

EDUCACIÓN, CREATIVIDAD E INTELIGENCIA ARTIFICIAL: NUEVOS HORIZONTES PARA EL APRENDIZAJE. ACTAS DEL VIII CONGRESO INTERNACIONAL SOBRE APRENDIZAJE, INNOVACIÓN Y COOPERACIÓN, CINAIC 2025

María Luisa Sein-Echaluce Lacleta, Ángel Fidalgo Blanco y Francisco José García Peñalvo (coords.)

1ª Edición. Zaragoza, 2025

Edita: Servicio de Publicaciones. Universidad de Zaragoza.



Servicio de
Publicaciones
Universidad Zaragoza

EBOOK ISBN 978-84-10169-60-9

DOI 10.26754/uz.978-84-10169-60-9



Esta obra se encuentra bajo una licencia Creative Commons Reconocimiento – NoComercial (ccBY-NC). Ver descripción de esta licencia en <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Referencia a esta obra:

Sein-Echaluce Lacleta, M.L., Fidalgo Blanco, A. & García-Peñalvo, F.J. (coords.) (2025). *Educación, Creatividad e Inteligencia Artificial: nuevos horizontes para el Aprendizaje. Actas del VIII Congreso Internacional sobre Aprendizaje, Innovación y Cooperación. CINAIC 2025 (11-13 de Junio de 2025, Madrid, España)*. Zaragoza. Servicio de Publicaciones Universidad de Zaragoza. DOI 10.26754/uz.978-84-10169-60-9

Evaluación Continua en Bolonia: la Autoevaluación en la enseñanza de Programación y sus efectos en las Calificaciones de los Exámenes

Continuous Evaluations in Bologna: Self-Assessment in Programming Education and its effects on Exam Scores

Javier Rodríguez-Vidal, Francisco Javier Del Álamo, Ascensión López-Vargas, Pablo Manuel Vigara Gallego, Ángel García-Beltrán

[javier.rodriguez.vidal, franciscojavier.delalamo, a.lvargas, pm.vigara, angel.garcia]@upm.es

Departamento de Automática, Ingeniería Eléctrica y Electrónica e
Informática Industrial
Escuela Técnica Superior de Ingenieros Industriales, UPM
Madrid, España.

Abstract- The Bologna process has revolutionized university studies by promoting continuous assessment throughout the academic period. However, this shift has increased the workload for teaching teams, making it essential to employ automatic tools for evaluating student assignments. In computer science education, the use of automatic code assessment tools (e.g., Java, C, C++) has proven particularly effective. This study investigates the correlation between self-assessment exercises in C programming and improvements in exam performance. Data from self-assessment exercises conducted on the AulaWeb platform for the Grado en Ingeniería de Organización, Grado en Ingeniería Química and Grado en Ingeniería en Tecnologías Industriales at the Universidad Politécnica de Madrid over the past three years were analyzed, involving of 2413 students. Key findings indicate that include that self-assessment exercises significantly influence final grades in the first call exams. However, varying correlations were observed for the second call exam, with some students neglecting complex programming problems.

Keywords: *Self-assessment exercises, continuous evaluation, C programming, exam performance.*

Resumen- Bolonia ha revolucionado los estudios universitarios al promover evaluaciones continuas a lo largo del período académico. Sin embargo, esto ha incrementado la carga de trabajo de los equipos docentes, lo que hace necesario el uso de herramientas automáticas para corregir tareas. En el ámbito de la informática, el uso de evaluadores automáticos de código (por ejemplo, Java, C, C++) se considera ideal. Este estudio investiga la relación entre los ejercicios de autoevaluación en C y la mejora del rendimiento en los exámenes. Se analizaron ejercicios de autoevaluación realizados en la plataforma AulaWeb para Grado en Ingeniería de Organización, Grado en Ingeniería Química y Grado en Ingeniería en Tecnologías Industriales en la Universidad Politécnica de Madrid durante los últimos tres años. Un total de 2413 estudiantes completaron estas tareas. Los hallazgos clave incluyen la influencia de los ejercicios de autoevaluación en la calificación del primer examen y diversas correlaciones con las notas del segundo, con casos de estudiantes que evitaron problemas de programación complejos.

