

Diferencias regionales en el pensamiento creativo: Un análisis psicométrico transcultural en América

Regional differences in creative thinking: A cross-cultural psychometric analysis in America

Urruticoechea Alar¹

Vernazza Mañan Elena²

Del Callejo Canal Diana Donaji³

Álvarez-Vaz Ramón⁴

¹Departamento de Educación, Universidad Católica del Uruguay, Av. 8 de Octubre 2738, 11600, Montevideo, Uruguay, alar.urruticoechea@ucu.edu.uy, <https://orcid.org/0000-0001-6229-2633>

²Facultad de Ciencias Económicas y de Administración. Universidad de la República, Av. Gonzalo Ramírez 1926, 11200, Montevideo, Uruguay, elena.vernazza@fcea.edu.uy, <https://orcid.org/0000-0003-3123-2165>

³Instituto de Investigación de Estudios Superiores, Económicos y Sociales de la Universidad Veracruzana, Luis Castelazo Ayala s/n. Instituto de Investigaciones y Estudios Superiores económicos y sociales (IISES--UV), 91190, Xalapa, Veracruz. México, ddelcallejo@uv.mx, <https://orcid.org/0000-0003-4753-6577>

⁴Facultad de Ciencias Económicas y de Administración. Universidad de la República, Av. Gonzalo Ramírez 1926, 11200, Montevideo, Uruguay, ramon.alvarez@fcea.edu.uy, <https://orcid.org/0000-0002-2505-4238>

Alar Urruticoechea, email: alar.urruticoechea@ucu.edu.uy

Elena Vernazza Mañan, email: elena.vernazza@fcea.edu.uy

Diana Donaji Del Callejo Canal, email: ddelcallejo@uv.mx

Ramón Álvarez-Vaz, email: ramon.alvarez@fcea.edu.uy

RESUMEN

Este estudio analiza las diferencias regionales en el rendimiento psicométrico de los ítems del dominio de pensamiento creativo de PISA 2022 en América. Mediante el modelo de



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

respuesta graduada multigrupo en el marco de la Teoría de Respuesta al Ítem (TRI), se examinaron los parámetros de discriminación, dificultad e información de 29 ítems en tres regiones culturalmente diferenciadas: América del Norte (Canadá y México), América Central (Costa Rica, Panamá y República Dominicana) y América del Sur (Brasil, Chile, Colombia, Perú y Uruguay). Los resultados muestran variaciones significativas en el funcionamiento de los ítems entre regiones. Por ejemplo, el ítem DT550Q02C2 presenta una mayor discriminación en América Central ($\alpha = 1.13$) que en América del Sur ($\alpha = .81$) o América del Norte ($\alpha = 0.47$). En cuanto a la dificultad, se observan variaciones notables, como en el ítem DT300Q01C2, que alcanza $b = 4.82$ en América Central frente a $b = .03$ en América del Sur. Algunos ítems, como DT400Q01C, mantienen niveles elevados de discriminación en todas las regiones. Otros, como DT620Q01C2, exhiben consistentemente niveles bajos de información y discriminación. Estos hallazgos subrayan la influencia del contexto cultural y educativo en la medición del pensamiento creativo y cuestionan su supuesta invarianza transcultural. Se destaca la necesidad de diseñar evaluaciones culturalmente sensibles, considerando la equivalencia lingüística y las diferencias de interpretación curricular y contextual entre regiones.

Palabras Clave: PISA, Pensamiento Creativo, Teoría de Respuesta al Ítem, Psicometría, Comparación Transcultural, Invarianza de Medición.

Códigos JEL: C10, C38, I21

ABSTRACT

This study analyzes regional differences in the psychometric performance of items from the PISA 2022 creative thinking domain in the Americas. Using the multigroup graded response model within Item Response Theory (IRT), the discrimination, difficulty, and information parameters of 29 items were examined in three culturally distinct regions: North America (Canada and Mexico), Central America (Costa Rica, Panama, and the Dominican Republic), and South America (Brazil, Chile, Colombia, Peru, and Uruguay). The results show significant variations in item functioning across regions. For example, item DT550Q02C2 shows greater discrimination in Central America ($\alpha = 1.13$) than in South ($\alpha = .81$) or North America ($\alpha = 0.47$). Regarding difficulty, notable variations are observed, such as in item DT300Q01C2, which reaches $b = 4.82$ in Central America versus $b = .03$ in South America. Some items, such as DT400Q01C, maintain high levels of discrimination across all regions. Others, such as DT620Q01C2, consistently exhibit low levels of information and discrimination. These findings underline the influence of the cultural and educational context in the measurement of creative thinking and call into question its supposed cross-cultural invariance. The need to design culturally sensitive assessments is highlighted, considering linguistic equivalence and curricular and contextual interpretation differences between regions.

Keywords: PISA, Creative Thinking, Item Response Theory, Psychometrics, Cross-cultural Comparison, Measurement Invariance

JEL Codes: C10, C38, I21

Recibido: 5 de abril de 2026

Aceptado: 28 de mayo de 2026

Publicado: 1 de junio de 2026

1. Introduction

Since the mid-20th century, the concept of "creative thinking" has generated increasing interest, especially since the introduction of divergent thinking as an alternative to rational, logical, or convergent thinking (Guilford, 1950). This approach proposed that human intelligence should be understood as a multidimensional set of skills, with creativity playing a significant role due to its ability to generate multiple ideas from a single stimulus. This ability, linked to cognitive flexibility, allows people to be innovative in problem-solving, adapt to new or uncertain contexts, and produce original solutions in different knowledge domains (Runco & Acar, 2012; Beghetto & Kaufman, 2014). More recent research has confirmed that creative thinking constitutes a key transversal competence for the 21st century, both in educational and professional contexts, due to its role in complex problem-solving, decision-making, and social and technological innovation (Lucas & Spencer, 2017; OECD, 2019).

