6th International Conference on
New Trends in Teaching and Education

26 - 28 May 2023 Brussels, Belgium

Design and Validation of a Rubric to Assess Problem-Solving Skills of Students in a Chemistry Context

Andrés Garcés*, Luis F. Sánchez-Barba, Damián Pérez-Quintanilla, Miguel Ángel González, Gonzalo Martínez, Esther M. Sánchez-Carnerero, Óscar R. Montoro and Carolina Vargas

Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, Móstoles (Spain)

**andres.garces@urjc.es*

Abstract

Interest in developing 21st Century Skills, particularly Problem-Solving (PS) abilities in Science Education, has grown in recent years. This study presents the design and validation of an analytic rubric to quantitatively assess PS performance in 23 students enrolled in the Water Resources (WR) degree program at Rey Juan Carlos University in Madrid, Spain. The rubric evaluated students' PS abilities through conveniently designed ill-structured problems within a General Chemistry course. The rubric consists of three dimensions, each representing different levels of PS mastery, with a total possible score of 60 points. By employing indicators ranked on a three-level scale, the rubric enables a comprehensive evaluation of students' problem-solving efforts in written solutions. Categorizing students into five stages of competence acquisition allows effective assessment of their PS skills within each level.

Findings from the study indicate a preliminary decrease in PS performance as students progress through the different levels, although a significant grade is achieved at the initial stage. Notably, a positive correlation ($r = 0.778$, $p < 0.05$) exists between students' final exam marks and the assigned rubric scores, reinforcing the validity of the rubric as an accurate instrument to measure of PS skills. The designed analytic rubric serves as a robust and practical tool for assessing PS abilities, facilitating personalized instruction, and fostering continuous improvement in students' problem-solving capabilities within the General Chemistry course. Moreover, reliability assessments, including Cronbach's alpha (0.983) and inter-rater reliability (Intra-Class Correlation Coefficient = 0.935), confirm the effectiveness and consistency of the rubric for objective PS skills assessment.

In summary, the analytic rubric offers valuable insights into students' problem-solving abilities and real-world applications. Instructors can adapt methods and provide targeted feedback to enhance performance. Incorporating student explanations can improve comprehensive

evaluation. The rubric is a valuable educational tool, empowering continuous improvement and excellence in PS skills.

Keywords: 21st Century Skills, Critical Thinking, Evaluation, Feedback, Soft skills.

1 Introduction

In today's rapidly advancing world of science and technology, there is a growing demand for a new model of education that ensures graduates possess the necessary skills to succeed in an ever-changing society. Employers now require individuals who are flexible and continuous learners, capable of keeping up with the challenges of our dynamic world (Stehle, S. M. and Peters-Burton, E. E., 2019). Education strives to foster independent thinking, self-directed learning, and workforce readiness, in addition to imparting basic content knowledge to learners (Lapek, J., 2018). As a result, educational policies worldwide are focused on promoting the cultivation of a set of core competencies known as 21st Century Skills, which are essential for effectively addressing contemporary challenges.

These 21st Century Skills have been categorized and organized in various frameworks in the literature. For instance, Trilling and Fadel (2009), led by the Partnership for 21st Century Skills institution (P21, 2007), introduced a group of critical skills organized into three domains: Learning and Innovation Skills, Digital Literacy Skills, and Life and Career Skills. In a different research study, Wagner (2008) emphasized the need for students to acquire a series of seven competencies, referred to as "survival skills," to be well-prepared for life and work in the 21st century. Other notable frameworks include those conducted by the Organization for Economic Co-operation and Development (OECD) (Ananiadou, K., and Claro, M., 2009), which identified three broad categories of skills including information, communication and ethics and social impact, as well as the learning standards published by the American Association of School Librarians (AASL, 2007), which presented four learning domains including cognitive, psychomotor, affective, and developmental skills.

Despite the differences in terminology and categorization across these frameworks, they all recognize Problem-Solving (PS) as a crucial skill for success in the modern world. This is particularly interesting in Science Education, where PS provides students with the ability to deepen their understanding of their own knowledge, thereby enhancing desired educational outcomes (National Research Council, 2012; Schunk, D.H., 2012).

