Research Article

Quantifying the Standard Definition of Creativity: A Weighted Geometric Mean Framework for Context-Sensitive Assessment

Minh Kim[1], **Mark A. Runco**[2,3]

1. Graduate School of Education and Human Development, University of Washington, United States; 2. Radical Creativity, Aalto University, Helsinki, Finland; 3. Global Institute of Creative Thinking – Director, Global Institute of Creative Thinking, Switzerland

At the time of writing, the article 'The Standard Definition of Creativity' remains the most-cited paper in the *Creativity Research Journal*—one of the leading journals in creativity research—and has accumulated thousands of Crossref-indexed citations[1]. While the qualitative standard definition—requiring both originality and effectiveness—has achieved wide acceptance , translating this qualitative definition into a quantitative scoring rule has proven to be methodologically challenging: common additive models allow compensation between criteria, while simple unweighted multiplicative models impose rigid importance assumptions and can introduce score-compression artifacts. This paper proposes a weighted geometric mean as a practical scoring framework for product-based creativity, preserving the intended joint-necessity ("veto") logic[2] while enabling context-sensitive weighting of component criteria. This weighted geometric mean framework is presented first as a two-component model combining originality and utility, and then generalized to an n-component formulation that can incorporate additional dimensions discussed in the literature (e.g., surprise, intentionality, authenticity) when warranted by evaluative goals. By separating (i) the definition of component indices from (ii) the weighting structure used to aggregate them, the proposed model offers a flexible, interpretable, and extensible approach to product-based creativity assessment across domains and contexts.

Corresponding author: Minh Kim, minh.kim@gwu.edu

Introduction: The measurement problem in creativity

The definition of creativity has long rested on the "Standard Definition", which evaluates creative products based on a bipartite requirement: originality (or novelty) and effectiveness (utility, fit, or value) [1]. While these qualitative criteria are commonly cited, accurately translating them into a functional quantitative formula has lacked widespread consensus[3][2][4]. This lack of consensus is severely exacerbated by the proliferation of scattered measurement tools, creating a methodological crisis where qualitative definitions are rarely translated into a unified mathematical model[5][6][7]. Consequently, the field remains fractured between additive and multiplicative scoring indices, each presenting distinct psychometric trade-offs[5]. While additive equations yield highly stable assessments, they theoretically fail to enforce the necessary coexistence of core criteria, allowing high originality to inappropriately compensate for low utility[5][3][2]. Conversely, multiplicative models enforce a strict “veto power”[2] that better captures the heavy-tailed variance of extreme creative outliers, but at the cost of some internal reliability[5]. Establishing a valid assessment foundation therefore requires a mathematical translation that effectively balances these competing psychometric properties. This undertaking is especially urgent today, as the rapid advancement of Artificial Intelligence complicates how we define and evaluate creative outputs[8]. Therefore, the primary aim of this paper is to establish the precise mathematical relation between the components of the standard definition, specifically for assessing creative products. Because a standardized, quantitative scale is necessary to establish the validity of empirical findings[9], this paper proposes a weighted geometric formula to deliver a universally reliable metric for modern assessments.

Attempts to mathematically formalize creativity scores have evolved from additive models to simple multiplicative combinations[3][2][10], both of which possess mathematical and conceptual flaws:

The failure of additive models

Traditional additive models (e.g., *C (Creativity) = O (Originality*[1]*) + U(Utility)*) fail because they cannot account for the necessary coexistence of novelty and utility[3][2]. Under an additive model, a highly original but completely useless product (e.g., an airplane made of cinderblocks) or a highly useful but completely unoriginal product (e.g., a reinvented wheel) would erroneously achieve a high creativity score simply by maximizing either originality or utility[3]. In such cases, the other variable loses its explanatory meaning.

 doi.org/10.32388/MQEOUV

Limitations of simple multiplicative creativity scores

Simonton and Ting[11] initially modeled creativity (C) using a multiplicative combination of novelty (N) and usefulness (U): $C = N \times U$. Simonton[3] later formalized a three-criterion definition by incorporating surprise into the standard novelty–utility framework ($C = N \times U \times S$, where \(C\\)is creativity, N is novelty, \(U\\)is utility, and \(S\\)is surprise, each ranging from 0 to 1). This expanded view has also been supported by other scholars. For example, Boden[12][13] emphasized a threefold conception of creative ideas in which they must be novel, valuable, and surprising, and work by Acar and colleagues[14] likewise identifies surprise as an important dimension of creativity.