Since the creation of the Torrance Test of Creative Thinking (TTCT) in 1966 as a tool to measure creative thinking in children and adults (Torrance, 1966; 1974), creative thinking has not only been considered a desirable and trainable individual quality but also a key educational competency for 21st-century students (Kim, 2006). Several studies have supported this conceptual evolution, highlighting that creative thinking can and should be actively fostered in school settings as an essential part of education oriented toward innovation and problem-solving in diverse contexts (Kim, 2006; Lucas & Spencer, 2017). Creative thinking generates original and valuable ideas, allowing problem-solving, expressing concepts, or producing innovations. This cognitive process involves both the generation of novel ideas and their evaluation and improvement to achieve practical solutions (Runco & Jaeger, 2012).

Among the components of creative thinking, the following stand out:

- Fluency: the ability to generate numerous ideas or solutions in response to a single stimulus.
- Flexibility: the ability to approach a problem from different perspectives or conceptual frameworks.
- Originality: the tendency to propose uncommon, unusual, or unique responses.
- Elaboration: the ability to develop, enrich, and elaborate on generated ideas in depth.

In education, the importance of these dimensions lies in their ability to foster deep, transferable, and motivating learning. Fluency allows students to explore multiple alternatives to a problem; flexibility allows them to change strategies to address challenges in innovative ways; originality allows them to move away from conventional solutions; and elaboration allows them to structure ideas with clarity and depth (Plucker, Beghetto, & Dow, 2004). These skills contribute to academic development, critical thinking, and students' preparation for educational, social, and work contexts (Beghetto & Kaufman, 2014).

It is important to note that this skill does not depend solely on the academic field; its acquisition is also related to the social and cultural context to which the student is linked (Craft, 2005; Crossley & Watson, 2003). Therefore, it is important to frame the study of creative thinking within comparative education. In this sense, creative thinking can be conceptualized as an individual cognitive ability and a product of diverse institutional, curricular, and cultural contexts (Crossley & Watson, 2003; Bray, Adamson & Mason, 2014).

1.1 Creative Thinking Framework

There are multiple approaches to measuring and assessing "creative thinking." In particular, the PISA 2022 assessment is based on a conceptual model defining creativity as generating, evaluating, and improving original and practical ideas. This framework recognizes that creativity is not innate but a developable competence shaped by individual and contextual factors. The competency model proposed by the OECD organizes creative thinking into three observable cognitive processes: generating diverse ideas, generating creative ideas, and evaluating and improving ideas. These processes are assessed in four domains: written expression, visual expression, social problem-solving, and scientific problem-solving (Figure 1), all designed to reflect authentic forms of creative thinking in 15-year-old students (OECD, 2021).

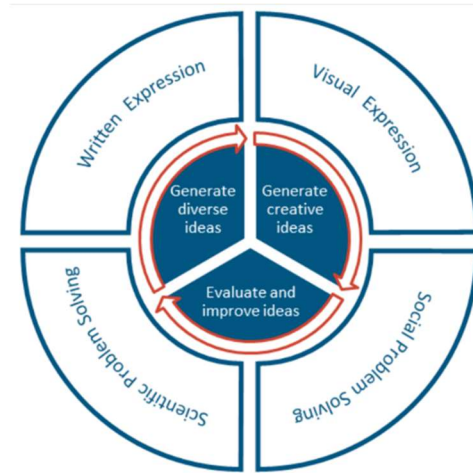


Figure 1. Competency model for the PISA test of creative thinking
Source: OECD, 2021.

Through this approach, PISA does not seek to identify "creative geniuses" but rather to understand how students can mobilize their creative thinking in real-life contexts and how the educational and cultural environment influences this. This broad perspective allows for analyzing performance and the conditions that enable or hinder creativity in different educational systems (OECD, 2021).

1.2 Literature review

Cultural comparability is a central issue in international standardized tests such as PISA and TIMSS, which assess students from diverse linguistic and cultural backgrounds. This heterogeneity of languages, curricula, and contexts has significant consequences for achieving equivalent and fair measurements (Rutkowski & Rutkowski, 2018). Previous research warns that comparing results across countries without ensuring measurement invariance can lead to misinterpretations (Alatlí, 2022). For example, in PISA 2018, certain reading comprehension items were found to show differential functioning (DIF) across countries; some questions favored students from one culture due to greater familiarity with the content, differences in translation, or differences in item format (Alatlí, 2022). These findings underscore the need to carefully adapt items and apply psychometric analyses—such as invariance and bias tests—to ensure that assessments measure the same skills equitably across cultures (Rutkowski & Rutkowski, 2018).

