

Toward Universal Creativity Assessment by Untrained Judges

Thibaut Batal

LINEACT, Cesi, Nanterre, France

tbatal@cesi.fr

Abstract: The purpose of a creativity session being first to generate many ideas, how to fast and reliably assess each one's creativity, even with non-creativity experts? Creative people appearing not only to be good at generating ideas but also at implicitly evaluating them, improving people's assessment would further improve their own creativity. This paper investigates canonical components of originality to identify universal creative performance subscales assessable by untrained judges. Three originality criteria emerged from theoretical research: (1) Feelings, emotions; (2) Imagination, fantasy; (3) Logical complexity. One criterion was used to assess appropriateness: (4) Consistency as defined by four rules. A corpus of 100 ideas were distributed for assessments by untrained judges (n=41). All ideas were assessed by 4 expert judges on a single criterion: creativity. Surprisingly the criterion (4), which had four clear rules to define it, showed insufficient reliability and had to be disregarded for the analysis. The three other criteria proved consistent with one another and were hence reduced to an assessment of originality. They showed poor consistency with expert judges' ratings. These results are discussed and call for further research to better understand creativity.

Keywords: idea, scale, evaluation, originality, appropriateness, novice

1. Improving creativity knowledge

Lack of innovation is a potential failure factor for businesses. New products are almost half of a company's income, but 35 to 45% of new product releases would fail (O'Quin & Besemer, 2006). Creativity is key to innovation and is a potentially significant predictor of purchase intentions and customer satisfaction (Horn & Salvendy, 2009), but most people misunderstand creativity (O'Quin & Besemer, 2006). Creativity can qualify either a *potential* or a *performance*. Creative *potentials* lead to creative outcomes (persons, influences from the environment, and processes). Creative *performances* are the outcomes themselves (ideas, processes, physical objects, communicating an idea). Creativity is commonly defined as producing a *performance*, or having the *potential* to produce a *performance*, which is both original and appropriate, i.e. which is unusual and responds to a need for its audience. Originality and appropriateness are subjective and context-dependent (e.g., Beghetto & Corazza, 2019; Tang & Gruszka, 2017). Ideas are implicitly favored when they are socially beneficial and immediately effective, and they can hardly be judged as both original and appropriate even by domain experts (Beghetto & Corazza, 2019; Ceh et al., 2022). Highly original ideas' expected usage and benefits would be harder to communicate and understand, increasing the uncertainty relative to ideas' effectiveness and workability, while possibly triggering defensive reactions (O'Quin & Besemer, 2011, 2006; MASTRIA et al., 2019; Cropley, 2023). To rightly assess creative ideas and thus be creative, one must understand *what* is creative avoiding implicit rules appearing behind the term "original" (Ceh et al., 2022; Forthmann et al., 2017; Runco et al., 2005). Rater's cognitive workload increases with ideas and scales complexity, leading to misjudgments (Forthmann et al., 2017). Simple and absolute subscales assessable by anyone would ease creative understanding, rating, and generation, relieving cognitive load and sharing the evaluation task. They would allow non-creativity experts to faster understand what is creative *performance*, what ability to train to increase their *potential*, and what to discuss and tweak in concepts to achieve high creativity. Besides, creativity science needs to identify a *universal aesthetic* to quantify creativity kinds (Cropley and Kaufman, 2013), a common language and definitions (Amabile, 1982) shared by different artifacts (e.g., knowledge, products, processes, art), and by different audiences (e.g., consumers, entrepreneurs). It would facilitate communication across domains (O'Quin & Besemer, 2011) and possibly remove assessment's contextualization needs. Otherwise, (1) neither cross-field *performances* nor performers can be compared, hindering interdisciplinary collaboration and *potentials* comparison; and (2) performances should be randomly assessed by experts (Amabile, 1982), decreasing fluidity in creativity sessions where ideas are constantly generated, combined, and discussed. These drawbacks could be overcome by identifying common creativity scales, which could be used as material to consider generated concepts' interesting sides, full discussions about their values and (pre-)sorting them fast. Besides, quality scoring may have a different meaning for assessing potential and performance (Reiter-Palmon et al., 2019). Identifying universal criteria would fuel both this debate and the one about whether creativity is general or domain-specific (Baer, 2012).

