AHP-BASED SOFTWARE FOR STUDENTS' SELF-ASSESSMENT OF CRITICAL THINKING CAPACITY

Jorge Salas[1], Leonardo Sierra[2], Víctor Yepes[3]

[1]Universitat Politècnica de València (SPAIN)
[2]Universidad de La Frontera (CHILE)
[3]ICITECH, Universitat Politècnica de València (SPAIN)

Abstract

Students' capacity for eliciting valid and accurate assessment is crucial for their self-assessment skill. Despite the central role that self-assessment plays in current conceptions of formative and classroom assessment, little attention has been paid for measuring the students' ability for carrying out valid and accurate assessment. This ability, connected with critical thinking, can be evaluated by comparing a students' assessment to performance on tests. Recently, the consistency index has been proposed as a metric for measuring students' critical thinking capacity, affording the observation of poor performance in the judgment capacity of a sample of 23 civil engineering students. This stresses the need of improving this aspect of students' education, which is required not only as a transversal competence, but also for the attainment of effective self-assessment capacity.

In our study, we present an educational software that implements the Analytic Hierarchy Process (AHP) method for providing students with self-assessment of their ability to issue consistent judgment, as well as for training their critical thinking capacity. This way we aim at making students aware of eventual lacks in their skill for valid and consistent assessment, as well as at giving them the opportunity of improving this aspect through self-assessment. The educational software embodies a three-step process, in which the self-assessment takes place in the last stage: first, the student is asked to carry out a conventional teacher assessment based on a set of criteria, which are then weighted by him/her through AHP without consistency control. Second, students are asked to turn on the software's consistency index control, which will reveal eventual inconsistencies, and accordingly revise his/her judgments until they are acceptable. Finally, the student is invited to self-assess his/her capacity to issue consistent judgment by analysing differences between their assessment with (step 2) and without (step 1) the aid of the consistency control, to reflect upon the differences between consistent and inconsistent assessment, and to think of potential causes driving to inconsistent judgment. The whole process can be performed by the students in the presented software, which serves as a platform for effective self-feedback and training of their assessment accuracy and critical thinking capacity.

Keywords: Self-assessment, critical thinking capacity, Analytic Hierarchy Process, educational software, consistent judgment.

1 INTRODUCTION

To future engineers and architects, the acquisition of transversal skills such as critical thinking critical, systems thinking or self-awareness during their education is essential for their further development as fully proficient professionals [1]. The development of such skills, therefore, is an important issue that can be addressed via university curricula.

In recent years, the evaluation of the degree of acquisition of transversal competences has become a key objective for the Universidad de La Frontera. Competency assessment is, however, a complex task, since there is no consensus neither as to what skills should be evaluated [2], nor as to how to evaluate them [1]. In this sense, autonomous learning [3] and self-assessment [4] have emerged as an approach for increasing students' capacity for critical thinking. The reliability, however, of this approach is subject to the objectivity and reliability of the student in their assessments [5]. In the case of effective self-assessment, students should deconstruct an event and make a judgement reflecting their understanding of the situation [6], which makes of students' capacity for eliciting valid and accurate judgment doubly important since it is a crucial aspect for both their self-assessment skill, and for their critical thinking competence [7].

Students' ability for making valid and accurate assessment can be evaluated by comparing their assessment to performance on tests [8]. Recently, the consistency index has been proposed as a metric for measuring students' critical thinking capacity [1]. The consistency index measures judgments consistency, which is regarded an essential skill for critical thinking [7]. In a research carried out over a sample of 23 civil engineering students in Spain, Navarro et al. [1] observed poor performance as to their capacity for eliciting consistent judgment, which stresses the need of improving this aspect of students' education.

In our study, we employed an educational software, available online [9], that implements the Analytic Hierarchy Process (AHP) [10] method for affording students self-assessment of their ability to issue consistent judgment, as well as for training their critical thinking capacity. The software has been designed for raising students' awareness on their eventual lack of skill for issuing valid and consistent assessment [11], as well as for training them into this capacity. By improving students' capacity for making consistent judgment, we aim not only at improving students' self-assessment and critical thinking capacity, but also to contribute to other key transversal competences such as decision-making, envisaging alternative futures, and the identification of complex connections between different criteria [12, 13]. At the same time, students become acquainted with AHP, a multi-criteria decision-making technique commonly employed in the field of civil engineering for addressing a number of problems in civil engineering and urban planning [14-25]

2   METHODOLOGY

In this research, we have resorted to a self-testing approach for providing students with a feedback of their capacity for issuing consistent judgment. Self-testing is a self-assessment technique in which students check their performance against provided test items upon request by the teacher or a software [26]. In this research, an educational software is employed to carry out a three-step process, in which the students are exposed to a judgment problem involving the assessment of the relative importance of the set of criteria. Then to check-out and correct the consistency of their judgments, and finally to reflect upon the impact that their inconsistency in judgment had over the assessment result.