While the multiplicative structure appropriately eliminates the compensation problem inherent in additive models, the simple unparameterized product introduces three measurement limitations that warrant consideration:

i. Rigid symmetry and the absence of empirical weighting

The plain product assumes symmetric importance across all dimensions – no component can be upweighted or downweighted – imposing a fixed structure that cannot be calibrated to context. Yet the relative valuation of novelty versus utility shifts drastically across cultures[15] and domains[10]. For instance, modern European individualistic cultures emphasize novelty, whereas traditional Chinese collectivistic cultures place a higher premium on usefulness[3]. Similarly, pure scientific research places a premium on originality and surprise, whereas applied research heavily weights utility[10]. Therefore, a fixed unweighted structure acts as an ungrounded normative constraint rather than an empirically defensible measurement choice.

ii. Geometric Compression on the Unit Square

When originality and utility are operationalized as bounded indices on [0,1][3], the simple product maps the unit square $[0,1]^2$ onto [0,1] with pronounced nonlinear compression. Even moderate component ratings shrink mechanically: for example, $O = U = 0.5$ yields $C = 0.25$. Interpreted as a composite rating on the same [0,1] scale, this is counterintuitive – two midscale component scores are converted into a bottom-quartile overall score, effectively halving the midpoint rather than preserving it. More generally, the iso-creativity sets $\{(O, U) : O \times U = k\}$ are rectangular hyperbolas that crowd toward the upper-right corner of the square, implying that only a small region of the input space attains high

doi.org/10.32388/MQEOUV

composite scores while a disproportionately large region maps to low scores. Under the illustrative baseline in which O and U are drawn independently and uniformly from [0,1], the expected value is $\mathbb{E}[C] = 0.25$ and the median is approximately 0.187, both well below the midpoint of the scale. *Indeed, 84.7% of all uniform input combinations produce $C \leq 0.5$, confirming that the compression is both severe and asymmetric.* This systematic compression reduces discriminability among typical cases and complicates calibration, making the unweighted product a poor psychometric scoring function when O and U are intended as midscale component ratings.

iii. Violation of Idempotence ("Fair Midpoint" Property)

A well-behaved aggregation function for same-scale component ratings is commonly expected to satisfy *idempotence*: if all inputs equal the same value v, then the aggregate should also equal v. The simple unweighted product produces $f(t,t) = t^2$ (and $f(t,t,t) = t^3$ for three components), implying systematic deflation except at the boundary values $t \in \{0,1\}$. This is not a superficial limitation – it provides an axiomatic explanation for the midpoint deflation illustrated above and shows that the product rule shifts the composite scale downward, thereby distorting the interval interpretability of scores and undermining calibration when component scores are meant to be combined into an overall creativity rating on a shared evaluative scale.

The Proposed Weighted Multiplicative Model

To correct the systemic artifacts of simple multiplication while retaining the crucial interaction between variables, we propose a normalized, weighted geometric mean:

The proposed framework formalizes creativity (C) as a weighted geometric mean of utility (U) and originality (O). By defining the relationship as: $C = U^{\alpha} \times O^{\beta}$, the model asserts that creativity is non-existent if either constituent dimension is absent; a high degree of utility cannot compensate for a total lack of originality, and vice versa.