Palabras clave: *Ejercicios de autoevaluación, evaluación continua, programación en C, calificaciones de exámenes.*

1. INTRODUCTION

The European Higher Education Process (Bologna process) (*European Commission, n.d.-a*) introduced a paradigm shift in university education emphasizing continuous assessment activities that monitor student progress throughout the academic term. Unlike traditional models that primarily rely on final exams, continuous assessment facilitates the gradual assimilation of theoretical and practical concepts in class.

Despite its pedagogical benefits, continuous assessment increases the burden on teaching staff requiring them to design, administer and evaluate additional assignments. Given the large number of students and limited teaching resources, manual assignment is often impractical. Moreover, delayed feedback can diminish the pedagogical effectiveness of these activities, especially given the relatively short academic terms.

To address these challenges, self-assessment strategies have emerged, enabling students to compare their performance against predefined evaluation criteria (Nieminen et al., 2021). In computer science education, self-assessment involves evaluating student's programming solutions predefined problems. Several studies have explored the impact of self-assessment in programming courses. For instance, Baruque et al. (2015) and García-Beltrán et al. (2006) present various platforms for self-assessment in programming. Cedazo et al. (2015) examine the effectiveness of self-assessment exercises over four academic years, while Chung & Hsiao (2020) investigates how these exercises influence student motivation and academic performance.

Given the complexity of implementing effective self-assessment systems for programming education, this study aims to examine how self-assessment results correlate with students' final exam performance. The research covers three academic

years and multiple degree programs, analyzing the role of self-assessment in student success.

This study analyzes the relationship between the results obtained in self-assessment exercises, completed by students throughout an academic year (from September to December or February to June), and their final exam grades (in January, June, or July) in a university programming course. To achieve this, data from multiple academic years of the same subject across three different degree programs will be used.

The remainder of the article is organized as follows. Section 2 introduces key concepts relevant to understanding this study, including: (i) the subject under study, (ii) the platform used for self-assessment, (iii) the final exams, (iv) data collection, (v) experimental design, (vi) preliminary analysis, and (vii) evaluation metrics. Section 3 presents and discusses the results obtained. Finally, Section 4 outlines the main conclusions of the study and suggests future research directions.

2. CONTEXT & DESCRIPTION

In this section, we describe the course under study, the self-assessment platform, the way to perform the final exams and the methodology.

A. Fundamentals of Programming Course overview

The study focuses on the course FP (Fundamentos de Programación or Fundamentals of Programming), which is part of the first-year curriculum for students in Grado en Ingeniería de Organización (GIO), Grado en Ingeniería Química (GIQ), and the second semester of Grado en Ingeniería en Tecnologías Industriales (GITI) at the Escuela Técnica Superior de Ingenieros Industriales (ETSII) of the Universidad Politécnica de Madrid (UPM). The course enrolls approximately 77, 80, and 655 students per year, respectively, and carries six ECTS (*European Credit Transfer and Accumulation System* (*European Commission, n.d.-b*)) credits.

B. AulaWeb Self-assessment platform

The AulaWeb platform, developed by UPM Computer Science Laboratory, has been in use since 1999 (García-Beltrán & Martínez, 2006). It serves as a comprehensive e-learning system, facilitating activities such as content distribution, assignment submission, virtual tutorials, and self-assessment exercises. AulaWeb includes a self-assessment module designed to support the creation of C programming exercises. This module offers teachers a question manager and a configuration system, allowing them to set up exercises and store them in the database. It also provides an exercise manager that presents the questions to students and records their responses. Once logged into the platform, students can access the exercises configured by their instructors and submit their answers. The system then automatically evaluates their submissions by comparing the students' results with the expected outputs defined by the teachers, providing immediate feedback by highlighting correct and incorrect answers. The platform is hosted on Windows Server 2016 with IIS and Microsoft SQL Server 2019, using ASP.NET and Java for development.

