Project-Based Learning for Assessing a Course on Computer Programming

Pedro J. Roig, Salvador Alcaraz, Katja Gilly, Cristina Bernad, Carlos Juiz
Miguel Hernández University
University of the Balearic Islands
Spain

Abstract

Project-Based Learning is a teaching method within the active learning paradigm where students are faced with complex problems in order for them to acquire the appropriate conceptual, procedural, and attitudinal competences. In this paper, we present the outcome of a group activity aimed at deploy a scenario made of a combination of logic gates coded in the Java programming language. Afterwards, a special session was set up in order for each group to make a pitch presentation so as to display what they did and how they did it. Each of those presentations was in turn assessed by all students with a specific questionnaire, whereas the level of engagement of the students while carrying on that activity was measured by means of a further construct. The results obtained showed an increase of academic performance of 11% and a increase of success rate of 10%, along with a high level of engagement.

1. Introduction

Active learning is an education paradigm which is usually viewed as a toolset to transform education, specially in the STEM domain [1]. Likewise, it is commonly seen as an educative framework leading to rise evaluation performance, while decreasing failure rates compared to traditional learning [2].

There are many evidences cited in the literature linking the increase in student interactions and engagement levels with gains in both learning and motivation [3], along with a positive impact in students' well-being and competence acquisition [4].

Different options are available to implement active learning, such as serious educational games, flipped classrooms or educational escape rooms [5]. However, project-based learning is an interesting option, as it also promotes the development of communication and teamworking skills [6].

In this sense, those features are essential when it comes to designing the skillset needed for students to adapt to the ever-changing technological world we are living in [7]. Actually, digital competence is one of the most sought-after feature in the current labor market, as it allows to successfully embrace change and being able to adapt effectively in response [8].

Digital competence include the knowledge, skills, and attitude to get by with technology in five key areas, such as managing information, communicating with others, creating content, keeping safe and secure, as well as problem solving [9]. Hence, digital literacy can be defined as an emerging and increasing priority when it comes to the education field, as digital competence has become a game changer for education in recent years [10].

Digital awareness could be introduced to students in different ways, such as screening digital resources in order to get the most convenient according to the learning objectives expected, boosting collaboration among students, offering innovative methods for evaluation and feedback, or personalizing the learning activities in order to adapt them to the competence level and learning needs of each individual student [11]. Moreover, a pedagogical approach as a theoretical framework is a key element in order to deliver digital competence properly [12].

One of the best ways to introduce digital competence to the students is to apply the 'Learning by Doing' paradigm, which allow them to get embedded into a lifelike practical environment [13]. Furthermore, the introduction of gamification in education contexts enhances the impact of that paradigm in the overall academic results [14].

Focusing on the application of project-based learning, it provides both cognitive outcomes, such as knowledge and cognitive strategies, as well as behavioral outcomes, such as skills and engagement [15]. Additionally, the implementation of an project-based learning approach not only furnishes students with hard skills, such as technical and professional ones, but also with soft skills, such as teamwork and communication, which are necessary in order to confidently face real-life problems [16].

In this paper, an instance of project-based learning is presented, which was undertaken by students in groups, and was in turn presented to the rest of students for evaluation.

The rest of the paper is organized as follows: first, section 2 introduces the methods, then, section 3 presents the results obtained, after that, section 4 displays the discussion, and finally, section 5 exposes the final conclusions.

2. Methods

We are based in a Spanish university and the course where we implemented an active learning approach is called "Introduction to Programming", which is clearly included into the STEM education. In the current academic year, namely 2023-2024, we had 33 students, whereas in the previous academic year, namely 2022-2023, we had 31 students, so the number of scholars was pretty much the same.

In the present academic year, we organized a project-based activity to assess the students within the course, where they got together in groups of 3 in order to build up a scenario composed of different logic gates, coded in Java, as this programming language is the guiding thread of the course.

Each team had to design a scenario with a constraint time interval. After that, a specific session was set up in the classroom where all groups had to make a pitch presentation of around 5 minutes each, where all team members had to take part in it. Actually, they needed to explain the model designed to the rest of school mates, with the following agenda: one of the team members had to expose the coding structures used, another one had to talk about the functionality of the design, and the other one had to execute some instances and explain the results.