Item Response Theory (IRT) has become a key tool for international assessments, especially when measuring complex or novel competencies. This approach allows items

to be calibrated and scores from different countries to be compared on a standard scale. PISA, for example, uses a generalized partial credit IRT model to construct its achievement scales (OECD, 2024), ensuring consistent measurements across non-traditional domains. In 2022, PISA incorporated creative thinking for the first time as a domain assessed globally, covering more than 140,000 students from over 60 countries (Barbot & Kaufman, 2025). This initiative demonstrated that IRT can model open-ended skills—such as generating and enhancing original ideas in written, visual, or problem-solving contexts—while maintaining international comparability. However, assessing these types of skills introduces methodological challenges (e.g., aggregated scores and complex sample designs that require careful analysis) (Barbot & Kaufman, 2025). Nevertheless, applying IRT models in international contexts allows for more reliable and equitable estimates of student performance in creative and other complex skills, enabling rigorous comparisons across countries (OECD, 2024).

1.2 Objectives and Significance of the Study

Understanding how creative thinking is expressed and measured in different cultural and educational systems is essential to ensuring the validity and equity of international assessments. Although PISA incorporated creative thinking as a domain assessed for the first time in its 2022 edition, few studies have yet to analyze how its items perform psychometrically in culturally diverse regions. This work aims to contribute to empirical and methodological development in this field, specifically focusing on the American continent, characterized by its significant linguistic, curricular, and cultural heterogeneity.

This research aims to analyze the psychometric comparability of items in the creative thinking construct in the Americas, assessing their invariance across regions and comparing discrimination, difficulty, and information parameters to detect potential cases of differential functioning that affect the cross-cultural validity of the measurement.

2. Methods

2.1 Participants

The sample analyzed is based on the results obtained in PISA 2022 (total = 499,843). These data were refined by eliminating all subjects with no responses to the variables of

interest (items associated with the construct "creative thinking"), resulting in a sample of 142,460 participants.

Of these, those belonging to the following three geographic regions were selected:

- South America, comprised of Brazil, Chile, Colombia, Peru, and Uruguay;
- North America, comprised of Canada and Mexico;
- Central America comprises Costa Rica, the Dominican Republic, and Panama.

Finally, the sample consisted of 24,681 participants, of whom 11,223 (45%) were from South America, 8,530 (35%) were from North America, and 4,928 (20%) were from Central America.

2.2 Measures

A total of 29 items were analyzed. Although the competency associated with each item is not yet known, it is understood that all 29 measure the construct of "Creative Thinking."

2.3 Procedure

The PISA 2022 creative thinking test was administered digitally, following standardized procedures defined by the OECD to ensure international comparability (OECD, 2025). The target population consisted of 15-year-old students enrolled at any level from seventh grade onward.

The sample had a multistage stratified cluster sampling design, with a random selection of schools in the first stage and students within each school in the second. This procedure ensures national representativeness within each participating country or economy (OECD, 2023).

However, due to its optional nature, the "creative thinking" assessment was not administered in all participating countries in PISA 2022 but only in 64 countries and economies. The OECD offered this assessment as an innovative domain, and its administration depended on each country's decision. Countries sometimes prioritized other domains or considered logistical, technical, or budgetary constraints (OECD, 2023). Of the total number of participants, three cultural regions were defined for this study for comparative purposes: South America (Brazil, Chile, Colombia, Peru, and Uruguay), North America (Mexico and Canada), and Central America (Costa Rica, Panama, and the Dominican Republic). This definition is based on criteria of geographic proximity and linguistic, cultural, or educational similarities.

2.3 Data analysis

The graded response model (GRM) proposed by Samejima (1969) was used, and it was appropriate for items with ordered categories. The model allows for estimating a discrimination parameter (a) and a series of difficulty thresholds (b) for each item, although only the first threshold (b) is reported in this analysis for simplicity.

The analysis was performed using a multigroup approach, using the mirt package (Chalmers, 2012) of the R software (R Core Team, 2023). A one-dimensional model was specified for each regional group without imposing invariance restrictions (a completely free model). The groups were defined according to the country's cultural region: South America, Central America, and North America. The parameters were estimated using maximum likelihood.

Once the models were estimated, the discrimination (a) and difficulty (b) values were extracted and compared between regions for each item. Descriptive comparisons and item characteristic curves (ICCs) assessed potential differential functioning between regions. No DIF statistical tests (such as DFIT or LRT) were applied, as the analysis was exploratory.

In addition, the maximum information provided by each item in each region was calculated and defined as the point on the ability scale where the item is most accurate.

3. Results

Table 1 presents the estimated discrimination parameter (a) values for each of the 29 creative thinking items included in the PISA 2022 test, differentiated by region: South America, North America, and Central America. This parameter indicates the extent to which an item distinguishes between students with different levels of creative thinking ability.

Table 1. Discrimination parameters (a) by item and region

Item	South America	North America	Central America
DT200Q02C2	.88	.60	1.42
DT240Q01C2	.52	.50	.60
DT420Q02C2	.72	.69	1.10
DT550Q01C	1.12	.96	1.00

DT550Q02C2	.81	.47	1.13
DT690Q01C	1.08	.98	1.16
DT690Q02C2	.83	.56	.99
DT300Q01C2	.51	.43	.12
DT300Q02C	.67	.63	.37
DT400Q01C	1.90	2.01	1.94
DT400Q02C2	1.99	1.79	1.78
DT400Q03C2	1.08	.67	.82
DT700Q01C	1.41	1.42	1.53
DT350Q01C2	1.38	1.15	1.05
DT350Q02C	1.39	1.10	1.38
DT350Q03C2	1.41	1.04	1.05
DT500Q01C	.94	.74	1.41
DT500Q02C2	.53	.53	.47
DT620Q01C2	.39	.30	.49
DT570Q01C	1.58	.33	1.11
DT570Q02C2	.86	.71	.75
DT570Q03C2	1.07	.81	.70
DT610Q01C	.81	.80	.83
DT680Q01C2	.52	.41	.60
DT520Q02C	.75	.69	.44
DT520Q03C2	.52	.39	.74
DT630Q01C2	.97	1.01	.72
DT370Q01C2	1.36	.82	1.15
DT360Q01C2	1.22	.85	1.06