C. Final Exams

Final exams take place in two sessions: January or June (first call) and July (second call). Exams are conducted in person,

except for the 2019/20 academic year when COVID-19 restrictions required online examinations. Students must complete programming tasks using C, either in a digital format (GIO & GIQ) or handwritten (GITI). Exam structures vary by program, with different grading weights assigned to self-assessment-related questions and programming problems.

a) GIO & GIQ: one S-A question and two large problems, with a weight of 20% of the exam grade for the S-A and the short problem, and 40% for each one of the long problems and

b) GITI where the exam involves 10 questions and each question can only be evaluated with a binary mark (0 or 1).

For the three grades, the total score is an integer between 0 and 10. Finally, the students pass the course in two cases: (i) if the total score obtained is greater than or equal to 5; (ii) if the score achieved in the first call is greater than or equal to 4, and the students satisfactorily completed the self-assessment exercises.

D. Data Collection

Data for this study include (i) self-assessment results and (ii) final exam grades, both collected through the AulaWeb platform. Exam grading records were retrieved from the Teaching Unit responsible for the FP course.

E. Experimental Design

During the course, self-assessment exercises were available to be completed for a certain period, usually 10 days per exercise. To complete these assignments, students needed to have a computer and an Internet connection that would allow them to connect to the website. The experiment was carried out on all students enrolled in the course (see Table 1 for the number of students per year). When the correlations were extracted, those students who were unqualified in each exam session were eliminated.

Table 1 S-A performance according to each academic year

G.	AY.	Stud.	One S-A	Half S-A	All S-A	Pass S-A
GITI	19/ 20	665	591 (88.87 %)	562 (84.51 %)	411 (61.80 %)	541 (81.35 %)
	20/ 21	610	554 (90.82 %)	467 (76.56 %)	247 (40.49 %)	521 (85.41 %)
	21/ 22	623	525 (84.27 %)	387 (62.12 %)	188 (30.18 %)	472 (75.76 %)
GIO	19/ 20	88	74 (84.09 %)	68 (77.27 %)	58 (65.91 %)	66 (75.00 %)
	20/ 21	82	64 (78.05 %)	58 (70.73 %)	56 (68.29 %)	63 (76.83 %)
	21/ 22	85	75 (88.24 %)	70 (82.35 %)	48 (56.47 %)	68 (80.00 %)
GIQ	19/ 20	72	64 (88.89 %)	56 (77.78 %)	51 (70.83 %)	57 (79.17 %)
	20/ 21	91	86 (88.66 %)	77 (79.38 %)	71 (73.20 %)	80 (82.47 %)
	21/ 22	91	81 (89.01 %)	69 (75.82 %)	58 (63.74 %)	68 (74.73 %)

F. Preliminary Analysis

This section presents an initial evaluation of students' academic performance in the FP course over three academic years (2019/20 to 2021/22) across the GITI, GIO, and GIQ programs.

Table 1 summarizes students' engagement with self-assessment exercises. The first column lists the program (GITI, GIO, or GIQ), followed by the academic year, the total number of students enrolled, and percentages of students who submitted at least one, half, or all self-assessment exercises. Additionally, the final column presents the percentage of students achieving at least 5 out of 10 in these exercises.

Findings reveal that while most students attempt at least one self-assessment exercise, engagement declines over time. The highest drop-off rates were observed in GITI (22.15% in 2020/21), GIO (7.32% in 2020/21), and GIQ (13.19% in 2021/22) by the midpoint of the assessment period. By the end of the academic year, participation dropped by as much as 54.09% (GITI), 31.77% (GIO), and 25.27% (GIQ). Moreover, not all students who engaged in self-assessment achieved passing scores, with pass rate discrepancies ranging from 8.51% to 1.22% across different programs.

These preliminary results highlight a significant decline in engagement over time, raising concerns about the long-term effectiveness of self-assessment exercises.

