



Controlling stimulus ambiguity reduces spurious creative ideation variance in a cyclic adaptation of the alternative uses task

Witczak, Olga; krzysik, Iga; Bromberek-Dyzman, Katarzyna; Thierry, Guillaume; Jończyk, Rafał

Scientific Reports

DOI:
[10.1038/s41598-024-63225-2](https://doi.org/10.1038/s41598-024-63225-2)

Published: 31/05/2024

Publisher's PDF, also known as Version of record

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Witczak, O., krzysik, I., Bromberek-Dyzman, K., Thierry, G., & Jończyk, R. (2024). Controlling stimulus ambiguity reduces spurious creative ideation variance in a cyclic adaptation of the alternative uses task. *Scientific Reports*, 14, Article 12492. <https://doi.org/10.1038/s41598-024-63225-2>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



OPEN

Controlling stimulus ambiguity reduces spurious creative ideation variance in a cyclic adaptation of the alternative uses task

Olga Witczak^{1,3}✉, Iga Krzysik^{1,3}, Katarzyna Bromberek-Dyzman¹, Guillaume Thierry^{1,2,3} & Rafał Jończyk^{1,3}

In the alternative uses task (AUT), a well-established creativity assessment, participants propose alternative uses for common items (e.g., a brick) within a 2–3 min timeframe. While idea evaluation is likely involved, the emphasis is strongly on idea generation. Here, we test the value of presenting a word overlapping an image compared to a word only prompt, and we introduce a cyclic adaptation of the AUT explicitly calling on participants to choose their best idea. In Experiment 1, as compared to word only, word + image prompts increased idea fluency but reduced idea originality and variability within a group of native Polish speakers. Thus, word + image prompts improve AUT baselining. In Experiment 2, different participants produced as many ideas as possible within two minutes (List) or their single best idea at the end of each of three 30 s ideation cycles (Cycle). Although originality did not differ between List and Cycle overall, the first three ideas in List were rated as less creative than the ideas in Cycle. Overall, we conclude that using disambiguating images reduces spurious interindividual variability in the AUT while introducing idea evaluation in the task allows us to assess creativity beyond idea generation.

Creativity is commonly defined as a distinctively human capacity to produce novel and context-appropriate outputs in response to prompts or problems^{1,2}. Given that creativity is a complex cognitive skill, dissecting and systematically evaluating it presents a considerable challenge³. Establishing a singular definition of creativity is also difficult because of its unseen, nonverbal, and unconscious characteristics. Consequently, creativity was once deemed one of “the vaguest, most ambiguous, and most confused terms in psychology”⁴.

The concept of creativity hinges on two prerequisites of originality and effectiveness; originality implies that generated ideas are unique and uncommon, and effectiveness means that they have to be appropriate and plausible in a given context^{5,6}. An unusual idea with no real-life application is useless; conversely, a conventional albeit practical solution is not creative, by definition. The hallmarks of creativity can be further extended to encompass such qualities as an element of surprise, novelty, and authenticity, as well as utility and aesthetics^{7,8}.

Although creativity is usually evaluated based on outputs, it must be kept in mind that the creative process is dynamic and elusive due to constant alternation between idea generation and evaluation⁹. Two key sub-processes involved are thus divergent and convergent thinking. Divergent thinking refers to the ability to generate novel ideas and consider answers to open-ended questions without much evaluation, while convergent thinking refers to evaluation and selection mechanisms aiming to identify the most suitable, effective, and practical solution(s) from an array of alternatives^{10–12}. Often, these two modes of thinking have been conceptualised as lying at opposite ends of a continuum, even though they can coexist and work synergistically to enhance creative thinking^{3,10,13}.

Measures conventionally used to assess divergent thinking require participants to produce a number of ideas in response to a verbal or visual prompt. The obtained production is then assessed for fluency (i.e., number of ideas), and originality, focusing on uniqueness¹⁴. Measures of convergent thinking test the ability to perceive and acknowledge similarities between words, geometrical figures, or pictures, and responses are generally scored based on accuracy and ability to solve a problem within a given time limit^{15–17}. One of the most established measures of creativity and, more specifically, divergent thinking, is the alternative uses task (AUT)¹¹. In the AUT, participants are asked to generate as many unconventional uses for a familiar item as possible (usually presented

¹Adam Mickiewicz University, Poznań, Poland. ²Bangor University, Bangor, UK. ³These authors contributed equally: Olga Witczak, Iga Krzysik, Guillaume Thierry, and Rafał Jończyk. ✉email: olga.witczak@amu.edu.pl

as a word^{6,18–22}), within a time limit (usually two minutes). Responses can then be evaluated for originality, flexibility, fluency, and elaboration by a pool of independent raters¹⁸.

One of the most robust phenomena observed in AUT studies is the serial order effect, that is, an increase in the creativity of generated ideas over time. This phenomenon could be linked to spreading of activation in semantic and episodic memory (but see Vartanian et al.,²³). First, close and well-entrenched memory associations are retrieved, then activation spreads to more distant, weakly related concepts, leading to more unconventional alternative uses²⁴. Gilhooly et al.²¹ further suggested that responses based on memory (pre-known, conventional, stored) tend to be given before responses associated with novelty (e.g., object properties, broad use, and disassembly). Moreover, they noted that memory responses are associated with lower cognitive load compared to other types of answers.