Right afterwards, all students in the class had to fill in a construct with a set of questions about the project presented, where each item had to be ranked in a 5-point Likert-type scale. It is to be noted that this construct was specifically hand-made for this activity and it was validated by a panel of experts before presenting it to the students. Additionally, another construct was presented to the students at the end of the presentation session in order to measure their level of engagement while developing the project and making the presentation about it.

In summary, three constructs have been employed herein. The first one was intended for a panel of five experts and was used to validate the questions or items to be presented to the students at a later stage. The second one was intended for the whole set of students and was used to evaluate all projects proposed on a peer review basis. The third one was also intended for all students and was used to assess their own level of engagement while undertaking their corresponding projects.

It is to be noted that a Likert-type scale consists of a unidimensional psychometric scale which comprises different items in order to collect the attitudes, opinions, or motivations of the respondents. Each item has to be rated within a range of answer options, thus moving from one extreme answer to the other, whilst passing through some intermediate answers, where the distance between any pair of neighboring answers is always the same.

Likert-type scales are typically rated from 1 to a given natural number, which depends on the number of intermediate answers allowed. The items included are usually highly correlated, aiming at achieving a high internal consistency among the set of answers to each item. Also, quantitative data is obtained from each item, as the answers allow for degrees of opinion, as opposed to binary or categorical answers.

Furthermore, the validation of the hand-made construct for students is going to be measured by the Aiken's V test according to the ratings given by a panel of experts. It is to be reminded that the value of the Aiken's test is obtained by subtracting the minimum value of the Likert scale considered from the average value of the whole construct, and dividing the result by the difference between the maximum and minimum values of the Likert scale.

On the other hand, the validation of the instrument to measure the level of engagement is not necessary, as the construct used for this matter is a standard instrument, which was specifically designed and validated before it was released.

Besides, the reliability of the data collected with the constructs aimed to students, namely, the instrument to assess them on a peer review basis and the instrument to assess their level of engagement, has been done by means of the Cronbach's alpha, which considers the number of items within an instrument, the variance of the results obtained for each item, and the variance of the overall values.

High values of the Cronbach's alpha imply high correlation among variables, which could also be reinforced by calculating Pearson's correlation coefficient and Spearman's rank correlation coefficient. Regarding the former, it is the most usual way to measure the lineal correlation between two given sets of data. However, it is aimed to continuous data, whereas the data collected with Likert-type scales are discrete data, even though it is quite often used with ordinal data as well because results are quite coherent, and this is the case of the outcome achieved with Likert-type scales.

With respect to the latter, it measures the monoticity of the relation between two given sets of data, which could be linear or otherwise. In other words, each set of data is sorted and substituted with its respective order number. Hence, this definition allows to study the correlation between two particular sets of data not only to discover the degree of linear correlation between them, but any type of it.

3. Results

3.1. Construct for item validation by judges

To begin with, a brand-new construct was designed for the peer review evaluation, which was originally built up with 8 questions or items. All of them were organized in two different categories, such as programming and presenting. Table 1 presents the original items for validation by judges.

This hand-crafted construct was then judged by a panel of 5 experts in the field with another construct

Table 1. Original items in the peer-review construct for validation by judges

Categories	Identifier	Item
Programming	I1	Coding skills
	I2	Code clarity
	I3	Code explanation
	I4	Coding structures
Presenting	I5	Functionality
	I6	Execution
	I7	Communication skills
	I8	Tools for presentation

where each of the items proposed was rated in a 4-point Likert type scale, according to 2 different dimensions, namely the construction of each item and its clarity. Hence, both dimensions were judged with separate ratings for all items, where a score of 1 corresponded to strongly disagree, a score of 2 to disagree, a score of 3 to agree and a score of 4 to strongly agree. It is to be noted that, in this case, there was no option for a neutral answer, as the goal was for the judges to clearly determine their views.

After the judgement was done, the average values (AVG) for each dimension related to each question were calculated, thus leading to the average of all ratings, which was eventually used to get the value of the Aiken's V test in order to assess the validity of the set of items [7]. Table 2 shows those calculations.

