THEORETICAL DEVELOPMENT OF CREATIVE THINKING:
THE ROLE OF INTEGRATIVE THINKING AND SOCIAL COGNITION

NUOYA TAN

Student Number: 16094599

A thesis submitted for the degree of PhD in Psychology and Language Science

Year 4

Department of Experimental Psychology

UCL

Declaration

I, Nuoya Tan confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgement

I want to express my gratitude to my parents, Rui and Yiping. Thanks for always having me and offering me trust, care, and understand for every aspect of my life. Also, I would like to offer special thanks to my supervisor, Dr Lasana T. Harris. Thanks for guiding and supporting me in the PhD project and my career. You are the best supervisor I could ever imagine. Finally, please allow me to extend my thanks to those who have assisted, contributed, and commented on this research project. Thanks for all the help.

Abstract

Creative thinking is the psychological mechanism underlying the descriptive process that produces real-life creative outcomes. However, the connection between individual creative thinking and real-life creativity remains unclear. For example, the widely employed psychometric tools for creative thinking showed limited predictive power towards real-life creativity. In addition, empirical evidence for the social psychology of creativity is inconsistent. Also, the links between creative thinking and social cognitive process are rarely validated in the field. Besides, some domains that require creativity lack guiding theories and empirical evidence. Therefore, this research project aimed to advance the understanding of creative thinking and its role in real-life situations.

To address the knowledge gap and fulfil the central purpose, we conducted four pilot and seven main studies using quantitative research methods. Accordingly, we created an integrative-thinking-based psychometric tool - Function Synthesis Task and validated its discriminate validity and predictive ability towards engineering students' creative product design. To understand the link between social comparison and creativity, we produced a new experimental paradigm that addressed existing methodological issues. We employed the paradigm and found that competition and star rating feedback altered speed or performance in creative thinking tasks. Besides, we produced a new product design task based on a hot topic at the time and found that ranking feedback benefited engineering students' creative performance in the task. Moreover, we designed a new un-stereotype intervention and found its effectiveness in improving marketers' divergent thinking. We also found that advertising stereotypes increased audiences' perceived creativity.

Our research shows that integrative thinking and social cognition might play essential roles in developing the theory of creative thinking and offers novel research tools for future studies. We also form practical advice to guide educators, organisational leaders, and policymakers to promote creativity, diversity, and inclusion in real-life situations.

Impact Statement

Until 2022, the psychological investigation of creativity has emerged and grown for more than half-century. However, the empirical evidence for creative thinking and real-life creativity remains disconnected. The disconnection could result in two side effects. First, the effectiveness of real-life creativity training may not be entirely fulfilled due to the lack of prerequisites. Second, the predictive power and measurability of real-life professional creativity may be limited due to the lack of appropriate measures. To address the issue, we created a new psychometric tool – Function Synthesis Task (FST), based on the concept of integrative thinking that has been repeatedly found in the process of creative product generation. We found that FST grasped different aspects of creative thinking from other tasks and showed more vital predictive power towards real-life creativity. Based on the effectiveness of FST in predicting real-life creativity, creativity training, coaching, or consultancy might incorporate integrative thinking in their programmes. Also, the educators may use FST to predict or evaluate students' creative thinking. Besides, our research may benefit the theoretical development of integration. Integration is an essential concept in the field of consciousness and metacognition. Thus, developing integrative creativity will allow researchers to see the overlap among these mental processes and develop an in-depth understanding of human beings.

Another unsolved problem is the unclear relationship between intelligence and creativity, and we proposed that social cognition may help us differentiate two cognitive abilities. In detail, intelligence tests usually have pre-determined objective answers so people can spend as many cognitive resources as possible on the task. However, the reference points of creative products are posterior and subjective. Due to the lack of objective pre-determined reference points, producing creative outcomes may automatically induce comparing with others' products and mentalising others' judgements. In other words, creativity may induce social cognition spontaneously. The different levels of automatic social information processing in intelligence and creativity offer a new perspective to understanding the relationship between human beings' two crucial higher-order cognitive abilities. In the long term, theoretical development may help gifted education to detect the gifts and unique potentials of as many students as possible and provide them with increasingly tailored and valuable education.