It is surprising, however, that most studies using the AUT have relied on (a small number of) words as prompts to refer to objects that can substantially vary in terms of shape, texture, colour, or size, and that can have different physical properties in terms of weight, breakability, elasticity, etc. depending on the materials from which they are made. Such under-specification likely results in spurious variability in the output for different participants, because default expectations about the nature and properties of objects introduced by the word prompt may vary substantially between individuals, contexts, and cultures. One way to mitigate this problem is to use a picture as prompt instead of a word. However, images can yield ambiguity of their own relating to low naming agreement and object prototypicality, as well as other kinds of under-specification relating to physical appearance.

Several studies have shown that images as prompts affect ideation and creativity in design sketching, word association, or alternative uses tasks^{25–27}. Chrysikou et al.²⁸, for instance, presented participants with either the name of an everyday object, an image depicting the object, or a combination of both (image overlaid with a word, henceforth word + image). Participants were asked to generate common uses, common alternative uses, or uncommon alternative uses. Images elicited more normative responses than words, with word + image prompts falling in between. Recently, George et al.²⁹ asked participants to predict their expected creative performance in the AUT when presented with either word + image or word prompts. Participants anticipated that they would be more creative in the word + image condition, but the group of participants who actually performed the AUT on the same items generated ideas that were less original in the word + image than the word only condition. However, it remains unclear whether it is originality that suffers when using word + image prompts or whether participants appear less variable in their output due to a more uniform starting point, that is a shared understanding of what the word prompt refers to.

Another dimension of measurement associated with the AUT is fluency, that is, the number of ideas generated within a given time window. Benedek and Neubauer³⁰ proposed that the more people are creative, the faster they are at providing responses thanks to greater associative fluency. Some studies have criticised fluency for its shallow explanatory potential. Beaty and Johnson³¹, for instance, referred to fluency as “a proxy of generative ability” and criticised its lack of reliability due to low inter-item correlations in AUT³². Such low correlation could be attributed to variability of object characteristics³³ as well as word frequency³⁴. We contend that such issues are less likely to arise and that correlations should be higher when words are presented together with a disambiguating image (see Experiment 1). Furthermore, the reason why fluency measures have poor reliability in the AUT may relate to the lack of consideration given to idea evaluation (see Experiment 2). If participants are unconstrained regarding idea quantity, then the value of fluency as a creativity measure is naturally diminished.

In the current study, we thus aimed to resolve two limitations of the classic AUT: (i) the ambiguity induced by a word only prompt; (ii) the lack of focus on idea evaluation. In Experiment 1, we compared the creative output in the form of lists of ideas from participants tested using word only and word + image prompts. Participants were randomly assigned to one of three experimental procedures: uniform with word only prompts, uniform with word + image prompts, or dual, consisting of half word only prompts and half word + image prompts (to allow for within-subject comparisons). In Experiment 2, we compared the classic list-based AUT with a new cycle procedure in which participants had to report their *single best* idea after each of three 30 s ideation cycles. Participants were randomly assigned to one of three experimental procedures: uniform with list trials (classic AUT task over two minutes), uniform with cycle trials (three cycles per item), or dual (consisting of half list and half cycle trials, allowing for within-subject comparisons). We hypothesised that:

- 1) Word + image prompts reduce inter-individual variability and thus provide a better baseline than word only prompts;
- 2) Word + image prompts tend to constrain creativity and thus reduce originality and/or fluency as compared to word only prompts;
- 3) Originality is higher for ideas generated in the cycle procedure than the list procedure;
- 4) Originality increases gradually with each consecutive cycle in the cycle procedure.

Results

Experiment 1: AUT items as word only or word + image

As regards idea fluency, we found a marginally significant effect of condition in the dual procedure (within-subject comparison), with higher number of ideas generated for word + image than word only items ($p = 0.051$, $\eta_p^2 = 0.144$, large effect size; Supplementary Table 1, Fig. 1).

When comparing between participants in the uniform procedures, we did not find significant differences in idea fluency between conditions ($p = 0.448$, $\eta_p^2 = 0.013$; Supplementary Table 2). When comparing across uniform procedures within each condition, we did not find any significant difference either, be it in the word

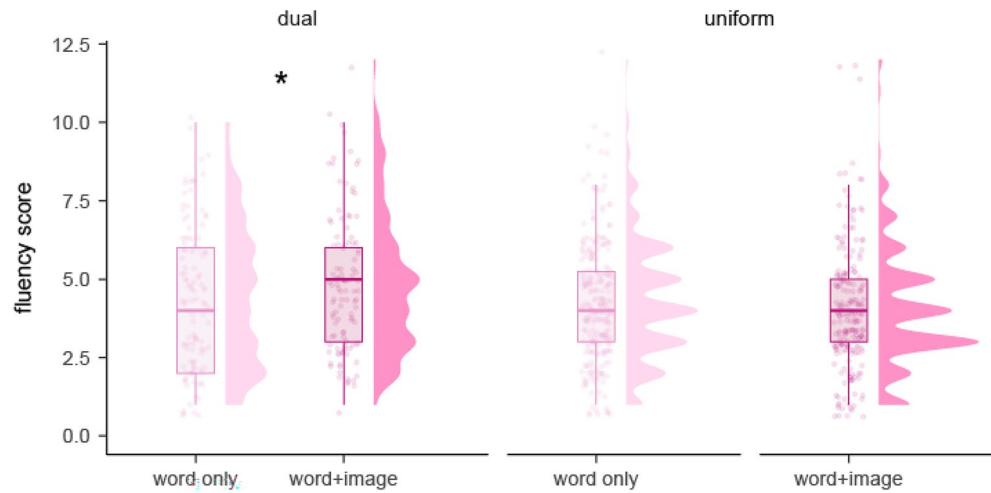


Figure 1. Idea fluency distribution in the three experimental procedures: dual (within-subject), uniform: word only (between-subject), and uniform: word + image (between-subject).

only condition ($p=0.757$, $\eta_p^2=0.002$) or the word + image condition ($p=0.071$, $\eta_p^2=0.067$, medium effect size; Supplementary Table 3 and 4).