2.1   Evaluation of consistent judgment capacity

2.1.1   Assignment of preferences through criteria pair-wise comparison

For the evaluation of the students' capacity for producing consistent judgment, we resorted to the AHP methodology [10]. This method is applicable to complex multi-criteria decision problems, and it affords the weighting of the criteria by judging their relative importance. This judgment is articulated through the pair-wise comparison of all the criteria being assessed. In this comparison, the experts indicate how important they consider a criterion with respect to the other by means of a scale ranging from equal preference for criteria A and B to total preference for criterion A to total ever criterion B, being A and B the criteria being pair-wise compared. This scale is usually is represented by means of the Saaty's fundamental scale [10], a numerical discrete scale ranging from 1 to 9, where 1 means equal preference, and 9 means total preference for A over B.

In the latter case, total preference of A over B also entails total wont of preference of B over A. Numerically, this is represented by the reciprocal value of 9, i.e., by the value 1/9. Conversely, it could happen that the expert totally prefers B over A, a fact that the Saaty's fundamental scale does not reflect since it this scale, the assignment is in one direction alone. To allow that the can assign a B over A preference, it would be necessary that for each pairwise comparison, he is offered two scales: one for the A over B direction, and another for the reciprocal direction, i.e. B over A, which certainly adds complexity to the judgment process both for the doubling in the number of possible comparisons, as to their understanding.

To overcome the above problem, and following prior work [27, 28], the presented educational software expands the range of the assessment scale for also including the possible preference of B over A (Figure 1). This expanded scale is graded in 10 levels, where the extreme values (1 and 10) represent total preference of one criterion over the other where the midpoint (value of 5) corresponds to equal preference. As to the translation of these values into those of the Saaty's fundamental scale, a value of 5 in the expanded scale corresponds with a value of 1 in that of Saaty, while the values 1 and 10 correspond to Saaty's values of 9 and of 1/9 (reciprocal of 9) respectively. This way, once the pair-wise judgment has been produced by the expert, it is possible to follow up the original AHP formulation for the calculation of each criterion's relative preference.
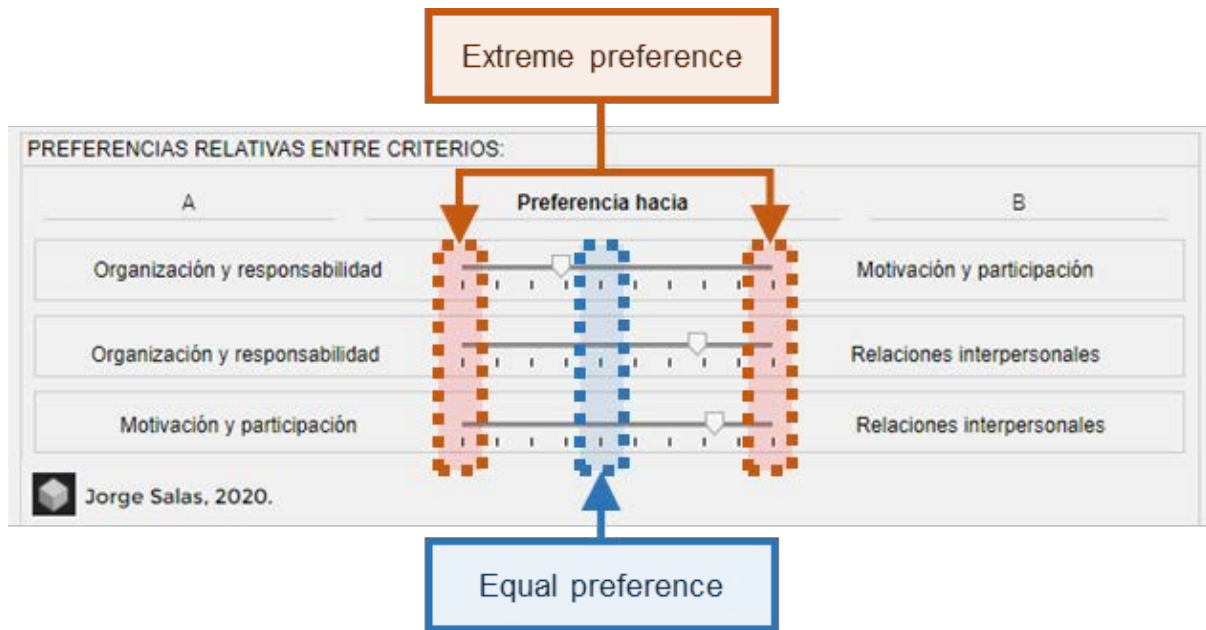
Figure 1. Implementation of the expanded pair-wise comparison scale into the software for the multi-criteria teacher multi-criteria assessment

2.1.2 Calculation of the criteria relative preferences

Once the values from the expanded scales have been translated in two those of Saaty's fundamental scale, AHP builds the A_{nxn} decision matrix, where "n" is the number of criteria whose relative importance is being assessed. The comparison matrix consists is a square matrix in which each a_{ij} element represents the numerical value "x" assigned in the comparison scale for the criteria "i" and "j". This matrix must be reciprocal, that is, if $a_{ij} = x$, then $a_{ij} = (1/x)$. From this matrix, the method affords the relevance of each criterion as the values of the eigenvector corresponding to the largest eigenvalue of the matrix (λ_{max}).