Moreover, our research can promote diversity, equality, and inclusion (DEI) in our society. For instance, we found a positive association between divergent thinking and stereotype avoidance in advertisement generation. We also discussed that divergent thinking might avoid the stereotyping rebound effect issue in traditional counter-stereotype approaches. Therefore, the policymakers and funding organisations may put increasing efforts into encouraging the supporting academic and industrial projects that enhance creativity and societal DEI at the same time.

Table of Contents

List of Figures and Tables

Chapter 1: Prologue

Creativity is a highly valued attribute for human beings as we consider it a human characteristic that drives our evolution, promotes us to the most intelligent species and differentiates us from other animals. We believe human creativity plays a crucial role in astronomy investigation, technology transformation, artistic expression, societal emergence, and social development. It is an interdisciplinary concept whose fundamental role is to forward our observable or perceivable world or its elements from a known form to an anonymous form.

After 70 years of investigation, creativity researchers now have widely accepted a two-criteria definition such that creative ideas and products are novel and effective (Barron, 1955; Bruner, 1962; Cattell & Butcher, 1968; Guilford, 1950; Jackson & Messick, 1965; Kneller, 1965; Newell et al., 1962; Runco & Jaeger, 2012; Simonton, 2011; Stein, 1953). Leading by this definition, psychologists are interested in the creative process that produces creative ideas and products from various perspectives. From the perspective of cognitive psychology, researchers look at creative thinking, which indicates the mental structures and processes that lead to creative ideas and products (Agnoli et al., 2020; Roger E Beaty, Paul J Silvia, et al., 2014; Finke et al., 1992; Joy Paul Guilford, 1967; Mednick, 1962; Mark A. Runco & Selcuk Acar, 2019) . Meanwhile, social psychologists investigate the effect of interpersonal factors (e.g., competition, evaluation, and surveillance) on the creative process (Amabile, 2011; Amabile, 1983; Amabile et al., 1990; Csikszentmihalyi, 2014; Sternberg & Lubart, 1996). Organisational psychology looks at the descriptive structures and process that drives creative products (Anderson et al., 2014; Andriopoulos, 2001; Eveleens, 2010; Martins & Terblanche, 2003; Mumford, 2000; Mumford & Gustafson, 1988; Shalley et al., 2004; Zhou & George, 2001). Finally, educational psychologists are interested in teaching structures and processes that bring creative ideas to students (e.g., curriculum and training) (Collard & Looney, 2014; Craft, 2003; Cropley, 2001; Hernández-Torrano & Ibrayeva, 2020). The four perspectives overlap and benefit the development of each other. This project covered and incorporated the above perspectives to theoretically develop the concept of creative thinking.

Central Purpose

The central purpose of this project was to advance our understanding of creative thinking and its role in real-life circumstances. In detail, we centred on creative thinking, related it to social cognition, and applied it to organisational and educational practice.

Key Concepts

In the field of creativity, we refer to creative thinking mainly as divergent thinking and convergent thinking since these are two essential cognitive styles that play key roles in producing novel thoughts (Mark A. Runco & Selcuk Acar, 2019). Specifically, divergent thinking indicates one's mental exploration in various directions, and one of the widely employed divergent thinking tasks is the alternative uses task (AUT) (Joy Paul Guilford, 1967). On the other hand, convergent thinking is finding a single correct answer or single best solution for a well-developed problem (Guilford, 1950; Runco, 2010). The most popular convergent thinking tasks are the remote associate test (RAT) (Mednick, 1962) and the compound remote associate test (CRAT) (Bowden & Jung-Beeman, 2003).