Turning to idea creativity, we found a significant fixed effect of condition in the dual procedure ($p=0.048$, $\eta_p^2=0.146$, large effect size), with lower creativity scores in word + image ($M=1.32$, 95% CI [1.26, 1.38]) than word only trials ($M=1.41$, 95% CI [1.34, 1.48]; Supplementary Table 5). We also found that variance in creativity dropped significantly for word + image items ($\sigma^2=0.50$) as compared with word only items ($\sigma^2=0.59$; $F(522,483)=0.86$, $p=0.043$; Fig. 2).

When comparing word only and word + image conditions across uniform procedures (between-subject comparison), there was no significant difference ($p=.189$, $\eta_p^2=0.034$; Supplementary Table 6) and variance did not differ between conditions either; $F(1001,817)=1.03$, $p=0.666$.

When considering differences between procedures, there was no significant difference for the word only condition ($p=0.221$, $\eta_p^2=0.034$) or the word + image condition ($p=0.362$, $\eta_p^2=0.019$; Supplementary Table 7 and 8).

Experiment 2: AUT procedure in the list and cycle format

In Experiment 2, we found no significant difference in creativity overall between the list and cycle conditions of the dual procedure (within-subject comparison; $p=0.118$, $\eta_p^2=0.089$; Supplementary Table 9; Fig. 3).

Because participants could only generate one idea per cycle in the cycle condition, we also compared the first three ideas generated in the list condition with the three ideas generated in the cycle condition in the dual procedure. We found a significant fixed effect of condition, with more creative ideas in the cycle ($M=1.89$, 95%

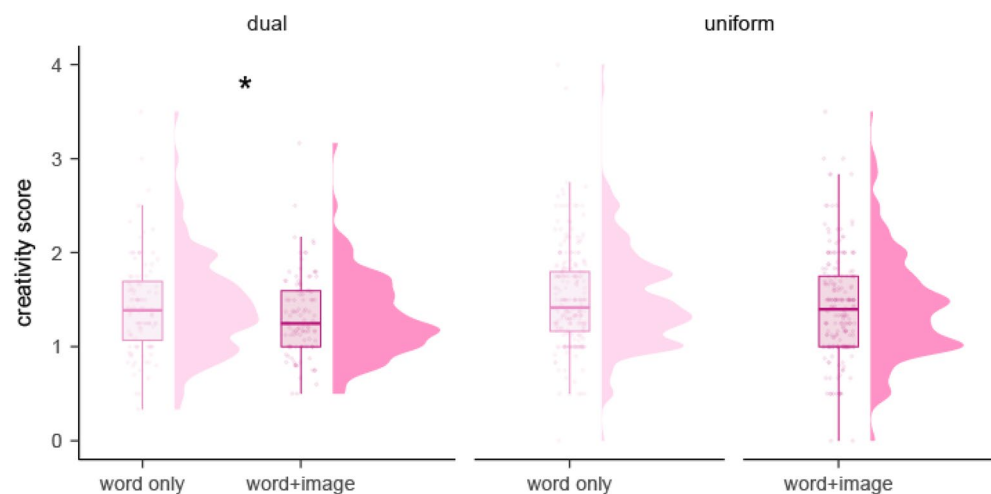


Figure 2. Distribution of creativity ratings in the three procedures of Experiment 1: Dual procedure: word only and word + image conditions in the same participants, uniform procedures: word only or word + image, in different participants.

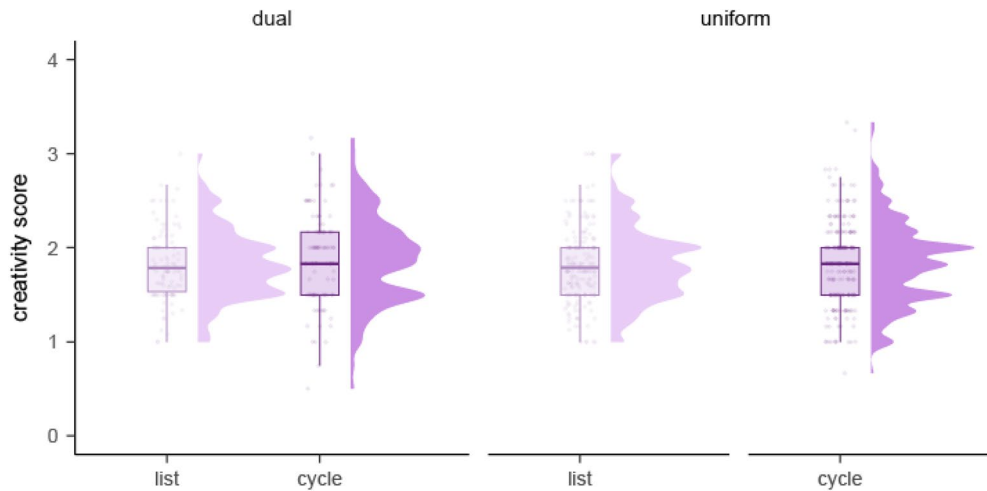


Figure 3. Distribution of creativity ratings in the three versions of Experiment 2: list, cycle, and dual (within-subject).

CI [1.82, 1.96]) than the list ($M = 1.73$, 95% CI [1.66, 1.80]) condition ($p = 0.026$, $\eta_p^2 = 0.165$, large effect size; Supplementary Table 10 and Fig. 4).