2.1.3 Evaluation of the judgment consistency

Saaty's method requires that the decision matrix be consistent, that is, that the pair-wise comparisons made by the expert for the assignment of relative preferences are consistent with each other. For example, in the case of a 3-criteria assessment, namely A, B and C, if A is preferred to B and B is preferred to C, C cannot be preferred to A. The AHP allows some inconsistency by measuring it and by providing a limit for discerning whether each comparison matrix, i.e. each set of pair-wise judgments, is acceptable or not. To provide a measure of the inconsistency in each set of judgments, AHP calculates the Consistency Index (CI) of each comparison matrix as:

$$CI = \frac{\text{Pricipal eigenvalue - size of matrix}}{\text{size of matrix - 1}} = \frac{\lambda_{max} - n}{n - 1}$$

Once calculated, the CI is compared to RI, a value derived by generating random reciprocal matrices of the same size, to give a consistency ratio (CR).

$$CR = \frac{CI}{RI}$$

Accordingly to Saaty [10], sets of judgments with CR lower than 10% can be regarded valid.

2.2 Self-testing of the consistent judgment capacity

To provide students with a feedback on their consistent judgment capacity, the presented software embodies a three-step process in which the student produce their judgment, then check-out their judgments consistency and correct them if necessary, and finally they compare the assessment results before and after the checking-out of their judgments' consistency (Figure 2).
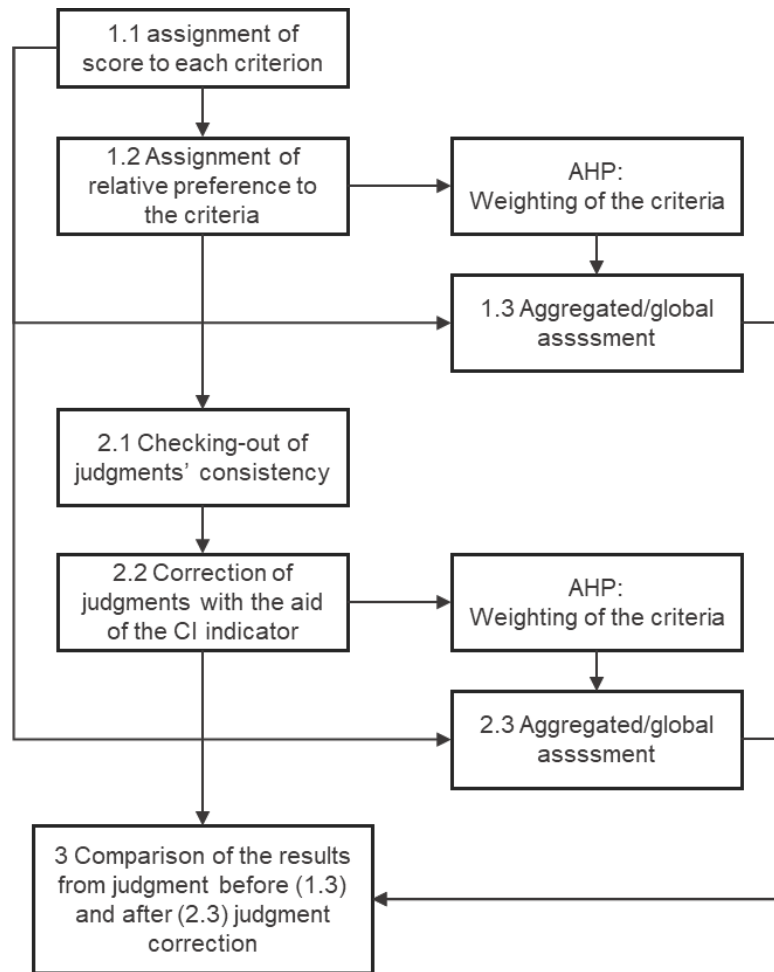
Figure 2. Flow-chart of the self-testing process of the consistent judgment capacity via comparison of judgments' results before and after making use of the consistency index control

