Metrics for Assessing Creativity

Creative learning in computer science is one of the ways of learning computer science. OBJECTIVE: objective of this article to unveil a metric where creative computer science learning is calculated in order to quantitatively distinguish one student from another. METHOD: To meet this objective, a documentary analysis was carried out with the results of the Scopus search. RESULTS: At first, the need for a system of empirical methods for its study and how to triangulate them is determined, taking into account their weight as determined by consultation with experts. In a second moment, the metric to evaluate this learning is established, and in a third moment, another metric to change the scale depending on the evaluation system in each country where it is measured. Finally, the validation of the system of methods is approached from its introduction in a case study to diagnose the creative learning of computer science in the computer engineering career.

Keywords: creative learning, creative learning in computer science, computer science.

ONLY FOR REVIEW

Introduction

The evaluation of creativity is a contentious issue.Some authors (Katz, Mann, Shen, Goncalo, & Ferguson, 2022) propose a scale for this purpose, but without a formally defined formula. In contrast, Romo, Alfonso-Benlliure, and Sanchez-Ruiz (2016) suggest a test of children's creativity that primarily involves manipulating images. This test opposes the use of divergent thinking as a diagnostic tool. The research conducted by (Díaz Gómez & Martínez, 2013) on the theory of creative learning as an expression of creativity in learning does not provide a description of the levels or calculation of this intrinsic process of the human personality.

It is important to note that the learning of computer science is currently a national priority, and training creative professionals in this field is even more crucial. The objective of this article is to propose a metric for quantitatively distinguishing one student's creative computer science learning from another. The paper aims to offer a quantitative method for evaluating creativity on a country-wide scale of evaluators.

Methodology

The methodology employed was documentary analysis, based on the collection of documents found on the subject in Scopus. The search terms used were 'evaluate' or 'metric' and 'creativity', which yielded over 100,000 documents. However, only 590 open access documents were selected for downloading, as they are accessible to all researchers and can be used to compare the results obtained. 70% of the downloaded documents use the Torrance test or other methods.

However, these methods only measure creative thinking and cannot be used to assess other conceptions of creativity, such as the cultural historical approach or the humanistic approach. This limits the scope of their application to the measurement of creativity. Additionally, these methods employ a Likert scale, which may not always be understood by the user. This is another significant limitation of the proposed assessment methods. Out of the 590 downloaded documents, only 14 explicitly refer to some form of creativity assessment or related scales, and 78% of them are from 2021 onwards, ensuring that they are up-to-date. Furthermore, access is provided to the documents referenced in the texts where metrics or ways of assessing creativity are obtained, resulting in a total of 15 additional documents.

Results

Only 14 of the retrieved documents that used a method to measure creativity provided a comprehensive explanation. Table 1 presents the authors and their proposals for measuring creativity. The source of this information is the authors' own research. It is important to note that the language used in the table is clear, concise, and objective, adhering to the conventional structure and formal register required in academic writing.

Table 1. Authors whose investigate in creativity related to evaluate

Authors	Metric	Scale	Method	System of methods	Instrument	Indicators
Görlich (2023)		X	X			X
Vasylenko et al. (2022)			X	X	X	
Saenna and Phusee-orn (2022)		X			X	
Roth, Conradty, and Bogner (2022)	X	X			X	X
Pei, Han, Zhao, Liu, and Pang (2022)			X		X	
Katz et al. (2022)			X		X	
Havârneanu (2022)					X	
Apaydin and Güven (2022)			X			X
Furtwengler		X				

ONLY FOR REVIEW

(2021)						
Leutner, Yearsley, Codreanu, Borenstein, and Ahmetoglu (2017)			X		X	
Hsu-Chan, Burnard, McLellan, and Ying-Yao (2017)			X		X	
Romo et al. (2016)					X	
Chesnokova (2014)			X		X	
Ott and Pozzia (2010)						X

Source: Authors' elaboration

Now we will analyse all of the measuring methods used by the 14 documents. One of these methods Romo et al. (2016) The author proposes the children's creativity test as an alternative to tests that rely solely on divergent thinking. It is important to note that the test is not intended to measure divergent thinking. No changes in content have been made. However, the author does not provide a metric for quantitatively differentiating between the subjects being evaluated. The proposed test "… involves manipulating images and is aimed at evaluating creative potential using drawing as a means of expression" (Romo et al., 2016, p. 96). This question cannot be generalised for all children as it depends on the skills developed for this activity. Nor do they establish observable quantitative relationships between the dimensions that are proposed.

However, Hsu-Chan et al. (2017) propose a set of indicators to qualitatively evaluate creativity. This study employs various methods and statistical analyses to develop a questionnaire that measures creative education, considering its dimensions. However, it does not provide explicit instructions on how to use the questionnaire or the metric for expressing the relationships between the five dimensions when assessing creativity using the Likert scale.

The study by Katz et al. (2022) outlines criteria for evaluating creativity. Vasylenko et al. (2022) propose several methods for assessing creativity, but they do not reveal the interrelationships between them or their integrated application. Each method assesses creativity without being compared to the others, despite measuring different aspects. Common indicators for applying the instruments and extracting information are not revealed. This information, although different, could provide a more holistic view of creativity.

Görlich (2023) proposes an evaluation scale for the creative process based on 24 items distributed across 8 phases. The scale is determined solely by these items,

ONLY FOR REVIEW

making it less extensible and limiting its operability in cases where additional factors need to be considered. Furthermore, the text does not provide a metric that expresses mathematical relationships to obtain a single quantitative value characterising the process.