Unfortunately, students often become familiar with routine problems or algorithmic questions, which can suppress their ability to develop a meaningful understanding. Expanding upon the significance of PS in the context of Science Education, it is essential to highlight its multifaceted nature, which transcends mere content knowledge acquisition. In order to foster the development of a deep understanding, learners must not only possess knowledge, but also effectively apply it through the use of appropriate strategies and organize it systematically.

Additionally, the conventional evaluation processes used to assess students' performance in these tasks often lack the ability to fully capture their PS skills or their capacity to make nuanced PS decisions, which are essential for achieving true mastery (Price, A. *et al.*, 2022). Therefore, there is an urgent need for innovative instructional methods that emphasize the development of students' PS abilities, along with a fresh approach to accurate performance assessment. In essence, the current shift in learning towards a model based on skills' development must be accompanied by a corresponding change in assessment methods, as both learning, and assessment are attached processes (Sánchez Ramírez, J. *et al.*, 2022). Therefore, the design of appropriate assessment instruments becomes critical in accurately measuring students' improvement in PS, which will promote effective learning.

The methods of assessment used to evaluate the level of acquisition of some skill such as PS clearly should differ from other grading tools traditionally employed in education (Jayasinghe, U., 2015). Methods for estimating the level of achievement in PS place a greater emphasis on the process rather than the result (Karakus, G. and Ocak, G., 2022). One common method for investigating PS skills in learners is conducting interviews with students where they can verbalize their thinking strategies for solving a problem (van Someren, M. W. *et al.*, 1994). Although interviews can provide more detailed information about a student's PS abilities, they take a significant amount of time to be prepared and conducted, and even with small groups, a great volume of information is collected (Docktor, J. L. *et al.*, 2016). A more common and effective tool for performance assessment to determine whether a student has acquired an expected skill is a rubric. Several benefits have been associated with the use of rubrics in Higher Education such as a greater consistency of scoring across respondents and, also between evaluators, in addition to its utility for promoting learning (Jonsson, A. and Svingby, G., 2007).

A rubric for assessing the level of competency acquisition is an instrument that is organized in matrix form, and it is used to qualitatively rate students' performance based on specific criteria. These criteria are conveniently scaled according to different descriptions that characterize the students' responses on each specific criterion. The user sets the scoring criteria, guided by the quality of definitions provided by the rubric, in order to evaluate each response. In addition, the evaluative criteria chosen by the rubric developer can be organized into different stages or dimensions based on the level of command that individuals show in that particular skill.

Two main categories of rubrics can be distinguished: *holistic* and *analytic*. Holistic rubrics are effective for evaluating overall performance on a task, while analytic rubrics evaluate separately each criterion of a task using descriptive ratings. The analytic rubric has been described as the more appropriate assessment tool for identifying strengths and weaknesses of a given task (Nsabayezu, E. *et al.*, 2022). Additionally, it helps students self-assess their level of achievement. Both types of rubrics result useful for removing misunderstandings and increasing students' performance since they are acting in response to teachers' expectations as shown in the provided instrument (Jescovitch L. N., 2019).

Rubrics have become widely used in STEM disciplines, particularly in chemistry, over the last few decades. These assessment instruments are used to evaluate students' performance in a wide range of skills, including critical thinking (Oliver-Hoyo, M., 2003), PS (Shadle, S. E., 2012), and other process skills (Reynders, G. *et al.*, 2019). Many of these studies have been focused on experimental courses where it is generally easier to assess soft skills such as PS compared to regular classroom settings. This can be attributed to the practical laboratory setting, which fosters hands-on and experimental learning. This environment encourages students to actively engage in PS activities such as conducting experiments, collecting data, and analyzing results.

This manuscript presents the development and validation of a rubric designed to assess freshman students' progress in their PS abilities after completing a series of original Multi-Concept Linked Problems (MCLPs) in a first-grade General Chemistry university course. The instructors previously designed these sorts of novel problems, and their effectiveness in improving students' learning and satisfaction was recently presented and discussed (Garcés-Osado, A. *et al.*, 2023).

Following a similar approach to Villa and Poblete (2007), the instrument was initially created by the instructors and structured into three levels of performance. The first level addresses the beginning features of PS, the second level focuses on maturing and developing PS skills, and the third level represents mastery of the competency. Each stage includes four questions that relate to specific levels of command of the skill, and the assessment is ranked as low, medium, or high in each case.