Knowledge Gaps

However, in at least three aspects, existing creative thinking tasks and creativity in real-life situations disconnect (creative thinking-reality disconnection). First, existing divergent thinking tasks (Barron, 1955; Batey & Furnham, 2006; Drevdahl, 1956; Hocevar, 1981; Lowenfeld & Beittel, 1959; Merrifield et al., 1964; Okuda et al., 1991; Zeng et al., 2011) and convergent thinking tasks (Andrews, 1967; Roger E Beaty, Paul J Silvia, et al., 2014; Brougher & Rantanen, 2009; Davis & Belcher, 1971; Worthen & Clark, 1971) exhibited low predictive power towards real-life creative performance and accomplishment.

Second, the findings on the impact of real-life social feedback on creative thinking are inconsistent (Amabile, 1982a; Balietti et al., 2016; Bittner & Heidemeier, 2013; Clark & Goldsmith, 2006; Conti et al., 2001; De Vet & De Dreu, 2007; Eisenberg & Thompson, 2011; Erat & Gneezy, 2016; Landers et al., 2019; Michinov & Primois, 2005; Raina, 1968; Redifer et al., 2021; Shalley & Oldham, 1997; Strong & Gray, 1972; Van de Ven et al., 2011; Van Knippenberg et al., 1981; Van Leeuwen & Baas, 2017).

Moreover, creativity researchers find it challenging to recruit sufficient marketers to join an experiment due to the lack of communication between academia and industry (Amabile, 1996; West et al., 2019), , which may result in the lack of scientific applications of creative thinking. For example, existing findings showed that creative thinking could be linked with stereotype avoidance. However, we do not have enough field studies to validate how creative thinking helps solve these social issues in real-life situations.(Gaither et al., 2015; Gocłowska & Crisp, 2013, 2014; Kharkhurin, 2011; Prior & MacWhinney, 2010; Sassenberg & Moskowitz, 2005; Sassenberg et al., 2017; Wen et al., 2019; Zuo et al., 2019).

Beyond the above three aspects, the disconnection between creative thinking and creativity in real-life situations has adverse side effects. Here, let us take the development of creative cognition as an example. With the development of technology, cognitive psychologists have been generating biological evidence of creative mental structures and processes and have made fruitful contributions (Beaty et al., 2020; Benedek & Fink, 2019; Benedek et al., 2014; Benedek & Neubauer, 2013; Gilhooly et al., 2012; Radel et al., 2015; Vartanian, 2016; Zabelina, 2018b). Nonetheless, the creative cognition and creative neuroscience research have been overwhelmed by AUT and RAT, which exhibited low predictive validity[1] toward real-life creativity. Accordingly, the value of creative cognition, neuroscience, and physiology studies may be harshly diminished if they cannot detect the mechanisms responsible for real-life creativity (Runco & Acar, 2012).

Furthermore, the creative thinking-reality disconnection exists in specific domains. For instance, recent studies reported that creative thinking and engineering practice disconnect, and existing creative thinking skills could not guarantee high-quality engineering solutions (Hirshfield & Koretsky, 2020). The situation may be one factor that results in the slow development of creativity in engineering education (Carpenter, 2016; Cropley, 2012; Cropley & Cropley, 2010; Richards, 1998).

Research Questions

To address the knowledge gaps, we generated a new psychometric tool – the Function Synthesis Task (FST), based on the widely mentioned but rarely validated concept called integrative thinking (also called integration and synthesis in the literature). Employing AUT, CRAT, and FST, we proposed four research questions that addressed this project's central purpose and the knowledge gaps in existing research. First, to advance the understanding of creative thinking, we investigated the statistical relationship between AUT, CRAT, and FST (Chapter 2). We also

---

[1] Predictive validity indicates the creative thinking tasks' predictive power towards real-life creative performance rather than the performance at a time point towards the performance at a later time point. In our thesis, predicative validity and predictive power are interchangeable words.

examined the effect of social comparison on AUT, CRAT, and FST (Chapter 3). These studies helped us identify the relationship between divergent, convergent, and integrative thinking and clarified the effect of social feedback on different kinds of creative thinking. To understand creative thinking in real-life situations, we assessed the impact of stereotypes on marketers' creative thinking and audiences' perceived advertising creativity in a real-life context in collaboration with industrial organisations (Chapter 4). In addition, we tested the predictive power of AUT, CRAT, and FST toward creative engineering design for COVID-19 prevention. The investigation helped us link creative thinking with real-life engineering problem-solving and addressed the role of creative thinking in engineering education (Chapter 5).

