación y Ejecución*; y *Monitoreo y Reflexión*. El estudio siguió un diseño instrumental y se llevó a cabo en Chile. Se aplicó el instrumento a 939 estudiantes matriculados en educación superior, que lo respondieron tras completar una tarea de resolución colaborativa de problemas de tres semanas de duración. La tarea fue diseñada por el equipo de investigación en colaboración con los docentes. En cuanto a las evidencias de validez referidas a la estructura interna del instrumento, este mostró índices de ajuste apropiados para un modelo de cuatro factores (RMSEA y SRMR por debajo de .05; CFI y TLI por encima de .95). Además, el instrumento mostró índices de fiabilidad óptimos (alfa ordinal por encima de .80 en todos los factores). Se describen los diferentes procesos grupales observados. Por último, se discuten las características del instrumento, así como los resultados obtenidos y sus implicaciones para el desarrollo de habilidades de resolución colaborativa de problemas en la educación superior.

## Palabras clave

resolución de problemas, resolución colaborativa de problemas, procesos grupales, educación superior, educación terciaria, medición.

## Cómo citar [APA]:

López-Cruz, M., Sánchez-Gómez, V., & Cabello, P. (2025). Development and psychometric properties of an instrument for assessing group processes in collaborative problem-solving in higher education. *Acta Colombiana de Psicología*, 28, 1-23. <https://doi.org/10.14718/ACP.2025.28.1>

- 
- 1 **Correspondence:** Victoria Sánchez-Gómez, Institute for Community Inclusion (INICO), Faculty of Psychology, University of Salamanca. Address: Ave. de la Merced 109-131, 37005, Salamanca, Spain. Email: [vsanchezgomez@usal.es](mailto:vsanchezgomez@usal.es).

### Authors' note:

Research project: This study was conducted as part of the project titled "Critical Factors Associated with Group Processes for Collaborative Problem-solving in the Context of Online Technical Education." Data availability statement: Data unavailable; Disclosure statement: The authors report no competing interests.

### Acknowledgments:

Funding provided by ANID/PIA/Basal Funds for Centers of Excellence FB0003 and ANID/Support 2024 AFB240004 is gratefully acknowledged.

## Introduction

Collaborative problem-solving (CPS) is defined as the ability 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” (Organisation for Economic Co-operation and Development [OECD], 2017a, p. 7). Problem-solving is essential for dealing with an increasingly complex world (Graesser et al., 2018) in which information and communications technologies play a more central role (Care & Griffin, 2017). Given the speed of technological advances and the interdisciplinary nature of our many challenges, modern-day problems can often only be solved successfully when individuals with different skill sets, training, and perspectives work together. As a result, CPS skills are frequently listed as 21st-century skills (Graesser et al., 2020; Griffin et al., 2012).

Although individual problem-solving has been widely studied for decades in many fields, including psychology, academic interest has recently turned toward the collaborative facet of this dimension. For example, studies have looked at some variables that can influence performance in CPS tasks, such as personality, the diversity of perspectives among team members, team members’ knowledge, mutual perception, and beliefs regarding the abilities of others (Graesser et al., 2018). Elsewhere, researchers have looked into group learning processes, group decision-making (Slavin, 2017), and the processes behind computer-supported collaborative learning (Kirschner & Erkens, 2013). However, more understanding is needed of the group processes and dynamics involved in specific CPS tasks, especially in tasks that include information and communication technologies (Scoular & Care, 2020).

Furthermore, while problem-solving has traditionally been studied in laboratory settings (Graesser et al., 2018), recent years have seen an increase in interest in understanding CPS skills studied in real-life contexts. Thus, it is relevant to highlight the development of instruments to evaluate and monitor these skills and the components they comprise, which is vital for the design of assessments, the development of indicators, and the planning of teaching interventions. This has brought with it a search for methodologies that can satisfy the varied relevant evaluation criteria, which are often mutually exclusive: increased ecological validity, increased internal validity, the possibility for large-scale

evaluation, and reasonable time for analysis, among other criteria (Graesser et al., 2018). Several authors agree that there is a need to generate knowledge around CPS skills, especially to improve the capacity of schools and teachers to design curricula that allow students to acquire and develop these skills through deliberate teaching in educational contexts (Graesser et al., 2020).

For example, since 2015, the Program for International Student Assessment (PISA) exam, thanks to new digital evaluation technologies, has included the evaluation of the CPS skills of 540,000 students from 72 countries using tests based on human-non-human interaction (OECD, 2017b). One of the results of this evaluation, for example, shows that non-migrant students perform better in CPS tasks in schools with a higher number of migrant students, likely because diversity favors varied perspectives in the planning of problem-solving strategies (OECD, 2017b).

Notwithstanding its contributions and its allowing for evaluation on a large scale, the analysis of human-non-human interactions has its limitations. For example, it lacks an ecologically valid setting and does not allow researchers to study discourse patterns (Graesser et al., 2018). The Assessment and Teaching of 21st-Century Skills (ATC21S) project aimed to formulate standardized measurements for several CPS skills in human-human interactions to allow for comparison in future studies (Hesse et al., 2015). Graesser et al. (2018) point out that this reference framework will enable researchers to analyze relevant discourse mechanisms for breaking CPS skills, including participation, perspective-taking, and social regulation. Its main limitation, however, is that it requires trained observers and a prolonged period for analysis, which limits its applications for generating formative assessments for students and teachers in educational contexts. Scoular and Care (2020) try to address this limitation. Using an inference model based on human-to-human language patterns, the authors identified skills from the ATC21S framework, such as negotiation, information gathering, perseverance, transactive memory, and systematicity, which offer teachers indicators of what their students can do through registered items (Scoular & Care, 2020).

In the literature, it is possible to find some works that focus on assessing CPS in higher education. For

instance, Pazos et al. (2010) designed an observational instrument for evaluating collaborative learning in peer-led groups in higher education contexts considering two dimensions: problem-solving approach and group interaction style. To design the instrument, they previously observed groups of 1st- and 2nd-year university students over a trimester in a weekly activity as they tackled problems created by faculty professors, which were characterized by being complex, poorly structured, requiring the integration of information, and their solution was not immediately apparent or unique. The instrument identifies four types of groups: simple instruction, supported discussion, elaborated instruction, and guided discussion.

Visschers-Pleijers et al. (2005) assembled a self-reporting questionnaire to evaluate the quality of group learning-focused interactions in problem-based learning tasks in higher education in three dimensions (exploratory questions, cumulative reasoning, and conflict management). Starting from the focus on the quality of interactions in problem-based learning, they study tutorial groups, in which students interact with each other to synthesize individually acquired information. They conclude that exploratory questioning is the most critical aspect of tutorial group productivity.

Dindar et al. (2020) evaluated 25 groups of three higher-education students through a simulated computer-based task. They analyzed the relationship between metacognitive experience and its perceived impact and objective in the performance of a CPS task. What was innovative about these authors' proposal was the assignment of significant value to self-reporting based on the role of metacognition in CPS performance.