2. Performance assessment

Several product creativity scales are experimentally supported. Besemer and colleagues developed the Creative Product Analysis Model (CPAM), then refined it to the Creative Product Semantic Scale (CPSS, O'Quin & Besemer, 2006). Cropley and colleagues refined the CPAM to a Creative Solution Diagnosis Scale (CSDS) and then to a reduced Revised CSDS (RCSDS). These scales use 24 (RCSDS) to 125 (CPAM) subscales (Brown, 2014; Cropley, 2023). The CPAM considers three dimensions: *novelty* (any element in the product that is original, incentive or enables new perspectives), *resolution* (the product meets a need, is feasible and understandable), and *synthesis and elaboration* (how the solution is implemented and displayed) (Tang & Gruszka, 2017; O'Quin & Besemer, 2011; Cropley, 2023). The CPSS renames *Elaboration and synthesis* to *Style* (novelty perception enhancement) (O'Quin & Besemer, 2011, 2006). The CSDS rates five dimensions: *Problematization* (definition of the problem) and *Propulsion* (new perspectives given to the problem, combined as *Novelty* in the RCSDS; *Relevance & Effectiveness*; *Elegance* (external, i.e. product's looking, or internal, i.e. how the solution is implemented); and *Genesis* (or *Generalizability*; workability for other needs, perspectives extension) (Cropley & Kaufman, 2013; Brown, 2014; Cropley, 2023). Only the CPSS and CSDS enable ratings close to experts' (Cropley, 2023), outside of technologically-oriented domains. CSDS' *Elegance* and *Genesis* could not be assessed even by quasi-experts (Cropley & Kaufman, 2013).

These scales are mostly concerned with functional relations (Reiter-Palmon et al., 2019) and knowledge-oriented, being relative to the knowledge brought to the user and seemingly loosely considering experience, feelings. "*Pleasingness*", *gracefulness*, and "*harmoniousness*" are some of the *External elegance's* subscales, but more hedonic scales could miss. Indeed, creativity is sometimes defined by the responses elicited by the product in an observer or user, like surprise (Han et al., 2021). Besides, to study the link between product creativity and purchasers' evaluations, Horn and Salvendy (2009) tested a model with seven dimensions, mostly based on the CPAM and notably adding *Emotion*, *Centrality* (interest in the product), *Importance* (how suitable and crucial the product is for the customer), and *Desire*.

Getzels and Csikszentmihalyi (1969) proposed three criteria targeting fine art: *originality/imaginativeness*, *overall aesthetic value*, and *craftsmanship/technical skill* (O'Quin & Besemer, 2011). Nevertheless, I argue technical aspects might be avoided. Indeed, *technicity* could be different from creativity (Amabile, 1982) and users appear sometimes to produce more original and valuable ideas than professionals (Kristensson et al., 2004). Moreover, Lewis (2005) argues to pursue "more subjective and elusive goals" to promote creative insight, and including technical aspects in criteria would hinder reaching criteria generalizable to all domains.

3. Potential assessment

Performances cannot arise without creative thinking, a *potential* composed of two main thinking: divergent (DT; generating many ideas) and convergent (CT; evaluating the ideas). DT can be measured through *fluency* (number of ideas), *originality* (rarity of ideas), *flexibility* (semantic variety of ideas), and *elaboration* (level of detail) (Reiter-Palmon et al., 2019; Runco et al., 2005). Their operationalization lacks agreement. They all require subjective decisions (e.g., what aspects are similar?) and are relative to a sample (Silvia et al., 2008; Reiter-Palmon et al., 2019). *Fluency* is generally preferred over other criteria because it is objective and fast computed, whereas it "is not directly tied to the creativity definition" (Forthmann et al., 2017; see also Reiter-Palmon et al., 2019). *Originality* is generally rated as one criterion but is sometimes measured by three subscales: *uncommonness* (infrequency), *remoteness* (far link from existing artifacts), and *cleverness* (smart nature, encompassing "imaginativeness, ingenuity, funniness, and cunning aptness" (Reiter-Palmon et al., 2019)). Not all ideas' facets are *remoteness*-assessed (e.g., either features overlap or part-whole relations) and high *cleverness* can compensate for the others (Reiter-Palmon et al., 2019; Silvia et al., 2008). Creativity's *appropriateness* is rarely assessed. It is the amount to which an idea is feasible and solves the problem at hand without being viewable as inappropriate by others (Runco et al., 2005).

4. Toward universal creativity scales

To ease the identification of universal creativity scales efficient among untrained judges, appropriateness is excluded, being defined by four specific rules imposed during the creativity task (see 7.1). Moreover, reliability of assessing one's own ideas and others' ideas being different (Mastria et al., 2019), this paper focuses on assessing exogenous/extrapersonal ideas. Besides, the scales are expected to be usable at any stage of the design process. Therefore, I suggest subscales relative to the presentation/communication (e.g., *Style*) of ideas should not be criteria. They should be later tweaked, when all ideas are gathered into concepts. At least, they

should be mediators of the assessments. Indeed the clearer the novelty, the more reliable the creativity assessment is. In the same way, level of detail should not matter.