Table 2. Aiken's V test with the original items

	Construction	Clarity
AVG I1	3.8	3.6
AVG I2	3.4	3.8
AVG I3	4.0	3.6
AVG I4	2.2	2.4
AVG I5	3.2	3.8
AVG I6	3.6	3.4
AVG I7	4.0	4.0
AVG I8	2.6	2.4
AVG for each DIMENSION	3.35	3.325
AVG for whole CONSTRUCT	3.338	
AIKEN'S V TEST	0.779	

The value obtained in this case for the Aiken's V test could be considered as acceptable with some of the cutoff marks stated in the literature, such as the 0.70 proposed by Charter [8] or the 0.50 established by Cicchetti [9]. Those benchmarks are thought to permit some degree of discrepancy among judges, thus leading to looser agreements. However, the original cutoff mark was established as 0.87 by Aiken [10], which imposes a tougher level of agreement among judges.

Hence, as we wanted to stick with Aiken's benchmark, then we decided to drop the items with much lower rating, so we deleted items I4 and I8 out of our construct. This way, questions I1 to I3 remained in the programming category, although they were renamed as items Q1 to Q3, whilst questions I5 to I7 where rebranded as questions Q4 to Q6, and stayed in the presenting category. Table 3 presents the definitive items for validation by judges.

Table 3. Final items in the peer-review construct for validation by judges

Categories	Identifier	Item
Programming	Q1	Coding skills
	Q2	Code clarity
	Q3	Code explanation
Presenting	Q4	Functionality
	Q5	Execution
	Q6	Communication skills

At that point, the average values (AVG) for each dimension referred to the remaining questions were recalculated, thus leading to the average of all rating, which finally was used to find out the value of the Aiken's V test. Table 2 exposes these calculations.

Table 4. Aiken's V test with the final items

	Construction	Clarity
AVG Q1	3.8	3.6
AVG Q2	3.4	3.8
AVG Q3	4.0	3.6
AVG Q4	3.2	3.8
AVG Q5	3.6	3.4
AVG Q6	4.0	4.0
AVG for each DIMENSION	3.667	3.7
AVG for whole CONSTRUCT	3.683	
AIKEN'S V TEST	0.894	

At this point, the value of Aiken's V test is higher than the cutoff mark established by Aiken, which was our benchmark. Therefore, it could be said that the construct with the remaining items was indeed validated according to the panel of experts. Hence, after the validation of the construct, it was ready to be presented to the students in the session dedicated to deliver their pitch presentations.

3.2. Construct for project evaluation by students

Once the different teams of students were done with their projects, then the pitch presentation session was held. At that point, after each group delivered its own presentation, the validated construct was presented to the whole set of students within the class in order for them to evaluate each project exposed on a peer review basis. This construct was used for content assessment by students, and the categories exposed in the construct for validation by judges were herein relabeled as dimensions. In other words, this construct for students so as to evaluate their peers was composed of 2 dimensions, such as programming and presenting, where each one contained 3 questions.

The evaluation of the items was undertaken in a 5-point Likert type scale, where a grade of 1 corresponded to strongly disagree, a score of 2 to disagree, a score of 3 to neutral, a score of 4 to agree and a score of 5 to strongly agree. Table 5 displays the most common descriptive statistics [11], organized by both dimensions and overall.

Table 5. Descriptive statistics of the results achieved, organized by dimensions and overall

Statistics	Programming	Presenting	Overall
Average	4.65	4.51	4.58
Mode	5	5	5
25th Percentile	4	4	4
Median	5	5	5
75th Percentile	5	5	5
Variance	0.29	0.31	0.31
Standard Deviation	0.54	0.56	0.55
Coefficient ofVariation	0.12	0.12	0.12

Furthermore, reliability of the results obtained was calculated by the Cronbach's Alpha, which measures the internal consistency of the data attained [12]. Table 6 exhibits those values for each dimension and overall.

Table 6. Cronbach's Alpha for the results achieved, organized by dimensions and overall

Programming	Presenting	Overall
0.741	0.723	0.780

Cronbach's Alpha also gives information about the correlation of the data attained in both dimensions [13], even though this was also done through the Pearson's correlation coefficient and the Spearman's rank correlation coefficient. Table 7 exposes the values attained for the correlation between the data included in both dimensions.