As the study was conducted with freshman students, the majority was expected to fall within the first or second level of the skill. Nonetheless, the information provided by the instrument can serve as effective feedback for instructors to adjust their future teaching and for students to enhance their performance in the course.

Additionally, the reliability of the rubric has been verified by comparing the different scores obtained by several instructors using the same instrument. Finally, a good agreement between students' marks on the subject final exam and the rubric scores assigned is also presented.

2 Objectives

The main objective of this study is to develop and implement an analytical rubric as an effective tool for evaluating PS skills in freshman General Chemistry students.

To achieve this aim, we have established the following specific objectives:

1. Design a comprehensive rubric with scoring descriptors categorized as low, medium, or high, aligned with multiple performance levels and criteria, to assess students' proficiency in comprehending and applying concepts within MCLPs in a freshman General Chemistry course.

2. Assess the rubric's efficacy in measuring the improvement of students with limited prior knowledge in chemistry throughout the learning process.
3. Provide valuable information to instructors and students through the rubric, emphasizing the evaluation process and encouraging self-assessment and self-directed learning skills.
4. Validate the rubric's reliability by involving four different instructors and assessing the consistency and accuracy of assessment results for evaluating PS skills in freshman General Chemistry students.
5. Contribute to the field of Chemistry education by offering a robust and effective assessment tool for evaluating PS skills in MCLPs.

3 Methodology

3.1 Context of the study

In recent years, the low success rate in General Chemistry courses at Rey Juan Carlos University prompted instructors to explore new instructional approaches, including ill-structured problems within real-life contexts. These PS exercises, known as MCLPs, proved to be highly beneficial for students with diverse academic backgrounds, including Environmental Sciences, Water Resources, and Industrial Technology Engineering programs (Garces-Osado, A. *et al*). Throughout the course, students faced progressively more complex MCLPs, fostering higher-order thinking skills and improved performance.

However, although MCLPs were developed to provide learners with appropriate skills to be more successful in their learning, it is also important to track conveniently their progress while they are facing this task. This information will not only serve instructors to understand how their students are improving but also to identify weak points and provide adequate feedback during instruction. Additionally, students can utilize this information to identify the areas they need to reinforce. To achieve this, a trustworthy instrument that facilitates the measurement of their progress is essential, and the gathered information should be easily understandable by both tutors and learners.

With this goal in mind, our study focuses on the design and evaluation of a rubric that serves as an appropriate tool for measuring PS abilities. To ensure a more meaningful and comprehensive analysis, we have narrowed our research scope exclusively to students enrolled in the Water Resources (WR) degree program. By doing so, we target individuals from less advantaged backgrounds, who possess relatively limited prior knowledge of chemistry. This deliberate selection allows us to assess the impact and evolution of the MCLPs more effectively on their PS skills. The study group consists of 23 students or individuals, without gender distinction, aiming to conduct a general analysis of their abilities and performance in the study.

3.2 Description of the instrument

The analytic rubric developed for the General Chemistry course assesses students' PS skills and establishes an easy-to-use evaluation system that accurately, reliably, and efficiently measures students‘ progress. It evaluates not only the final result of the problem but also students' ability to develop a "mental map" of the problem and seek alternative approaches to obtain the results, enabling them to solve subsequent related questions. Importantly, the rubric also assesses students' ability to apply their PS skills to real-world scenarios that they may encounter in their future careers. By connecting the problem, course material, and proposed real-world situation, students learn the authenticity of their learning experience, making it, more engaging, and motivating. Therefore, the rubric serves not only as an evaluation tool for measuring students' problem-solving skills but also as a valuable means for enhancing their ability to apply these skills to real-world scenarios. This, in turn, can contribute to their overall professional development and prepare them for success in their future careers.

To create the rubric, we followed the following steps:

1. *Criteria identification.* We began by clarifying the purpose of the instrument and desired outcomes for the students. As we created the rubric for assessing PS skills in a chemistry course for science and engineering degrees, our criteria also included other aspects that students need to cultivate, such as critical thinking and communication and collaboration skills - all of which we aimed to develop with the application of our developed MCLPs.
2. *Performance level establishment*. By dividing the subject matter to be assessed into levels can make easier to standardize the results. For this reason, we developed a rubric that is divided into three different levels of achievement, following the approach used by Villa and Poblete (2007).