From an evolutionary viewpoint, the brain is triune. Two of them are the limbic system, primarily responsible for emotional responses, and the neocortex, primarily responsible for logical reasoning. Both have major roles in creativity (Carson, 2010). Therefore, emotions and logic seem two major components constituting creative performances. In the continuity of Getzels and Csikszentmihalyi's (1967) conclusions that "creative achievement [...] depends upon a union of the intuitive and the rational, of the imaginative and the analytic, of fantasy and control", I suggest at least two dimensions of universal creativity exist: *feelings/emotions* and *logic*.

Feelings/emotions might be an indirect estimator of any idea's originality and value. Indeed, highly novel ideas are accompanied by a surprise effect (see 2). With repeated exposures to a stimulus, whether it is positive (e.g., food) or negative (e.g., a shock), the central nervous system gets *habituated*; the elicited responses, both cognitive and behavioral, decrease (Thompson, 2009). Therefore, I suggest the feeling subscale would be rated higher for original ideas. Moreover, according to the Conservation of Resources Theory (COR), people value their resources (objects, conditions, personal characteristics, and energies) (see Hobfoll & Shirom, 2019), and would naturally feel emotions to conserve them (Park et al., 2014). Consequently, ideas would be valuable when they are perceived to prevent losses of resources, or to enable gains, generating *feelings* (e.g., a loss prevention can relieve, a gain anticipation can elicit joy). Meaningfulness of an idea could increase the elicited feelings, making *feelings* an indirect estimator of any idea's value.

To assess opportunities given by an idea, the observer must be conscious enough of the problem and imagine the idea's impacts (Cropley & Cropley, 2010). Otherwise, *feelings/emotions* would not be elicited. However, the idea might be viewed as too complex or absurd, as highly imaginative, fanciful. Therefore, I suggest *fantasy/imagination* and *logical complexity* would be two dimensions constituting creativity.

That is, *feelings/emotions* are the emotions reflected by the idea, or the feelings it elicits. *Logical complexity* is the amount of involved systems interacting with each other, referring then to the depth and sophistication of the idea. *Fantasy/imagination* is the amount of push-backed limits of reality and appeal to the imagination, defying conventional expectations and proposing concepts that don't exist in reality. By combining these three criteria, an original idea or creation would be emotionally powerful, intellectually stimulating and capable of inspiring the imagination. These criteria are assumed to be absolute because when an idea does not meet any criterion enough, the related component must simply be added (e.g., if it idea does not reflect emotions enough, emotional components must simply be added). These criteria might seem too subjective to reach a reliable consensus among raters. Nevertheless, Amabile (1982) argued judgments about creativity, even with subscales, "can ultimately only be subjective." Moreover, the CPSS' *Style* rose the same concern, but studies finally showed great consistency among raters (O'Quin & Besemer, 2011).

5. Hypothesis

Originality of ideas will be assessed through three subscales: (1) feelings, emotions, (2) imagination, fantasy, and (3) logical complexity (see 4). Appropriateness through one item defined by four rules (see 7.1). Suggesting the criteria would rate the degree of creativity of ideas, I hypothesized:

H1: Assessment by untrained judges of creativity, based on the abovementioned criteria, will be consistent with the assessment of creativity by expert judges.

6. Participants

Participants were 45 French engineering students from the same promotion and engaged in an innovation project. They were aged 20-27 years (M = 22.28; SD = 1.45). Three were women. All were divided into groups (n=6) and evaluated 20 ideas generated by another group. Four expert judges from a research team in Innovation were recruited to score each idea's overall creativity.

7. Material

7.1 Experimental ideas sample

Ideas assessed by participants were generated the day before during a creativity task inspired by the causality effect (Michotte, 1963) and by three types of typical tasks: instantiation, consequence, and bridge-the-gap

(Silvia, 2011). The task requests explaining a hidden part in a 5-second animation made of geometrical shapes seemingly interacting. Ending frames were partially hidden so that the result of the interaction could be seen, but not the interaction itself, giving participants the opportunity to imagine as many possible interaction(s) as possible that occurred between the allegorical shapes (see Figures 1 and 2). Each participant individually completed idea sheets with a drawing and a textual description (see Figure 3), because Torrance's figural and verbal tasks appeared to measure almost orthogonal creative abilities (Baer, 2012). Ideas had to follow four rules:

1. When they are invisible (in a black area), the existing elements can be anywhere (visible or invisible area). Specify their location in the drawing/description.
2. You can add one or more new elements anywhere, but not delete existing ones.
3. Size and time are flexible (smaller/larger existing elements, slower/faster animation sections).
4. The whole animation must remain coherent.