Table 7. Correlation coefficients for the results achieved between both dimensions

Pearson's correlation coefficient	0.579
Spearman's rank correlation coefficient	0.582

3.3. Construct for engagement assessment by students

At the end of the pitch presentation session, the level of engagement of the students when carrying out their corresponding projects was measured by means of a standard construct called the ISA engagement scale [14]. This is a 7-point Likert type instrument, where three dimensions are assessed, namely Intellectual, Social, and Affective, with three items associated to each of those. The available ratings are a mark of 1 to strongly disagree, a mark of 2 to disagree, a mark of 3 to partially disagree, a mark of 4 to neutral, a mark of 5 to partially agree, a mark of 6 to agree and a mark of 7 to strongly agree.

The goal in this survey is to achieve at least an average of 6 in all dimensions and overall, which accounts for a high level of engagement [15]. Table 8 displays the standard questions in the ISA engagement scale, which is actually the acronym of the three dimensions considered.

Table 8. The ISA engagement scale

Dimensions	Items
Intellectual Engagement	I focus hard on my work.
	I concentrate on my work.
	I pay a lot of attention to my work.
Social Engagement	I share the same work values as my colleagues.
	I share the same work goals as my colleagues.
	I share the same work attitudes as my colleagues.
Affective Engagement	I feel positive about my work.
	I feel energetic about my work.
	I am enthusiastic in my work.

Moreover, Table 9 exposes the most common statistics for centralization and dispersion in the outcome of the ISA engagement scale, grouped by both dimensions and overall.

Table 9. Descriptive statistics of the outcome achieved, organized by dimensions and overall

Statistics	Intell.	Soc.	Affect.	Overall
Average	6.37	6.49	6.53	6.46
Mode	7	7	7	7
25th Percentile	6	6	6	6
Median	7	7	7	7
75th Percentile	7	7	7	7
Variance	0.69	0.46	0.27	0.47
Standard Deviation	0.83	0.68	0.52	0.69
Coefficient ofVariation	0.13	0.10	0.08	0.11

Additionally, reliability of the outcome achieved was calculated by the Cronbach's Alpha, which measures the internal consistency of the data attained. Table 10 exhibits those values for each dimension and overall.

Table 10 Cronbach's Alpha for the outcome achieved, organized by dimensions and overall

Intellectual	Social	Affective	Overall
0.965	0.872	0.975	0.922

Table 11. Correlation coefficients for the results achieved among dimensions

Corr.coef.	Int-vs-Soc	Soc-vs-Aff	Int-vs-Aff
Pearson	0.661	0.723	0.625
Spearman	0.634	0.753	0.602

The correlation among dimensions was calculated through Pearson's correlation coefficient and the Spearman's rank correlation coefficient. Table 11 shows those values for the correlation between each pair of dimensions.

3.4. Comparing the results obtained in the current and the previous academic year

The academic performance achieved last year was measured by means of a traditional evaluation, carried out by a written exam, which delivered an average score of 8.25 out of 10. In this sense, it is to be noted that our college is based in Spain, whose academic grading system ranges from 0 to 10 [16].

On the other hand, the academic performance attained this year was measured by means of a peer review rating, which delivered an average value of 4.58 out of 5, as stated above. Hence, in order to compare those values of academic performance, both must be measured with the same scale.

As 5 is a multiple of 10, such that 10/5=2, then the average grade corresponding to this year in base 10 is easily found out by just doubling the value in base 5, so it results in an average mark of 9.16 out of 10.

Table 12 shows the average variation in marks between the previous academic year and the current one, whereas Table 13 depicts the success rate variation between both years.

Table 12. Average academic performance in the current and the previous academic years

2022-2023 average mark	2023-2024 average mark	Average mark variation
8,25	9,16	11%

Table 13. Success rate in the current and the previous academic years

2022-2023 success rate	2023-2024 success rate	Success rate variation
25/31 = 0,81	31/33 = 0,91	10%

4. Discussion

In order to clarify the discussion about the different results obtained, they are organized in the same subsections as above.

4.1. Construct for validation by judges

The number of questions originally proposed for this construct was 8, such that each category (programming and presenting) was composed of 4 items each. However, after the panel of experts judged each question according to two dimensions, namely its construction and its clarity, it occurred that items I4 and I8 got significantly lower rates than the rest of the items. Specifically, neither of them reached an average value of 3 in any dimension, considering that the number 3 was assigned to "agree" in this particular construct.

This was the main reason why the Aiken's V test yielded 0.779, which is not considered as valid according to the Aiken's benchmark, that being 0.87. It is to be reminded that other cutoff marks are below this the value obtained, such as 0.70 established by Charter, or 0.50 proposed by Cicchetti. However, as the Aiken's cutoff mark was the benchmark expected, then we decided to drop both items out of the construct. This way, the new construct had only 6 questions, where three of them remained into the programming category, while other three stayed into the presenting category.