For affording students self-testing of their consistent judgment capacity, the educational software allows students to turn on the consistency index control (CIC) once they have completed step 1.3 (Figure 2). By turning on this indicator, the software identifies the inconsistent sets of judgments, which the students can correct with the help of the CIC. The CIC is an indicator that provides them with real-time feedback of their judgments' consistency, so that students have immediate feedback of their corrections, which they have to perform until the set of judgments has become consistent (Figure 3).
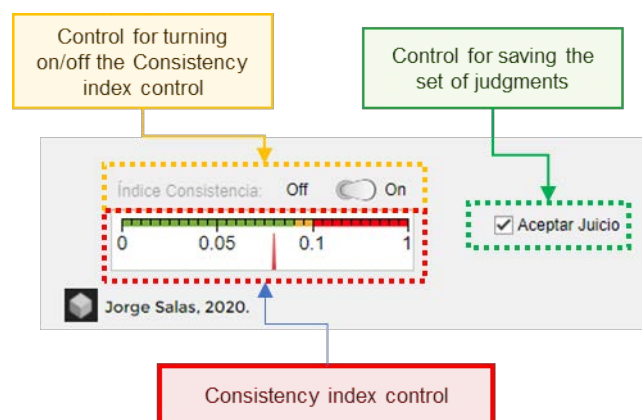

Figure 3. Consistency index control CIC implemented in the presented software. In step 1, judgments can be saved at any moment, while in step 2, judgments can only be saved if their CI is below 10%

Finally, the software displays the results of the assessment before and after introducing corrections in the original judgment in order to make them consistent. By comparing these results, the students become aware not only of their capacity for producing consistent judgment, which they experienced in the correction of their judgments in step 2.2, but also of the impact that inconsistency may have over the assessment results.

2.3 The judgment problem: Teacher multi-criteria assessment

As the judgment problem the students were exposed to, we selected the teacher assessment case due to students familiarity with this evaluation problem. This way, we avoided eventual inconsistency due to lack of expertise by the respondents in the domain being assessed. Since the research was carried out with the help of students from the Career of Engineering Construction at Universidad de La Frontera, we included the criteria usually employed in this university for teacher assessment.

Table 1, Criteria for teacher assessment, differences between criteria assessment before and after using the aid of the consistency-control

Assessment criteria being judged	HL (*)	Mean Score	MWS(**)		ACC-BCC (%)		
			BCC (***)	ACC (***)	Mean	Min	Max
Methodological Conditions:	1		0.473				
Up-to-date knowledge and clarity	2	7.525	0.79	1.09	27.5%		
Learning achieved	2	7.932	1.49	1.21	-23.1%		
Evaluations	2	5.678	0.29	0.45	35.6%		
Instrumental and Personal Conditions:	1		0.5266				
Comprehensive education	2	7.504	0.96	0.73	-31.5%		
Organization and responsibility	2	8.379	1.19	1.35	11.9%		
Motivation and participation	2	7.971	1.93	1.85	-4.3%		
Inter-personal relationships	2	7.256	1.03	0.96	-7.3%		
Global evaluation			8.11	8.11	2.8%	0.12%	8.6%

(*) HL: Hierarchical level

(**) MWS: Mean Weighted Score

(***) BCC: Before consistency-index control ; ACC: After consistency-index control

3 RESULTS

The proposed methodology was tested thanks to the participation of the degree of Construction Engineering, academic course 2020, at the Universidad de La Frontera, Chile. After the finishing of the workshop "SEMINARIO DE ESTUDIO INTERNACIONALES EN LA GESTION DE PROYECTOS DE CONSTRUCCION 2020: Introducción al Riesgo y la Incerteza en la Programación de Obras", the participants were asked to assess the teacher's assessment by using the presented software.

3.1 Assignment of scores and relative importance to the assessment criteria

After receiving basic instruction on how to employ the software, they first scored the teacher's performance in each criteria, and then judged the relative importance of each criterion (Figure 4).
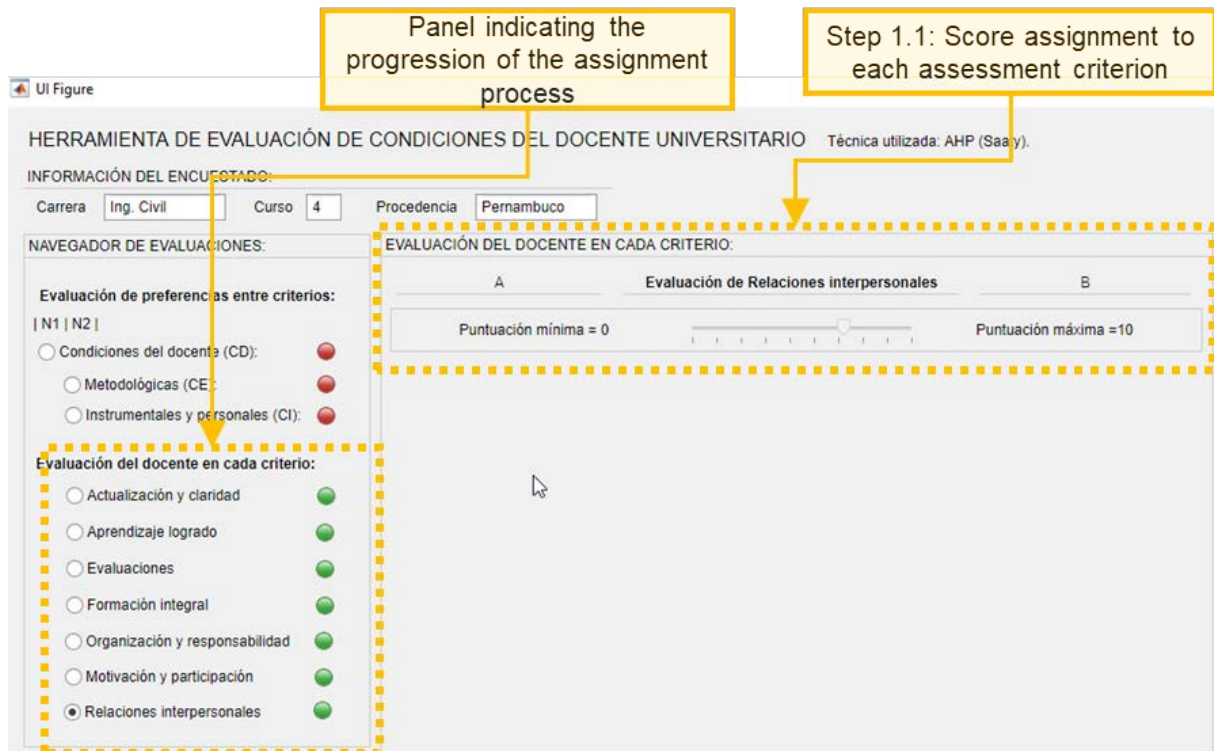
Figure 4. Process for the completion of the scoring of the teacher's performance accordingly to the assessment criteria