   Level 1 – Problem Identification and Information Utilization.
   Level 2 – Competently identification of causes and selection of effective approaches.
   Level 3 – Proficiency in developing innovative solutions with broad applicability.
3. *Criterion definition*. To ensure a comprehensive assessment, we have established specific criteria for each performance level. These criteria will serve as the basis for evaluating students' problem-solving skills within each level. For each criterion, we have included clear descriptors that aid in accurately determining the level of achievement. The descriptors align with the low, medium, and high proficiency levels, providing specific guidelines for assessing the performance of students at each level of PS ability. By incorporating these defined criteria and descriptors, our rubric enables a precise evaluation of students' problem-solving skills across different levels of achievement. This approach ensures standardized and consistent assessment results, facilitating a fair evaluation of students' abilities in PS.

4. *Scoring system development*. To convert the qualitative results into numerical values, we have assigned a score to each descriptor within each criterion. Specifically, we have assigned a score of 1 to descriptors indicating a low proficiency level, a score of 3 to descriptors representing a medium proficiency level, and a score of 5 to descriptors indicating a high proficiency level. This scoring system allows for a quantitative representation of the assessment results, enabling easier analysis and comparison of students' problem-solving skills. To categorize students based on their scores, we have established five stages of competence acquisition for each level. Each level has a maximum of 20 points and a minimum of 4 points. Depending on the score obtained within each level, students are placed into the following categories:

    1. Expert: Scores above 16 points.
    2. Proficient: Scores ranging from 12 to 16 points.
    3. Competent: Scores ranging from 8 to 12 points.
    4. Beginner: Scores ranging from 4 to 8 points.
    5. Novice: Scores below 4 points.

Considering that the rubric consists of 3 levels and four questions per level, the minimum possible score is 12 points (one point per question), while the maximum score is 60 points (five points per question). By categorizing students into these stages of competency acquisition, we can effectively evaluate their PS skills and gain insights into their proficiency levels within each level of the rubric.

4 Results and Discussion

4.1 Rubric Application for Assessing Problem-Solving Skills. Statistical analysis

This study investigates the effectiveness of utilizing a rubric as an invaluable tool for assessing students' PS skills, which also offers instructors a systematic approach to evaluate their performance on specific problems. By analyzing the application of this rubric in assessing PS abilities, we gain insights into its utility and impact on evaluating students' performance in the context of General Chemistry.

Through the application of the rubric, instructors accurately analyze students' abilities at each level, considering the outlined criteria detailed in the rubric (see Appendix). In the initial level, students are evaluated based on their capacity to identify and employ relevant data, utilize appropriate laws or theories, and accurately cite bibliographic references. The results reveal an average score of 8.902 points (see Table 1), indicating a competent level among students in this introductory stage.

Advancing to the second level, instructors assess the clarity of students' PS strategies, including their consideration of alternative approaches and justification for selecting the most suitable path. However, the evaluation demonstrates that students achieved an average score of

5.728 (see Table 1), placing them in the beginner category and indicating challenges in utilizing appropriate strategies.

Lastly, the evaluation of the third level involves a supervised class where teachers guide students individually to assess their ability to apply the presented problem within the context of their respective knowledge areas. The results from this supervision process indicate an average score of 4.956 points (see Table 1), classifying students as novices at this level.