Then, the Aiken's V test was recalculated with the ratings assigned to the remaining 6 questions, and this time, the value attained was 0.894. Hence, as the value yielded was greater than the Aiken's benchmark, namely 0.87, then it was concluded that the construct with 6 questions got validated by the panel of five experts.

4.2. Construct for content assessment by students

First of all, it is to be noted that the categories considered in the validation construct are renamed as dimensions in this content assessment construct. Hence, the results herein are going to be exposed grouped by both dimensions, namely programming, and presenting, as well as for the overall construct.

The descriptive statistics about the results obtained regarding the different projects exhibited similar values for both dimensions and overall, where both were higher than 4, which represents a rate of "agree". Also, this happens in both centralization and dispersion statistics. However, it seems that the values achieved in the programming dimension are slightly higher than those in the presenting counterpart, which might be associated to the fact that STEM students are often supposed to be better at technical skills than at soft skills, even though it might also be caused by some slight mistakes occurred when making the pitch presentations.

Additionally, the internal consistency of the data collected is considered as acceptable because the value of the Cronbach's Alpha for the overall construct is 0.780, which is higher than the benchmark considered for acceptable reliability of the data collected, namely 0.7. Moreover, the Cronbach's Alpha values calculated for the data related to each dimension are also higher than 0.7, specifically 0.741 for programming and 0.723 for presenting.

Those values of Cronbach's Alpha also stand for a high correlation among dimensions. This point has been reinforced by the values obtained for the Pearson's correlation coefficient and the Spearman's rank correlation coefficient, where both values are greater than 0.5, which implies a high level of correlation among the dimensions considered.

Additionally, the values obtained for both coefficients are quite similar, which reinforces the idea that the Pearson's correlation coefficient could be used to assess correlations between two data sets when such data are not continuous values, but ordinal ones.

4.3. Construct for engagement level assessment by students

The statistical values attained in the ISA engagement scale account for a high level of motivation in the students when undertaking the project, as the average values corresponding to the dimensions considered and the overall construct were all higher than 6, which represents a rate of "agree". Moreover, the centralization and dispersion statistics are really similar in all cases. Specifically, the average values attained are 6.37 for the intellectual dimension, 6.49 for the social one, 6.53 for the affective one, which accounts for an average value of 6.46 for the overall construct.

Those values show a bit lower level of engagement for the intellectual dimension than those related to its counterparts, whilst the level of the affective dimension is just slightly higher than that of the social one. Those differences may be due to the fact that students appreciated a bit more the fact of working with other colleagues than the work itself.

In other words, those results lead to the conclusion than students were a bit more motivated by teaming up with their colleagues and the feelings they got about it. Besides, those results also imply that students were slightly more motivated by the feelings they got during the experience. However, it is to be reminded that all values overcame the value representing the 'agree' rate, and the difference among those values is pretty small. Hence, it could be said that the level of engagement was high in all three dimensions considered.

Furthermore, the internal consistency of the data collected is excellent as the values obtained for the Cronbach's Alpha in all dimensions and overall are greater than 0.9. Those values also stand for a high correlation among dimensions, which has been reinforced by the values of the correlation coefficients among each pair of dimensions. Specifically, the Pearson's correlation coefficient and the Spearman's rank correlation coefficient have been found out for each pair of dimensions, where all correlation values were above 0.5, which accounts for a high correlation among dimensions. Also, the values attained in both coefficients for each correlation carried out are so similar that it proves that the Pearson one could be used instead of the Spearman one without a significant loss of accuracy.

4.4. Comparing the results obtained in the current and the previous academic year

To begin with, it is to be reminded that the number of scholars registered within the course in the current academic year was 33, whereas there were 31 within the same course in the last academic year, so both values are certainly comparable.

Furthermore, it is to be considered that the evaluation system in the previous year was carried out through a traditional written exam, whilst the adoption of active learning led to an active evaluation approach by means of a peer review assessment based on the team-based projects done.