When comparing creative ratings between participants across the uniform procedures, there was no significant difference between the list and cycle conditions ($p = 0.721$, $\eta_p^2 = 0.003$; Supplementary Table 11).

When comparing creativity ratings for the first three ideas in the cycle and list procedure, we found a significant effect of idea sequence in the cycle ($p = 0.016$) but not the list ($p = 0.081$) procedure (Fig. 5, Supplementary Table 12 and 13). Post hoc pairwise comparisons revealed significantly higher creativity ratings for ideas generated in cycle three ($M = 1.91$, 95% CI [1.77, 2.05]) than cycle one ($M = 1.76$, 95% CI [1.63, 1.90]), $t(589.89) = -2.78$, $SE = 0.05$, $p = 0.015$.

When considering differences across the dual and uniform procedures, there was no significant difference either in the list ($p = 0.979$, $\eta_p^2 < 0.001$) or cycle ($p = 0.545$, $\eta_p^2 = 0.007$) conditions ($p_s > 0.1$; Supplementary Table 14 and 15).

Methods

Experiment 1: AUT items as (written) word or word + image

Participants

Sixty-seven participants ($M_{age} = 21$, $min_{age} = 18$, $max_{age} = 37$; 48 women, 14 men, 5 nonbinary) gave informed consent to take part in the experiment. All participants were Polish native speakers, they had normal or corrected-to-normal vision and hearing, and no neurological or language-related disorders.

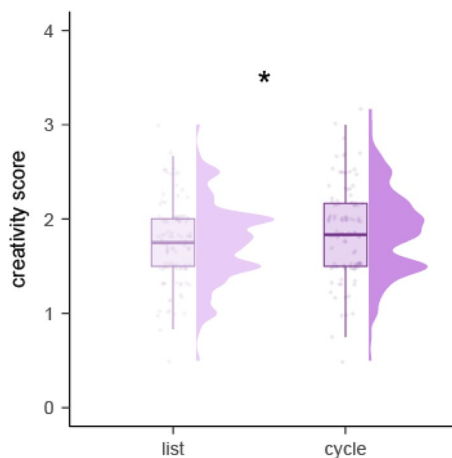


Figure 4. Distribution of creativity ratings for the first three ideas generated in the list and cycle conditions of the dual procedure (within-subject) of Experiment 2.

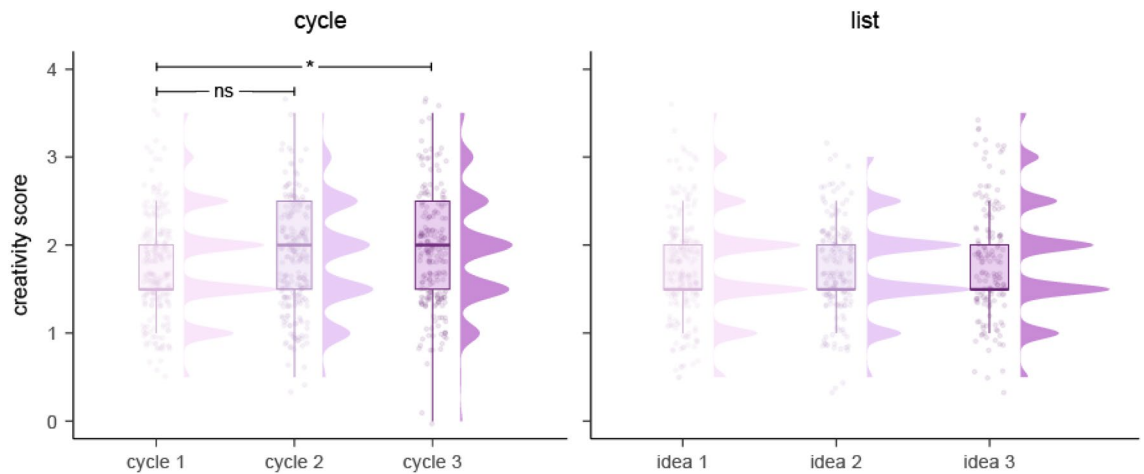


Figure 5. Creativity ratings generated in for the first three ideas in the cycle and list procedure in Experiment 2.

Stimuli

Thirty everyday objects were selected from previous studies on the AUT to be presented either as word only or word + image (image overlaid with a word). Item number was in the middle of the range used in the literature (one¹⁹ to 60²⁹), so as to allow item variation between participants (ten items randomly selected for each participant) and measure inter-item variability in our analyses.

Words, mostly selected from previous publications using the AUT as the core task, were presented in Arial font, in size 28 (for word + image stimuli) or 32 (for word stimuli) points and 2-point character spacing, in white outlined in black (for word + image stimuli only, text glow option in PowerPoint, to maximise readability when superimposed on images in the word + image condition). Stimuli are listed on the Open Science Framework server (link in the Data availability statement).

For each item, three images deemed “a standard depiction” by all researchers in our group were collated based on Google image searches/BOSS³⁵ and Linguapix³⁶ databases. Items were then presented within two norming sessions to 20 (27 items) and 24 participants (5 items) participants. Participants (1) provided the first label that came to mind upon viewing each image presented in isolation, (2) selected one photo out of three that in their opinion best depicted the object keyword, and (3) rated the degree of relatedness between each picture and word on a scale of 0 to 4 (see Supplementary Table 16; Fig. 6).