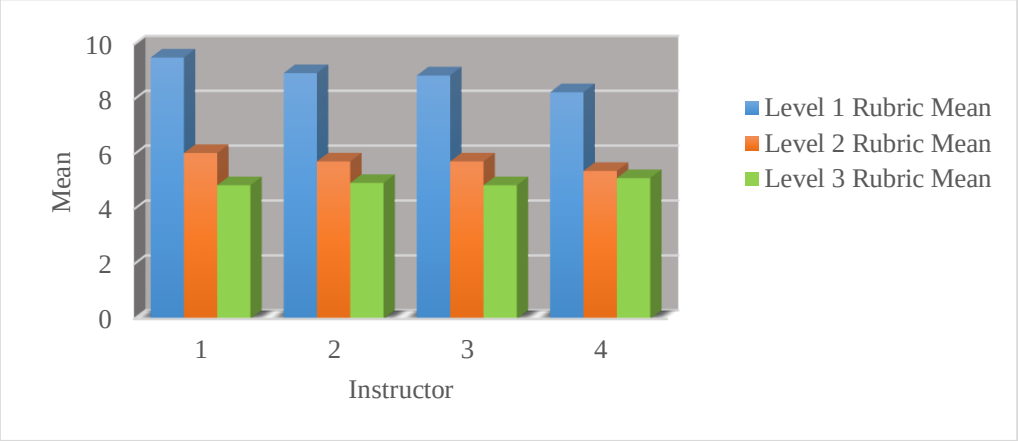
As students progress through the levels, it is expected that their PS performance may decrease due to the increasingly demanding criteria that require a deeper understanding of the subject matter. Nevertheless, it is noteworthy that students were able to attain a competent grade in the first level, which is a desirable outcome for students in their early years of study.

Table 1. Mean and standard deviation of the score for each instructor. The data are separated on levels. Therefore, the average of score by level and the category has been given.

Instructor	Level 1 Rubric Mean	Standard Deviation	Level 2 Rubric Mean	Standard Deviation	Level 3 Rubric Mean	Standard Deviation	N
1	9.523	4.907	6.043	1.942	4.869	1.010	23
2	8.956	4.117	5.739	2.200	4.956	1.022	23
3	8.870	3.924	5.739	2.200	4.869	1.014	23
4	8.261	3.825	5.391	1.644	5.130	1.325	23
Mean	**8.902**		**5.728**		**4.956**		
Category	**Competent**		**Beginner**		**Novice**		

Figure 1 illustrates the mean scores for each instructor across the different levels as indicated by the rubric. The rubric scores show a consistent similarity among instructors, with the most noticeable variation observed in the first instructor's evaluations. However, despite these slight differences, all instructors exhibit a similar downward trend in the rubric scores. Remarkably, the scores assigned by the four instructors were highly comparable, even when evaluated independently through peer review. The main distinction among the instructors was found in the standard deviations of the generated data sets, although these variances remained within a similar range. These findings demonstrate the robustness and reliability of the rubric developed in this study, independent of which instructor is using it.

Figure 1. Mean of the rubric for each level and collected by the four instructors.



4.2 Reliability and validation of the rubric. Cronbach method

Statistical analysis was performed using SPSS Statistics software (IBM, 2021). Cronbach's alpha, a measure of internal consistency reliability, was used to assess the reliability of the rubric presented in this study. We consider that a study using the total rubric scores is necessary to provide an overview of the study carried out with the new rubric presented in this work. To conduct a comprehensive analysis, the average total score of the rubric was computed based on the 23 results from each instructor. The resulting alpha value for the global analysis was 0.983, indicating high reliability since the alpha values are close to 1. It is worth noting that values above 0.8 are considered good for research purposes. In both global and level-by-level analysis, the estimation of alpha was performed using the correlation matrix of the analyzed data, as shown in Table 3 for the global study. The high inter-item correlations, with values close to 1, indicate a strong monotonic and linear correlation between the data generated by each instructor. The lowest correlation value of 0.945 was observed between instructor 1 and instructor 2, still demonstrating a nice correlation between their data sets.

Additionally, a level-specific analysis was conducted. Table 2 presents the alpha values for each level and instructor. The alpha values for the levels were all close to 1 and higher than 0.8: 0.904 for level 1, 0.961 for level 2, and 0.950 for level 3. The correlation matrices corresponding to the Cronbach's alphas for this part of the study are provided in Table 3. The lowest values in the correlation matrices were higher than 0.7, indicating a clear correlation between the results of the instructors. The strong correlation matrices, combined with the high Cronbach's alpha values, provide further evidence of the reliability of the rubric developed in this study.

Table 2. Values of Cronbach´s alpha for each level and for the global data set.