Where:

- $\alpha + \beta = 1$
- α represents the weight of Utility
- β represents the weight of Originality
- U and O are measured along continuous ratio scales on [0, 1].

 doi.org/10.32388/MQEOUV

- *Special Case:* Equal Weighting (α = β = 0.5)

When no prior empirical basis exists for privileging one dimension over the other, the natural default is equal weighting: *α = β = 0.5*. In this case, the formula reduces to:

$$C = U^{0{,}5} \times O^{0{,}5} = \sqrt{(O \times U)}$$

This is the geometric mean of *O* and *U*, which carries a clear interpretation: each dimension exerts equal proportional influence on the composite. Under this parameterization, the formula satisfies *idempotence* – when *O = U = t*, the score is $\sqrt{(t \times t)} = t$ – and it preserves the critical veto effect: when either dimension is zero, the score collapses to zero. Compared to the simple product (which yields v^2 for uniform inputs), the geometric mean returns the intuitively correct midpoint. For instance, a product with *O = U = 0.5* scores $\sqrt{(0.25)} = 0.5$ rather than *0.25*. This default provides a practical starting point for assessment contexts where the relative priority of originality and utility has not been empirically calibrated.

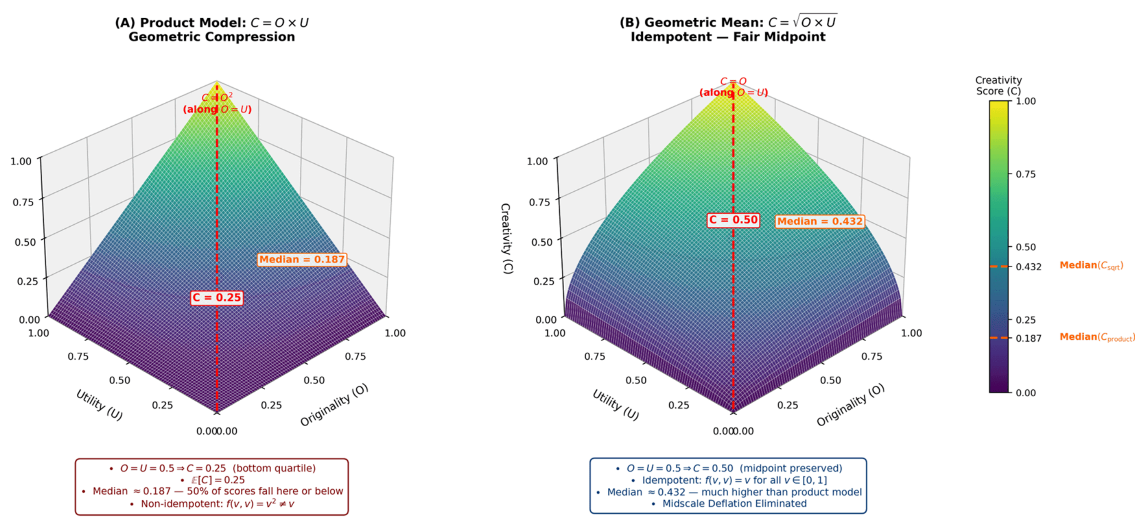

Figure 1. *Comparison of Two Multiplicative Creativity Scoring Functions: Product Model Versus Geometric Mean Model.*

Note. Three-dimensional surface plots of (A) the product model C = O × U and (B) the geometric mean model C = √(O × U) over the unit square $[0,1]^2$. Both panels use identical axes, viewing angle, and color scale. The dashed diagonal shows model behavior along O = U. Red markers indicate the midscale input O = U = 0.5, and orange contours mark the median score under each model. Under the illustrative baseline where O and U are independently drawn from a uniform distribution on [0,1], the median creativity score

doi.org/10.32388/MQEOUV

is approximately 0.187 for the product model but 0.432 for the geometric mean model, illustrating how the geometric mean substantially reduces the midscale compression produced by the simple product.

Discussion

Shift toward consensual product-based creativity in cross-cultural context

Although Simonton's original multiplicative formulation explicitly accommodated both subjective and objective evaluations[3], his subsequent refinements are often strictly bounded to the psychological parameters of personal creativity[2][10]. In contrast, the present adaptation shifts the focus toward a consensual, product-centered definition (which is "clearly most useful for empirical research in creativity"[16]) within a cross-cultural framework: The application of exponents (α, β) as weights allows for the integration of specific cultural or industrial priorities–such as an East Asian context that might weight Utility at 70% ($\alpha = 0.7$) and Originality at 30% ($\beta = 0.3$). The geometric structure ensures the final score remains scale-invariant and representative of the synergistic interaction between the two variables.