Evaluating CPS skills in higher education is particularly important as these skills are essential to good performance in professional and work-related contexts. Likewise, despite the importance of these skills, current development levels are low, and there is no clarity on the curricular strategies and proposals in educational contexts that would encourage their development (OECD, 2017a). This is why it is necessary to continue developing instruments to evaluate group processes in different problem-solving tasks where student interactions are encouraged. Therefore, instruments based on the experience and metacognition of participants are required

to incorporate dimensions of analysis of the quality of group interactions and processes, such as the establishment of shared understanding, planning, monitoring, and execution of personal and group performance, exploratory communication, and task regulation. In turn, the literature has identified the need to generate instruments that provide information and indicators on aspects of group functioning that allow teachers to offer timely feedback to students. Finally, such instruments must be easy to apply and analyze and do not necessarily require experienced evaluators.

This study aimed to develop a self-report instrument for evaluating group processes in CPS tasks in a higher-education context and provide evidence for the instrument's validity and reliability. The instrument developed in this work aims to evaluate the group processes during CPS tasks. This tool assesses group processes in CPS from the conceptualization described below.

### ***Theoretical Framework***

In line with recent conceptualizations of CPS, we can understand that this is a complex skill composed of many sub-skills (Scoular & Care, 2020). More specifically, this study uses the definition drawn up for the 2015 PISA test, which proposes that the CPS skill is the individual's ability to effectively involve themselves in a process wherein two or more agents are attempting to solve a problem by sharing understanding and the effort required to reach a solution and add their knowledge, skills, and efforts to the search for that solution (OECD, 2017a). This definition assumes the interdependence between individual group members' skills and the group's performance in the problem-solving task. Within this framework, CPS processes are defined as the interrelationship between two axes: (i) the individual axis of problem-solving processes (i.e., Exploring and Understanding, Representing and Formulating, Planning and Executing, Monitoring and Reflecting) and (ii) the collaborative axis of CPS skills (i.e., establishing and maintaining shared understanding, taking appropriate action to solve the problem, and establishing and maintaining group organization). As shown in Table 1, this interrelationship gives rise to a matrix of 12 specific abilities (Graesser et al., 2018; OECD, 2017a) that are relevant for exploring group processes that arise during CPS tasks.

**Table 1.** Initial Framework for Conceptualizing Group Processes in CPS

Collaborative problem-solving skills		Collaborative problem-solving skills		
		Understanding (Establishing and maintaining shared understanding)	Action (Taking appropriate action to solve the problem)	Organization (Establishing and maintaining group organization)
Problem-solving processes	Exploring and understanding	Discover team members' perspectives and skills	Discover the type of collaborative interaction to solve the problem according to the objectives	Understand the roles needed to solve the problem
	Representing and formulating	Build a shared representation and negotiate the meaning of the problem	Identify and describe the tasks to be performed	Describe team roles and organization (communication protocols/ participation rules)
	Planning and executing	Communicate with team members about actions to be taken	Execute the agreed plan	Follow the rules of participation (e.g., encourage other team members to perform their tasks)
	Monitoring and reflecting	Monitor and correct shared understanding	Monitor the results of the actions and evaluate the success in solving the problem	Monitor, provide feedback, and adapt team organization and roles

From this analysis network, we can conclude that many CPS skills require high levels of metacognition, i.e., knowledge regarding one's cognition and the ability to regulate cognitive processes with a specific objective in mind (Dindar et al., 2020). Subjective judgments regarding this kind of group task and its results can intervene in the development of CPS tasks (Dindar et al., 2020). Moreover, the collaborative nature of group problem-solving implies that points of view and opinions must be made explicit, justified, and debated (OECD, 2017a), meaning many of these processes are observable (Hesse et al., 2015). Many of these skills are communicative, i.e., the ability to correctly communicate information and report actions that a given person has taken at a given time, allowing group members to construct a shared understanding of the task at hand (OECD, 2017a).

These group processes in CPS are defined as any dynamic, interpersonal processes implied in CPS among

group members competent in this skill. The following passage presents the conceptual definition of each dimension of group processes in CPS considered in the developed tool. Each dimension considers the CPS skills understanding, action, and organization:

- (1) Exploring and Understanding (EU): This process includes exploring the problem by observing it, interacting with it, searching for information, identifying limitations and obstacles, and demonstrating an understanding of the information provided and new information. It implies discovering team members' perspectives and skills, the type of collaborative interaction needed to solve the problem according to the objectives, and the roles needed to solve the problem.
- (2) Representing and Formulating (RF): This involves using tables, graphs, symbols, or words to represent aspects of the problem and formulating hypotheses around the factors relevant to the problem and the

relationships between them to build a coherent mental representation of the problem. It implies a shared representation and negotiation of the meaning of the problem, identification of the task to be performed, and description of team roles and organization.

- (3) Planning and Executing (PE): This process involves devising a plan or strategy to solve the problem, executing that strategy, and defining general and specific objectives. It includes communication between team members about actions to be taken, executing an agreed-upon plan, and following the rules.
- (4) Monitoring and Reflecting (MR): It comprises monitoring progress, acting on feedback, and reflecting on the solution, the information provided in the

problem, or the strategy adopted. This implies a shared understanding and monitoring of the actions and results, providing feedback, and adapting the team organization and roles if needed.

This study reports on developing the Scale for the Assessment of Group Processes in Collaborative Problem Solving (GROUPS), which assesses the degree to which specific group processes are present in a CPS-centered task. This tool considers the conceptualization described above, i.e., it assesses four dimensions of problem-solving processes, and each contemplates the three types of CPS skills. All the characteristics of the instrument designed and the process by which it was developed are explained below.

## Method

### Study Design

This is an instrumental study as it aimed to design, develop, and evaluate the psychometric properties of an instrument (Montero & León, 2007). The study followed the steps that Fonseca and Pedrero (2019) suggested for the development of psychometric instruments.

### Participants

#### Pilot Test Administration

A non-probabilistic sample of 120 students participated in the pilot test, all in higher education at a professional institute (technical college) that offers courses lasting two to four years. These students were enrolled in several programs but took a general education course in different areas of Santiago de Chile, Chile. Participating students in this general education course completed a CPS task before completing the instrument. The mean age of the students was 23.7, and 35% of them were female ( $n = 42$ ) and 65% male ( $n = 78$ ).

#### Revised Test Administration

Non-probabilistic purposive sampling was used to select the participants for the revised test administration. This sampling method was used since the instrument was administered after a CPS task in already-formed classes. The only inclusion criterion was to be part of these courses and complete the CPS task (described in Procedure). Thus, 939 students completed the CPS task and answered the

instrument in its revised version and its entirety. As in the pilot sample, all the participants were pursuing higher education at a professional institute (technical college) and were enrolled in several programs. The mean age of the students in this sample was 24.09 ( $SD = 12.80$ ), and 56% of them were female ( $n = 526$ ), 42.4% were male ( $n = 402$ ), and 1.2% ( $n = 11$ ) identified with another gender.