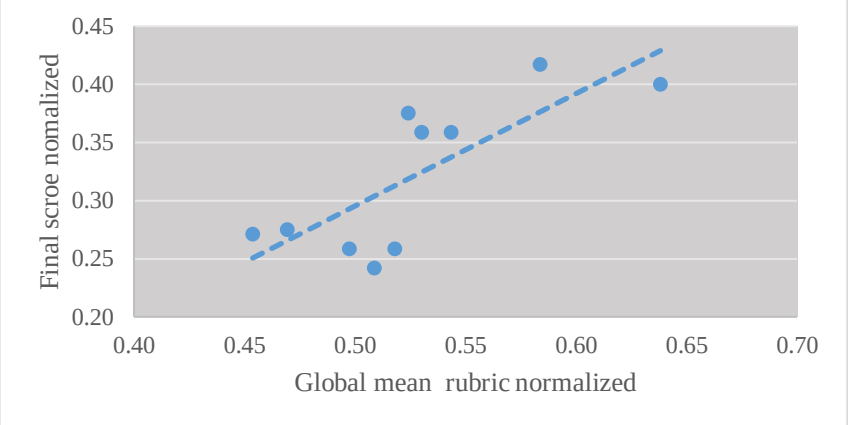
Level	Cronbach´s alpha	Number of items
1	0.904	4
2	0.961	4
3	0.950	4
Global	**0.983**	**4**

Table 3. Correlation´s matrices among items for the global data set, for level 1, level 2 and level 3.

Correlation Global	I1	I2	I3	I4
I1	1.000	0.945	0.955	0.969
I2	0.945	1.000	0.963	0.960
I3	0.955	0.963	1.000	0.947
I4	0.969	0.960	0.947	1.000
Correlation Level 1	**I1**	**I2**	**I3**	**I4**
I1	1.000	0.910	0.874	0.875
I2	0.910	1,000	0.938	0.823
I3	0.874	0.938	1.000	0.777
I4	0.875	0.823	0.777	1.000
Correlation Level 2	**I1**	**I2**	**I3**	**I4**
I1	1.000	0.875	0.833	0.834
I2	0.875	1.000	0.962	0.859
I3	0.833	0.962	1.000	0.859
I4	0.834	0.859	0.859	1.000
Correlation Level 3	**I1**	**I2**	**I3**	**I4**
I1	1.000	0.787	0.859	0.739
I2	0.787	1.000	0.916	0.911
I3	0.859	0.916	1.000	0.859
I4	0.739	0.911	0.859	1.000

To ensure comparability, both rubric scores and students' final grades were normalized by dividing each result by the maximum possible score: 60 for the rubric and 10 for the final grade, restricting all data sets to a range of 0 to 1. The correlation analysis involved plotting the final grades, as well as the rubric scores, after normalizing them to examine their relationship. Only the results of students who passed the course were considered, as the dispersion of results among the remaining students eliminated the correlation. Figure 2 provides a visual representation of these data.

Figure 2. Plot of normalized rubric versus normalized final scores for each student.



The graphic illustrates a modest, although statistically significant, correlation between rubric scores and students' final grades, with a Pearson's correlation coefficient of 0.778. This finding suggests a potential association between PS abilities, as evaluated through the rubric, and students' overall performance. However, it is important to recognize that this correlation alone does not offer a comprehensive measure of the strength or quality of PS skills. Other factors, such as content knowledge, study habits, and individual learning styles, can also impact in students' performance. Therefore, while our results suggest a promising connection, it is important to consider additional aspects and conduct further investigations to gain a more comprehensive understanding in the relationship between PS skills and subject performance.

To enhance the assessment process for PS skills, potential improvements could involve the incorporation of sections where students explain their PS approaches and expanding the rubric to encompass other subject competencies such as comprehension, idea development, and writing conclusions. These adjustments would provide a more suitable evaluation of students' abilities and contribute to a more holistic understanding of their performance in the subject.

5 Conclusion

This study describes the design and validation of an effective rubric for assessing PS skills in a freshman General Chemistry course. The rubric incorporates real-life embedded MCLPs and consists of four items included in three different levels. Previous related work on MCLPs (Garcés et al., 2023) showed their effectiveness in improving students' problem-solving achievement. The present study successfully assesses PS students' abilities using the designed rubric, supported by reliability and internal consistency measurements. Our results align with Price et al.'s findings (2022), presenting a template for assessing authentic real-world advanced problem-solving skills across science and engineering disciplines.