Source: Own elaboration

From the values presented, variations can be identified both between items and between regions. Some items show high discrimination values in all three regions, such as DT400Q01C with $a_{\text{south}} = 1.90$, $a_{\text{north}} = 2.01$, and $a_{\text{center}} = 1.94$, indicating high discriminative capacity in all contexts. Similarly, other items, such as DT700Q01C or DT350QxxC, also have high and similar levels of discrimination across regions.

Conversely, some items show low discrimination values. For example, DT620Q01C2 shows a $a_{\text{south}} = 0.39$, $a_{\text{north}} = 0.30$, and $a_{\text{center}} = 0.49$, suggesting a lower ability to distinguish levels of creative thinking ability, regardless of region.

Also, differences in discrimination are observed between regions for the same item, which could indicate differential behavior. For example, DT550Q02C2 shows a higher value in Central America (1.13) than in South America (0.81) and North America (0.47). A similar pattern can be observed in DT570Q01C and DT300Q01C2.

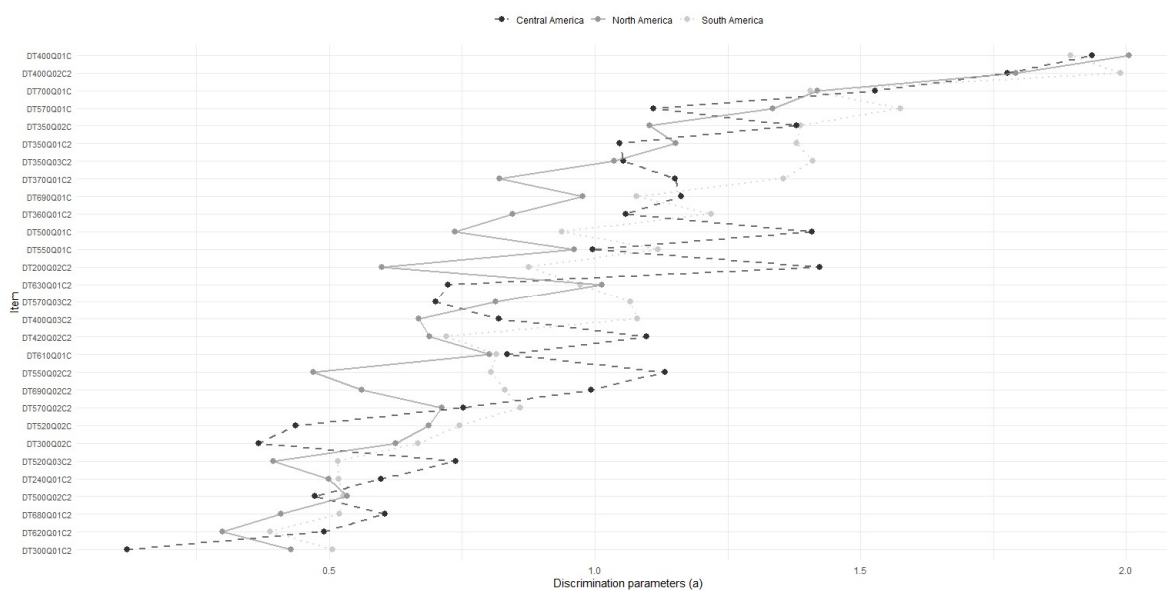


Figure 2. Discrimination parameters (a) by item and region

Source: Own elaboration

Figure 2 shows that items DT300Q01C2 and DT620Q01C2 have low discrimination values in South America, North America, and Central America. This indicates they will not distinguish well between low- and high-performing students in any region. Therefore, their diagnostic or predictive value is limited, and they could be considered for revision or even exclusion, especially if the instrument's objective is to assess performance differences accurately.

At the other extreme, items DT400Q01C, DT400Q02C2, DT700Q01C, DT350Q01C2, and DT350Q02C show high discrimination values in all three regions, indicating that they are highly effective at differentiating between students with different ability levels. Their regional stability suggests they perform robustly and reliably, making them strong candidates for cross-sectional interpretation or international assessment design. Finally, the articles that register the most significant differences by region are DT570Q01C (very

low in North America, high in South America and Central America), DT200Q02C2 (very high in Central America), and DT500Q01C (high in Central America, moderate in others).

Table 2 shows the values of the difficulty parameter (b) for the 29 items in the creative thinking domain assessed in PISA 2022, estimated for each of the three regions analyzed: South America, North America, and Central America. This parameter indicates the skill level required to achieve an average level of performance on each item; that is, the higher the b value, the greater the skill required to respond successfully.