The analysed works discuss instruments such as the Torrance test or those developed by researchers. However, they do not propose general measures that are applicable to different situations, dimensions, and indicators for measuring creativity. It is important to note that the Torrance test can only measure creative thinking, which is a serious limitation. Similarly, there are scales available that measure specific aspects, but they do not clarify how they relate to assessment systems in different educational contexts. Clarifying this would help teachers better understand research results by presenting them on a scale that is comprehensible to them. It is recommended that teachers use a national evaluation scale that they are familiar with.

Discussion

Metrics to assessing Creativity

Assessing the development of IT creativity can be challenging for novice researchers. Quantitative measures that depend on the importance of each dimension for the organization would be helpful in these cases, as well as for personnel selection for a project. The metric for determining the level of creativity assessed from one method (C1M) is shown in the expression $C1M = \sum_{i=1}^{n} P_i D_i$ (1.1)

Where: h is the number of dimensions, D_i is the evaluation of the i-th dimension and P is the weight

$$D_i = \frac{1}{m_i} \sum_{j=1}^{m_i} P_{ij} \frac{1}{n_{ij}} \sum_{k=1}^{t_{ij}} I_{ijk} \qquad (1.2)$$

Where:
m_i : total indicators of dimension i
t_{ij} : total number of aspects to be assessed for indicator j of dimension i
I_{ijk} : assessment given to aspect k of indicator j in dimension i

To determine the weight of each dimension, it is recommended to use the paired comparisons' method. Although this method is classified as subjective weighting, specialists can quantify the intensity of their preferences between criteria using the rating scale proposed by Saaty (1987) instead of relying solely on their subjective evaluations (Martínez, Gómez, Ibarra, & Moncada, 2018).

The AHP Online System (Gopel, 2018) will be used to determine these weights. The software tool is designed for the hierarchical analytical process (HAP). The first step is to identify specialists who will evaluate the relative importance of each dimension. According to Zeballos (2015), an expert is someone who has acquired knowledge and skills in a particular field over the years

ONLY FOR REVIEW

and stands out for their competencies in understanding and problem-solving. The selection criteria for these specialists were specified as experience as a computer science teacher and competence in educational research. The number of experts chosen will depend on the research and criteria.

The table displays the consultation process for the experts. They must fill in the blanks according to the order in the first row of the table, taking into account the given scale.

Table 2. Expert Consultation

If the criterion in the row is more important than the one in the column, the more important it is according to the scale. If it is less important, the reciprocal of the value of the scale is used.					
	$Dimension_1$	$Dimension_2$	$Dimension_3$	....	$Dimension_n$
$Dimension_1$	1				
$Dimension_2$		1			
$Dimension_3$			1		
......				1	
$Dimension_n$					1

Source: Authors' elaboration

Table 3. Escala de Saaty

Value	Definition	Comments
1	Equal Importance	Criterion A is equally as important as criterion B.
3	Moderate Importance	Experience and judgement slightly favour criterion A over B
5	High Importance	Experience and judgement strongly favour criterion A over B
7	Very High Importance	Criterion A is much more important than criterion B
9	Extreme Importance	The greater importance of criterion A over B is beyond doubt
2,4,6,8	Intermediate values between the above when there is a need for nuancing	

Source: (Penades Pla (2017)

The results of the expert judgement are aggregated using the geometric mean to arrive at a new consensus priority vector (Table 3).

Table 4. Weighting of Dimensions

Criterion		**Comment**	**Weights**
1	$Dimension_1$		$\mathbf{Weights_1}$
2	$Dimension_2$		$\mathbf{Weights_2}$
3	$Dimension_3$		$\mathbf{Weights_3}$
	....	………….	……….
n	$Dimension_n$		$\mathbf{Weights_n}$

Source: Authors' elaboration

Once the creativity assessment metric has been determined, it is necessary to obtain the measurement scale and to do so it is necessary to obtain the range

ONLY FOR REVIEW

$[VMínimo, VMáximo]$. The maximum value obtained $VMáximo$ is the value obtained by substituting I_{ijk} with the maximum value given by the researcher for each aspect in the formula (1.2) and $VMínimo$ is the initial value of the scale.

Subsequently, Medina, González, Robayo, López, and T. Freire (2021) formula is then applied to determine equivalence, but it is necessary to correct the maximum and minimum of the interval. This is because the way the intervals are calculated includes the maximum of the previous interval as the minimum in the upper interval, which is an error. To correct this, a value that is small enough to be greater than the maximum of the previous upper limit and less than the one following this number should be added. This small number can be obtained using the formula. $o = \frac{1}{10^{r+1}}$ where r is the number of decimal places that the maximum value of the previous interval has. This leads to transforming the above author's formula 3 into the one presented as 3.1. In such a way that $escala_{indi} = (VMínimo - ValMaxInter_1, VMínimoInter_2 - ValMaxInter_2, \ldots, VMínimoInter_{n-1} \text{-} VMáximo)$ (2.1).

The distance between each of the maximum and minimum values of each interval is calculated according to the expression $Dist = \frac{VMáximo - VMínimo}{NumInt - 1}$ where $NumInt$ is the number of intervals in which the variable is measured. Having this distance, we then proceed to calculate the $VMaxInter_1 = VMínimo + Dist$. Once this maximum value is obtained, the minimum value of the second interval is obtained in the following way $VMínimoInter_2 = VMaxInter_1 + o$ y and the maximum value of the second interval is obtained as follows $VMaxInter_2 = VMaxInter_1 + Dist$. In the case of the maximum of the second interval the distance to the minimum value is not added to avoid the progressive addition and that in the last interval it is reflected. In this way, the value of the interval n $VMínimoInter_n = VMaxInter_{n-1} + o$ and the maximum value is $VMáximo$. $escala_{indi}$ is the vector indicating the scales, $ValMaxInter_1$, $ValMaxInter_2, \ldots, ValMaxInter_n$ the maximum values of each of the scales, is the maximum value of the weighting of the course indicators, $VMáximo$ is the maximum value on the country scale, $TotalEscala$ is the maximum value on the country scale, $valmxinte$ is the maximum value of each interval and $valmxinte$ is the minimum value.