Procedure

Participants were randomly assigned to one of three experimental procedures: (i) uniform with word only condition items ($n = 24$), (ii) uniform with word + image condition items (image overlaid with a word, $n = 21$), or (iii) dual, consisting of half with word only and half with word + image trials ($n = 22$). Disparity in counts of occurrence is due to random allocation of procedures to participants. Each experimental procedure featured ten items selected randomly from a pool of 30. In the word only condition, items were single words presented at the centre of the screen. In the word + image condition, the same words were presented with their best matching image selected from norming, as a backdrop. The dual procedure was implemented to increase statistical power (within-subject comparisons) and investigate possible contextual effects of one condition on the other.

All procedures had the same internal structure. On each trial, a randomly selected AUT item was presented at the centre of the screen. Images were presented as background to words on a 23-inch LCD monitor with a resolution of 1920×1080 . Selection was random without repetition (i.e., no item could be selected both as word only and word + image). When prompted, participants had 15 s to type in a common use for that item. Next, participants had two minutes to type in unusual yet possible uses for that item. In the dual procedure, items were blocked (word only, image + word) with block order counterbalanced between participants. Participants completed the experiment in about 30 min and received 50 PLN or credit points for taking part in the experiment.



Figure 6. Three pictures of bricks used in the norming procedure.

Behavioural data analysis

Data preprocessing was done using the *tidyverse* package³⁷ in R³⁸. Participants' creativity scores were analysed with linear-mixed models (LMM) using *lme4*³⁹ and *lmerTest* packages⁴⁰. Tables with model summaries were computed with the *sjPlot* package⁴¹. Estimates of effect size were calculated using the *effectsize* package⁴².

For each LMM, we first computed a maximal model with a full random-effect structure, including subject- and item-related variance components for intercepts and by-subject and by-item random slopes for fixed effects⁴³. All fixed effects were coded using sum contrast coding. If maximal models did not converge and/or turned out to be too complex, we progressively selected more parsimonious LMMs following the recommendations from Bates, Kliegl, Vasishth, and Baayen (2018)^{44,45}. Small variance parameters were removed using the *lme4::rePCA* and *lme4::VarCorr*⁴⁶ functions until the LMMs were supported by the data. The final structure of each model is presented below:

Idea fluency:

- (1) Dual procedure (within-subject), word only vs. word + image: $fluency \sim word_only_vs_word_plus_image + (1 + word_only_vs_word_plus_image|participant) + (1|item)$
- (2) Uniform procedure (between-subject), word only vs. word + image: $fluency \sim word_only_vs_word_plus_image + (1|participant) + (1|item)$
- (3) Word only condition, uniform vs. dual procedure (between-subject): $fluency \sim uniform_vs_dual + (1|participant) + (1|item)$
- (4) Word + image condition, uniform vs. dual procedure (between-subject): $fluency \sim uniform_vs_dual + (1|participant) + (1 + uniform_vs_dual|item)$

Idea creativity:

- (1) Dual procedure (within-subject), word only vs. word + image: $Rating \sim word_only_vs_word_plus_image + (1 + word_only_vs_word_plus_image|participant) + (1|item)$
- (2) Uniform procedure (between-subject), word only vs. word + image: $Rating \sim word_only_vs_word_plus_image + (1|participant) + (1 + word_only_vs_word_plus_image|item)$
- (3) Word only condition, uniform vs. dual procedure (between-subject): $Rating \sim uniform_vs_dual + (1|participant) + (1 + uniform_vs_dual|item)$
- (4) Word + image condition, uniform vs. dual procedure (between-subject): $Rating \sim uniform_vs_dual + (1|participant) + (1|item)$

Estimates and significance of fixed effects and interactions (*p*-values) are based on the Satterthwaite approximation for LMM.

Assessment of AUT performance

Originality ratings were compiled and processed following the Consensual Assessment Technique (CAT)^{47–50}. Four raters judged the creativity of ideas produced by participants on a 5-point Likert scale (0—common, unoriginal use, 1—not very original but a bit uncommon, 2—quite original and quite uncommon, 3—original, uncommon, 4—highly original, rare). Raters were instructed that an answer considered creative needed to be novel and/or unique whilst being principally possible; principally impossible uses were given a score of 10 and were excluded from further analyses. For details about instructions for raters, see Supplementary Information. Raters were also encouraged to use the full-scale range. Inter-rater reliability was very high [ICC(C,4)=0.93]. Ratings were averaged across raters, resulting in one creativity rating per item and participant.

Experiment 2: AUT procedure as list or cycle

Participants

Seventy-two participants ($M_{age} = 22$, $min_{age} = 18$, $max_{age} = 41$; 49 women, 14 men, 8 nonbinary) gave informed consent to take part in the experiment. All participants were Polish native speakers, they had normal or corrected-to-normal vision and hearing, and no neurological or language-related disorders.

Stimuli

Stimuli were identical to the word + image stimuli described in Experiment 1.

Procedure

Participants were randomly assigned to one of three experimental procedures: (i) uniform with list trials ($n = 22$), (ii) uniform with cycle trials ($n = 25$), or dual (half list and half cycle trials, $n = 24$), disparity between counts being due to random allocation of procedures to participants. Each experimental procedure featured ten items selected randomly from a pool of 30 items and presented as word + image in all cases. In the list condition, participants had two minutes to list unusual uses for ten randomly selected AUT items. In the cycle condition, participants engaged in three successive 30 s cycles of idea generation-evaluation-selection per item. In each cycle, they were asked to think of unusual uses for the current item displayed at the beginning of each cycle and report their *single best* idea by typing into a blue frame displayed at the end of the cycle, so as to foster evaluation. In the dual procedure, participants engaged in a block of five list trials and a block of five cycle trials (with items randomly selected without repetition), and block order was counterbalanced between participants.