In addition, our research team's rubric provides an objective assessment method for evaluating PS skills. This is supported by the high correlation analysis between the four instructors involved in the study. The rubric's dimensions and levels reflect students' understanding of problem identification, effective approaches, and innovative solutions. Students' problem-solving proficiency tends to decrease as the rubric levels increase, consistent with first-year expectations. Our rubric is in line with Karakus et al.'s recent work (2022), providing the validity by taking into consideration the opinions of instructors, and with Shadle et al.'s findings (2012), who evaluates students' abilities to address real-world problems.

The rubric effectively identifies students' needs and informs for instructional adjustments. Further investigations are recommended to explore the rubric's efficacy in distinguishing between students with different experience levels, as highlighted by Shadle et al (2012). Additionally, other studies could explore the rubric's alignment with instructors' perceptions of the evaluated items, allowing the identification of additional criteria to enhance the instrument's

utility and effectiveness. The research team is actively working on implementing the rubric in Advanced Chemistry courses to evaluate PS skills in these contexts.

Acknowledgment

We gratefully acknowledge financial support from the Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, Spain and all students participating in this study.

References

American Association of School Librarians [AASL]. (2007). Standards for the 21st-century learner. Chicago, IL: American Library Association.

Ananiadou, K., and Claro, M. (2009). 21st Century Skills and Competences for New Millennium Learners in OECD Countries. OECD Education Working Papers, No. 41. Paris: OECD Publishing.

Docktor, J. L., Dornfeld, J., Frodermann, E., Heller, K., Hsu, L., Jackson, K. A., and Yang, J. (2016). Assessing student written problem solutions: A problem solving rubric with application to introductory physics. *Physical Review Physics Education Research*, 12(1), 010130.

Garcés Osado, A., Sánchez-Barba Merlo, L. F., Pérez Quintanilla, D., González González, M. A., and Martínez García, G. (2023). Context-based learning instructional approaches using multi-concept linked problems in a general chemistry course. In *INTED2023 Proceedings*. Valencia, Spain, pp. 266-275.

IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp.

Jayasinghe, U., Dharmaratne, A., and Atukorale, A. (2015). Students' performance evaluation in online education system Vs traditional education system. *Proceedings of 2015 12th International Conference on Remote Engineering and Virtual Instrumentation (REV)*, 131–135.

Jescovitch, L. N., Scott, E. E., Cerchiara, J. A., Doherty, J. H., Wenderoth, M. P., Merrill, J. E., Urban-Lurain, M., and Haudek, K. C. (2019). Deconstruction of holistic rubrics into analytic rubrics for large-scale assessments of students' reasoning of complex science concepts. *Practical Assessment, Research and Evaluation*, 24, 1-16.

Jonsson, A., and Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.

Karakus, G., and Ocak, G. (2022). The Implementation of Cooperative Problem-Solving Rubric Towards Turkish Fourth Grade Students. Mimbar Sekolah Dasar, 9(1), 1-23.

Lapek, J. (2018). Promoting 21st century skills in problem-based learning environments. *CTETE-Research Monograph Series*, 1(1), 66–85.

National Research Council. (2012). Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. Committee on Defining Deeper Learning and 21st Century Skills, James W. Pellegrino and Margaret L. Hilton, Editors. Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nsabayezu, E., Iyamuremye, A., Nahimana, J. P., Habimana, J. B., Munyemana, E., and Ntaganda, B. (2022). The progress in the application of rubric materials in chemistry teaching and students' learning enhancement during 21st century: A systematic review. *Discovery Education*, 1, 5.

Oliver-Hoyo, M. (2003). Designing a written assignment to promote the use of critical thinking skills in an introductory chemistry course. *Journal of Chemical Education*, 80, 899-903.

Partnership for 21st Century Skills. (2007). Framework for 21st-century learning. Available: https://static.battelleforkids.org/documents/p21/P21_Framework_Definitions_New_Logo_2015_9pgs.pdf

Price, A., Salehi, S., Burkholder, E., Kim, C., Isava, V., Flynn, M., and Wieman, C. (2022). An accurate and practical method for assessing science and engineering problem-solving expertise. *International Journal of Science Education*, 44(13), 2061-2084.

Reynders, G., Suh, E., Cole, R., and Sansom, R. (2019). Developing student process skills in a general chemistry laboratory. *Journal of Chemical Education*, 96, 2109-2119.