Main Findings

To answer the research questions, we conducted seven quantitative studies (along with additional four pilot studies). The results showed that 1) FST grasped different aspects of creative thinking from AUT and CRAT, 2) social comparison feedback enhanced the creative performance in FST and engineering product design, and the time participants spent into solving the CRAT. However, the social feedback did not affect AUT performance. Also, 3) stereotype avoidance benefitted marketers' divergent thinking and caused a U-shape effect on the audience's perceived creativity of advertisements. In addition, 4) FST and AUT, instead of CRAT, predicted engineering creativity. The integrated main findings, limitations, and implications were addressed in the General Discussion (Chapter 6).

We summarised the essential information of this PhD project in Figure 1. Please refer to the figure to track the reading progress.

Figure 1

Project overview.

Central purpose: advance the understanding of creative thinking and its role in real-life circumstances				
Chapter No.	**Research questions**	**Addressed knowledge gaps**	**Literatures scope**	**Key findings**
Chapter 2	Generate FST based on the concept of integrative thinking Investigate the statistical relationship among AUT, CRAT, and FST	Creative thinking research has been overwhelmed by AUT and (C)RAT	Creative thinking and psychometric tools The link between creative product, creative process, and creative thinking	FST grasped qualitatively different aspects of creative thinking from AUT and CRAT
Chapter 3	Examine the effect of social comparison on AUT, CRAT, and FST	Unclear impacts of social feedback on creative thinking	The link between social comparison and creative thinking	Social comparison benefitted efforts in FST and CRAT but did not affect AUT
Chapter 4	Assess the impact of stereotypes on marketers' creative thinking and audiences' perceived creativity	The lack of field study that validates a robust link between creativity and stereotypes	The link between stereotypes and creativity Advertising creativity	Stereotype avoidance benefitted marketers' divergent thinking and caused a U-shape effect on audiences' perceived creativity
Chapter 5	Determine the predictive validity of AUT, CRAT, and FST toward engineering product design creativity	Low predictability of dominant creative thinking tasks towards real-life creativity The lack of engineering creativity research	Engineering creativity	FST and AUT, instead of CRAT, showed predictive power towards engineering creativity Social comparison benefitted engineering creativity
Chapter 6	General discussion: integrated main findings, limitations, and implications.			

Chapter 2: Creative Thinking and Function Synthesis Task

Chapter 2 starts with a review of the psychometric tools (see Appendix A for *Creativity Psychometric Tools Searching Process*) and two creative thinking concepts - divergent and convergent thinking. After that, we address and analyse the reasons for the low predictive power of existing psychometric tools. Beyond that, we discuss how integrative thinking helps solve the low-predictive-power issues and why existing psychometric tools are insufficient. Finally, based on integrative thinking, we introduce a new psychometric tool, FST. We also demonstrate a correlational study that examined the statistical relationship between AUT, CRAT, and FST.

Literature Review

Divergent thinking.

Divergent thinking is one of the most important indicators of individual creative thinking. Its origin is rooted in Guilford’s structure-of-intelligence model, in which Guilford divided intelligence into 120 abilities such as cognition, memory, divergent production, and evaluation. In detail, Guilford proposed that divergent thinking ability refers to the variety of the ideas one can produce in problem-solving, and the level of variety is determined by the number, the category number, and the level of details of the produced ideas (Joy P Guilford, 1967; Joy Paul Guilford, 1967). This classic definition of divergent thinking ability has been piloting the theoretical and methodological development of divergent thinking since its emergence and is still widely accepted and employed.

Psychometric tools.