Then, the students judged the relative importance of the criteria corresponding to each of the groups of criteria, namely methodological criteria and instrumental and inter-personal criteria (Figure 5).
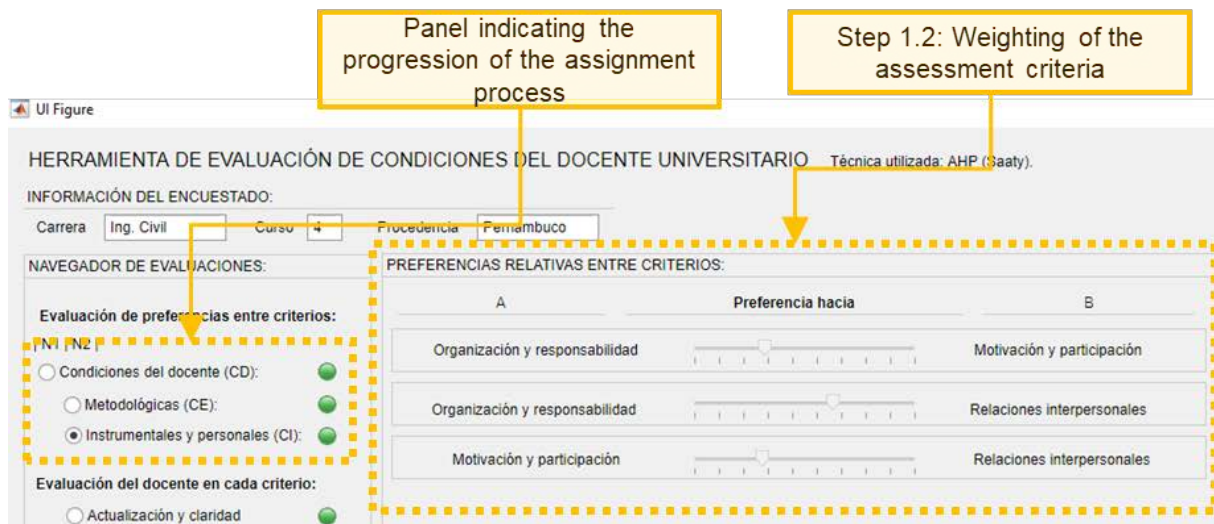

Figure 5. Assessment process of the relative importance of the assessment criteria

3.2   Checking out of judgments consistency and judgment correction

In this stage, students activated the consistency index control for checking out their performance as to their capabilities for consistent judgment, and corrected their inconsistent judgments until they became consistent. The educational software employed facilitated the identification of consistency problems, which the students were enabled to adjust in order to reduce the inconsistency down to acceptable level (Figure 6). The results table to the right of the software interface, kept the assessment results updated for providing the students with real-time feedback (Figure 6, Results table, P.R. CC column). These results were presented together with those of the assessment without the aid of the consistency

(Figure 6, Results table, P.R. SC column) in order to provide students with a reference of their initial assessment so that they can identify the impacts of the corrections over the global assessment and minimize them.