3. RESULTS & DISCUSSION

Tables 2 and 3 summarize student performance across both exam calls. The first column specifies the grade (GITI, GIO or GIQ), followed by the academic year, mean exam score, variance, standard deviation, skewness and kurtosis. Table 2 indicates that average first-call exam scores were consistently below the passing threshold across all years. The highest scores were recorded in the GITI program, while GIO and GIQ displayed peak performance in the 2020/21 academic year. Most score distributions exhibited right-skewed asymmetry, reflecting overall poor performance.

Table 3 reveals that second-call exam scores were even lower, both in mean values and distribution ranges. Similar to the first-call results, scores were concentrated around the mean, with most distributions skewed to the right.

These findings confirm that self-assessment exercises play a more significant role in first-call exam preparation than in second-call exams. One possible explanation is that first-call students have less time to prepare, making self-assessment exercises crucial for consolidating knowledge. Conversely, second-call students tend to focus on alternative study methods rather than repeating self-assessment exercises, despite the presence of self-assessment questions in the GIO and GIQ exams.

Despite the pedagogical benefits of self-assessment exercises, which encourage students to take an active role in their own learning, this methodology also presents certain challenges. The immediate feedback students receive on their correct and incorrect answers makes this approach particularly valuable for the early detection and correction of misinterpretations. However, as the results show, there is a steady decline in student participation in these activities as the course progresses. This reduced engagement limits their ability to identify conceptual gaps in time, which ultimately affects their performance, as reflected in the final exam results. Additionally, differences in exam formats across degree programs seem to lead to higher participation among GIO and GIQ students. However, as shown in Tables 2 and 3, these self-assessment exercises still do not receive much attention from

students when preparing for the course, limiting their potential impact.

Table 2 Statistics for the 1st call (January GIO & GIQ; June-GITI)

G.	AY.	Av.	Vr.	SD.	Sk.	Kt.
GITI	19/20	3.99	8.49	2.91	0.24	-0.97
	20/21	4.23	7.42	2.72	-0.02	-0.98
	21/22	3.82	5.78	2.40	0.22	-0.70
GIO	19/20	3.19	4.56	2.13	0.67	-0.12
	20/21	3.83	9.49	3.08	0.49	-0.96
	21/22	4.84	5.68	2.38	-0.50	-0.63
GIQ	19/20	2.72	5.53	2.35	0.80	-0.03
	20/21	4.13	7.65	2.77	0.29	-1.03
	21/22	4.39	6.23	2.50	-0.02	-0.70

Table 3 Statistics for the 2nd call (June GIO & GIQ; July-GITI)

G.	AY.	Av.	Vr.	SD.	Sk.	Kt.
GITI	19/20	2.46	5.27	2.30	0.69	-0.53
	20/21	2.09	3.51	1.87	0.76	-0.18
	21/22	3.65	4.52	2.13	-0.02	-0.97
GIO	19/20	3.96	7.02	2.65	0.63	-0.53
	20/21	3.20	3.21	1.79	-0.07	-1.10
	21/22	4.11	3.67	1.19	-0.07	-0.62
GIQ	19/20	2.56	4.36	2.09	0.37	-0.78
	20/21	3.25	3.23	1.80	0.52	-0.63
	21/22	3.10	4.25	2.06	0.27	-1.10

4. CONCLUSIONS

This study examined the relationship between self-assessment exercises and final exam performance in the FP course, covering three programs (GITI, GIO, and GIQ) over three academic years. The key findings include:

- The use of self-assessment throughout the course is essential for two main reasons: it helps reduce the burden of teaching workload for instructors, and it provides students with immediate feedback on their learning, allowing them to take an active role in their own educational process.
- The results from the first final exam suggest more influence of self-assessment exercises on students' preparation, likely due to the short interval between the end of classes and the exam date, which encourages the use of short-term evaluation tools. In contrast, when preparing for the second final exam, students tend to focus on practicing and repeating problem-solving tasks rather than consolidating the concepts they did not fully acquire in the first exam. These concepts, however, remain essential for the second assessment and are more effectively reinforced through

self-assessment activities. This outcome is particularly noteworthy in the GIO and GIQ degree programs, where such types of exercises are explicitly included in the exam format.