Researchers developed a test battery for individual differences in originality (i.e., Wilson Test Battery) in the early 1950s. The Unusual Uses Task (UUT) in the Wilson Test Battery is a widely employed psychometric tool for divergent thinking (Robert C Wilson et al., 1953; R. C. Wilson et al., 1953). In the UUT, we provide participants with a common object and a common use it can serve and ask them to list six other uses. For example, we provide participants with “newspaper” and a stereotypical use such as “reading”, and participants might list other uses such as “start a fire”.

Subsequent studies adjusted UUT, and AUT is one of the adjusted versions (Joy P Guilford, 1967). In AUT, we provide people with a household object (e.g., “paperclip”) and ask them to come up with as many uses as possible. The AUT can be found in many test batteries, including the Wilson Test Battery (R. C. Wilson et al., 1953), Getzel-Jackson Test Battery (GJTB) (Getzels & Jackson, 1961, 1962), Test of Creativity in Engineering (TCE) (Harris, 1960a, 1960b), Wallach-Kogan Creativity Test (WKCT) (Wallach & Kogan, 1965b), and Torrance Test of Creative Thinking (TTCT) (Torrance, 1972).

In addition to UUT and AUT, other divergent thinking tasks provide people with one concept, such as “summer”, and ask them to produce associations differently. The Free Association Task asks people to list as many associations relate to the concept as possible (e.g., “beach, holiday, and ice cream”). The Association Chain Task asks people to produce chains of associations. Only the first association should relate to the concept, and the following should be related to the last associative response (e.g., “beach, sand, castle, war, weapon, and technology”). The Dissociation Task asks people to think of many unrelated or remotely related concepts (e.g., “banana, button, and key”) (Benedek et al., 2012). Besides, the Instance Task in WKCT asks people to list as many instances of a concept as possible. For “a round thing”, one might come up with “ball”, “earth”, and “stone” (Wallach & Kogan, 1965a).

Some divergent thinking tasks provide people with a visualised or written item and ask them to think about the item’s meaning (e.g., cause and effect). For example, some tasks provide people with a visualised object and ask people to list as many “what is this” as possible. The example tasks are a Visual Object Association in TCE (Harris, 1960a, 1960b), a Pattern Meaning Task and a Line Meaning Task in WKCT, and the Line Meaning Task (Wallach & Kogan, 1965a). A Questioning Task in TTCT provides people with a contextual drawing and has them ask as many questions as possible to figure out what is happening in the drawing. A Causes Guessing Task and a Consequence Guessing Task in TTCT provide people with a drawing and ask them to think of the causes (consequence) that can result in (from) the action shown in the drawing (Torrance, 1972). A Causes Guessing Task and a Consequence Guessing Task in C-SAT are more specific (Sak & Ayas, 2013). The former task provides participants with a food chain figure and a graph of the changes in this chain. Participants are asked to list as many chain causes as they can. In the later task, participants are provided with a figure of a biological experiment and asked to list related research hypotheses.

In addition, a Number Association Task in Wilson Test provides people with one number and asks them to list as many synonyms, uses, and things as possible for which the number can associate. For example, given the number four, participants might come up with “coach-and-four” and “quartet” (Robert C Wilson et al., 1953; R. C. Wilson et al., 1953). A World Associations Task in GJTB provides people with one common word, such as “bark”, and asks them to think of as many definitions as possible for “bark” (Getzels & Jackson, 1961). A Product Improvement Task in TTCT provides people with one product, such as “a stuffed toy”, and asks people to come up with many unusual ways to change the toy to make it more fun to play with (Torrance, 1972). Additionally, an Extracurricular Activity Ideation Task asks people to list as many events as possible that relate to one group of people. For example, given college students, one of the answers might be organising an event to promote an on-campus university organisation dedicated to the cinema (Forgeard & Benson, 2019).