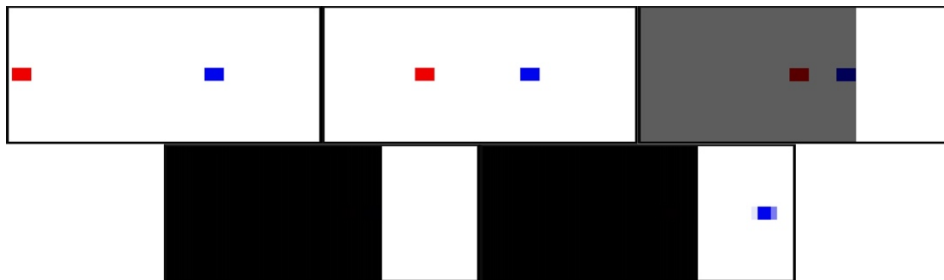


Figure 1: First task's animation

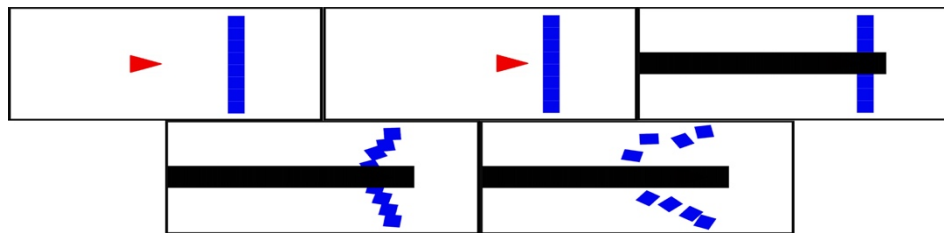


Figure 2: Second task's animation

Fiche création Id participant :

Thème/catégorie : Sélection :

(en lettres capitales)

Explications

Un footballeur avait lancé le carré, qui est retombé au bon moment: lorsque l'obstacle était ouvert.

Figure 3: Idea sheet presented as example

From 374 generated ideas, ideas sheets were randomly selected (n=120) for this study. Duplicates were replaced. Each group assessed 20 ideas, ten addressing the first animation and ten the second one.

7.2 Expert judges' warm-up sample

For the expert judges to appropriate the new creativity task, they trained on ideas generated with the same task through a pilot study (n=95), and by participants from the same engineering school's promotion.

7.3 Measures

Each participant completed a numerical form. Each page displayed the idea number and the four items to be assessed on a 7-point Likert-type scale. Each item was followed by an example illustrating its meaning. Expert judges assessed each idea by a single creativity score.

8. Procedure

8.1 Expert judges' assessment

Expert judges appropriated the task rating the *warm-up sample*, before rating all the ideas (n=374) including those examined by the participant (n=120). The task lasted three hours with a 15-min break. Judges were reminded of the two creativity dimensions - originality and appropriateness -, and were given the four used appropriateness rules (see 7.1). Judges first watched three times the first animation then scored the related ideas, then watched three times the second animation and scored the related ideas. Idea sheets were sequentially displayed on a screen. For each idea, the title and description were collectively read, before individually scoring it. Judges remained neutral to minimize social influence.

8.2 Participants' assessment

All participants performed the task in the same room, each group around a table. The experimenter presented the items to be scored through examples illustrating each one and reminded the four rules of appropriateness. These rules remained displayed to all throughout the task. On each group table, the stack of idea sheets was upside down, to hide the next idea. The experimenter showed the first animation three times and all groups simultaneously assessed the ten first ideas. In each group, one participant took the first sheet, read aloud its title and description, then showed the idea to everyone. The idea sheet could be circulated among the group. Participants could speak to understand the descriptions and the drawings. When all groups had assessed the ten first ideas, the experimenter showed the second animation three times, and the same process was followed for the ten next ideas.