In all experimental procedures, participants first saw a randomly selected AUT item and typed in its common use (within 15 s). Next, they were presented with the item again and asked to come up with unusual yet possible uses for each object (within a list or cycle context). Thus, apart from the registration of the common use, the same item was presented three times in a row in the cycle but only once in the list condition. Participants completed the experiment in about 30 min. They received 50 PLN or credit points for taking part in the experiment.

Behavioural data analysis

Preprocessing and data analysis followed the same approach as in Experiment 1. The final structure of each model in Experiment 2 is presented below.

- (1) Dual procedure (within-subject), list vs. cycle condition:
- (2) $Rating \sim list_vs_cycle + (1 + list_vs_cycle|participant) + (1|item)$
- (3) Uniform procedures, list vs. cycle condition (between-subject): $Rating \sim list_vs_cycle + (1|participant) + (1 + list_vs_cycle|item)$
- (4) Cycle condition, uniform vs. dual procedure (between-subject): $Rating \sim uniform_vs_dual + (1|participant) + (1|item)$
- (5) List condition, uniform vs. dual procedure (between-subject): $Rating \sim uniform_vs_dual + (1|participant) + (1|item)$

Assessment of AUT performance

Three raters evaluated the creativity of ideas produced by participants following the same protocol as in Experiment 1. Inter-rater reliability was very high ($ICC(C,3) = 0.91$).

Ethics statement

The study was designed and performed in accordance with the Declaration of Helsinki. The Ethics Committee for Research Involving Human Participants at Adam Mickiewicz University, Poznań approved the study (Resolution no. 23/2021/2022). Participants were informed about the study protocol and gave their informed consent to participate in the study.

Discussion

In two experiments, we set out to test for differences in the AUT (i) elicited by words presented in isolation or together with an image, and (ii) related to the format of idea production, list or cycle.

In Experiment 1, the picture was used to disambiguate the word in word + image trials. Whilst fluency was greater in the word + image condition, we found the expected slight drop in originality ratings for word + image compared to word only trials. We also found the expected lower variance in originality ratings for word + image than word only prompts. In both cases, however, the differences were only found in the within-subject, not the between-subject comparisons.

In Experiment 2, as expected we found that creativity was significantly higher in the cycle than list condition, albeit only when the output was quantitatively matched, that is, for the first three ideas of the list condition. Furthermore, idea creativity increased with cycle, although a significant difference was only found between cycle 3 and cycle 1. We found no equivalent difference in the list condition. Interestingly, as was the case in Experiment 1, both differences found here concerned within-subject comparisons, not between-subject comparisons.

As mentioned in the introduction, Chrysikou et al.²⁸ and George et al.²⁹ showed that inclusion of images in the AUT can affect the quality of generated ideas, resulting in normative and less original answers. However, one should not deduct from this result, which is consistent with ours in Experiment 1, that introducing images alongside words is undesirable (e.g., because it reduces idea originality at face value). Indeed, the relatively lower originality measured in the word + image condition is likely linked to reduced variability between participants, since the disambiguating picture would have harmonised participant expectations regarding the objects referred to by the prompts. However, we note that idea fluency was significantly greater in the image + word condition than the word only condition, which can be seen as a trade-off effect (more productive creativity in the image + word condition). This observation might relate to the increased level of constraint realised by the introduction of a disambiguating picture. Haught⁵¹ showed that presenting a picture as a prompt led to more creativity in a sentence production task than presenting words, in line with predictions of the Integrated Constrains in Creativity model⁵². Taken together, the disambiguating effect of images and the implicit constraint imposed by them may have compensated each other to result in only mild differences between the two experimental conditions in Experiment 1.

A review of the literature on the AUT as a classic paradigm to quantify creativity shows that most studies used word only prompts^{6,18–22} which we believe might have inflated inter-individual variability at the outset of each trial, due to the lack of specification inherent to word labels. Thus, if images appear to lower originality in the AUT, this may be a necessary sacrifice to homogenise the conceptual plane from which participants start ideation. Furthermore, it must be noted that improving the baseline not only benefits participant data but also enhances the reliability of the rating procedure. If, as participants do, raters start from a common ground when being exposed to images as well as labels, they are more likely to deliver consistent ratings. The effect of presenting images together with words thus contributes to reducing spurious variability (that can be mistaken for originality) on both the level of participants and raters.

In Experiment 2, we attempted to increase the contribution of evaluation in order to enhance the potency of the AUT to measure creativity as opposed to merely idea originality. As expected, we found that the first

three ideas of the list procedure were less original than the three ideas generated during the cycle procedure. This observation is *convergent* with the serial order effect often reported in AUT, which has been linked to progressively greater involvement of executive function as time passes on task and resulting in greater originality for ideas generated later¹⁹. This is further compatible with the observation of greater originality in cycle 3 compared to cycle 1 observed here, which can be construed as a serial order effect across cycles in our adaptation of the task. These results are also consistent with the idea that increasing the level of constraint in a creative idea generation task like the AUT leads to greater idea originality⁵². We note, however, that this difference did not apply overall, which also suggests, *a contrario*, that the originality of ideas produced in the list procedure did not significantly go beyond that of the cycle procedure. Thus, the cycle procedure appears to have captured most of the originality to be expected in the AUT in conditions where participants are more readily comparable with one another, because the number of items is constant.