Divergent thinking tasks measure the novelty dimension of ideas via five aspects of participants’ responses. The measurements are fluency, flexibility, originality, frequency, and elaboration. Let us take AUT as an example. When we ask one to come up with as many alternate uses of bricks as possible, one’s creativity level is positively related to several measures. They are the number of different uses (fluency), the number of different categories of the uses (flexibility), the originality score of the uses given by several judges (originality), and the uncommonness level of the uses based on the occurrence of the uses in a dataset (frequency). The choice of measures differs across the experimenters. Some experimenters employ one measure while others measure several. The higher scores one receives for the measure(s), the better the divergent thinking (Mark A. Runco & Selcuk Acar, 2019) (see Figure 2a & 2b).

Figure 2a

Origin, task input, task output, and measures of divergent thinking tasks (continue).

Task	Origin	Input	Output	Measures
Unusual Uses / Alternative Uses	Wilson Test	A common object and its common use.	List six other uses for which the objects could serve.	Frequency
	Getzel-Jackson Test Battery	A common object that has a simple stereotyped function	List as many uses as possible	Fluency and originality
	Test of Creativity in Engineering	A visualised object	List as many uses as possible	Flexibility and originality
	Wallach-Kogan Creativity Tests	A common object	List as many uses as possible	Fluency and frequency
	Torrance Test of Creative Thinking	An object	List as many uses as possible and ask as many questions as possible	Fluency, flexibility, and originality
Number Associations	Wilson Test	A number	List as many synonyms, uses, and things for which the number can associate	Frequency
Word Associations	Getzel-Jackson Test Battery	A common stimulus-word	List as many definitions as possible to the stimulus-word	Frequency and flexibility
Visual object association / Pattern meaning / line meaning	Test of Creativity in Engineering	A visualised object	List as many "what is this" as possible	Fluency and originality
	Wallach-Kogan Creativity Tests	A visualised pattern	List as many things the drawing could be as possible	Fluency and frequency
	Wallach-Kogan Creativity Tests	A visualised line	Guess as many meanings of the drawing of line as possible	Fluency and frequency
Instances	Wallach-Kogan Creativity Tests	A classic concept	List as many possible instances as possible	Fluency and frequency
Asking for a drawing	Torrance Test of Creative Thinking	A drawing	Ask as many questions as possible to know what is happening in the drawing	Fluency, flexibility, and originality

Figure 2b

Origin, task input, task output, and measures of divergent thinking tasks.

Task	Origin	Input	Output	Measures
Causes guessing	Torrance Test of Creative Thinking	A drawing	Guess as many causes that result in the action shown in the drawing as possible.	Fluency, flexibility, and originality
	C-SAT	A figure of a food chain and a graph of the changes in this chain	List as many causes as they can of the changes.	Fluency, flexibility, and creativity.
Consequence guessing	Torrance Test of Creative Thinking	A drawing	Guess as many consequences that result from the action shown in the drawing as possible.	Fluency, flexibility, and originality
	Torrance Test of Creative Thinking	An improbable situation	List the things that would happen in the situation	Flexibility and originality
	C-SAT	A figure of experiment	List as many hypotheses as possible that the researchers might come up with.	Fluency, flexibility, and creativity.
Fables endings	Getzel-Jackson Test Battery	A fable in which the last line was missing.	Provide a moralistic ending, a humorous ending, and a sad ending to the fable.	Fluency, and originality
Product Improvement	Torrance Test of Creative Thinking	A product	List the unusual ways of improving a product.	Fluency, flexibility, and originality
Free association	Association Tasks	A concept	List as many associations as possible.	Fluency
Association Chain	Association Tasks	A concept	List chains of associations. Only the first association relate to the concept and the following relate to the respectively last associative response.	Fluency of discriminable concepts
Dissociation task	Association tasks	A concept	List as many concepts as possible that the concepts are unrelated to each other.	Fluency of unrelated concepts
Extracurricular activity ideation	Extracurricular activity ideation	A group of people	List as many creative ideas as possible for a real-life situation relevant to these people.	Subjective creativity score

Convergent thinking.