The modification to the proposed formula enables more precise determination of the boundaries for each interval on the scale. Additionally, global quantitative evaluation systems typically only use two decimal places, meaning that the sum of the proposed values would not significantly impact the results and would allow for clear establishment of the range limits. In this way, it is possible to evaluate computer creativity independently of the evaluation scale and the criteria used in the different countries, homogenising the scale and making it comprehensible to any evaluator.

(Peralta Hernández & Segura, 2023; Zada, Khan, Saeed, Zada, & Jun, 2023). (Görlich, 2023; Vasylenko et al., 2022).

The concept of creativity is a complex and multidimensional phenomenon that is widely discussed in literature (Peralta Hernández & Segura, 2023; Zada, Khan, Saeed, Zada, & Jun, 2023) Different approaches are taken into account, and

ONLY FOR REVIEW

multiple dimensions and indicators are used to measure it. It is important to ensure that any instrument used to measure creativity is sufficiently objective to do so. This fragment discusses the unexplored aspect of creativity assessment due to the limited focus of previous research on a single instrument to measure creativity (Görlich, 2023; Vasylenko et al., 2022). To improve clarity, the language has been made more objective and value-neutral, and the sentence structure has been simplified. Technical term abbreviations have been explained, and the text has been checked for grammatical correctness and spelling mistakes. No changes in content have been made. It highlights the discrepancy between a student's self-assessment of their creativity and the assessment of their classmates and teachers. Technical terms should be explained when first used, and the language should be formal and free from biased or emotional language. To achieve greater objectivity in evaluating a student's creativity, it is necessary to integrate various methods that measure different aspects.

The integration of these methods depends on the assumed concept of creativity, its dimensions, and indicators. Each research project must integrate different methods, making it impossible to assume a standard set of methods for all. Therefore, it is necessary to structure a configuration of methods consistent with the concept of creativity assumed by the researchers.

In the case that all measurement criteria are assessed in each of the instruments then the indicators are calculated using the following formula: $Ind = \sum_{i=1}^{I}(\sum_{j=1}^{m} P_{j*}C_{ij})$ (5.1), where P_j is the value of the weight of each method whose sum is equal to 1, C_{ij} is the value of the measurement criterion and m is the number of methods applied. In case there are measurement criteria that are not applied in any method then the following expression is proposed for the configuration: $Ind = \sum_{i=1}^{I}\left(\frac{1}{M_{ijh}}\sum_{m=1}^{M_{ijh}}\frac{P_m}{\sum_{l=1}^{M_{ijh}} P_l}C_{ijhm}\right)$ (5.2), where: M is the number of methods, C_{ijhm} is the assessment of the measurement criterion of the indicator that is in each dimension, M_{ijh} is the number of methods that measure criterion i of indicator j of dimension h, P_m weight of each method because each of the methods will have a different weight depending on the expert judgement that will be taken into account using the panel for that purpose for the case of Ind. The weight of each method is obtained in the same way as in the case of the dimensions so that their sum is equal to 1 and a table # 4 should be obtained as shown below:

Table 5. Weight of Methods

Criterion		Comments	Weight
1	Method1	Comments on the method	Weight1
2	Method2	Comments on the method	Weight2
3	Method3	Comments on the method	Weight3
	....	....	....
n	Methodn	Comments on the method	Weightn

Source: Authors' elaboration.

ONLY FOR REVIEW

In the case of dimension assessment, the metric is: $D_h = \frac{1}{J_n}\sum_{j=1}^{J_n} P_j I_j$ (6), where J_n would be the number of indicators in the dimension, P_j would be the weight of each indicator and I_j is the value of each indicator obtained in expression 5.

For the evaluation of the variable the expression is: $Var = \sum_{h=1}^{H} P_h D_h$ (7), where H: total dimensions, P_h: is the weight of the dimension.

The obtained metrics meet the criteria for a system of methods as they evaluate the relationships between methods and are not dependent on the number of methods used. The relationship between them is determined by the average product of the weights and the sum of criteria, which homogenises the differences in measurement criteria and the number of dimensions and indicators of each type of method. Each metric measures different indicators and calculates a distinct aspect of the same variable, highlighting its configurational nature. If the metric does not break down into dimensions, indicators, or measurement criteria and arrives at a single value, it can adapt by simply providing a value in the expression to be calculated.

Although originally developed to measure creativity, this metric can be applied to any variable that is operationalised by breaking it down into dimensions and indicators with their respective measurement criteria. If the variable is assessed using a single method, use expressions 1.1 through 5.1. If multiple methods are used, use expression 5.2 for 5.1.

Validation of Metrics

One method of validating metrics is through expert judgement (Molina Hernández, González Hernández, & Cruz Lemus, 2021). Expert judgement is a qualitative method that aims to verify the validity of scientific results that, for various reasons, should not be implemented in practice. One common reason for testing is to evaluate pedagogical results before implementing them in practice, in order to ensure their educational potential (Kanhime-Kasavube & González-Hernández, 2017). This involves applying empirical methods to determine the number of experts in a group of candidates. Once experts are identified, the scientific result to be validated is sent to them along with a voting scale. The researcher can include as many items as they deem necessary in the voting scale, although a Likert-type scale is commonly used (Molina Hernández et al., 2021).