Having said that, the average mark in the previous academic year was 8.25 out of 10, whereas it was 9.16 out of 10 in the current academic year. Therefore, it results in an average mark increment of 11% for the academic performance in the current year. According to the literature, the use of an active learning approach in STEM courses rises the academic performance around 15% [17]. Hence, the increment obtained is pretty close to that benchmark.

On the other hand, the success rate in the previous academic year was 25 out of 31, namely an 81%, whilst in the current academic year it has been 30 out of 33, namely a 91%. Therefore, it results in an increment for the success rate of 10% in current academic year. According to the literature, the implementation of an active learning approach in STEM courses increases the success rate around 20% [18]. Hence, the increment obtained may be seen as somehow similar, as it follows the same monotonic tendency, although its magnitude is just half of it.

5. Conclusions

In this paper, a project-based learning initiative is exposed, which was taken in a course called Introduction to Programming at a higher education level. In fact, this course is intended for students to learn the basics of programming in Java at college. In the previous academic year, the evaluation was done through a written exam, although in the current academic year we adopted an active learning approach for the evaluation.

The assessment was to be undertaken in groups of 3 students, where each team had to build up a scenario composed of different logic gates, which had to be coded in Java. After the deadline to finish the projects, we organized a specific session where each team had to make a pitch presentation so as to expose the different models to their class mates.

Right after each presentation, the audience had to assess the projects proposed with a hand-made construct, whose items had to be rated in a 5-point Likert type scale. Actually, this construct was composed of 6 questions, where there were considerations about the programming part of the project, along with its presentation to the rest of scholars. Hence, the evaluation was carried out on a peer review basis. Additionally, another construct was presented at the end of that session in order to measure the level of engagement of the students during the development of the project.

The hand-crafted construct to assess the projects was first validated by a panel of experts, who judged the questions proposed according to two dimensions, such as the construction and the clarity of each item. Afterwards, the average of those judgements were used to calculate the Aiken's V test, which got passed according to Aiken's benchmark, thus resulting in the validation of the construct.

This construct was used for students to evaluate the projects presented for all the teams, which resulted in high ratings in all questions, where there were no significant differences between the descriptive statistics applied to the data collected. Nonetheless, values related to programming were slightly higher than those related to presenting, even though the difference between them was pretty small. Moreover, the reliability of the data collected was measured through the Cronbach's Alpha, leading to an acceptable value for the internal consistency of data, as well as a high correlation between data related to programming and presenting.

Eventually, the ISA engagement scale was used to measure the level of engagement of the students during the development of the project, resulting in high levels of motivation regarding the intellectual dimension, the social one and the affective one, as well as overall. Besides, the values of Cronbach's Alpha showed an excellent reliability of the data collected and a high correlation among dimensions.

Eventually, the evaluation results obtained in this course for the current academic year with respect to the previous academic year reveal an increment of 11% in academic performance, as well as an increment of 10% in success rate. Those percentages are in line with those established in the literature when adopting an active learning approach in courses related to STEM education.

6. References

[1] D. Lombardi and T. F. Shipley, "The Curious Construct of Active Learning", Psychological Science in the Public Interest, vol. 22, no. 1, pp. 8-43, 2021.

[2] S. Freeman et al., "Active learning increases student performance in science, engineering, and mathematics", Psychological and Cognitive Sciences, vol. 111, no. 23, pp. 8410-8415, 2014.

[3] M. Stains et al., "Anatomy of STEM teaching in North American universities", Science, vol. 359, no. 6383, pp. 1468-1470, 2018.

[4] E. Ribeiro-Silva et al., "Trends of Active Learning in Higher Education and Students' Well-Being: A Literature Review", Frontiers in Psychology, vol. 13, 844236, 2022.

[5] P.J. Roig et al, "Using Escape Room as Evaluation Tool in Active Learning Contexts", Education Sciences, vol. 13, no. 6, 0535, 2023.

[6] P. Crespí, J. M. García-Ramos, and M. Queiruga-Dios, "Project-Based Learning (PBL) and Its Impact on the Development of Interpersonal Competences in Higher Education", Journal of New Approaches in Education Research, vol. 11, no. 2, pp. 259-276, 2022.

[7] A. Çebi, T. B. Özdemir, I. Reisoglu, and C. Çolak, "From digital competences to technology integration: Re-formation of pre-service teachers' knowledge and understanding", International Journal of Educational Research, vol. 113, 101965, 2022.