### Instrument

The GROUPS scale assesses four dimensions of group processes in CPS: EU, RF, PE, and MR. In the GROUPS scale, students assess the extent to which specific group processes were present during the CPS tasks. Each item is a statement that can be answered on a 5-point scale of agreement (1 = *Not at all*, 2 = *A little*, 3 = *Partially*, 4 = *Mostly*, 5 = *Totally*). For example, the translation of one of the items included in the test is "We spent some time getting to know each other's opinions and points of view about the problem." The GROUPS scale in its pilot version contained 24 items, while the revised version contained 28 items. The final version of the instrument, which contains 24 final items, can be found in Appendix 1.

Although the instrument was administered online, it meets the criteria for in-person administration (i.e., it is brief, the answers are simple). Self-reporting was chosen over other methodologies used in the literature, such as observing interactions (e.g., Pazos et al., 2010) or the computer-based evaluation of human-non-human

interaction (OECD, 2017a). This was because, in contrast to the former, self-reporting allows researchers to collect data on subjective judgments regarding collaborative activities and, in contrast to the latter, it is administered by asking participants about their experience in a CPS task that involves negotiation and the alignment of representations of the task and group objectives among human agents (Dindar et al., 2020). This is why the items on the instrument we presented here describe activities that groups typically partake in when conducting CPS tasks, designed to be applied once the task is over. Since collaboration in problem-solving involves high levels of metacognition (Dindar et al., 2020), self-reporting was chosen as an appropriate measure for these processes and tasks.

### **Procedure**

The GROUPS scale was developed in two stages. The first stage comprised the processes of instrument design, while the second involved the test administration and subsequent analyses. Each stage considered the steps Muñiz and Fonseca-Pedrero (2019) suggested to develop psychometric instruments. The first stage included conceptual and operational definition, item construction, expert judgment, instrument specifications, editing, and assembly of the pilot version. The second stage included the pilot study, editing and assembly of the revised version, application of the test in its revised version, and evaluation of the psychometric properties. The procedure followed for item construction, expert judgment, and assembling of the pilot version is explained in more detail below. Since the instrument was administered, both in its pilot and revised versions, after a CPS task, this task is also described.

#### *Item Construction*

For a construct to be measurable, it must be made operative with observable indicators that can later be reflected as items. To this end, the matrix of 12 specific abilities for exploring group processes during the CPS task detailed above in Table 1 was considered. Therefore, three indicators were taken into account for each dimension corresponding to problem-solving processes (EU, RF, PE, MR). Each dimension considered one indicator for each CPS skill: one for understanding, one for action, and one for organization. Two items were constructed based on each indicator. Thus, 24 items were generated. Appendix 2 details the process of psychometric scaling that took place, showing the indicators of each

dimension and the items developed in each case. This construction resulted in the initial test version presented to the judges for evaluation.

#### *Expert Judgment and Assembly of the Pilot Version*

Expert judgment is essential to provide evidence about the content validity of an instrument. Three experts evaluated the instrument described herein. Two of them were experts in psychometry and data analysis in educational contexts. The third was a professional with experience in teaching and student assessment in higher education. Each judge reviewed the instrument and rated each item against the following criteria: item relevance (*relevant/irrelevant*) and item composition (*adequate/inadequate*). Besides rating each item based on these dimensions, the experts made observations on each item and then filled out a form in which they were asked to make comments on the following aspects: (1) conceptual definitions of the construct and its dimensions; (2) selected indicators (operationalization); (3) instrument format; (4) instrument instructions; and (5) any other comments.

The consistency between pairs of judges was assessed using Cohen's kappa (Cohen, 1960). Observed kappa ranged from moderate to perfect for item relevance (Cohen's kappa: Judges 1-2 = .467;  $p = .007$ ; Judges 1-3 = .467;  $p = .007$ ; Judges 2-3 = 1;  $p < .001$ ) and was adequate for item composition (Cohen's Kappa: Judges 1-2 = .700;  $p < .001$ ; Judges 1-3 = .500;  $p = 0.013$ ; Judges 2-3 = .500;  $p = .013$ ). Krippendorff's alpha ( $\alpha$ ) was calculated to evaluate the simultaneous consistency among the three judges (Hayes & Krippendorff, 2007). Observed alpha indicated moderate agreement for relevance ( $\alpha = .576$ ) and item composition ( $\alpha = .577$ ). The judges, on average, rated 21.7 items as "relevant" and 21.3 items as "adequate" and made observations (at least one judge) on 19 items, all of which were incorporated. After incorporating the judges' recommendations, the instrument was assembled in its pilot version, which consisted of 24 items.

**The CPS Task.** The test, both in its pilot and revised versions, was administered after the same CPS task described here.

The regular teaching staff implemented and oversaw a CPS task within the studied subject. The activity was as follows: 1) teachers formed random groups of students; 2) each group was given the same task, defined as a collaborative problem-solving



activity. The task consisted of discussing the opinion of a fictitious expert about global warming and its consequences (see Appendix 3). After the session moderated by the teacher, students had between

one and two weeks to prepare a text in the form of a newspaper column, responding to the position of the supposed expert. Table 2 describes the procedure for the implementation of the task.

**Table 2.** Task Implementation Procedure

Step	Description
<b>Preparation</b>	Ninety-seven teachers were trained to participate in the process and provide support in evaluating and implementing the CPS task during this phase.
<b>Week 1</b>	Information on student demographics was gathered, random groups were formed, and the CPS task was delivered alongside instructions for completing it.
<b>Week 2</b>	Teachers monitored whether the students had doubts regarding the task without directly mediating them and encouraged the autonomous organization of individuals and groups.
<b>Week 3</b>	The students handed in their solutions to the task and answered the GROUPS scale.

## Analysis

### *Pilot Test Analysis*

In this stage, it was verified that all answer options were eligible through the frequencies observed for each answer category. Although no issues were identified in that analysis, items with possible compositional issues (4 items) and test items were suggested as potential replacements (i.e., Items 25, 26, 27, and 28). Thus, during the pilot stage, four more items were added. For example, the EU dimension initially had two items for the understanding indicator, “Discovering team members’ perspectives and abilities,” and an additional item (Item 25) was created after the pilot test. As a result, the revised instrument contained 28 items.

### *Revised Test Analysis*

Although the instrument was administered to 939 students (who completed the task), the research team filtered the database to eliminate any answers that did not optimally respond to the instrument (i.e., presented variability). Thus, 112 cases were eliminated for null variability ( $SD = 0$ ) (11.92 % of total task participants). Of the remaining 827 cases, 108 were eliminated due to very low variability (i.e., only one item was answered differently). Finally, 722 cases (76.89 % of total task

participants) were validated for analysis. The analysis presented below was carried out using the software suites RStudio (RStudio Team, 2023), jamovi (The jamovi project, 2023), and SPSS v. 28 (IBM Corp., 2021).