By restricting this geometric framework specifically to the evaluation of creative products, this model adheres to Green et al.'s[17] directive to explicitly parse the '4 P's' of creativity. As Green et al. note, products and processes are ontologically distinct constructs that must be operationalized separately; consequently, a mathematical model defining the attributes of a creative product must not be conflated with the cognitive processes that generated it.

Extending the Formula for Additional Product Criteria

While the two-component model captures the core of the standard definition, several scholars have proposed additional criteria that may be relevant when assessing creative products. This section shows how the weighted geometric framework naturally accommodates such extensions without altering its mathematical properties.

As noted above, several scholars have proposed *surprise* as an important component alongside originality and utility, including Boden[12][13], Simonton[3], and Acar et al.[14]. Runco[8] proposed that *intentionality* and *authenticity* be considered when evaluating products, particularly to distinguish genuinely human-originated works from AI-generated outputs. Weisberg[18] also argued for *intentional novelty* as the sole definitional criterion. Each of these represents a potential additional dimension for product assessment.

doi.org/10.32388/MQEOUV

To incorporate such dimensions, the formula generalizes cleanly. For example, we can use *S* to represent Surprise and *I* to represent Intentionality (or Authenticity). The expanded formula becomes:

$$C = O^{\alpha} \times U^{\beta} \times S^{\gamma} \times I^{\delta}$$

Where $\alpha + \beta + \gamma + \delta = 1$ and $\alpha,\beta,\gamma,\delta > 0$. The constraint that exponents sum to 1 preserves idempotence and scale invariance regardless of how many components are included.

The specific weights assigned to each component would depend on the evaluative context, for instance:

- In pure scientific discovery contexts, originality and surprise may carry greater weight, whereas in applied research, creative combinations may load much higher on utility (e.g., the invention of a new vaccine)[10].
- A survey of the fine arts might heavily weight Originality (α) and Intentionality/Authenticity (δ) while reducing Utility to near zero.
- Different cultural contexts may assign different priorities to each component[15].

It is important to note that evaluators are not required to adopt all four components. The two-component model (*O* and *U*) remains the default, faithful to the standard definition. Additional dimensions are available when the assessment context warrants them.

Generalizing to n Components

Because debates surrounding the definition of creativity continue to evolve, a functional quantitative formula must be infinitely extensible. Generalizing to n components, let the components be $x_1, x_2, \ldots, x_n$ (e.g., utility, originality, surprise, intentionality), and let their importance weights be $w_1, w_2, \ldots, w_n$ (each weight \(\left. \ w_{i} \in (0,1 \right\rbrack,\\) and they are not necessarily summing to 1).

Define the sum of the weights as:

$$W = \sum_{i=1}^{n} w_i$$

Then the normalized multiplicative (weighted geometric mean) creativity score is:

$$C = \left(\prod_{i=1}^{n} {x_i}^{w_i}\right)^{\frac{1}{W}}$$

Equivalently, normalize the weights $\alpha_i = \frac{w_i}{W}$ so $\sum \alpha_i = 1$, giving the finalized operational formula:

$$C = \prod_{i=1}^{n} {x_i}^{\alpha_i}$$

This formulation ensures that the three desirable mathematical properties hold for any number of components: (1) the veto power—if any $x_i = 0$, then $C = 0$; (2) idempotence—if all $x_i = t$, then $C = t$; and (3) scale invariance—the score remains on [0, 1] regardless of how many components are included.

Practical Implications

Application Across Rating Modalities

The weighted geometric model is versatile and can be applied across different product-evaluation contexts. Human raters, for instance, can evaluate each component using tools like the Consensual Assessment Technique (CAT)[16]. The CAT is based on the assumption that a panel of independent raters—persons who have not had the opportunity to confer with one another and who have not been trained by the researcher—are best able to make such judgments[16].