In both experiments, between-subject comparisons yielded almost no difference between conditions, and this is presumably due to uncontrolled variance stemming from individual differences and likely superseding subtle effects that are detected in within-subject comparisons. We thus argue that measures of creativity using the AUT as the core paradigm should be compared within-subject wherever possible. Furthermore, we acknowledge both widely recognised limitations of the AUT as a tool to measure creative ideation and more subtle limitations relating to the kind of cognitive processes at work in this task that significantly depart from what can be expected in more ecologically valid contexts. First, we acknowledge that the AUT is a mostly artificial task since people are rarely asked to come up with as many creative uses for an object as possible within a limited time period (except perhaps in engineering design education). However, our goal was not to propose a more ecologically valid adaptation of the AUT. It was rather to improve the baseline from which participants start and increase the evaluative dimension of the task. This was made possible by comparing our manipulations to standard AUT measures established in the field.

In conclusion, using pictures in combination with words may disambiguate the cue in tasks measuring creativity and participants' performance may be more comparable across conditions, whilst incurring minimal loss of originality. This is even more important considering that the same manipulation is likely to have had the same disambiguating effect on raters, and may thus have contributed to the high inter-rater reliability in our experiments. We believe that the cycle procedure employed here is a valid way of fostering idea evaluation without incurring a loss of originality whilst making participants' output more comparable than that of a list procedure. Finally, the lack of difference observed here in between-group comparisons argues in favour of using within-subject designs whenever possible. Although this is true of essentially all comparisons in experimental psychology, this seems to be particularly relevant in the domain of creativity testing.

Data availability

Stimuli, raw data, ratings, and R scripts are listed on the Open Science Framework server (https://osf.io/q5hfz/?view_only=c4e85e49224848f5bd56f0d4949521ba).

Received: 29 November 2023; Accepted: 27 May 2024

Published online: 31 May 2024

References

1. Beaty, R., Benedek, M., Silvia, P. J. & Schacter, D. L. Creative cognition and brain network dynamics. *Trends Cogn. Sci.* **20**, 87–95 (2016).
2. Schubert, E. Creativity is optimal novelty and maximal positive affect: A new definition based on the spreading activation model. *Front. Neurosci.* **15**, 612379 (2021).
3. Cortes, R. A., Weinberger, A. B., Daker, R. J. & Green, A. E. Re-examining prominent measures of divergent and convergent creativity. *Curr. Opin. Behav. Sci.* **27**, 90–93 (2019).
4. Ausubel, D. P. Creativity, general creative abilities, and the creative individual. *Bull. Council. Res. Music Educ.* **3**, i–v (1964).
5. Runco, M. A. & Jaeger, G. J. The standard definition of creativity. *Creat. Res. J.* **24**, 92–96 (2012).
6. Madore, K. P., Jing, H. G. & Schacter, D. L. Divergent creative thinking in young and older adults: Extending the effects of an episodic specificity induction. *Mem. Cogn.* **44**, 974–988 (2016).
7. Kharkhurin, A. V. Creativity.4in1: Four-criterion construct of creativity. *Creat. Res. J.* **26**, 338–352 (2014).
8. Simonton, D. K. Taking the U.S. Patent office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creat. Res. J.* **24**, 97–106 (2012).
9. Finke, R. A., Ward, T. B. & Smith, S. M. *Creative Cognition: Theory, Research, And Applications* (MIT Press, 1996).
10. Eysenck, H. J. Creativity, personality and the convergent-divergent continuum. In *Critical Creative Processes* (ed. Eysenck, H. J.) 95–114 (Hampton Press, 2003).
11. Guilford, J. P. *The nature of human intelligence* (McGraw-Hill, 1967).
12. Palmiero, M., Nori, R., Piccardi, L. & D'Amico, S. Divergent thinking: The role of decision-making styles. *Creat. Res. J.* **32**, 323–332 (2020).
13. Gabora, L. The neural basis and evolution of divergent and convergent thought. In *The Cambridge Handbook of the Neuroscience of Creativity* (eds Jung, R. E. & Vartanian, O.) 58–70 (Cambridge University Press, 2018). <https://doi.org/10.1017/9781316556238.005>.
14. Leon, S. A., Altmann, L. J. P., Abrams, L., Gonzalez Rothi, L. J. & Heilman, K. M. Divergent task performance in older adults: Declarative memory or creative potential?. *Creat. Res. J.* **26**, 21–29 (2014).
15. Mednick, S. A. & Mednick, M. T. *Examiner's Manual, remote Associates test: College and Adult Forms 1 and 2* (Houghton Mifflin Boston, 1967).
16. Bowden, E. M. & Jung-Beeman, M. Normative data for 144 compound remote associate problems. *Behav. Res. Methods Instrum. Comput.* **35**, 634–639 (2003).
17. Wu, C.-L., Huang, S.-Y., Chen, P.-Z. & Chen, H.-C. A systematic review of creativity-related studies applying the remote associates test from 2000 to 2019. *Front. Psychol.* **11**, 573432 (2020).
18. Alhashim, A. et al. Work in Progress: Assessing Creativity of Alternative Uses Task Responses: A Detailed Procedure. in *2020 ASEE Virtual Annual Conference Content Access Proceedings* 35612 (ASEE Conferences, 2020). <https://doi.org/10.18260/1-2--35612>.