After the experts have cast their votes, a voting matrix is created with the voting items in the columns and the results for each expert in the rows. Additionally, the experts provide criteria regarding the results, which the researcher considers for improvement. Statisticians are used to evaluate the variation, deviation, and agreement among experts on a given topic. This process can be repeated if the researcher believes that the criteria provided by the experts have altered the results and require further analysis.

Additionally, metrics can be validated through a case study to confirm their relevance. The case study should exhibit all the features of the different scenarios in which the metric can be applied.

ONLY FOR REVIEW

Validation by Expert Judgement

The application of the criterion of experts goes through a first phase which is the detection of the experts and for this purpose they are sent a questionnaire that allows them to evaluate their coefficient of competence through the relationship between the coefficient of argumentation and the coefficient of knowledge, which is expressed in the following expression: $K_{comp} = \frac{K_c + K_{arg}}{2}$, donde K_c is the knowledge coefficient and K_{arg} is the argumentation coefficient. Once the coefficients of each person have been obtained, they are then selected as experts or not. In the case of this metric, it was sent to the 15 coordinators of the Computer Engineering career in Cuba and to the 30 managers of the main software development companies in the country and their offices in each province. In the case of the career coordinators, all of them are full professors and 10 of them have PhDs in Technical Sciences in the area of Computer Science and 5 have PhDs in Pedagogical Sciences with subjects related to the teaching of Computer Engineering, while 25 of the managers have records of software marketed inside and outside Cuba. Of the 45 respondents, 40 were considered experts.

The voting scale selected was from 1 to 10 where true (1), almost true (0.9), quite true (0.8), somewhat true (0.7), more true than false (0.6), as true as false (0.5), more false than true (0.4), somewhat false (0.3), quite false (0.2), almost false (0.1), false 0. The 40 selected experts voted for a mean of more than 0.85 with a rating of almost true in 90% of the cases. The table below expresses the measured statistics in which the columns represent the key metrics:

Table 7. Applied Statisticians

Media	9.57	9.64	9.53	9.50	9.72
Typical Deviation	0.534006933	0.5660916	0.50741626	0.50854763	0.4794633
Variance	0.254022989	0.21724138	0.25747126	0.25862069	0.22988506

Source: Authors' elaboration

The mean represents an assessment closer to true than almost true and the experts' ratings were very positive, although they led to a variation of the formula for the calculation of $C1M = \sum_{i=1}^{n} P_i D_i$ originally proposed $C1M = \frac{1}{n}\sum_{i=1}^{n} P_i D_i$. Of the concordance coefficients, Kendall's coefficient is used to determine whether the experts agree on the voting of all the proposed metrics. To determine it, the free software PSPP was used and the final result is 0.85, which is much higher than that accepted for approving the scientific results.

Case of Creative learning in Computer Science (APCI) Fourth Year Computer Engineering University of Matanzas

Vygotsky's cultural-historical approach has been developed in three main strands: activity theory, personality theory, and, in the last 10 years, the Theory of Subjectivity (Rey, 2019). The first strand assumes the reflexive category of Marxist-Leninism, and the fundamental critique is the excessive objectification of psychic processes. The second strand represents a continuity of the most fruitful

ONLY FOR REVIEW

period of Vygotsky's work but does not transcend it. While the last strand manages to open new spaces of theoretical analysis with the categories of subjective meaning and subjective configuration (González-Rey, 2019). These categories make it possible to describe human development by focusing on individuality and its interrelation with society; they also make it possible to explain human development by taking into account its historicity.

According to the Theory of Subjectivity, “creativity is a configuration that arises from singular subjective processes and productions of the individual, which are related to their current context and life trajectory” (de-Almeida & Mitjáns Martínez 2020, p. 97). This individualised character of creativity is a result of each person's social and historical context. However, not every socio-historical context facilitates the development of creativity. This passage does not consider the affective process as isolated from thinking, but rather as an integral part of the symbolic process. According to this conception, the symbolic and affective processes are inseparable, and the Torrance test is not an adequate measure of creativity.

Torres Oliveira and Mitjáns Martínez (2020, p. 129) state that creative learning "... involves the subjective meanings of the individual's life history and the subjective meanings that the individual produces in the context in which his or her action takes place by the way in which he or she subjectivises it. This context is characterised, among other things, by its social subjectivity". One of the authors in a previous article (Mitjáns Martínez, 2013) defines it as "... a form of learning that differs from the common forms of learning in the school environment, and is characterised by the type of production that the learner makes and by the subjective processes involved in it (...). This learning has different forms of expression and involves a set of subjective resources and is expressed in the configuration of at least three processes: the personalisation of information, the confrontation with what is given and the production of new ideas of one's own" (p. 317).

In line with this definition, we assume the proposal González Hernández, Petersson Roldán, and García Moreno (2022a) as creative learning of computer science as "...the subjective, social or individual configuration of three main processes: the personalisation of information, the confrontation with the given and the production of own and new ideas resulting from the emergence of subjective meanings associated with learning computer science in a project context" (p. 130).

The authors have derived three dimensions and 15 indicators from this definition, which are unevenly distributed among the dimensions and express the unity of symbolic and affective processes. It is important to note that the Torrance test should not be used as it does not analyse the affective processes that are present in unity with the symbolic processes.