To account for Abraham's[19] observation that external evaluations cannot legitimately substitute for the internal experience of the creator, assessments aiming to capture personal fulfillment can seamlessly replace the term 'Utility' with 'Satisfying'. This substitution does not compromise the quantitative model; in fact, Simonton[4] explicitly replied to Abraham by noting that his utility parameter (u) inherently calculates 'total satisfaction' and is therefore 'not incompatible' with her framework. Consequently, the geometric formula accommodates distinct evaluation matrices where each component can be independently assessed from either the creator's internal frame of reference or society's external frame of reference[19][3][4]

Computational Assessment of Products

AI and computational tools are well-suited to evaluate specific components of this formula objectively:

i. Automated Scoring of Originality (O)

As Weisberg[20] noted, computer algorithms can objectively quantify novelty by mapping the similarity space of thousands of digitized works (e.g., paintings or musical compositions) to calculate a novelty score for a new product based on its deviation from a historical database. This objective novelty score can then be directly integrated into the formula proposed herein as the parameter for Originality (O).

doi.org/10.32388/MQEOUV

ii. Computational assessment of the Utility (U) and Originality (O) parameters

To automate the assessment of creativity in open-ended, constraint-based mathematical tasks, Kim and Nguyen[21] proposed a computer-implemented system that independently computes variables for Utility (U) and Originality (O) on a scale. Utility is evaluated through a strict two-stage process: first, a "feasibility gate" uses a computer algebra system (CAS) or constraint solver to check if the submission violates any required mathematical constraints; if it does, the expression is deemed infeasible and strictly assigned . If the expression is feasible, is computed as the percentile rank of its optimization performance relative to a baseline of other feasible solutions. Conversely, Originality (O) is scored by normalizing the submission into a canonical mathematical representation and comparing it against a customizable reference baseline (e.g., a historical corpus or a time-stamped classroom snapshot). If the canonical expression perfectly matches an existing solution in the corpus, it receives ; otherwise, the system calculates a statistical rarity or structural dissimilarity metric to assign a novelty score. Finally, the model mathematically enforces the standard definition of creativity by combining these variables via a geometric mean ($C = \sqrt{(O \times U)}$), ensuring that an output must be both fundamentally viable and mathematically distinct to yield a positive creativity score.

iii. Distinguishing Human, AI, and Collaborative Creativity

When the extended formula is applied to a purely AI-generated product, the Intentionality/Authenticity (I/A) parameter drops to zero, as computational algorithms fundamentally lack autonomous creative intent and authentic expression. Because the geometric model is multiplicative, this zero triggers a complete veto effect, collapsing the overall creativity score to 0. This mathematical outcome perfectly operationalizes Runco's[8] theoretical conclusion: no matter how novel or effective a machine's output may be, its lack of authentic intentionality classifies it strictly as 'artificial creativity' rather than genuine creativity. In contrast, in human-AI collaborative products—where a human provides the authentic direction and problem-finding initiative while the AI merely assists with execution—the score remains intact. Because the human collaborator contributes the autonomous, mindful decisions that characterize genuine intentionality, the overall score does not collapse to zero, accurately reflecting the presence of true creative agency.

doi.org/10.32388/MQEOUV

Conclusion

The proposed weighted geometric mean offers a practical, mathematically disciplined scoring tool for assessing product-based creativity within the standard definition. By correcting the systematic compression and idempotence violations of simple multiplicative models, the formula provides scores that are both interpretable and well-behaved on the unit interval. The equal-weighting special case ($C = \sqrt{(O \times U)}$) supplies a transparent default, while the parametric flexibility of the exponents ($C = U^{\alpha} \times O^{\beta}$) allows empirical calibration for specific domains and cultural contexts. Generalizing to n components, the framework can accommodate additional criteria – such as surprise, satisfaction, intentionality, and authenticity – as the field's consensus on creative product definition and assessment continues to evolve. This paper does not claim to resolve the broader theoretical debates surrounding the definition of creativity; rather, it contributes a workable quantitative formula that researchers and practitioners may find useful when a numerical creativity score for products is required.