Table 2. Difficulty parameters (b) by item and region

Item	South America	North America	Central America
DT200Q02C2	3.45	3.32	3.23
DT240Q01C2	.86	-.38	1.61
DT420Q02C2	-.44	-1.06	.22
DT550Q01C	1.65	.74	3.68
DT550Q02C2	1.85	2.01	2.39
DT690Q01C	.62	.21	1.82
DT690Q02C2	1.67	2.18	1.97
DT300Q01C2	.03	-.52	4.82
DT300Q02C	-.14	-1.42	.44
DT400Q01C	1.48	.62	1.95
DT400Q02C2	1.54	.74	2.22
DT400Q03C2	1.03	.70	1.79
DT700Q01C	2.18	1.22	2.41
DT350Q01C2	1.08	.39	2.50
DT350Q02C	.30	-.65	.83
DT350Q03C2	.95	.44	1.72
DT500Q01C	.21	-.63	1.19
DT500Q02C2	3.29	2.31	3.80
DT620Q01C2	.24	-.67	.92
DT570Q01C	-.10	-.93	-.04
DT570Q02C2	.10	-.64	.90
DT570Q03C2	.36	-.32	.94
DT610Q01C	1.49	.63	2.11

DT680Q01C2	1.08	.42	1.33
DT520Q02C	.49	.08	.72
DT520Q03C2	.61	-.99	1.08
DT630Q01C2	1.18	.48	2.37
DT370Q01C2	.66	-.44	.96
DT360Q01C2	1.07	-.12	1.94

Source: Own elaboration

Differences in difficulty levels can be observed both between items and between regions. Some items have high b values in all three regions, such as DT200Q02C2 with $b_{\text{south}} = 3.45$, $b_{\text{north}} = 3.32$, and $b_{\text{center}} = 3.23$, indicating that they require high levels of creative thinking skill to be answered correctly. Similarly, DT500Q02C2 shows high values in all regions (between 2.31 and 3.80).

In contrast, several items have low levels of difficulty, such as DT570Q01C ($b_{\text{south}} = -.10$, $b_{\text{north}} = -.93$, $b_{\text{center}} = -.04$), suggesting that they can be answered correctly by students with low skill levels. This can also be observed in items such as DT300Q02C or DT520Q03C2 in the North American region, whose b values are negative.

Furthermore, it can be observed that there are items with unequal difficulty across regions, which could indicate differences in functioning or sensitivity depending on the context. For example, DT300Q01C2 shows a significant difference between South America ($b = .03$), North America ($b = -.52$), and Central America ($b = 4.82$), suggesting that the same task may have very different demands depending on the region. Another case is DT550Q01C, where $b_{\text{south}} = 1.65$, $b_{\text{north}} = .74$, and $b_{\text{center}} = 3.68$, indicating greater relative demand for the item in Central America.

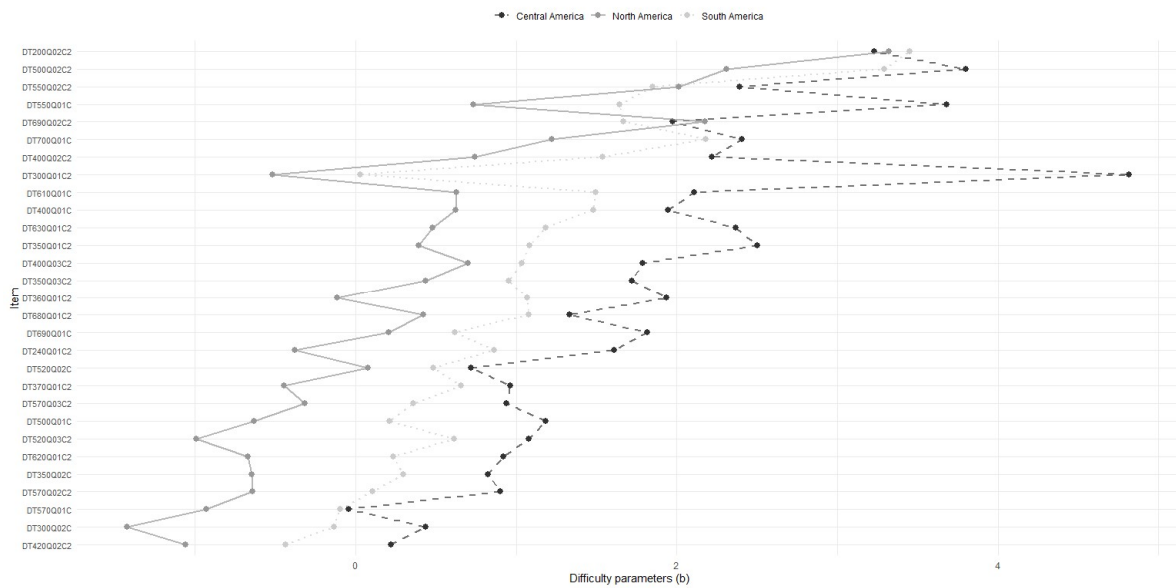


Figure 3. Difficulty parameters (b) by item and region

Source: Own elaboration

Figure 3 provides a visual representation for rapidly identifying regional patterns in item difficulty distribution. Positive difficulty levels predominate in Central America, indicating that, in general, the items are more challenging for this region. South America presents a more balanced distribution, with a greater presence of items of medium difficulty and some of low difficulty. In contrast, North America has more items with relatively low difficulty, suggesting that specific tasks were, on average, more accessible for students in that region. This visualization complements the previous analyses and facilitates the detection of possible asymmetries in item functioning across different educational and cultural contexts.

As previously mentioned, this work also analyzes the maximum information provided by each item in the creative thinking domain.