[8] L. Ilomäki, S. Paavola, M. Lakkala, and A. Kantosalo, "Digital competence – an emergent boundary concept for policy and educational research", Education and Information Technologies, vol. 21, pp. 655-679, 2016.

[9] M. Spiteri, and S. N. Chang-Rundgren, "Literature Review on the Factors Affecting Primary Teachers' Use of Digital Technology", Technology, knowledge and learning, vol. 25, pp. 115-128, 2020.

[10] A. Pérez-Escoda, R. García-Ruiz, and I. Aguaded, "Dimensions of digital literacy based on five models of development", Culture and Education, vol. 31, no. 2, pp. 232-266, 2019.

[11] J. Cabero-Almenara, R. Romero-Tena, and A. Palacios-Rodríguez, "Evaluation of Teacher Digital Competence Frameworks Through Expert Judgement: the Use of the Expert Competence Coefficient", Journal of New Approaches in Educational Research, vol 9, no. 2, 2020.

[12] P. Colás-Bravo, J. Conde-Jiménez, and S. Reyes-de-Cózar, "The development of the digital teaching

competence from a sociocultural approach", Comunicar, vol. 62, no. XXVII, pp. 19-30, 2019.

[13] M. Bacigalupo, "Competence frameworks as orienteering tools", Revista Interuniversitaria de investigación en tecnología educativa, vol. 12, pp. 20-33, 2022.

[14] D. Corsi, F. Domínguez, and M. Rodríguez, "Adquisición de competencias emocionales mediante el desarrollo y uso de Serious Games en Educación Superior", Píxel-Bit - Revista de Medios y Educación. vol. 56, pp. 95-112, 2019.

[15] P. Guo, N. Saab, L. S. Post, and W. Admiraal, "A review of project-based learning in higher education: Student outcomes and measures", International Journal of Educational Research, vol. 102, 101586, 2020.

[16] M. Ríoseco-País, J. Silva-Quiro and C. Carrasco-Manríquez, "Development of Digital Competences in Students of a Public State-Owned Chilean University Considering the Safety Area," Education Sciences, vol. 13(7), 710, 2023.

[17] P.J. Roig, "Formal algebraic modelling for fog computing network architecture", PhD Thesis, University of the Balearic Islands, Palma de Mallorca, Spain, 2022.

[18] R. A. Charter, "A breakdown on reliability coefficients by test type and reliability method, and the clinical implications of low reliability", Journal of General Psychology, vol. 130, no. 30, pp. 290-304, 2003.

[19] D. V. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology", Psychological Assessments, vol. 6, 284-290, 1994.

[20] L. R. Aiken, "Three Coefficients for Analyzing The Reliability And Validity of Rating", Educational and Psychological Measurement, pp. 131-142, 1985.

[21] Z. Ali, S. B. Bhaskar, and K. Sudheesh, "Descriptive Statistics: Measures of Central Tendency", Dispersion, Correlation and Regression, Airway, vol. 2, no. 3, pp. 120-125, 2019.

[22] K. S. Taber, "The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education", Research in Science Education, vol. 48, no. 1, 1-24, 2018.

[23] M. Amirrudin, "Effect of Variability on Cronbach Alpha Reliability in Research Practice", Matematika, Statistika & Komputasi, vol. 17, no. 2, pp. 223-230, 2021.

[24] E. Soane et al., "Development and application of a new measure of employee engagement: The ISA Engagement Scale", Human Resource Development International, vol. 15, pp. 529-547, 2012.

[25] M. A. Mañas-Rodríguez et al., "Validation of the Spanish version of Soane's ISA Engagement Scale", Revista de Psicología del Trabajo y de las Organizaciones, vol. 32, no. 2, pp. 87-93, 2016.

[26] Spanish Grading System. https://www.upv.es/entidades/OPII/infoweb/pi/info/1147768normali.html (Access Date: July 17 2024).
[27] G. Hacisalihoglu et al., "The use of an active learning approach in a SCALE-UP learning space improves academic performance in undergraduate General Biology", PLoS ONE, vol. 13, e0197916, 2018.

[28] O. Nurbavliyev, S. Kaymak, and B. Sydykov, "The Effect of Active Learning Method on Students' Academic Success, Motivation and Attitude Towards Mathematics", Journal of Language and Linguistic Studies, vol. 18, no. 2, pp. 701-713, 2022.