Firstly, the team analyzed the distribution of answers for each item, as well as their means and standard deviation. This ensured variability across all items and no input or lost data errors. Then, as a preliminary step, a global analysis of the items was conducted based on indicators derived from CTT (i.e., discrimination index and reliability if the item is eliminated). This was done to identify any items with deficient functioning at this preliminary stage. The criteria were item-test discrimination indices and item-dimension discrimination indices above .30, considered adequate, while items that improved reliability when removed were considered deficient. Given that all items showed discrimination indices above .30 for both the total scale and the dimension and that none significantly improved reliability if removed (see Appendix 4), no preliminary removals were made in this step. After contrasting necessary assumptions (i.e., covariance matrix) and evaluating the validity based on the internal structure of the instrument, the research team performed confirmatory factor analysis (CFA). It assessed the observed adjustment between the participants’ responses and the theoretical structure of the instrument.



The team used the WLSMV estimator to conduct the CFA. This analysis contrasted a factorial model consistent with the theoretical model upon which the instrument was based, i.e., a first-order model comprising four related factors. An alternative model with a second-order factor was also tested. Both models were contrasted, considering all the items and eliminating some deficient items. CFA was used to analyze the models, and the following fit indices were considered: (i) the ratio between Chi-square and degrees of freedom ( $\chi^2/df$ ), adequate if its value was less than 2 ( $\chi^2/df < 2$ ); (ii) the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR), both with values below .08 indicating an acceptable fit and values under .06 indicating a good fit; and (iii) the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), with values above .90 being adequate and above .95 being optimal (Hu & Bentler, 1995).

Once a satisfactory factorial solution was obtained, evidence of reliability was estimated using the ordinal alpha, suitable for variables of an ordinal nature (Domínguez-Lara, 2012). These values were reported alongside Cronbach's alpha. Values above .70 were considered satisfactory indices (Prieto & Delgado, 2010).

Finally, descriptive statistics were calculated for the final composition of the scale. The mean, standard deviation, minimum and maximum observed, and standardized skewness coefficient (z-skewness) were

reported for dimensions and total score. The z-skewness, calculated as the ratio between the skewness coefficient and its standard error, was used to assess the skewness of the distribution (Doane & Seward, 2011) and determine whether the responses clustered in low, medium, or high scores. Considering a 5 % significance level and a 95 % range, values equal to or greater than 1.96 indicate significant positive skewness (suggesting a clustering of data points at the lower end of the scale), and values equal to or less than -1.96 indicate significant negative skewness (suggesting a clustering of data points at the higher end of the scale). In contrast, values between -1.96 and 1.96 suggest that the distribution is approximately symmetrical.

### ***Ethical Considerations***

A Research Ethics Board of the University of Chile approved the study (no. 015). All participants signed an informed consent form in which they were assured confidentiality in handling information. Although personal data were recorded for follow-up purposes, these were eliminated to anonymize the analysis database. Following the ethical guidelines for psychological research outlined in the code of ethics of the American Psychological Association (APA, 2017), this research would be considered low risk as it is framed within "normal educational practices, curricula, or classroom management methods conducted in educational settings" (p. 11).

## **Results**

This project aimed to develop an instrument for evaluating group processes in CPS in higher educational contexts and to provide evidence for its validity and reliability. To this end, the results are presented in two subsections: one discusses results associated with the psychometric properties of the final instrument and the other focuses on describing the group processes observed.

### ***Psychometric Properties of the Instrument***

This subsection describes the results obtained from administering the revised instrument and the characteristics of the final proposed version.

#### ***Distribution of Answers in Each Item***

The revised instrument contained 28 items organized into four dimensions and was answered by 722 higher education students. Table 3 details the distribution of the answers observed in each item applied in the revised instrument and their mean and standard deviation. A higher frequency in the five response categories implies a higher score (*Not at all* = 1 to *Totally* = 5). Likewise, the means per item maintain the same metric, from 1 to 5, in which a higher mean implies a more significant presence of the group process. As seen in the table, all response options were considered eligible, and none of the items accumulated a percentage frequency of more than 80 % in a given answer category.

**Table 3.** Distribution of Frequencies, Means, and Standard Deviation of Each Item

Item	Not at all (%)	A little (%)	Partially (%)	Mostly (%)	Totally (%)	Mean	SD
1	4.99	10.11	22.44	35.18	27.29	3.70	1.12
2	7.76	9.70	25.35	36.98	20.22	3.52	1.15
3	6.37	7.76	18.56	34.49	32.83	3.80	1.17
4	5.26	10.11	21.75	35.04	27.84	3.70	1.13
5	7.76	11.63	18.98	29.50	32.13	3.67	1.25
6	3.60	8.73	18.14	31.72	37.81	3.91	1.11
7	5.82	7.76	18.70	31.72	36.01	3.84	1.17
8	7.48	13.02	21.61	33.66	24.24	3.54	1.20
9	3.74	6.79	18.42	39.89	31.16	3.88	1.04
10	2.77	6.37	16.90	39.61	34.35	3.96	1.01
11	7.48	8.31	20.08	33.80	30.33	3.71	1.20
12	15.79	17.59	27.29	26.59	12.74	3.03	1.26
13	6.37	9.56	20.08	34.77	29.22	3.71	1.17
14	6.51	7.76	16.48	34.77	34.49	3.83	1.17
15	4.29	5.96	14.54	33.93	41.27	4.02	1.09
16	6.09	9.97	21.61	31.58	30.75	3.71	1.18
17	7.76	10.11	17.04	31.86	33.24	3.73	1.24
18	7.34	9.70	17.45	32.83	32.69	3.74	1.22
19	5.82	5.68	14.54	33.10	40.86	3.98	1.14
20	9.70	9.97	23.13	34.35	22.85	3.51	1.22
21	4.99	5.40	13.57	30.47	45.57	4.06	1.12
22	6.93	7.34	17.87	32.27	35.60	3.82	1.19
23	11.63	11.36	23.41	31.16	22.44	3.41	1.27
24	10.39	10.67	17.73	26.87	34.35	3.64	1.33
25	10.11	11.08	20.64	31.16	27.01	3.54	1.27
26	5.96	8.45	18.01	34.07	33.52	3.81	1.16
27	6.79	7.48	17.87	33.80	34.07	3.81	1.18
28	11.22	10.39	20.02	27.98	28.39	3.52	1.30

*\*In Spanish, the original response options were "Para nada", "Poco", "Parcialmente", "En gran medida" y "Totalmente".*

### Item Selection and Validity Evidence Based on Internal Structure

Given that no item presented issues in the preliminary analysis according to CTT indices (as reported in Appendix 4), the results presented here regarding the CFA began by considering all items. The first model contrasted using CFA contained four interrelated factors corresponding to the theoretical dimensions described previously (EU, RF, PE, and MR) and grouped the 28 items. This model was contrasted as a first-order model (Model A) and a second-order model (Model B).

The four worst-performing items were then identified, and a theoretical model containing four

related factors with six items in each factor was then contrasted. This considered eliminating four items, specifically 25, 26, 27, and 18. The model was contrasted as a first-order model (Model C) and a second-order model (Model D).