Table 3 presents the maximum information values for the 29 items in the creative thinking domain assessed in PISA 2022, broken down by region: South America, North America, and Central America. This parameter indicates the information an item provides to estimate students' skill level accurately. The higher the value, the more useful the item is for discriminating between students with different proficiency levels.

Table 3. Maximum information per item and region

Item	South America	North America	Central America
------	---------------	---------------	-----------------

DT200Q02C2	.19	.09	.51
DT240Q01C2	.07	.06	.09
DT420Q02C2	.13	.12	.30
DT550Q01C	.31	.23	.25
DT550Q02C2	.16	.06	.32
DT690Q01C	.29	.24	.34
DT690Q02C2	.17	.08	.25
DT300Q01C2	.06	.05	.01
DT300Q02C	.11	.10	.03
DT400Q01C	.90	1.01	.94
DT400Q02C2	.99	.80	.79
DT400Q03C2	.29	.11	.17
DT700Q01C	.49	.50	.58
DT350Q01C2	.48	.33	.27
DT350Q02C	.48	.30	.48
DT350Q03C2	.50	.27	.28
DT500Q01C	.22	.14	.50
DT500Q02C2	.07	.07	.06
DT620Q01C2	.04	.02	.06
DT570Q01C	.62	.45	.31
DT570Q02C2	.19	.13	.14
DT570Q03C2	.28	.17	.12
DT610Q01C	.17	.16	.17
DT680Q01C2	.07	.04	.09
DT520Q02C	.14	.12	.05
DT520Q03C2	.07	.04	.14
DT630Q01C2	.24	.26	.13

DT370Q01C2	.46	.17	.33
DT360Q01C2	.37	.18	.28

Source: Own elaboration

Some items show high levels of informativeness across all three regions, suggesting they are especially effective at capturing variations in student performance. For example, item DT400Q01C stands out with values of .90 in South, 1.01 in North, and .94 in Central, making it one of the most informative items in the set. Similarly, DT400Q02C2 shows high levels of informativeness across all regions (.99 in the South, .80 in the North, and .79 in the Central), indicating high measurement precision. Other items, such as DT700Q01C and DT350Q01C2, also show moderately high values, especially in South America.

Conversely, some items provide relatively little information, limiting their usefulness in accurately measuring student ability. For example, DT300Q01C2 and DT300Q02C show very low values across all regions, particularly Central America (.01 and .03, respectively), suggesting that these items may not be well calibrated for that population. DT620Q01C2 and DT680Q01C2 also present low levels of information ($\leq .09$), which reduces their contribution to the measurement. Of note is item DT500Q02C2, which, despite being highly difficult, presents low levels of information across all regions (around .06–.07), suggesting that it does not adequately discriminate between students.

Regional differences are also observed in the information provided by some items, which could reflect contextual or cultural variations in their functioning. For example, DT500Q01C shows a higher value in Central America (.50) compared to South America (.22) and North America (.14), which could indicate a greater adaptation of the item to the educational content or experiences of that region. Item DT520Q03C2 provides more information in Central America (.14) than North America (.04), suggesting differential sensitivity. Other items, such as DT370Q01C2 and DT360Q01C2, are also more informative in South and Central America than in North America, which could be related to differences in familiarity, relevance, or solution style.