Footnotes

[1] Originality(O) and novelty (N) are used interchangeably; however, this paper primarily uses originality (O)

References

1. a, b Runco MA, Jaeger GJ (2012). "The Standard Definition of Creativity." Creat Res J. **24**(1):92–96. doi:10.1080/10400419.2012.650092.
2. a, b, c, d, e, f, g Simonton DK (2018). "Defining Creativity: Don't We Also Need to Define What Is Not Creative?" J Creat Behav. **52**(1):80–90. doi:10.1002/jocb.137.
3. a, b, c, d, e, f, g, h, i, j, k Simonton DK (2012). "Taking the U.S. Patent Office Criteria Seriously: A Quantitative Three-Criterion Creativity Definition and Its Implications." Creat Res J. **24**(2–3):97–106. doi:10.1080/10400419.2012.676974.
4. a, b, c Simonton DK (2023). "Review of: Why the Standard Definition of Creativity Fails to Capture the Creative Act." Qeios. doi:10.32388/XWL0MQ.
5. a, b, c, d Atakaya MA, Sak U, Ayas MB (2024). "A Study on Psychometric Properties of Creativity Indices." Creat Res J. **36**(2):348–364. doi:10.1080/10400419.2022.2134550.

6. ^Said-Metwaly S, Kyndt E, Van den Noortgate W (2017). "Methodological Issues in Measuring Creativity." *Creat Theories – Res – Appl.* **4**(2):276-301. doi:10.1515/ctra-2017-0015.
7. ^Thys E, Sabbe B, De Hert M (2013). "The Assessment of Creativity in Creativity–Psychopathology Research: A Systematic Review." *Cogn Neuropsychiatry.* **18**(4):359-377. doi:10.1080/13546805.2012.714799.
8. a, b, cRunco MA (2023). "Updating the Standard Definition of Creativity to Account for the Artificial Creativity of AI." *Creat Res J.* **37**(1):1–6. doi:10.1080/10400419.2023.2257977.
9. ^Chen L, Song Y, Sun L, Childs P, Yin Y (2024). "Development of an Exploratory Creativity Assessment Scale." *Int J Des Creat Innov.* **12**(2):101–117. doi:10.1080/21650349.2024.2319772.
10. a, b, c, d, eSimonton DK (2021). "Scientific Creativity: Discovery and Invention as Combinatorial." *Front Psychol.* **12**:Article 721104. doi:10.3389/fpsyg.2021.721104.
11. ^Simonton DK, Ting S-S (2010). "Creativity in Eastern and Western Civilizations: The Lessons of Historiometry." *Manag Organ Rev.* **6**(3):329–350. doi:10.1111/j.1740-8784.2010.00188.x.
12. a, bBoden MA (2004). *The Creative Mind: Myths and Mechanisms.* 2nd ed. Routledge.
13. a, bBoden MA (2012). *Creativity and Art: Three Roads to Surprise.* Oxford University Press.
14. a, bAcar S, Burnett C, Cabra JF (2017). "Ingredients of Creativity: Originality and More." *Creat Res J.* **29**(2):133–144. doi:10.1080/10400419.2017.1302776.
15. a, bKim M (2025). "Applying Simonton's Definition to Japanese Creativity in Cross-Cultural Analysis." *Qeios.* doi:10.32388/QFFD8E.
16. a, b, cAmabile TM (1982). "Social Psychology of Creativity: A Consensual Assessment Technique." *J Pers Soc Psychol.* **43**(5):997–1013. doi:10.1037/0022-3514.43.5.997.
17. ^Green AE, Beaty RE, Kenett YN, Kaufman JC (2023). "The Process Definition of Creativity." *Creat Res J.* **36**(3):544–572. doi:10.1080/10400419.2023.2254573.
18. ^Weisberg RW (2015). "On the Usefulness of “Value” in the Definition of Creativity." *Creat Res J.* **27**(2):111–124. doi:10.1080/10400419.2015.1030320.
19. a, bAbraham A (2023). "Why the Standard Definition of Creativity Fails to Capture the Creative Act." *Qeios.* doi:10.32388/LS88G9.
20. ^Weisberg RW (2018). "Response to Harrington on the Definition of Creativity." *Creat Res J.* **30**(4):461–465. doi:10.1080/10400419.2018.1537386.
21. ^Kim M, Nguyen LPH (2026). "Computer-Implemented Creativity Assessment for Open-Ended Constraint-Based Mathematical Expression Tasks" [Provisional Patent] (U.S. Patent Application No. 63/964,410). U.S. Pat

ent and Trademark Office.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.