9. Results

9.1 Validation of the expert judges task appropriation

To validate the expert judges appropriated the new creativity task, we analyzed with Cronbach's alpha the inter-rater agreements for the warming-up sample and the experimental one, for each animation. With the first animation, inter-rater agreement was low with the *warm-up sample* ($\alpha=.550$; n=44) and high with the first ideas from the experimental one ($\alpha=.811$; n=44). With the second animation, inter-rater agreement was high both with the *warm-up sample* ($\alpha=.707$; n=51) and the first ideas from the experimental one ($\alpha=.796$; n=51). Merging assessments of both animations, agreement was moderated with the *warm-up sample* ($\alpha=.653$; n=95) and high with the ideas the experimental sample comes from ($\alpha=.752$; n=373). The reached inter-rater agreements suggest the experts satisfactorily appropriated the task.

9.2 Creativity criteria validation

Inter-rater agreement for each of the four assessment criteria was computed. Each group but the fourth ($\alpha \in [.579; .678]$), which had the lower number of judges (n=4), had an acceptable inter-rater agreement for each criterion ($\alpha \in [.664; .930]$) (cf. Table 1). We removed group 4 from further analysis. Appropriateness appeared to lead to lowest agreements.

Table 1: Inter-rater agreements per criterion

Assessing group	1	2	3	4	5	6
n	8	8	6	4	9	10
α (feelings, emotions)	.786	.805	.708	.678	.924	.825
α (imagination, fantasy)	.816	.738	.767	.579	.926	.904
α (logical complexity)	.764	.764	.678	.631	.930	.881
α (appropriateness)	.664	.810	.718	.589	.849	.762

Inter-criteria reliability proved high ($\alpha=.839$). The highest agreement decrease was by removing *logical complexity* ($\alpha=.747$). The highest agreement increase was by removing *appropriateness* ($\alpha=.899$) (see Table 3). *Appropriateness* ratings were removed from further analysis.

Table 2: Inter-criteria agreements

Criterion	Cronbach's alpha removing the criterion
Feelings, emotions	.766
Imagination, fantasy	.755
Logical complexity	.747
Appropriateness	.899

We computed the consistency of the assessments between participants and expert judges. Consistency was low ($\alpha=.511$), which invalidates H1.

10. Discussion

The purpose of this study was to validate universal creativity criteria intended to untrained judges, by comparing *untrained judges'* assessments of ideas with *expert judges'* assessments of ideas. Untrained judges' inter-rater-reliability was good with groups of more than 5 individuals, consistent with Ceh et al.'s (2022) results. Experts inter-judge agreement increased with the study sample relatively to the warm-up one and was high, suggesting they reliably assessed the ideas. They may have got a better grasp either of what original ideas are for this task and participants population, and/or of what appropriateness is through implicit rules underlying ideas (Forthmann et al., 2017). Besides because of the long assessment session, intuitive feelings may have gradually overcome appropriateness checking by the trained judges, feelings elicited by originality gaining then more importance than defined rules.

10.1 Appropriateness seems harder to agree upon

Ideas' *appropriateness*, defined as maintaining consistency of animations by following four rules, was assumed to be a more objective-oriented measure than other subscales. Therefore we could expect *appropriateness* to reach more agreement of *untrained judges* than other subscales. Surprisingly, results invalidated this assumption, *appropriateness* being the criterion that achieved the less satisfactory inter-rater agreement. This suggests that: (1) although the rules were simple, some participants may instead have used their own implicit understanding to assess the consistency of animations (Forthmann et al., 2017), leading to a defined appropriateness not as objective as it seems. (2) Appropriateness is hard to assess probably because it requires to well understand the end-user needs and constraints, and creativity renders existing solutions to problems obsolete (Cropley, 2023). Change brought by an idea's reflected originality might then have sometimes overcome rater's conception of appropriateness, at different degrees among raters. (3) Assessing consistency of animation requires envisioning what the whole animation would look like with the idea inserted inside. This capability may have been owned to different degrees by participants. (4) Non-verbal cues may have influenced appropriateness ratings. Indeed, non-verbal cues influence judgments occurring in groups (Milliken et al., 2003)

and participants could speak to understand what was written as description of the ideas, but they were not required to remain neutral. To improve the appropriateness criterion, maintaining raters' neutrality and both the COR and Schwartz's notion of values (Taylor & Kaufman, 2021) might be good leads.