Figure 6. Checking out of judgment consistency, revision of the allocation of relative importance in inconsistent sets of judgments, and comparison of results from steps 1 and 2
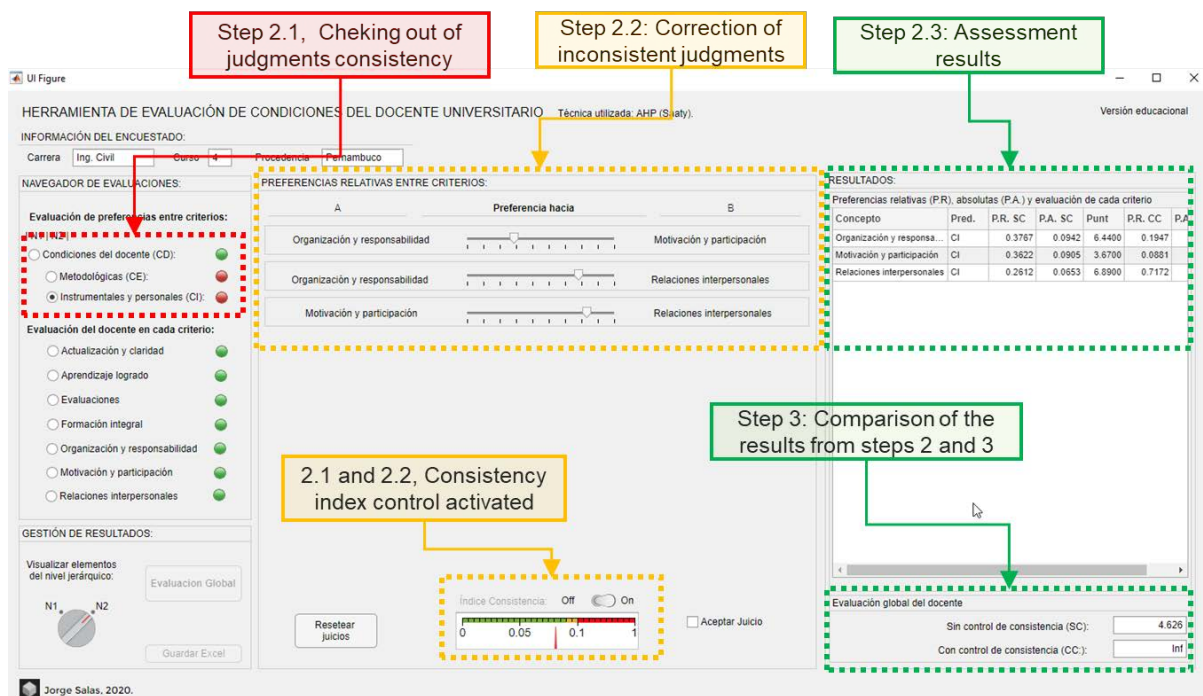
Table 1 shows that the mean deviation between the assessment of relative importance before and after the correction of judgments was quite heterogeneous by criteria and groups of criteria. While in the methodological criteria the differences are uniform, ranging from 23.1% to 35.6%, in the case of the instrumental and inter-personal criteria the differences ranges from 4.3% to 31.5%. This means that for the reduction of inconsistency to an acceptable level, the students needed to introduce more changes in the methodological group of criteria than in the instrumental, and more changes in the comprehensive education criterion than in the other criteria of the instrumental and inter-personal group of criteria.

The heterogeneous distribution of changes in the correction of judgments of the instrumental and inter-personal group of criteria (Table 1), also suggests either a selective, rather than a random behaviour of students as to what criterion should be corrected.

The differences between the relative importance assigned by the students before and after judgment correction in the case of the methodological group of criteria (Table 1), are indicative of the importance of the corrections they had to work out in order to make their judgments acceptable, and therefore of their practicing on multi-criteria assessment.

Finally, the great differences between the relative importance assigned by the students before and after judgment correction in the case of the methodological group of criteria, compared to those of the instrumental and inter-personal group of criteria, reflects the special difficulty that judgments with high preference for some criterion entails for the issuing of consistent judgment.

3.3   Comparison of the results before and after judgment correction

As to the differences between the teacher global assessments carried out by the students before and after judgment correction, Figure 7 shows that 30% of the students initial assessment was very close to the corrected assessment, which indicates that they had to introduce little correction into their judgment. This idea is consistent with the conclusion reached by Navarro et al. [1], who in their research on students consistency, found out that nearly 30% of then attained a consistency ratio below 15%, which they assumed to be acceptable for the purpose of their research.