19. Beaty, R. & Silvia, P. J. Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychol. Aesthet. Creat. Arts* **6**, 309–319 (2012).
20. Benedek, M. *et al.* To create or to recall original ideas: Brain processes associated with the imagination of novel object uses. *Cortex* **99**, 93–102 (2018).
21. Gilhooly, K. J., Fioratou, E., Anthony, S. H. & Wynn, V. Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *Br. J. Psychol.* **98**, 611–625 (2007).
22. Hass, R. W. & Beaty, R. Use or consequences: Probing the cognitive difference between two measures of divergent thinking. *Front. Psychol.* **9**, 2327 (2018).
23. Vartanian, O. *et al.* The relationship between methods of scoring the alternate uses task and the neural correlates of divergent thinking: Evidence from voxel-based morphometry. *NeuroImage* **223**, 117325 (2020).
24. Bai, H., Leseman, P. P. M., Moerbeek, M., Kroesbergen, E. H. & Mulder, H. Serial order effect in divergent thinking in five- to six-year-olds: Individual differences as related to executive functions. *J. Intell.* **9**, 20 (2021).
25. Cardoso, C. & Badke-Schaub, P. The influence of different pictorial representations during idea generation. *J. Creat. Behav.* **45**, 130–146 (2011).
26. George, T. & Wiley, J. Need something different? Here's what's been done: Effects of examples and task instructions on creative idea generation. *Mem. Cogn.* **48**, 226–243 (2020).
27. Saffran, E. M., Coslett, H. B. & Keener, M. T. Differences in word associations to pictures and words. *Neuropsychologia* **41**, 1541–1546 (2003).
28. Chrysikou, E. G., Motyka, K., Nigro, C., Yang, S.-I. & Thompson-Schill, S. L. Functional fixedness in creative thinking tasks depends on stimulus modality. *Psychol. Aesthet. Creat. Arts* **10**, 425–435 (2016).
29. George, T., Mielicki, M. K. & Wiley, J. Great expectations: Misleading effects of images in the alternate uses task. *Psychol. Aesthet. Creat. Arts* **17**, 56–67 (2023).
30. Benedek, M. & Neubauer, A. C. Revisiting Mednick's model on creativity-related differences in associative hierarchies. Evidence for a common path to uncommon thought. *J. Creat. Behav.* **47**, 273–289 (2013).
31. Beaty, R. & Johnson, D. R. Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behav. Res.* **53**, 757–780 (2021).
32. Barbot, B. The dynamics of creative ideation: Introducing a new assessment paradigm. *Front. Psychol.* **9**, 2529 (2018).
33. Beaty, R., Kenett, Y., Hass, R. W. & Schacter, D. L. 2019 *A Fan Effect for Creative Thought: Semantic Richness Facilitates Idea Quantity but Constrains Idea Quality*. 10.31234/osf.io/pfz2g.
34. Forthmann, B. *et al.* The be-creative effect in divergent thinking: The interplay of instruction and object frequency. *Intelligence* **57**, 25–32 (2016).
35. Brodeur, M. B., Guérard, K. & Bouras, M. Bank of standardized stimuli (BOSS) phase II: 930 new normative photos. *PLoS One* **9**, e106953 (2014).
36. Krautz, A. E. & Keuleers, E. LinguaPix database: A megastudy of picture-naming norms. *Behav. Res.* **54**, 941–954 (2022).
37. Wickham, H. *et al.* Welcome to the Tidyverse. *JOSS* **4**, 1686 (2019).
38. R Core Team. *R: A Language and Environment for Statistical Computing*. (2021).
39. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using **lme4**. *J. Stat. Soft.* **67**. <https://doi.org/10.18637/jss.v067.i01> (2015).
40. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. **lmerTest** package: Tests in linear mixed effects models. *J. Stat. Soft.* **82**. <https://doi.org/10.18637/jss.v082.i13> (2017).
41. Lüdtke, D. *sjPlot: Data Visualization for Statistics in Social Science*. (2022).
42. Ben-Shachar, M., Lüdtke, D. & Makowski, D. effectsize: Estimation of effect size indices and standardized parameters. *JOSS* **5**, 2815 (2020).
43. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
44. Bates, D., Kliegl, R., Vasishth, S. & Baayen, H. Parsimonious Mixed Models. Preprint at <http://arxiv.org/abs/1506.04967> (2018).
45. Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. & Bates, D. Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* **94**, 305–315 (2017).
46. Pinheiro, J. C. & Bates, D. M. *Mixed-Effects Models in Sand S-PLUS* (Springer, 2000). <https://doi.org/10.1007/978-1-4419-0318-1>.
47. Amabile, T. M. Social psychology of creativity: A consensual assessment technique. *J. Personal. Soc. Psychol.* **43**, 997–1013 (1982).
48. Amabile, T. M. The social psychology of creativity: A componential conceptualization. *J. Personal. Soc. Psychol.* **45**, 357–376 (1983).
49. Cseh, G. M. & Jeffries, K. K. A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychol. Aesthet. Creat. Arts* **13**, 159–166 (2019).
50. Kaufman, J. C., Lee, J., Baer, J. & Lee, S. Captions, consistency, creativity, and the consensual assessment technique: New evidence of reliability. *Think. Skills Creat.* **2**, 96–106 (2007).
51. Haught, C. The role of constraints in creative sentence production. *Creat. Res. J.* **27**, 160–166 (2015).
52. Tromp, C. Integrated constraints in creativity: Foundations for a unifying model. *Rev. Gen. Psychol.* **27**, 41–61 (2023).

Acknowledgements

The study was a part of the project “Unleashing the Power of Second Language Mastery: Cognitive Foundations of the Foreign Language Effect and Strategic Language Use”, financed by the National Science Centre of Poland (2021/43/B/HS6/00184) and Polish National Agency for Academic Exchange (PPN/PRO/2020/1/00006).

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: K.B.D., R.J., G.T., O.W. data collection: K.B.D., I.K., O.W.; analysis and interpretation of results: K.B.D., R.J., I.K., G.T., O.W.; draft manuscript preparation: K.B.D., R.J., I.K., G.T., O.W. Editing and revision of manuscript: K.B.D., R.J., I.K., G.T., O.W. All authors reviewed the results and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-63225-2>.

Correspondence and requests for materials should be addressed to O.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024