This definition summarises the characteristics of creative IT learning and places it in the fundamental context of IT creation: the project. Simultaneously, it establishes the conditions for creative IT learning to occur. It offers researchers a definition and a set of dimensions and indicators to evaluate its progress. To evaluate this multifaceted variable, the dimensions, indicators, and measurement criteria outlined in González Hernández, Petersson Roldán, and García Moreno (2022b) were adopted. The students are required to evaluate themselves on a scale ranging from 0 to 1 using the following linguistic labels: true (1), almost true (0.9), quite true (0.8), somewhat true (0.7), more true than false (0.6), as true as false (0.5), more false than true (0.4), somewhat false (0.3), quite false (0.2), almost false (0.1), and false (0). The students are required to evaluate themselves on a scale ranging from 0 to 1 using the following linguistic labels: true (1), almost true (0.9), quite true (0.8), somewhat true (0.7), more true than false (0.6), as true as false (0.5), more false than true (0.4), somewhat false (0.3), quite false (0.2), almost false (0.1), and false (0). The evaluation criteria were established by (Canjongo Daniel, González Hernández, & Becalli Puerta, 2022; Molina Hernández et al., 2021).

To apply the metrics for the student survey, each expert was asked to fill in the blanks according to the order shown in the first row of Table 9, taking into account the given scale.

Table 9. Expert Consultation

If the criterion in the row is more important than the one in the column, the more important it is according to the scale. If it is less important, the reciprocal of the value of the scale is used.			
	Customisation of IT processes	Confrontation with already existing IT processes that enable the emergence of subjective meanings.	Production, generation of own and "new" ideas during the implementation of an IT project
Customisation of IT processes	1		
Confrontation with already existing IT processes that enable the emergence of subjective meanings.		1	
Production, generation of own and "new" ideas during the execution of an IT project			1

Source: Authors' elaboration.

Table 10. Saaty Scale

Value	Definition	Comments
1	Equal Importance	Criterion A is equally as important as criterion B.
3	Moderate Importance	Experience and judgement slightly favour criterion A over B
5	High Importance	Experience and judgement strongly favour criterion A over B
7	Very High Importance	Criterion A is much more important than criterion B
9	Extreme Importance	The greater importance of criterion A over B is beyond doubt
2,4,6,8	Intermediate values between the above when there is a need for nuancing	

Source: Penades Pla (2017)

ONLY FOR REVIEW

Figure 1 shows the paired comparison matrices for each expert and the preference vectors.

Figure 1. Paired Comparison Matrices for each Expert

Experts 1, 2 and 4

	A - wrt AHP priorities - or B?		Equal	
1	◉ Dimensión 1	○ Dimensión 2	◉ 1	○ 2 ○ 3
2	○ Dimensión 1	◉ Dimensión 3	◉ 1	○ 2 ○ 3
3	◉ Dimensión 2	○ Dimensión 3	◉ 1	○ 2 ○ 3

CR = 0% OK

Cat		Priority	Rank
1	Dimensión 1	33.3%	1
2	Dimensión 2	33.3%	1
3	Dimensión 3	33.3%	1

Expert 3

	A - wrt AHP priorities - or B?		Equal	
1	◉ Dimensión 1	○ Dimensión 2	◉ 1	○ 2 ○ 3
2	○ Dimensión 1	◉ Dimensión 3	○ 1	◉ 2 ○ 3
3	◉ Dimensión 2	○ Dimensión 3	◉ 1	○ 2 ○ 3

CR = 5.6% OK

Cat		Priority	Rank
1	Dimensión 1	26.0%	3
2	Dimensión 2	32.7%	2
3	Dimensión 3	41.3%	1

Expert 5

	A - wrt AHP priorities - or B?		Equal	H
1	○ Dimensión 1	◉ Dimensión 2	◉ 1	○ 2 ○ 3
2	○ Dimensión 1	◉ Dimensión 3	○ 1	○ 2 ◉ 3
3	○ Dimensión 2	◉ Dimensión 3	○ 1	◉ 2 ○ 3

CR = 1.9% OK

Cat		Priority	Rank
1	Dimensión 1	21.0%	3
2	Dimensión 2	24.0%	2
3	Dimensión 3	55.0%	1

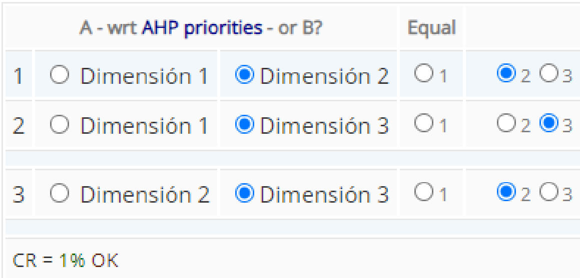
Expert 6

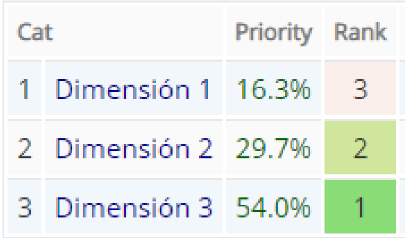
	A - wrt AHP priorities - or B?		Equal	How
1	○ Dimensión 1	◉ Dimensión 2	○ 1	○ 2 ◉ 3 ○ 4
2	○ Dimensión 1	◉ Dimensión 3	○ 1	○ 2 ○ 3 ◉ 4
3	○ Dimensión 2	◉ Dimensión 3	◉ 1	○ 2 ○ 3 ○ 4

CR = 1% OK

Cat		Priority	Rank
1	Dimensión 1	12.6%	3
2	Dimensión 2	41.6%	2
3	Dimensión 3	45.8%	1

Expert 7





Source: Authors' elaboration.

These results are aggregated using the geometric mean to arrive at a new consensus priority vector in table 11.

Table 11. Weighting of Dimensions

Criterion		Comment	Weights
1	Dimension 1	Personalisation of information	16.3%
2	Dimension 2	Confrontation with the given	29.7%
3	Dimension 3	Idea generation	54.0%

Source: Authors' elaboration.