Despite these insights, the study has limitations. The findings are specific to the FP course in the grades of GITI, GIO and GIQ of the ETSII-UPM and may not generalize to other subjects, particularly those outside STEM fields, i.e. arts, where is impossible to evaluate automatically a sculptural work.

This work aligns with the United Nations Sustainable Development Goals (SDG) (United Nations, n.d.): SDG4-Quality Education.

5. FUTURE WORK

According to the results there is a need to better understand the differences between student groups in order to improve FP subject accordingly.

From the point of view of the software development, the goal is to enhance AulaWeb's self-assessment module by introducing personalized exercise recommendations and automatic selection based on each student's learning progress, as well as historical performance data from previous cohorts. Additionally, there are plans to improve the way results are presented to students by incorporating Generative Artificial Intelligence (GAI) models trained on feedback provided by the teachers on exam exercises. These models would offer students quick and insightful explanations of their achieved results.

On the other hand, from the point of view of theoretical studies, there is a need to better understand the profile of students taking the FP course. The idea is to explore differences between repeat and first-year students, both in how they engage with the self-assessment exercises and in the outcomes of their final exams. Also, another topic of interest is to study the use of self-assessment exercises in other subject areas.

ACKNOWLEDGEMENTS

This work was supported by the project: "Hackathon educativo basado en escape rooms inversos como actividad de gamificación motivadora para el aprendizaje de la programación de ordenadores" (IE25.0504) through the Universidad Politécnica de Madrid, Proyectos de Innovación Educativa 2024-2025 program.

REFERENCES

Baruque, B., & Herrero, Á. (2015). Self-assessment web tool for Java programming. *International Joint Conference:*

CISIS'15 and ICEUTE'15, 583–592. https://doi.org/10.1007/978-3-319-19713-5_51

Cedazo, R., Garcia Cena, C. E., & Al-Hadithi, B. M. (2015). A friendly online C compiler to improve programming skills based on student self-assessment. *Computer Applications in Engineering Education*, 23(6), 887–896. <https://doi.org/10.1002/cae.21660>

Chung, C. Y., & Hsiao, I. H. (2020). Investigating patterns of study persistence on self-assessment platform of programming problem-solving. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 162–168. <https://doi.org/10.1145/3328778.3366827>

European Commission. (n.d.-a). *Bologna Process*. European Education Area. Retrieved April 14, 2025, from <https://education.ec.europa.eu/education-levels/higher-education/inclusive-and-connected-higher-education/bologna-process>

European Commission. (n.d.-b). *European Credit Transfer and Accumulation System*. European Education Area. Retrieved April 14, 2025, from <https://education.ec.europa.eu/education-levels/higher-education/inclusive-and-connected-higher-education/european-credit-transfer-and-accumulation-system>

García-Beltrán, A., & Martínez, R. (2006). Web assisted self-assessment in computer programming learning using Aulaweb. *International Journal of Engineering Education*, 22(5), 1063–1069.

García-Beltrán, A., Martínez, R., Jaén, J. A., & Tapia, S. (2006). La autoevaluación como actividad docente en entornos virtuales de aprendizaje/enseñanza. *Revista de Educación a Distancia (RED)*, 2006, 1–14. <http://dx.doi.org/10.6018/red/50/14>

Nieminen, J. H., Asikainen, H., & Rämö, J. (2021). Promoting deep approach to learning and self-efficacy by changing the purpose of self-assessment: A comparison of summative and formative models. *Studies in Higher Education*, 46(7), 1296–1311. <https://doi.org/10.1080/03075079.2019.1688282>

United Nations. (n.d.). *The 17 goals*. United Nations Sustainable Development Goals. Retrieved April 14, 2025, from <https://sdgs.un.org/goals>