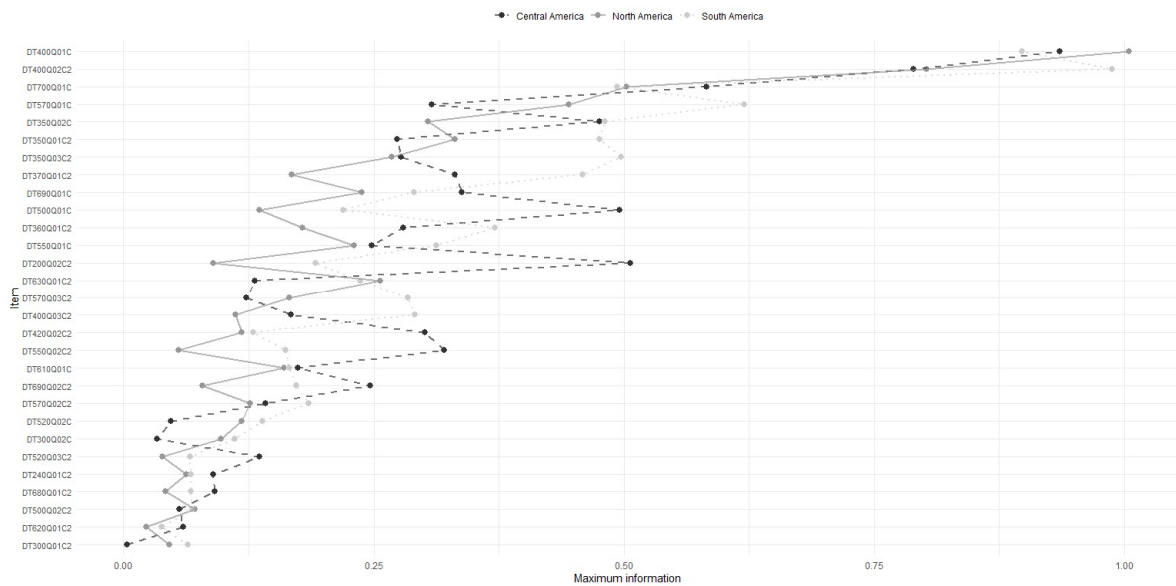


Figure 4. Maximum information per item and region

Source: Own elaboration

From Figure 4, three important patterns can be identified. First, the items that consistently exhibit low values across all three regions—indicating a poor ability to provide helpful information in any context—are DT300Q01C2, DT300Q02C, DT620Q01C2, DT520Q02C, DT520Q03C2, DT680Q01C2, and DT240Q01C2, all with values below 0.15 in at least two regions, and very low in the third. This indicates that these items cannot discriminate effectively across the ability continuum, making them candidates for revision or removal. In contrast, items demonstrating high and consistent informativeness in the three regions stand out, including DT400Q01C, DT400Q02C2, and DT700Q01C, with values equal to or greater than 0.5 in all contexts. These items are particularly robust and functional in cross-regional assessments, as they maintain their informative capability regardless of the geographic context. Their inclusion in international tests would be appropriate and reliable.

Finally, items exhibiting significant regional differences are noted, suggesting sensitivity to cultural or curricular context. For example, DT200Q02C2 has a high value in Central America (0.51) but a much lower value in North America (0.09); DT550Q02C2 shows a disparity between Central America (0.32) and North America (0.05); and DT500Q01C is highly informative in Central America (0.50) but considerably less so in North America (0.13). This unequal behavior may compromise comparative equity between regions and

indicates that these items require specific calibration or differentiated interpretation depending on the context in which they are applied.

4. Discussion

The creative thinking assessment implemented in PISA 2022 was administered to 15-year-old students in 64 countries using 32 open-ended tasks distributed across domains such as written expression, visual expression, social problem-solving, and scientific problem-solving (OECD, 2024). These tasks were designed to measure processes such as generating original and diverse ideas and evaluating and improving those ideas (Barbot & Kaufman, 2025). The psychometric analysis carried out in this research showed that the parameters analyzed (difficulty, discrimination, and information) vary across regions of the American continent. This suggests that the items do not function equivalently in all cultural contexts and may reflect the influence of local social, educational, and/or linguistic factors on the measurement of creative thinking (OECD, 2023).

Some items were easier or more informative in regions where the cultural and pedagogical context is more aligned with the content and statements of those specific tasks. In contrast, in other regions, these items showed lower discriminatory power, suggesting possible differences in the interpretation of the instructions, familiarity with the assessment formats, or the cultural value attributed to the proposed activities (Rutkowski & Rutkowski, 2018).

The OECD attempted to maximize cross-national comparability by defining creative thinking based on cognitive processes considered universal (generating, evaluating, and improving ideas) and by designing grading systems based on the originality and diversity of responses rather than their subjective quality (OECD, 2021 OECD, 2024). However, the results of this study suggest that the methodology used does not eliminate differences derived from the cultural context. Previous research has warned that educational environments with a strong emphasis on standardization tend to offer fewer opportunities for students to develop and express creative thinking in the classroom, which can negatively influence their performance on tasks of this nature (OECD, 2024). In contrast, in educational systems that explicitly promote originality, autonomy, and experimentation, identical items tend to show higher levels of discrimination and provide greater information. This pattern is consistent with the idea that cultural orientation (individualism vs. collectivism), teaching practices, or creative activities in the

curriculum significantly influence how students approach and solve these types of items (Barbot & Kaufman, 2025).

In line with this evidence, it has been pointed out that reporting a single aggregate score for creative thinking can hide substantive differences in the manifestation of the construct between countries or regions. Barbot and Kaufman (2025) warn that this strategy, although useful for public communication, may be insufficient to capture the complexity of the construct and its cultural dimensions. In this sense, the results obtained in this study support the need to move toward a more exhaustive analysis of each item to improve understanding of how creative thinking manifests itself in different contexts.

These results align with the methodological caveats noted by Barbot and Kaufman (2025), who emphasize the need to conduct item-level analyses and consider the cultural context when interpreting results. Along these lines, the present study follows the methodological recommendations proposed by the OECD (2024), using Item Response Theory models and reporting key parameters such as discrimination, difficulty, and information. However, as recent literature suggests, the importance of moving toward more robust designs that include formal invariance tests and more comprehensive reporting strategies is recognized.

4.1. Limitations

This study has some methodological limitations that should be considered when interpreting the results. First, the analysis was exploratory, and it was not statistically verified whether the items mediate creative thinking in the same way across regions. Therefore, it cannot be assured that the observed differences are due solely to variations in ability and not to differences in how the items function in different contexts. Furthermore, like any study based on secondary data, the analysis depends on the methodological decisions made by PISA, including sample design and how open-ended responses are coded. Barbot and Kaufman (2025) caution that these factors can influence the results if not adequately controlled. Finally, although PISA applies rigorous linguistic adaptation and cross-cultural review processes (OECD, 2024), cultural biases may persist in some items, especially tasks requiring creative interpretation or specific cultural references. In addition, PISA has not released all the items used in the assessment, so it is not possible to know with certainty which specific dimensions each item loads onto. This limits the structural analysis of the construct and restricts the possibility of interpreting the results at the subdimension level of creative thinking.

4.2. Future Steps

This exploratory study opens up several lines of research to further analyze creative thinking in international assessments. Applying formal tests of invariance and differential item functioning (DIF) would be an important first step. This would allow us to verify whether the differences between regions reflect true variations in abilities or item biases. Likewise, it is necessary to link psychometric results with item content and the domains assessed to identify task characteristics that could favor or hinder the expression of creativity depending on the cultural context.