Using formula 3, the scale is determined according to the Cuban evaluation system for Higher Education, as shown in the following table:

Table 12. Quantitative and Qualitative Scale for assessing Creative Learning

Qualitative Scale	Quantitative Scale	
	Initial Value	Final Value
Excellent	0.753	1
Well	0.502	0.752
Regular	0.251	0.501
Bad	0	0.25

Source: Authors' elaboration.

The scale defined enables the assessment of creative learning and determination of the level of development, taking into account the assumed qualitative and quantitative scales. Once these elements are obtained, creative learning can be diagnosed.

In 2018-2019, Cuban higher education underwent a curricular reform that reduced the number of study years to four, known as Curriculum E. In 2023, students graduated from the University of Matanzas and were selected to validate the metric measuring creative learning in computer science. The assessment of creative learning in computer science for the Computer Engineering course is conducted through a questionnaire that is administered to the entire population of 19 students in the year 2023. The fourth year of the degree course was chosen as it is the final year and students are expected to have the highest level of development in creative learning in computer science. This questionnaire was derived from Moreno García (2019) self-assessment questionnaire for creative learning in

ONLY FOR REVIEW

Mathematics, which has been adapted for Computer Science. It takes into account all the aspects that determine the level of the indicator for all students in the Computer Engineering program.

Table 13. Assessment of creative learning in the fourth year

Student	Dimension 1	Dimension 2	Dimension 3	APCI	Qualitative Scale
Est 1	1	1	1	1	Excellent
Est 2	0.5	0.297	0.54	0.343785714	Regular
Est 3	0.5	0.297	0.54	0.571357143	Well
Est 4	0.5	0.297	0.54	0.512728571	Well
Est 5	0.5	0.297	0.54	0.301357143	Regular
Est 6	0.833333333	0.297	0.54	0.710547619	Well
Est 7	0.5	0.297	0.54	0.528928571	Well
Est 8	0.5	0.297	0.54	0.166357143	Bad
Est 9	0.5	0.297	0.54	0.428642857	Regular
Est 10	0.166666667	0.297	0.54	0.154452381	Bad
Est 11	0	0.571428571	0.25	0.304714286	Regular
Est 12	0.333333333	0.714285714	1	0.80647619	Excellent
Est 13	0.333333333	0.571428571	0.5	0.494047619	Regular
Est 14	0.5	0.571428571	0.75	0.656214286	Well
Est 15	0.166666667	0.714285714	0.75	0.644309524	Well
Est 16	0.166666667	0.714285714	0.75	0.644309524	Well
Est 17	0.333333333	0.678571429	0.25	0.390869048	Regular
Est 18	0.333333333	0.571428571	0	0.224047619	Bad
Est 19	0.166666667	0.428571429	0	0.154452381	Bad

Source: Authors' elaboration.

Comment on the number of students
Regularities

- Comprehensive learning is emphasised, where individuals are taught to understand and apply the content they are given.
- However, there is a lack of appreciation for learning methods that foster creativity, such as idea generation and problem-solving.
- The comments provided do not express a sense of confidence, trust, or commitment to learning computer science. Rather, they focus on the usefulness of computer science for other purposes.
- Additionally, project and role-based work is not valued, and mistakes are not seen as a natural part of the learning process. Communication with others is also lacking, and few students create their own projects.
- Furthermore, there is no evidence that the teacher is guiding comprehensive learning towards creative learning.
- There is currently no evidence to suggest that the organisational structures of the degree programme are working towards a transformation from comprehensive learning to creative learning.

ONLY FOR REVIEW

It is important to note that complex variables, such as creativity, are generally measured using multiple instruments to triangulate results and increase reliability. (Agnoli & Mastria, 2022; Betti Frare, Beuren, & da-Silva, 2022). In this case, three methods were used: teacher interview, student survey and participant observation of work practice.

Using the expert method in the same way as for the weights, the weights for each method were obtained as shown in table 11:

Table 14. Weight of Methods

	Criterion	Comments	Weight
1	Interview	Teachers are interviewed to assess students' creative learning.	0.25
2	Survey	The survey allows students to self-assess themselves.	0.25
3	Observation	The observation will be carried out by one of the researchers who teaches the group, so it will be of the participant type and sufficient trust is guaranteed with the students so that they do not feel self-conscious about the presence of outsiders.	0.50

Source: Authors' elaboration

As all measurement criteria are evaluated in each method, the expression 5.1 is used and the results are shown in the following table:

Table 15. Assessment of Creative Learning in the Fourth Year using Three Methods

	Dimension 1	Dimension 2	Dimension 3	APCI	Qualification
Student 1	0.39375	0.244166667	0.31875	0.3088	Regular
Student 2	0.272222222	0.415	0.325	0.3431	Regular
Student 3	0.33125	0.575	0.664583333	0.5836	Well
Student 4	0.706944444	0.575	0.426041667	0.5161	Well
Student 5	0.491666667	0.3775	0.3375	0.3745	Regular
Student 6	0.510416667	0.74	0.6125	0.6337	Well
Student 7	0.61875	0.53	0.395833333	0.472	Regular
Student 8	0.252083333	0.405	0.25	0.2964	Regular
Student 9	0.465972222	0.35	0.5	0.4499	Regular
Student 10	0.2625	0.1425	0.25	0.2201	Bad
Student 11	0.29375	0.3425	0.625	0.4871	Regular
Student 12	0.822222222	0.8075	0.725	0.7653	Excellent
Student 13	0.557638889	0.5	0.391666667	0.4509	Regular
Student 14	0.5	1	0.75	0.7835	Excellent
Student 15	0.622222222	0.19	0.75	0.5629	Well
Student 16	0.333333333	0	0.5	0.3243	Regular

Source: Authors' elaboration.