Sánchez Ramírez, J., Íñigo Mendoza, V., Marcano Lárez, B., and Romero-García, C. (2022). Design and validation of an assessment rubric of relevant competencies for employability. *Journal of Technology and Science Education*, 12(2), 426-439.

Schunk, D.H. (2012). *Learning Theories: An Educational Perspective*, 6th ed. Boston, MA. Pearson Education Inc.

Shadle, S. E., Brown, E. C., Towns, M. H., and Warner, D. L. (2012). A rubric for assessing students' experimental problem-solving ability. *Journal of Chemical Education*, 89, 319-325.

Stehle, S.M., Peters-Burton, E.E. (2019). Developing student 21st Century skills in selected exemplary inclusive STEM high schools. *International Journal of STEM Education*, 6, 39.

Trilling, B., and Fadel, C. (2009). *21st century skills: Learning for life in our times*. Jossey-Bass/Wiley.

van Someren, M. W., Barnard, Y. F., and Sandberg, J. A. C. (1994). *The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes*. Academic Press.

Villa, A., and Poblete, M. (Dirs.). (2007). *Aprendizaje basado en competencias. Una propuesta para la evaluación de las competencias genéricas*. Editorial Mensajero, Colección Estudios e Investigación del ICE. Universidad de Deusto.

Wagner, T. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need-and what we can do about it*. New York: Basic Books.

6 APPENDIX. Analytic Rubric for assessing problem-solving skills

Level 1: Problem identification and information utilization

	Application of theoretical laws and concepts	Identification of problem causes	Search for additional information	Recognition of irrelevant data
Low	The student is not able to relate the problem to any theoretical principle that regulates it	The student is not able to identify the ultimate goal of the problem or any goal prior to its achievement	The student does not use data from bibliography and relies on the data from the problem	The student is not able to identify irrelevant data of the problem
Medium	The student is able to relate the problem to some theoretical principle that regulate it	The student is not able to identify the ultimate goal of the problem, but rather the objectives prior to achieving it	The student uses data from bibliography, but not all necessary to solve the problem	The student identifies some of the irrelevant data of the problem
High	The student is able to relate the problem to all theoretical principle that regulate it	The student is able to identify the ultimate goal of the problem and all the goals prior to its achievement	The student uses all the necessary bibliographic data to solve the problem	The student identifies all irrelevant data of the problem

Level 2: Competently identifies causes and selects effective approaches

	Reasoned problem-solving strategy	Reasonable conclusion	Alternative solution strategies	Evaluation of appropriate alternatives
Low	The student is not able to communicate in a simply way the process followed to solve the problem	The student is not able to draw conclusions from problem solving	The student is not able to present more than one solution of the problem	The student is not able to reason the difference between the provided solutions
Medium	The student explains the process followed, but in a confusing and disorganized manner	The student comes to conclusions, but they are not entirely accurate	The student presents more than one solution, but some of them are not effective	The student can distinguish between solutions, but is not able to explain rationally the advantage of one over the another
High	The student is able to describe in a clear and concise manner the process followed to solve the problem	The student is able to draw accurate conclusions from the problem resolution	The student presents several solutions and all of them are effective	The student understands the difference between solutions and is able to explain the convenience of one solution over the another

Level 3: Proficiency in developing innovative solutions with broad applicability

	Identification of major drawbacks	Referencing and approach support	Identification of real-world applicability	Consideration of factors for solution application
Low	The student has difficulty anticipating problems when their effects is not evident	The student is not capable of obtaining rigorous support for his arguments	The student has difficulties applying the problem to other situations	The student is not able to list any key factors to consider
Medium	The student anticipates the emergence problems, but is not able to explain the reason in a simple way	The student finds some arguments to support his decisions, but they are not entirely rigorous and varied	The student is capable to proposing situations where the same solutions could be applied, but his arguments are not logical	The student mentions some factors, but does not mention the most important ones
High	The student anticipates the appearance of problems and is able to explain in a simple way the reason why it occurs	The student proposes several rigorous sources that support his arguments	The student rigorously proposes the application of solutions to problem in other areas	The student mentions all the important factors that would affect the implementation of his solutions in a real-life environment