10.2 The suggested creativity criteria are not sufficient to assess creativity

I suggested creativity assessment by *untrained* judges through the four criteria would correlate with a single-scale creativity assessment by *expert* judges. The *appropriateness* criterion rated by untrained judges did not meet sufficient reliability and overall results invalidated H1. Nevertheless, this may not be the only reason for the invalidation. Indeed, (1) because of the feelings elicited by originality, the *experts* might have implicitly placed more weight on it (Diedrich et al., 2015), theoretically reducing the impact of the appropriateness removal; (2) if the appropriateness criterion did not reach agreement among untrained judges, we can suspect it was also less reliable among trained judges. That is, if the expert judges' warm-up improved their rater agreement, it may be through an increase in originality rating agreement instead of appropriateness rating; (3) originality and appropriateness are not necessarily related (Runco et al., 2005) and an inappropriate but original idea would still be viewed as creative (Runco & Charles, 1993). As a result, a few criteria might be missing for a universal originality assessment by untrained judges. They might be related to germinal aspect (e.g., the extent to which an idea inspires thoughts), to the COR (arousal, relaxation), to the *remoteness* criterion (see 3; omitted because DT enables flexibility measurement), or the third triune brain layer, the reptilian (e.g., movements or energies). Furthermore, novices' assessments should probably include an understandability criterion similar to CPSS' *Style*, to weight their ratings: if they estimate the idea was unclear, their ratings should be less considered. In the same way, raters' emotional states would influence their assessments (Mastria et al., 2019); this could be due to a conflict with the emotion the performance rises, perhaps assessable by the *feelings/emotions* criterion. Therefore, understandability and the studied criteria could be not creative *dimensions*, but creative *perceptions*, that is, mediators of creativity assessment. Finally, the criteria might not be enough because they intend to be universal. Indeed if creativity is domain-specific, according to Baer and Kaufman's Amusement Park Theory, the proposed criteria would fit only between the first and the second level (respectively initial requirements such as motivation, and domain into which creativity is applied) (Baer, 2012; see also Taylor & Kaufman, 2021), leaving the next levels unrated by the untrained judges.

10.3 Limits

This study constitutes only a first step toward the development of a creativity assessment tool. Firstly, to further challenge our creativity criteria, it would be necessary to test them in creative contexts varying in purpose and approach, covering a broad spectrum from e.g., technical creativity responding to a functional brief, to artistic creativity relying more on an introspective process. Different tasks may lead to differential matching with each creativity criterion: for example, ideas in the current task did not have to meet functional requirements or achieve any performance, which may partly explain why appropriateness was particularly difficult to assess. Secondly, our experimental population holds several limitations: it was quite a homogeneous sample (same school, same study level, mostly males) and they may have been influenced to different degrees by the instructions like performers do (Runco et al., 2005), questioning the generalizability of our findings. Finally, regarding the study protocol, experts rated all ideas in one session and could then endure cognitive workload undermining their ratings (Forthmann et al., 2017).

11. Conclusion

Four subscales for universal creativity assessment by untrained judges were tested. Surprisingly *appropriateness*, the criteria expected to be the more objective because it was defined by four rules the ideas had to follow, reached lower inter-rater agreement than the others. The reasons could be that assessing appropriateness requires envisioning capabilities and leads to implicit subjective rules, which could be mediated by the perception of originality. Furthermore, non-verbal cues could have wrongly influenced appropriateness ratings whereas they would have been effective for the other criteria. These other criteria were aggregated but they did not match *experts'* ratings. It is argued the mismatch was not due to the removed appropriateness, therefore a few subscales might be missing. Some proposals to identify them were given. Furthermore, the studied criteria could be not creative *dimensions*, but creative *perceptions*, that is, mediators of creativity assessment. The criteria deserve to be completed, to work for any field at least as fast pre-assessment, and to nurture the debates about creativity's nature. Identifying them all and which ones lead to easier consensus would help to orient debate when discussing creative concepts across domains, enabling cross-communication

between experts. Further studies are required to identify all canonical creativity dimensions/perceptions, for untrained judges to discover creativity and assess it more affordably and engagingly than with existing thorough scales.

Acknowledgements

Thanks to Professor Stéphanie BUISINE (LINEACT, Cesi) for support in statistical and limits analysis and in synthesis. Thanks to Doctor Beatrice BIANCARDI (LINEACT, Cesi) for support in research practices. Thanks to both of them and Doctor Andréa BOISADAN (LINEACT, Cesi), for participation as expert judges. To all three, many thanks for the foolproof presence when needed. Finally, thanks to the anonymous reviewer for the interesting notes.