Finally, a theoretical model of four factors, each containing five items, was contrasted. This considered eliminating eight items, respecting the balance of content and the representation of indicators, namely Items 25, 26, 27, 18, 22, 3, 7, and 13. This model was contrasted as first-order (Model E) and second-order (Model F). The observed fit indices for each model are shown in Table 4.

**Table 4.** Fit Indices Observed in CFA

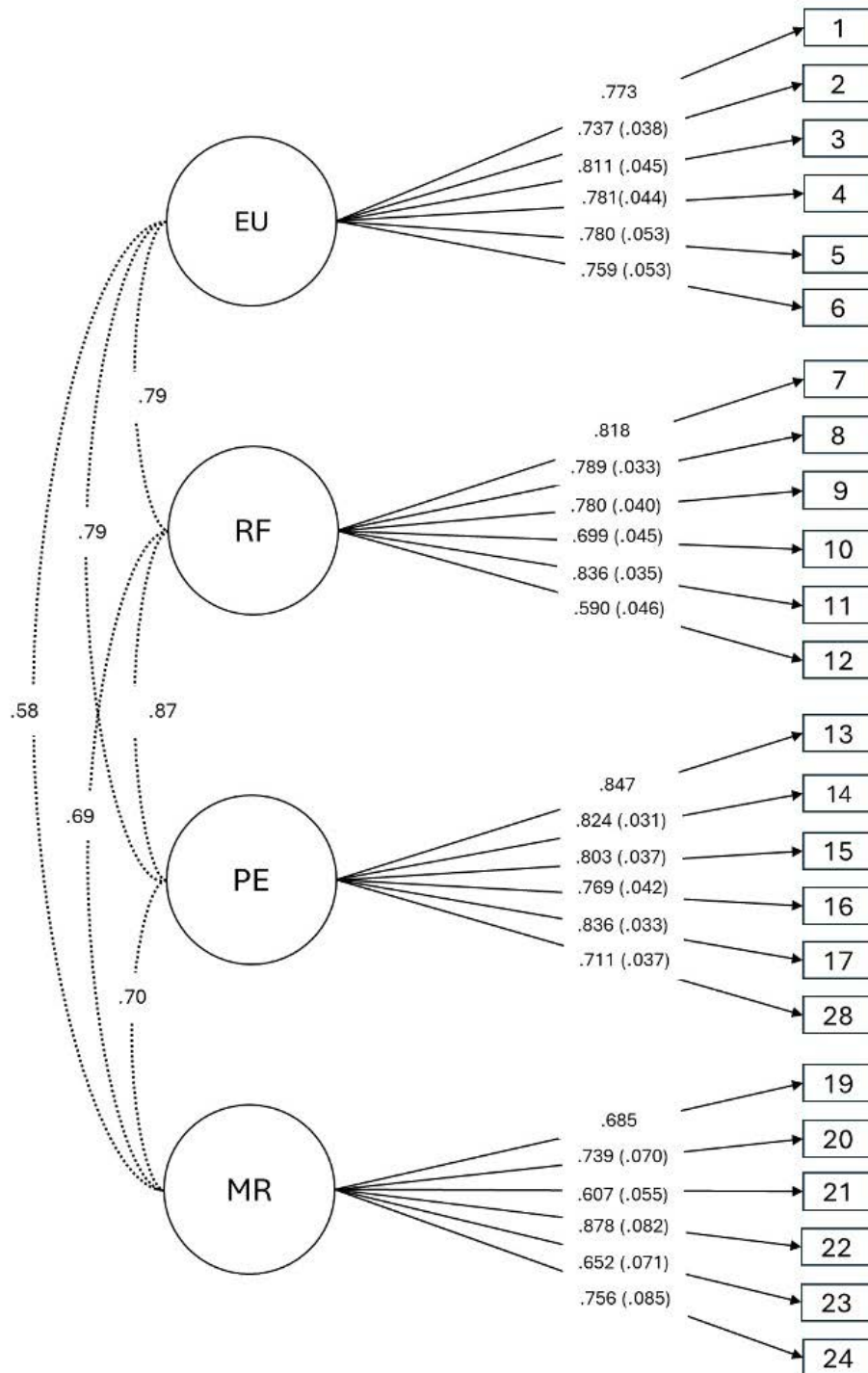
	$\chi^2/df$	<i>p</i>	RMSEA (IC 90 %)	SRMR	CFI	TLI
<b>Model A</b> 4F 28 items	2.142	<.001	.024 (.022–.027)	.034	.997	.997
<b>Model B</b> 1F 4f 28 items	2.222	<.001	0.032 (.029–.035)	.043	.995	.995
<b>Model C</b> 4F 24 items	2.051	<.001	0.024 (.021–.027)	.033	.997	.997
<b>Model D</b> 1F 4f 24 items	2.424	<.001	.034 (.030–.037)	.044	.995	.994
<b>Model E</b> 4F 20 items	2.532	<.001	.026 (.023–.029)	.035	.997	.996
<b>Model F</b> 1F 4f 20 items	2.546	<.001	.034 (.030–.038)	.044	.994	.994
<b>Reference values, according to Hu and Bentler (1995)</b>	< 2	>.05	<.08 adequate <.05 optimal		>.90 adequate >.95 optimal	

**Note:**  $\chi^2/df$  = ratio between Chi-square and degrees of freedom;  
RMSEA = Root Mean Square Error of Approximation;  
SRMR = Standardized Root Mean Square Residual;  
CFI = Comparative Fit Index; TLI = Tucker-Lewis Index.

Given its optimal adjustment according to Hu and Bentler's previously specified criteria and its balance

of content and number of items per dimension, Model C was chosen as the most appropriate.

**Figure 1.** Factor Loadings, Estimation Errors, and Covariances between Factors in Model C (Four Factors and 24 Items)



**Note.** The factor loading of the first item in each dimension is fixed to 1 as part of the model estimation process; hence, the estimation error is not calculated for these items (Items 1, 13, 7, and 19).

Figure 1 shows the factor loadings observed for each item in this model, estimation errors, and the covariance between factors.

### **Evidence of Reliability**

Table 5 shows the ordinal alpha and Cronbach's alpha reliability indicators. All the dimensions indicated optimal internal consistency ( $>.80$ ).

**Table 5.** Reliability Indices

Dimension	Number of items	Items	Cronbach's alpha	Ordinal alpha
EU	6	1, 2, 3, 4, 5, 6	.899	.916
RF	6	7, 8, 9, 10, 11, 12	.884	.908
PE	6	13, 14, 15, 16, 17, 28	.912	.931
MR	6	19, 20, 21, 22, 23, 24	.867	.894

**Note:** EU = Exploring and Understanding; RF = Representing and Formulating; PE = Planning and Executing; MR = Monitoring and Reflecting.