Furthermore, future research could incorporate contextual variables such as gender, socioeconomic status, self-efficacy, or curricular exposure to creative activities to understand better which factors influence performance. Comparative analyses across educational systems are also suggested, considering the role of pedagogical policies and practices in fostering creative thinking. Together, these advances will contribute to a more valid, equitable, and culturally informed assessment of creativity in diverse educational contexts-curricular and contextual interpretation differences between regions.

4.3. Conclusions

The results suggest that creative thinking is not manifested equally in all the regions analyzed. The differences observed in the items could indicate that the creative thinking construct may be influenced differently depending on each region's cultural, educational, and linguistic context.

These results are consistent with the arguments made by Barbot and Kaufman (2025), who warn that relying solely on a global index of creative thinking can be misleading when analyzing a multidimensional skill. Therefore, culturally, the items in the creative thinking test must be adapted. The statements should be reviewed with the participation of local experts to ensure their clarity, relevance, and contextual relevance. At the same time, the scoring criteria could be enriched with examples tailored to each culture. Beyond linguistic adaptation, moving toward a more contextualized assessment of transversal competencies is necessary. This would entail recognizing that educational systems value and teach creativity differently and that these differences should be reflected in the

assessment instruments. Only in this way will it be possible to measure creative thinking fairly, meaningfully, and comparatively internationally.

Ultimately, these findings could offer valuable insights for teacher training and the development of educational policy. Incorporating formative creativity assessment strategies into professional development programs, along with nurturing learning environments that cover various creative domains—such as verbal, visual, scientific, and social expression—could help align teaching practices with a broader and more culturally relevant view of creativity.

Referencias Bibliográficas

- Alathi, B. (2022). An investigation of cross-cultural measurement invariance and item bias of PISA 2018 reading skills items. *International Online Journal of Education and Teaching*, 9(3), 1047–1073. <https://files.eric.ed.gov/fulltext/EJ1351913.pdf>
- Barbot, B., & Kaufman, J. C. (2025). *PISA 2022 creative thinking assessment: Opportunities, challenges, and cautions*. *The Journal of Creative Behavior*, 59(1), e70003. <https://doi.org/10.1002/jocb.70003>
- Barbot, B., & Kaufman, J. C. (2025). PISA 2022 creative thinking assessment: Opportunities, challenges, and cautions. *Journal of Creative Behavior*, 59(1), e70003. <https://doi.org/10.1002/jocb.70003>
- Beghetto, R. A., & Kaufman, J. C. (2014). Classroom contexts for creativity. *High Ability Studies*, 25(1), 53–69. <https://doi.org/10.1080/13598139.2014.905247>
- Bray, M., Adamson, B., & Mason, M. (2014). *Comparative education research: Approaches and methods* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-319-05594-7>
- Chalmers, R. P. (2012). *mirt: A multidimensional item response theory package for the R environment*. *Journal of Statistical Software*, 48(6), 1–29. <https://doi.org/10.18637/jss.v048.i06>
- Craft, A. (2005). *Creativity in schools: Tensions and dilemmas*. Routledge.

- Crossley, M., & Watson, K. (2003). *Comparative and international research in education: Globalisation, context and difference*. Routledge.
<https://doi.org/10.4324/9780203452745>
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5(9), 444–454.
<https://doi.org/10.1037/h0063487>
- Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, 18(1), 3–14.
https://doi.org/10.1207/s15326934crj1801_2
- Lucas, B., & Spencer, E. (2017). *Teaching creative thinking: Developing learners who generate ideas and can think critically*. Crown House Publishing.
- OECD. (2019). *OECD learning compass 2030: A series of concept notes*. OECD Publishing. <https://www.oecd.org/education/2030-project/>
- OECD. (2021). *PISA 2021 creative thinking framework* (Third draft). OECD Publishing.
<https://www.oecd.org/pisa/publications/pisa-2021-creative-thinking-framework.pdf>
- OECD. (2023). *PISA 2022 assessment and analytical framework*. OECD Publishing.
<https://doi.org/10.1787/dfe0bf9c-en>
- OECD. (2024). *PISA 2022 results (Volume III): Creative minds, creative schools*. OECD Publishing. <https://doi.org/10.1787/765ee8c2-en>
- OECD. (2025). *Creative minds in action: Students' imagination and ideas in storytelling, design and problem-solving tasks on the PISA test* (OECD Education Policy Perspectives No. 120). OECD Publishing. <https://doi.org/10.1787/5b718eb2-en>
- Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39(2), 83–96.
https://doi.org/10.1207/s15326985ep3902_1

- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, 24(1), 66–75. <https://doi.org/10.1080/10400419.2012.652929>
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, 24(1), 92–96. <https://doi.org/10.1080/10400419.2012.650092>
- Rutkowski, L., & Rutkowski, D. (2018). Improving the comparability and local usefulness of international assessments: A look back and a way forward. *Scandinavian Journal of Educational Research*, 62(3), 354–367. <https://doi.org/10.1080/00313831.2016.1261044>
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *ETS Research Bulletin Series*, 1968(1), i–169. <https://doi.org/10.1002/j.2333-8504.1968.tb00153.x>
- Torrance, E. P. (1966). *Torrance tests of creative thinking: Norms technical manual research edition — Verbal tests, forms A and B; figural tests, forms A and B*. Personnel Press.