References

- Amabile, T. (1982) Social Psychology of Creativity: A Consensual Assessment Technique. *Journal of Personality and Social Psychology*. 43, pp 997–1013. [doi:10.1037/0022-3514.43.5.997](https://doi.org/10.1037/0022-3514.43.5.997).
- Baer, J. (2012) Domain Specificity and the Limits of Creativity Theory. *The Journal of Creative Behavior*. 46 (1), pp 16–29. [doi:10.1002/jocb.002](https://doi.org/10.1002/jocb.002).
- Beghetto, R.A. & Corazza, G.E. (2019) *Dynamic perspectives on creativity : new directions for theory, research, and practice in education*.
- Brown, D.C. (2014) Let's not get too Creative. *Design Computing and Cognition* 14.
- Carson, S. (2010) *Your creative brain: Seven steps to maximize imagination, productivity, and innovation in your life*. Jossey-Bass.
- Ceh, S.M., Edelmann, C., Hofer, G. & Benedek, M. (2022) Assessing Raters: What Factors Predict Discernment in Novice Creativity Raters? *Journal of Creative Behavior*. 56 (1), pp 41–54. [doi:10.1002/jocb.515](https://doi.org/10.1002/jocb.515).
- Cropley, D. & Cropley, A. (2010) Functional Creativity: “Products” and the Generation of Effective Novelty. In: J.C. Kaufman & R.J. Sternberg (Eds.). *The Cambridge Handbook of Creativity*. First edition. Cambridge University Press. pp 301–317. [doi:10.1017/cbo9780511763205.019](https://doi.org/10.1017/cbo9780511763205.019).
- Cropley, D.H. (2023) *A Review of Approaches to the Assessment of Product Creativity (Preprint)*. [doi:10.31234/osf.io/7d4av](https://doi.org/10.31234/osf.io/7d4av).
- Cropley, D.H. & Kaufman, J.C. (2013) Rating the creativity of products. In: K. Thomas & J. Chan (eds.). *Handbook of Research on Creativity*. Edward Elgar Publishing Ltd., pp 196–211. [doi:10.4337/9780857939814](https://doi.org/10.4337/9780857939814).
- Cropley, D.H., Kaufman, J.C. & Cropley, A.J. (2011) Measuring Creativity for Innovation Management. *Journal of Technology Management & Innovation*. 6 (3), pp 13–30. [doi:10.4067/s0718-27242011000300002](https://doi.org/10.4067/s0718-27242011000300002).
- Diedrich, J., Benedek, M., Jauk, E. & Neubauer, A.C. (2015) Are creative ideas novel and useful? *Psychology of Aesthetics, Creativity, and the Arts*. 9 (1), pp 35–40. [doi:10.1037/a0038688](https://doi.org/10.1037/a0038688).
- Forthmann, B., Holling, H., Zandi, N., Gerwig, A., Çelik, P., Storme, M. & Lubart, T. (2017) Missing creativity: The effect of cognitive workload on rater (dis-)agreement in subjective divergent-thinking scores. *Thinking Skills and Creativity*. 23, pp 129–139. [doi:10.1016/j.tsc.2016.12.005](https://doi.org/10.1016/j.tsc.2016.12.005).
- Getzels, J.W. & Csikszentmihalyi, M. (1967) Scientific Creativity. *Science Journal*, 3(9), pp 80–84.
- Han, J., Forbes, H. & Schaefer, D. (2021) An exploration of how creativity, functionality, and aesthetics are related in design. *Research in Engineering Design* 2021 32:3. 32 (3), pp 289–307. [doi:10.1007/s00163-021-00366-9](https://doi.org/10.1007/s00163-021-00366-9).
- Hobfoll, S.E. & Shirom, A. (2019) Conservation of Resources Theory: Applications to Stress and Management in the Workplace. In: R.T. Golembiewski (ed.). *Handbook of Organizational Behavior*. 2nd edition. New York, NY, Routledge. pp 57–78.
- Horn, D. & Salvendy, G. (2009) Measuring consumer perception of product creativity: Impact on satisfaction and purchasability. *Human Factors and Ergonomics In Manufacturing*. 19 (3), pp 223–240. [doi:10.1002/hfm.20150](https://doi.org/10.1002/hfm.20150).
- Kristensson, P., Gustafsson, A. & Archer, T. (2004) Harnessing the Creative Potential among Users. *Journal of Product Innovation Management*. 21 (1), pp 4–14. [doi:10.1111/j.0737-6782.2004.00050.x](https://doi.org/10.1111/j.0737-6782.2004.00050.x).
- Lewis, T. (2005) Creativity - A framework for the design/problem solving discourse in technology education. *Journal of Technology Education*. 17 (1), pp 35–52. [doi:10.21061/jte.v17i1.a.3](https://doi.org/10.21061/jte.v17i1.a.3).
- Mastria, S., Agnoli, S. & Corazza, G.E. (2019) How does emotion influence the creativity evaluation of exogenous alternative ideas? *PLoS ONE*. 14 (7). [doi:10.1371/journal.pone.0219298](https://doi.org/10.1371/journal.pone.0219298).
- Michotte, A. (1963) The perception of causality. *Basic Books*. <https://psycnet.apa.org/record/1964-05029-000>.
- Milliken, F.J., Bartel, C.A. & Kurtzberg, T.R. (2003) Diversity and Creativity in Work Groups: A Dynamic Perspective on the Affective and Cognitive Processes That Link Diversity and Performance. In: P.B. Paulus & B.A. Nijstad (eds.). *Group creativity: Innovation through collaboration*. Oxford University Press. pp 32–62. [doi:10.1093/acprof:oso/9780195147308.003.0003](https://doi.org/10.1093/acprof:oso/9780195147308.003.0003).
- O’Quin, K. & Besemer, S. (2011) Creative Products. In *Encyclopedia of Creativity*. Runco, M.A. & Pritzker, S.R. (Eds.). 2nd edition. Elsevier. pp 273–281. [ISBN:978-0-12-375039-6](https://doi.org/10.1016/B978-0-12-375039-6).
- O’Quin, K. & Besemer, S.P. (2006) Using the Creative Product Semantic Scale as a Metric for Results-Oriented Business. *Creativity and Innovation Management*. 15 (1), pp 34–44. [doi:10.1111/j.1467-8691.2006.00367.x](https://doi.org/10.1111/j.1467-8691.2006.00367.x).