### **Descriptive Statistics Observed in the Dimensions of the Instrument**

The score of each student in a dimension was calculated from the sum of the points they assigned to each item (1 to 5, based on the answer options). Each dimension had a theoretical minimum of 6 points (six items with a minimum score) and a theoretical maximum of 30 (six items with a maximum score). The total score of each student was calculated as the mean of the four dimensions.

Table 6 details the main descriptive statistics summarizing the answers given by students for each of the instrument's dimensions and the whole instrument. Each dimension showed a markedly negative skewness, indicating a clustering of data points at the high scores. Nevertheless, there was variation between dimensions. The RF dimension had the lowest score and lowest variability. The PE dimension, meanwhile, had the highest score but also the highest variability among students, followed by the MR and EU dimensions, which had slightly lower means.

**Table 6.** Main Descriptive Statistics Observed in the Dimensions

	EU	RF	PE	MR	GROUPS
Mean	22.296	21.970	22.514	22.422	22.301
Standard deviation	5.654	5.486	5.969	5.643	5.295
Min.	6	6	6	6	6
Max.	30	30	30	30	29.75
Z-skewness	-9.934	-9.428	-10.374	-10.285	-11.010

**Note:** EU = Exploring and Understanding; RF = Representing and Formulating; PE = Planning and Executing; MR = Monitoring and Reflecting.

## Discussion

This paper focused on developing an instrument to evaluate group processes in CPS tasks in higher education. To that end, the research team designed, applied, and analyzed the GROUPS scale. This self-reporting instrument can be easily applied and analyzed, where students evaluate the presence of group processes relevant to CPS in a group problem-solving task. This section discusses the study in two subsections: the first concerns the instrument developed, and the second examines the group processes observed. Each case discusses strengths, limitations, and the project's potential applications.

Firstly, regarding the psychometric properties of the developed instrument, the observed adjustment indicators and the internal consistency indices provide favorable evidence of validity based on internal structure and reliability. Despite the instrument reported here being a new tool, we decided to perform CFA and not EFA, given the theoretical hypotheses underlying the work carried out. As Ferrando and Anguiano-Carrasco (2010) indicated, in a purely exploratory analysis, the researcher would analyze a data set without having any prior hypothesis about its structure and leave it to the analysis results to provide information about it (data-driven decisions). On the other hand, a CFA is preferable when the researcher has a series of well-specified hypotheses (based on a theoretical formulation or preliminary evidence) that will be tested by evaluating the fit of a model (i.e., the number of factors underlying the construct or pattern of relationships between variables). The same authors indicate that while an exploratory analysis would be acceptable in evaluating a new phenomenon, it is not the most appropriate when analyzing a test we have developed or adapted. This is because we know its formulation and composition. In the same sense, Lloret-Segura et al. (2014) state that an EFA is not appropriate when we want to test a base theory or a known composition of the tool, but when the researcher does not know the composition and wants to know what factors underlie the instrument.

As such, the instrument that was developed is a potentially helpful tool for analyzing activities in higher education and for use in research, for example, to analyze the relationships between its results and other variables of interest (e.g., sociodemographic factors, participants' study programs, and age, among others). However, as it is a new instrument, more research is required using other groups of students, for example,

from other universities. It will also be essential to apply the instrument alongside other data sources that evaluate related constructs to assess its evidence of validity regarding other variables. Likewise, it may be relevant to examine how this tool functions when used in other kinds of tasks related to CPS, as all the students took part in the same task in this study.

We can highlight strengths and limitations regarding the format and type of instrument designed. It is important to remember that this instrument was designed to evaluate a task involving group processes in CPS among groups of students (human-human). The evaluation reported in this study was conducted on a group task carried out online. Although this may be seen as a limitation, it can also be considered a strength that the instrument developed could be easily administered with an in-person group task for two reasons. Firstly, the instrument applies to group processes that can occur both in person and online. Secondly, the instrument's characteristics allow for its application in both contexts. One future challenge will be subjecting the instrument to evaluation in different contexts and conditions.

Furthermore, another limitation is that the instrument developed is a self-report style value scale, which may be affected by social factors or be subject to bias on the participant's part. However, this brings to mind two considerations. Firstly, the characteristics of this instrument make it easy to administer. It can be used without specific training, which makes its application, analysis, and reporting much more accessible, allowing teachers in higher education to use it autonomously in their classrooms. Secondly, the content of the items alludes to activities that groups can carry out (and that they ideally would carry out), which provides specific guidance for teachers when it comes to providing timely feedback based on group performance, which itself is central to the teaching of CPS skills (Graesser et al., 2018; Hesse et al., 2015). Another aspect that might be considered a limitation of the instrument is that it has been developed based on skills drawn from the PISA test. Nevertheless, this limitation is justified as it is a useful skills model for developing self-reporting items, in contrast to the ATC21S model, which is more appropriate for observation or interaction analysis instruments.

In terms of group processes observed across dimensions, students assessed the extent to which group

processes were present during a CPS-focused task on four relevant dimensions: EU, RF, PE, and MR. Like the others, the EU dimension showed an accumulation of high scores. However, it did not stand out as the best- or worst-rated area, with a moderate evaluation compared to the other dimensions. Based on exploratory conversation, this process has been highlighted in the literature as an effective type of collaborative learning (Mercer, 2019). Exploratory conversation occurs when participants interact critically and constructively with each other's ideas; suggestions are offered for joint consideration, and thus, reasoning becomes more visible in the conversation.

The RF dimension stood out in this study as the least frequently reported process among the participants. A possible interpretation of this result is that the efficiency of using representational systems is linked to the degree of skill and, therefore, the conceptual and procedural knowledge of those using these processes (Borsinger, 2009; Pérez-Echeverría et al., 2010). Following these authors, as students become more specialized in their fields, their skill in using different ways to represent and formulate the problems they face will increase.

The PE dimension had a higher valuation but more significant student variability. This collaborative process involves communicating with team members about the agreed actions, executing them, and following the rules of participation. According to Scoular and Care (2020), this dimension corresponds with the systematicity and task regulation indicators of the ATC21S model.

The MR dimension involved monitoring and correcting shared understanding, the results of actions, and success in solving the problem, as well as giving feedback and adapting team organization and roles. This was the dimension with the second-highest mean after PE. This result is striking, considering the task did not contemplate a role script specifying monitoring actions. In contrast, in the literature, role scripts have been identified as an essential component of monitoring in CPS learning. For example, Lu et al. (2023) argue that role scripts in higher education can guide individual behavior, regulate group interaction processes, coordinate the structure of assigned learning tasks, influence individual engagement with tasks and peers, facilitate students' understanding of the problem, structure the collaborative learning process, and clarify the tasks to be completed by group members. Although the students did not work with role scripts in this study, they tackled a complex, poorly structured problem that required the integration of information and no immediately apparent solution over several weeks of work (Pazos et al., 2009), which could have contributed to group engagement and reflection.

Finally, regarding this type of instrument in higher education, it is possible to conclude that this tool could become beneficial in evaluating implementation processes in CPS tasks. Indeed, no similar instruments currently allow for reporting these experiences, which is essential in a context where collaboration is increasingly central to developing skills.