On the other hand, it is noteworthy the range of the variation in teacher's global evaluation, ranging from 0.12% up to 8.59%, which gives an idea of the different capacity shown by the students in the elicitation of consistent judgment. This suggest the idea of the existence of factors influencing the students skill for consistent judgment. However, the lack of detailed information on the students' profile, and the limited number of observations, prevented us from carrying out a deeper analysis of the results, which nonetheless we expect to address through future research.
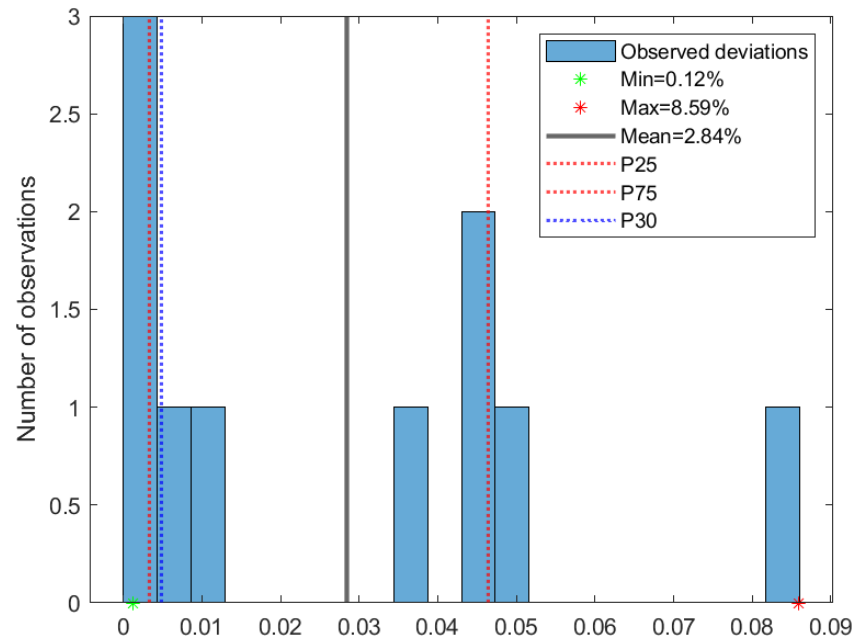

Figure 7. Histogram of the deviations of initial global assessment from corrected global assessment

4   CONCLUSIONS

Transversal skills such as critical thinking critical are essential in the education of future engineers and architects. The development of such skills, therefore, is an important issue to be addressed via university curricula. In this sense, autonomous learning and self-assessment have emerged as an approach for increasing students' capacity for critical thinking. This self-assessment, however, relies in the students' capacity for producing consistent judgment, and skill, which, at the same time, is also a driver of critical thinking capacity, and therefore doubly important for students' development. Notwithstanding the importance of this skill, there are but few studies devoted to its specific measuring and promotion. In this work, we present a software implementing a methodology for providing students with self-testing of their capacity for issuing consistent judgment, as well as for training them into this practice through the correction of their judgments until these become consistent. By exposing students to the problem of assessing their teacher's performance accordingly to a set of criteria, the presented software allows measuring students' skill for consistent judgment in terms of the impact produced by changes the students had to introduce in their own initial judgments to make them consistent. In doing so, the students on the one hand become aware of their eventual lack of skill, and on the other hand, they are trained into the practice of making consistent judgment. To this end, the software implements the consistency index indicator, which provides students with real-time evaluation of their judgments' consistency. The results of this study show that there is an important heterogeneity in the results attained by the students, both in terms of change in the assignment of relative preference as in terms of impact over global assessment. As to this latter, the proportion of students whose initial assessment was close to the corrected is near 30%, which is consistent with prior studies. Finally, by using this software the students became familiar with AHP, a qualitative multi-criteria decision-making approach widely employed in the field of civil engineering.

ACKNOWLEDGEMENTS

REFERENCES

[1] I.J. Navarro, V. Yepes, J.V. Martí, "Comparación pareada como método de evaluación de competencias transversales en materia de sostenibilidad", *VI Congreso de Innovación Educativa y Docencia en Red. IN-RED 2020*, pp. 670-679, 2020. http://dx.doi.org/10.4995/INRED2020.2020.12000

[2] M. L. Cruz, G. Saunders-Smits, P. Grone, "Evaluation of competency methods in engineering education: a systematic review", *European Journal of Engineering Education*, 2019. http://dx.doi.org/10.1080/03043797.2019.1671810.

[3] A. G. Ljungman, C. Silén, "Examination involving students as peer examiners". *Assessment & Evaluation in Higher Education*, vol. 33, no 3, pp. 289-300, 2008.

[4] A. G. Carter, D. K. Creedy, and M. Sidebotham, "Critical thinking skills in midwifery practice: Development of a self-assessment tool for students", *Midwifery*, vol. 50, pp. 184–192, 2017.

[5] K. Pond, "Student Experiences of Peer Review Marking of Team Projects", *International Journal of Management Education*, vol. 6, no. 1, pp. 30-43, 2007.

[6] W. D. Hendricson, S. C. Andrieu, D. G. Chadwick, J. E. Chmar, J. R. Cole, M. C. George, G. N. Glickman, J. F. Glover, J. S. Goldberg, N. K. Haden, C. Meyerowitz, L. Neumann, M. Pyle, L. A. Tedesco, R. W. Valachovic, R. G. Weaver, R. L. Winder, S. K. Young, and K. L. Kalkwarf, "Educational strategies associated with development of problem-solving, critical thinking, and self-directed learning.," *Journal of dental education*, vol. 70, no. 9, pp. 925–936, 2006.

[7] M. Kallet, *Think smarter: Critical thinking to improve problem-solving and decision-making skills*. Wiley, 2014.