- Park, H.I., O'Rourke, E. & O'Brien, K.E. (2014) Extending conservation of resources theory: The interaction between emotional labor and interpersonal influence. *International Journal of Stress Management*. 21 (4), pp 384–405. [doi:10.1037/a0038109](https://doi.org/10.1037/a0038109).
- Reiter-Palmon, R., Forthmann, B. & Barbot, B. (2019) Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*. 13 (2), pp 144–152. [doi:10.1037/aca0000227](https://doi.org/10.1037/aca0000227).
- Runco, M.A. & Charles, R.E. (1993) Judgments of originality and appropriateness as predictors of creativity. *Personality and Individual Differences*. 15 (5), pp 537–546. [doi:10.1016/0191-8869\(93\)90337-3](https://doi.org/10.1016/0191-8869(93)90337-3).
- Runco, M.A., Illies, J.J. & Reiter-Palmon, R. (2005) Explicit Instructions to Be Creative and Original: A Comparison of Strategies and Criteria as Targets With Three Types of Divergent Thinking Tests. *Korean Journal of Thinking & Problem Solving*, 15(1), pp 5–15. <https://psycnet.apa.org/record/2005-04061-001>
- Runco, M., Illies, J. & Eisenman, R. (2005) Creativity, Originality, and Appropriateness: What do Explicit Instructions Tell Us About Their Relationships? *The Journal of Creative Behavior*. 39. [doi:10.1002/j.2162-6057.2005.tb01255.x](https://doi.org/10.1002/j.2162-6057.2005.tb01255.x).
- Silvia, P.J. (2011) Subjective scoring of divergent thinking: Examining the reliability of unusual uses, instances, and consequences tasks. *Thinking Skills and Creativity*. 6 (1), pp 24–30. [doi:10.1016/j.tsc.2010.06.001](https://doi.org/10.1016/j.tsc.2010.06.001).
- Silvia, P., Winterstein, B., Willse, J., Barona, C., Cram, J., Hess, K., Martinez, J. & Richard, C. (2008) Assessing Creativity With Divergent Thinking Tasks: Exploring the Reliability and Validity of New Subjective Scoring Methods. *Psychology of Aesthetics, Creativity, and the Arts*. 2, pp 68–85. [doi:10.1037/1931-3896.2.2.68](https://doi.org/10.1037/1931-3896.2.2.68).
- Tang, M. & Gruszka, A. (2017) The 4P's Creativity Model and its application in different fields. In *Handbook of the management of creativity and innovation: Theory and practice*. pp 51-71. [doi:10.1142/9789813141889_0003](https://doi.org/10.1142/9789813141889_0003).
- Taylor, C.L. & Kaufman, J.C. (2021) Values Across Creative Domains. *The Journal of Creative Behavior*. 55 (2), pp 501–516. [doi:10.1002/jocb.470](https://doi.org/10.1002/jocb.470).
- Thompson, R.F. (2009) Habituation: A history. *Neurobiology of Learning and Memory*. 92 (2), pp 127–134. [doi:10.1016/j.nlm.2008.07.011](https://doi.org/10.1016/j.nlm.2008.07.011).