## References

- American Psychological Association (APA). (2017). *Ethical Principles of Psychologists and Code of Conduct*. <https://www.apa.org/ethics/code/ethics-code-2017.pdf>
- Borsinger, A. M. (2009). Writer development in the sciences: Expressing new meanings in research. In C. Andersen, N. Scheuer, M. P. Pérez Echeverría, & E. Teubal (Eds.), *Representational systems and practices as learning tools* (pp. 267–280). Sense. [https://doi.org/10.1163/9789087905286\\_016](https://doi.org/10.1163/9789087905286_016)
- Care, E., & Griffin, P. (2017). Collaborative problem solving processes. In B. Csapó & J. Funke (Eds.), *The nature of problem solving. Using research to inspire 21st century learning*. OECD Publishing.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Dindar, M., Järvelä, S., & Järvenoja, H. (2020). Interplay of metacognitive experiences and performance in collaborative problem solving. *Computers & Education*, 154, 103922. <https://doi.org/10.1016/j.compedu.2020.103922>
- Doane, D., & Seward, L. E. (2011). Measuring Skewness: A Forgotten Statistic? *Journal of Statistics Education*, 19(2), 1–18. <https://doi.org/10.1080/10691898.2011.11889611>
- Domínguez-Lara, S. (2012). Propuesta para el cálculo del alfa ordinal y theta de amor. *Revista de Investigación en Psicología*, 15(1), 213–217. <https://doi.org/10.15381/rinvp.v15i1.3684>
- Ferrando, J., & Anguiano-Carrasco, F. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del psicólogo*, 31(1), 18–33. <https://papelesdelpsicologo.es/pdf/1793.pdf>
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, 19(2), 59–92. <https://doi.org/10.1177/1529100618808244>
- 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. *Computers in Human Behavior*, 104, 106134. <https://doi.org/10.1016/j.chb.2019.09.010>
- Griffin, P., McGaw, B., & Care, E. (Eds.) (2012). *Assessment and teaching of 21st century skills*. Springer. <https://doi.org/10.1007/978-94-007-2324-5>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1), 77–89. <https://doi.org/10.1080/19312450709336664>
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A Framework for Teachable Collaborative Problem Solving Skills. In P. Griffin & E. Care, *Assessment and Teaching of 21st Century Skills*. Melbourne Graduate School of Education. <http://doi.org/10.1007/978-94-017-9395-7>
- Hu, L., & Bentler, P. M. (1995). Evaluating Model Fit. In R. H. Hoyle (Ed.), *Structural Equation Modelling: Concepts, Issues, and Applications* (pp. 77–99). SAGE.
- IBM Corp. (2021). *IBM SPSS Statistics for Windows, Version 28.0*.
- Kirschner, P. A., & Erkens, G. (2013). Toward a Framework for CSCL Research. *Educational Psychologist*, 48(1), 1–8. <http://doi.org/10.1080/00461520.2012.750227>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151–1169. <http://dx.doi.org/10.6018/analesps.30.3.199361>

- Lu, Y., Li, K. R., Sun, Z., Ma, N., & Sun, Y. (2023). Exploring the effects of role scripts and goal-orientation scripts in collaborative problem-solving learning. *Education and Information Technology*, 28, 12191–12213. <https://doi.org/10.1007/s10639-023-11674-z>
- Mercer, N. (2019). *Language and the Joint Creation of Knowledge: The Selected Works of Neil Mercer*. Routledge. <https://doi.org/10.4324/9780429400759>
- Montero, I., & León, O. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7, 847–862. [http://www.aepc.es/ijchp/GNEIPo7\\_es.pdf](http://www.aepc.es/ijchp/GNEIPo7_es.pdf)
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7–16. <http://doi.org/10.7334/psicothema2018.291>
- Organisation for Economic Co-operation and Development (OECD). (2017a). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*. OECD Publishing. <https://doi.org/10.1787/9789264281820-en>
- Organisation for Economic Co-operation and Development (OECD). (2017b). *PISA 2015 Results. Collaborative problem solving. Volume V*.
- Pazos, P., Micari, M., & Light, G., (2010). Developing an instrument to characterise peer-led groups in collaborative learning environments: assessing problem solving approach and group interaction. *Assessment and Evaluation in Higher Education*, 35(2), 191–208. <https://doi.org/10.1080/02602930802691572>
- Pérez-Echeverría, M. P., Martí, E., & Pozo, J. I. (2010). Los sistemas externos de representación como herramientas de la mente. *Cultura y Educación*, 22(2), 133–147. <https://doi.org/10.1174/113564010791304519>
- Prieto, G., & Delgado, A. R. (2010). Fiabilidad y Validez. *Papeles del Psicólogo*, 31(1), 67–74. <https://www2.papeles-delpsicologo.es/pdf/1797.pdf>
- RStudio Team. (2023). *RStudio: Integrated Development Environment for R*. RStudio. <http://www.rstudio.com/>
- 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. <https://doi.org/10.1016/j.chb.2019.01.007>
- Slavin, R. E. (2017). Instruction based on cooperative learning. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 388–404). Routledge Press.
- The jamovi project. (2023). *jamovi. (Version 2.3)* [Computer Software]. <https://www.jamovi.org>
- Visschers-Pleijers, A., Dolmans, D., Wolfhagen, I., & van der Vleuten, C. (2005). Development and validation of a questionnaire to identify learning-oriented group interactions in PBL. *Medical Teacher*, 27(4), 375–81. <http://doi.org/10.1080/01421590500046395>

**Appendix 1.** Final Version in Spanish

<b>Ítem</b>	<b>Para nada (1)</b>	<b>Un poco (2)</b>	<b>Parcialmente (3)</b>	<b>En gran medida (4)</b>
1. Dedicamos tiempo a compartir los puntos de vista de cada una/o sobre el problema planteado.				
2. Dedicamos tiempo a identificar las habilidades de cada una/o para la solución del problema planteado.				
3. Encontramos una forma de trabajo en la que todos contribuimos a la solución del problema planteado.				
4. Exploramos distintas posibilidades para organizar el trabajo.				
5. La distribución de tareas la realizamos mediante la discusión grupal.				
6. Cada uno comprendió las tareas que le correspondían.				
7. Logramos acordar una definición compartida del problema.				
8. Discutimos diferentes posibilidades antes de encontrar una definición compartida del problema.				
9. Logramos identificar las tareas necesarias para resolver el problema.				
10. Logramos describir en qué consistía cada tarea a realizar.				
11. La descripción de las tareas a realizar se logró mediante la discusión grupal.				
12. Realizamos diagramas, resúmenes (u otros) que nos ayudaron a organizar el trabajo.				
13. Conversamos sobre los pasos que teníamos que seguir para resolver el problema.				