It is interesting to assess the differences between the results obtained when measured by one instrument or by several instruments. The values would not be equal but the difference should not be significant enough to imply a jump from one qualitative rating to a higher or lower one. For this reason, we took 0.25 as the significant difference because it is the distance between the maximum and minimum value in each interval.

The difference is shown in the table below:

Table 16. Difference of Results with One Method and with nMethods

	APCI nMethods	APCI	Difference
Student 1	0.30882375	0.3438	-0.0349619
Student 2	0.343127222	0.5714	-0.2282299
Student 3	0.58364375	0.5127	0.07091518
Student 4	0.516069444	0.3014	0.2147123
Student 5	0.374509167	0.7105	-0.3360384
Student 6	0.633727917	0.5289	0.10479935
Student 7	0.47201625	0.1664	0.30565911
Student 8	0.296374583	0.4286	-0.1322682
Student 9	0.449903472	0.1545	0.29545109
Student 10	0.22011	0.3047	-0.0846042
Student 11	0.48710375	0.8065	-0.3193724
Student 12	0.765349722	0.494	0.2713021
Student 13	0.450895139	0.6562	-0.2053191
Student 14	0.7835	0.6443	0.13919048
Student 15	0.562852222	0.6443	-0.0814573
Student 16	0.324333333	0.3909	-0.0665357
Student 17	0.22434523	0.224	0.00029761
Student 18	0.224333547	0.1545	0.06988117
Student 19	0.154452381	0.1545	4.7619E-11

Source: Authors' elaboration.

The difference between the results obtained by the three methods and the survey was calculated by subtracting the first from the second. The fact that only 4 students exceeded the expected range suggests a 21% deviation in one evaluation compared to the others. It is hypothesized that this result may be due to students' self-assessment. For the three students with positive values, it suggests an overestimation of their abilities, while for the student with a negative value, it implies an underestimation of their potential. An additional interview with this student revealed that she is a highly demanding individual with ambitious goals and does not fully appreciate the knowledge she gains with the assistance of her teacher or peers, including creativity.

Out of the 19 students, 9 received negative results which suggest that they underestimated their creativity in computing. This is because their self-assessment scores were lower than the scores obtained during the teacher interview. Therefore, it can be concluded that students' self-assessment is often negative.

Conclusions

The proposed metrics enable the assignment of a quantitative value to creativity when expressed in dimensions, indicators, and measurement criteria.

Each dimension and indicator can be calculated independently, allowing for the identification of the most and least developed areas, and enabling targeted improvement. This is one of the advantages of these metrics over other evaluation systems.

The validation of the metrics using expert criteria enabled the collection of expert opinions and correction of formulation errors. This serves as an initial starting point for introducing the metrics to the terminal year group. Applying the metrics in a learning context allows for highly differentiated comparisons between individual students.

References

Agnoli, Sergio, & Mastria, Serena. (2022). Editorial Introduction to the Special Issue "Advances in the Measurement and Evaluation of Creativity". *Creativity Research Journal*, 1-6. doi:10.1080/10400419.2022.2152555

Apaydin, Sezen, & Güven, Sibel. (2022). Pre-Service Teachers' Evaluations of Creativity in Higher Education. *Educational Policy Analysis and Strategic Research, 17*(1), 58-87. doi:10.29329/epasr.2022.248.4

Betti Frare, Anderson, Beuren, Ilse Maria, & da-Silva, Evelise Souza. (2022). Performance Measurement System, Organizational Learning, and Creativity. *BAR − Brazilian Administration Review, 19*, 1-22. doi:10.1590/1807-7692bar2022210099

Canjongo Daniel, Eduardo Pinto, González Hernández, Walfredo, & Becalli Puerta, Laura Helena. (2022). La enseñanza-aprendizaje de la simbología química como parte del lenguaje químico en la Escuela Superior Pedagógica de Bié (ESPB). *Educación Química, 33*(2), 37-49. doi:10.22201/fq.18708404e.2022.2.76864

Chesnokova, Olga. (2014). *Social Creativity in Primary-school Children: How to Measure, Develop and Accept it*. Paper presented at the Third Annual International Conference «Early Childhood Care and Education».

de-Almeida, Pilar, & Mitjáns Martínez , Albertina. (2020). Emergencia del aprendizaje creativo. *Alternativas cubanas en Psicología, 8*(23), 95-111.

Díaz Gómez, Álvaro , & Martínez, Albertina Mitjáns. (2013). Creatividad y subjetividad: su expresión en el contexto escolar. *Diversitas: Perspectivas. Psicológicas, 9*(2), 427-434.

Furtwengler, Scott R. (2021). Development of a creativity orientation scale using EFA. *Journal of Creativity, 31*, 1-6. doi:10.1016/j.yjoc.2021.100004

González-Rey, Fernando. (2019). Methodological and Epistemological Demands in advancing the study of subjectivity. *Culture & Psychology, 0*(0), 1-16.

González Hernández, Walfredo, Petersson Roldán, Maritza, & García Moreno, Marcelina. (2022a). *Aprendizaje creativo de la Informática: conceptualización*. Paper presented at the Universidades 2022, La Habana-Cuba.

González Hernández, Walfredo, Petersson Roldán, Maritza, & García Moreno, Marcelina. (2022b). *El aprendizaje creativo de la informática: definición de una variable*. Paper presented at the XXIII Evento Internacional de la Enseñanza de la Matemática, la Computación y la Estadística, Matanzas-Cuba.

Görlich, Yvonne. (2023). Development of Creative Process Assessment Scale (CPAS). *Journal of Creativity, 33*, 1-12. doi:10.1016/j.yjoc.2023.100042

Havârneanu, Geanina. (2022). *The architecture of a standardised mathematical creativity test for developmental evaluation*. Paper presented at the International Conference on Innovations in Science and Education, Prague-Czech Republic.