[8] G. T. L. Brown, H. L. Andrade, and F. Chen, "Accuracy in student self-assessment: directions and cautions for research," *Assessment in Education: Principles, Policy and Practice*, vol. 22, no. 4, pp. 444–457, 2015.

[9] mgnesio, Educational Software, Accessed 28 December, 2020, Retrieved from https://www.mgnesio.es/servicios/software-educativo

[10] T. L. Saaty, "HOW TO MAKE A DECISION - THE ANALYTIC HIERARCHY PROCESS," *European Journal of Operational Research*, vol. 48, no. 1, pp. 9–26, 1990.

[11] C. T. McKevitt, "Engaging students with self-assessment and tutor feedback to improve performance and support assessment capacity," *Journal of University Teaching and Learning Practice*, vol. 13, no. 1, 2016.

[12] G. Cebrián, M. Junyent, "Competencies in Education for Sustainable Development: Exploring the Student Teachers' Views", *Sustainability*, vol. 7, no. 3, p. 2768-2786, 2015.

[13] W. Lambrechts, I. Mulà, K. Ceulemans, I. Molderez, V. Gaeremynck, "The integration of competences for sustainable development in higher education: an analysis of bachelor programs in management", *Journal of Cleaner Production*, vol. 48, p. 65-73, 2013.

[14] J. Salas, V. Yepes, "Enhancing Sustainability and Resilience through Multi-Level Infrastructure Planning," *International Journal of Environmental Research and Public Health*, vol. 17, no. 3, p. 962, 2020.

[15] J. Salas, V. Yepes, "MS-ReRO and D-ROSE methods: Assessing relational uncertainty and evaluating scenarios' risks and opportunities on multi-scale infrastructure systems," *Journal of Cleaner Production*, vol. 216, pp. 607-623, 2019.

[16] I. J. Navarro, V. Penadés-Plà, D. Martínez-Muñoz, R. Rempling, V. Yepes, "Life cycle sustainability assessment for multi-criteria decision making in bridge design: a review", *Journal of Civil Engineering and Management*, vol. 26, no. 7, pp. 690-704, 2020.

[17] I. J. Navarro, V. Yepes, J. V Martí, "A Review of Multicriteria Assessment Techniques Applied to Sustainable Infrastructure Design", *Advances in Civil Engineering*, vol. 2019, p. 6134803, 2019.

[18] E. Pellicer, L. A. Sierra, V. Yepes, "Appraisal of infrastructure sustainability by graduate students using an active-learning method," *Journal of Cleaner Production*, vol. 113, pp. 884–896, 2016.

[19] L. A. Sierra, E. Pellicer, V. Yepes, "Social Sustainability in the Lifecycle of Chilean Public Infrastructure," *Journal of Construction Engineering and Management*, vol. 142, no. 5, p. 5015020, 2016.

[20] T. García-Segura, V. Yepes, D.M. Frangopol, "Multi-Objective Design of Post-Tensioned Concrete Road Bridges Using Artificial Neural Networks", *Structural and Multidisciplinary Optimization*, vol. 56, no. 1, pp. 139-150, 2017.

[21] V. Penadés-Plà, J.V. Martí, T. García-Segura, V. Yepes, "Life-cycle assessment: A comparison between two optimal post-tensioned concrete box-girder road bridges", *Sustainability*, vol. 9, no. 10, p. 1864, 2017.

[22] I. Payá-Zaforteza, V. Yepes, F. González-Vidosa, A. Hospitaler, "On the Weibull cost estimation of building frames designed by simulated annealing", *Meccanica*, vol. 45, no. 5, pp. 693-704, 2010.

[23] C. Torres-Machi, A. Chamorro, E. Pellicer, V. Yepes, C. Videla, "Sustainable Pavement Management Integrating Economic, Technical, and Environmental Aspects in Decision Making", *Transportation Research Record*, vol. 2523, pp. 56-63, 2015.

[24] L.A. Sierra, E. Pellicer, V. Yepes, "Method for estimating the social sustainability of infrastructure projects", *Environmental Impact Assessment Review*, vol. 65, pp. 41-53, 2017.

[25] [1] J. Salas and V. Yepes, "Urban vulnerability assessment: Advances from the strategic planning outlook," *Journal of Cleaner Production*, vol. 179, pp. 544–558, Apr. 2018.

[26] E. Panadero, G. T. L. Brown, and J.-W. Strijbos, "The Future of Student Self-Assessment: a Review of Known Unknowns and Potential Directions," *Educational Psychology Review*, vol. 28, no. 4, pp. 803–830, 2016.

[27] J. Salas, V. Yepes, "A discursive, many-objective approach for selecting more-evolved urban vulnerability assessment models," *Journal of Cleaner Production*, vol. 176, pp. 1231–1244, 2018.

[28] J. Salas, V. Yepes, "VisualUVAM: A Decision Support System Addressing the Curse of Dimensionality for the Multi-Scale Assessment of Urban Vulnerability in Spain", *Sustainability*, vol. 11, no. 8, p. 2191, 2019.