Ítem	Para nada (1)	Un poco (2)	Parcialmente (3)	En gran medida (4)
14. Logramos ponernos de acuerdo en los pasos a seguir.				
15. Logramos ejecutar las tareas acordadas.				
16. Seguimos un plan de trabajo.				
17. Nos apoyamos mutuamente para realizar las tareas acordadas.				
28. Acordamos algunas reglas básicas de funcionamiento como equipo.				
19. Al menos un/a integrante del grupo se encargó de corroborar que todos/as entendiéramos bien el problema.				
20. Cambiamos de opinión sobre la forma de resolver el problema siempre que fue necesario.				
21. Al menos un/a integrante del grupo se encargó de chequear que las tareas acordadas se realizaran.				
22. Logramos decidir en conjunto si habíamos resuelto el problema.				
23. Hicimos cambios en la organización del trabajo las veces que fue necesario durante el desarrollo de la actividad.				
24. Todos/as pudimos opinar sobre cómo los demás integrantes desarrollaban sus tareas.				

**Appendix 2. Indicators by Dimension and Items Constructed**

<b>Exploring and Understanding (EU)</b>	
<b>Indicators by dimension</b>	<b>Items*</b>
<b>EU1. Discovering team members' perspectives and skills</b>	1. We spent some time sharing our perspectives and points of view regarding the problem.
	2. We spent some time identifying our individual skills that could help solve the problem.
	25. We identified the skills of each team member that could help solve the problem.
<b>EU2. Discovering the type of collaborative interaction that could solve the problem based on objectives</b>	3. We found a way to work collaboratively toward solving the problem.
	4. We considered different ways of organizing our work.
<b>EU3. Understanding the roles needed to solve the problem</b>	5. Discussing the distribution of roles helped us understand how each of us could contribute to solving the problem.
	6. Each team member understood their role in working towards the group's objective.
<b>Representing and Formulating (RF)</b>	
<b>Indicators by dimension</b>	<b>Items*</b>
<b>RF1. Building a collective representation and negotiating the meaning of the problem</b>	7. We could quickly agree on the definition of the problem we had to solve.
	8. Our group discussed different possibilities before coming to a shared definition of the problem.
	26. We each shared our understanding of the problem to come to a shared vision.
<b>RF2. Identifying and describing the tasks to be undertaken</b>	9. We were able to identify the specific tasks that were necessary to solve the problem.
	10. As a group, we found it easy to explain the specific tasks we had to carry out to solve the problem.
<b>RF3. Discovering roles and organizing the team</b>	11. Role and task distribution was achieved through group communication and reflection.
	12. We created diagrams and drawings that helped us organize our work.
	27. We wrote down what we had to do to facilitate communication and task completion.
<b>Planning and Executing (PE)</b>	
<b>Indicators by dimension</b>	<b>Items*</b>
<b>PE1. Communicating with team members regarding actions to be taken</b>	13. We discussed the steps we needed to follow to solve the problem.
	14. We agreed on the steps to solve the problem.

<b>PE2. Executing the agreed plan</b>	15. We were able to execute specific agreed-upon tasks without any trouble.
	16. We followed the agreed plan to solve the problem.
<b>PE3. Following the rules of participation</b>	17. We mutually supported and encouraged each other when completing agreed tasks and actions.
	18. We followed the rules we had set for ourselves as a team.
	28. We agreed on some basic rules for working as a team.
<b>Monitoring and Reflecting (MR)</b>	
<b>Indicators by dimension</b>	<b>Items*</b>
<b>MR1. Monitoring and correcting shared understanding</b>	19. At least one group member ensured the others understood the problem.
	20. We were willing to change our opinions regarding how the problem should be solved.
<b>MR2. Monitoring the result of actions and evaluating the success of problem-solving</b>	21. At least one group member checked that the agreed tasks were being carried out correctly.
	22. We collectively decided whether we had solved the problem.
<b>MR3. Monitoring, giving feedback on, and adapting role distribution and group organization</b>	23. During the task, we reflected on how we organized our work, making strategic changes when necessary.
	24. Each group member could comment and give opinions on how the rest of the team was working.

**Note:** (\*) This is a direct translation of the item provided for informational purposes. The scale has been designed and administered in Spanish and has not been adapted to English.

### Appendix 3. CPS Task

#### Global Warming and Its Consequences for Life

For many years, experts have warned us about global warming and its consequences on nature and human life. This has led to much information on the subject circulating on television, radio, the printed press, and social media. One day, while checking Instagram, you find a post where a highly regarded French economist states that “human life is destined for extinction, as it is no longer possible to create economic development without destroying the planet. Therefore, extinction is a matter of time, and we can do nothing about it.” When you read this, you feel the need to say something about it. That afternoon, you meet your classmates, and as the statement is still swimming around in your head, you show them the post. One of your classmates noticed an online newspaper published a piece in the post saying, “Send us your opinion about this news item, and we will publish it.” The group then decides to write an opinion together and send it to the newspaper.

Note: The newspaper only accepts reader-generated texts that are between 500 and 800 words long (Word shows you the word count in the bottom left corner of the window).

#### Appendix 4. Initial Item Analysis Based on the Discrimination Index and Reliability if the Item is Removed

Item	Dimension	Item-test			Item-dimension		
		Item-test discrimination index (>.30)	Reliability if the item is removed	Removal improves reliability (>.97)	Item-dimension discrimination index (>.30)	Reliability if the item is removed	Removal improves reliability
1	EU	.734	.971	No	.763	.892	No
2	EU	.702	.971	No	.721	.896	No
3	EU	.772	.971	No	.790	.889	No
4	EU	.742	.971	No	.730	.895	No
5	EU	.744	.971	No	.719	.897	No
6	EU	.723	.971	No	.673	.901	No
7	RF	.789	.971	No	.761	.898	No
8	RF	.761	.971	No	.734	.900	No
9	RF	.753	.971	No	.767	.898	No
10	RF	.675	.972	No	.674	.906	No
11	RF	.807	.971	No	.788	.896	No
12	RF	.575	.972	No	.551	.917	Slightly
13	PE	.806	.971	No	.782	.915	No
14	PE	.786	.971	No	.800	.913	No
15	PE	.768	.971	No	.777	.916	No
16	PE	.733	.971	No	.753	.918	No
17	PE	.805	.971	No	.820	.911	No
18	PE	.812	.971	No	.814	.912	No
19	MR	.648	.972	No	.644	.847	No
20	MR	.698	.971	No	.697	.838	No
21	MR	.577	.972	No	.602	.854	No
22	MR	.827	.971	No	.772	.825	No
23	MR	.615	.972	No	.622	.852	No
24	MR	.720	.971	No	.649	.847	No



Item	Dimension	Item-test			Item-dimension		
		Item-test discrimination index (>.30)	Reliability if the item is removed	Removal improves reliability (>.97)	Item-dimension discrimination index (>.30)	Reliability if the item is removed	Removal improves reliability
25	EU	.778	.971	No	.700	.899	No
26	RF	.814	.971	No	.766	.898	No
27	RF	.757	.971	No	.716	.902	No
28	PE	.693	.972	No	.660	.928	Slightly