ONLY FOR REVIEW

Hsu-Chan, Kuo, Burnard, Pamela, McLellan, Ros, & Ying-Yao, Cheng. (2017). The Development of Indicators for Creativity Education and a Questionnaire to Evaluate its Delivery and Practice. *Thinking Skills and Creativity, 24*, 186-198. doi:10.1016/ j.tsc.2017.02.005

Kanhime-Kasavube, Miguel, & González-Hernández, Walfredo. (2017). Estrategia metodológica para lograr la evaluación desarrolladora de la matemática en la escuela de formación de profesores de Kuando Kubango, Angola. *Revista Góndola, Enseñanza y Aprendizaje de las Ciencias, 12*(1), 73-91. doi:10.14483/udistrital.jour. gdla.2017.v12n1.a5

Katz, Joshua H., Mann, Thomas C., Shen, Xi, Goncalo, Jack A., & Ferguson, Melissa J. (2022). Implicit impressions of creative people: Creativity evaluation in a stigmatized domain. *Organizational Behavior and Human Decision Processes, 169*, 1-14. doi:10. 1016/j.obhdp.2021.104116

Leutner, Franzisk, Yearsley, Adam, Codreanu, Sonia-Cristina, Borenstein, Yossi, & Ahmetoglu, Gorkan. (2017). From Likert scales to images: Validating a novel creativity measure with image based response scales. *Personality and Individual Differences, 106*, 36-40. doi:10.1016/j.paid.2016.10.007

Medina, P., González, W., Robayo, D., López, G., & T. Freire. (2021). *Metric to evaluate virtual courses: case Ecuador*. Paper presented at the XI International Conference on Virtual Campus (JICV), Salamanca-España.

Mitjáns Martínez, Albertina. (2013). Aprendizaje creativo: desafíos para la práctica pedagógica. *CS*(11), 311 - 341.

Molina Hernández, Carlos Rafael, González Hernández, Walfredo, & Cruz Lemus, Gil. (2021). Habilidad modelar procesos dinámicos de control automático. *Educación Química, 32*(1), 100-111. doi:10.22201/fq.18708404e.2021.1.75429

Moreno García, Marcelina Caridad. (2019). El aprendizaje creativo en la matemática, su contribución a la formación del ingeniero industrial. *Atenas, 2*(46), 47-63.

Ott, Michela, & Pozzia, Francesca. (2010). *Towards a model to evaluate creativity-oriented learning activities*.

Pei, Yilai, Han, Jiantao, Zhao, Jingwen, Liu, Mengrong, & Pang, Weiguo. (2022). The Effects of the Creator's Situation on Creativity Evaluation: The Rater's Cognitive Empathy and Affective Empathy Matter in Rating Creative Works. *Journal of Intelligence*, *10*(75), 1-19. doi:10.3390/jintelligence10040075

Penades Pla, Vicent. (2017). *Aplicación de la toma de decisión multicriterio al diseño sostenible de puentes de hormigón.* (Máster Universitario en Ingeniería del Hormigón), Universitat Politècnica de València, Retrieved from https://riunet.upv.es/ handle/10251/89181

Peralta Hernández, Jesús, & Segura, Felipe Tirado. (2023). The Teaching of Collaborative Creativity, a Methodological Approach to Designing the Environment. *Journal of Formative Design in Learning*, -15. doi:10.1007/s41686-022-00073-1

Rey, Fernando González. (2019). Fifty years after L I Bozhovich s personality and its formation in childhood recovering her legacy and her historical role. *Mind, Culture, and Activity*, 1-13. doi:10.1080/10749039.2019.1616210

Romo, Manuela, Alfonso-Benlliure, Vicente, & Sanchez-Ruiz, Maria-Jose. (2016). El test de creatividad infantil (TCI): evaluando la creatividad mediante una tarea de encontrar problemas. *Psicología Educativa, 22*(2), 93-101. doi:10.1016/j.pse.2016.01.005

Roth, Tamara, Conradty, Cathérine, & Bogner, Franz X. (2022). Testing Creativity and Personality to Explore Creative Potentials in the Science Classroom. *Research in Science Education, 52*, 1293-1312. doi:10.1007/s11165-021-10005-x

Saenna, Watcharaporn, & Phusee-orn, Songsak. (2022). The Development of a Scientific Creativity Test in Science for High School Students of Northeastern Princess

ONLY FOR REVIEW

Chulabhorn Science High Schools. *Higher Education Studies, 12*(3), 105-113. doi:10.5539/hes.v12n3p105

Torres Oliveira, Carolina, & Mitjáns Martínez, Albertina. (2020). Expresiones de la subjetividad social en el aula y creatividad en el aprendizaje: un estudio de caso. *Alternativas cubanas en Psicología 8*(23), 126-144.

Vasylenko, Olena, Potapchuk, Yevhen, Karpova, Dariia, Podkorytova, Larysa, Kuleshova, Olena, Mikheieva, Liudmyla, & Varhata, Oksana. (2022). Diagnosis of Features of Creativity Development of Future Psychologists: Gender Aspect. *Revista Românească pentru Educaţie Multidimensională, 14*(4), 408-422. doi:10.18662/rrem/14.4/648

Zada, Muhammad, Khan, Jawad, Saeed, Imran, Zada, Shagufta, & Jun, Zhang Yong. (2023). Curiosity may have killed the cat but it has the power to improve employee creativity. *Current Psychology*. doi:10.1007/s12144-022-04171-y

ONLY FOR REVIEW