Convergent thinking is another crucial component of creative thinking. Unlike divergent thinking, which has a widely accepted definition, convergent thinking has been defined in at least three different ways. For example, some researchers defined *convergent thinking* as producing a new entity via the reinterpretation and reorganisation of one's experience or knowledge. This definition is derived from the *transformation* ability proposed by Guilford (Joy P Guilford, 1967) and the associative basis proposed by Mednick (Mednick, 1962). Guilford also argued that the readiness to be flexible might determine convergent thinking ability. Another kind of convergent thinking refers to a process of finding a single correct answer or single best solution for a well-developed problem. Specifically, it requires people to be logical and employ facts, principles, relations, rules, laws, and formulas to solve the problem where correct answers exist and leaves no room for ambiguity. Therefore, this definition of convergent thinking is consistent mainly with one's ability to solve a problem in that the correct answer exists. In addition, some research conceptualised convergent thinking as evaluative thinking, in which people use numerous criteria to select the best solution to existing options (Cropley, 2006). In this project, we employ the second definition of convergent thinking because it fits the thinking styles of the classic convergent thinking psychometric tools – RAT.

Psychometric tools.

Many convergent thinking tasks are in close-ended format. Some have one correct answer and ask participants to come up with this one piece of information. Let us take the most widely employed convergent thinking task – RAT, as an example. In RAT, participants encounter several word puzzles. In each puzzle, they are provided with three words. Participants should consider the fourth word that can serve a specific associative link between three stimulus words. For instance, the correct answer to the word puzzle “rat, clue, cottage” is “cheese”. The more word puzzles one can solve, the better the convergent thinking one exhibits (Mednick, 1962). The following studies produced adjusted versions of RAT by altering word associative logics (CRAT) (Bowden & Jung-Beeman, 2003) and languages (e.g., Chinese Compound Remote Associate Test, CCRAT) (Wu & Chen, 2017). A recent study proposed a Visual Remote Associates Test (VRAT) in which people are asked to think of a fourth object that visually co-occurs with three provided objects. Given the pictures of “glove”, “handle”, and “pen”, people must come up with “hand” (Olteţeanu & Zunjani, 2020).

In addition to RAT and CRAT, the Association Task in Wilson Test Battery is another commonly used convergent thinking task (Robert C Wilson et al., 1953; R. C. Wilson et al., 1953). The task has several correct answers and asks participants to come up with one. In detail, participants are asked to think of a word that connects a pair of words where the connection between them is not apparent. For example, “penny”, “copper”, and “wampum” are correct answers for a pair of words “Indian” and “Money”. An adjusted version of the Association Task has one correct answer and asks people to find it from several options. In detail, participants are asked to choose one letter from five letters where the selected letter is the first letter of the word that connects a pair of words. For example, given “Tree” and “Dog” and five letters “b, g, t, w, z”, the correct answer is “b” for “Bark”. Also, a Quick Responses Task provides people with 50 words at the rate of one every five seconds and asks them to report the first word that comes to mind.

Convergent thinking tasks mainly measure the appropriateness dimension of ideas. They measure this dimension via the number of correct responses of participants. Let us take the RAT as an example. When we provide participants with “rat, blue, cottage”, “cheese” is the correct response. Experimenters usually provide several RAT questions and measure how many questions they can correctly solve. One’s creativity level is positively related to the number of correct responses (see Figure 3).

Please allow us to remind readers that we link different thinking (divergent and convergent thinking) with different measuring dimensions (novelty and appropriateness) for the convenience of communication. We are not proposing a causal relationship between the information processing and the measuring dimension because we can still measure originality in convergent processing tasks. For instance, the convergent processing task - Quick Responses Task, measured the originality of people’s responses (Robert C Wilson et al., 1953; R. C. Wilson et al., 1953). Here, the link aims to bring readers’ attention to the fact that existing tasks may be able to grasp only one dimension of creative products. We illustrate this argument in detail in the following sections of this chapter.

Figure 3

Origin, task input, task output, and measures of convergent thinking tasks.

Task	Origin	Input	Output	Measures
Quick Responses	Wilson Test	50 words at the rate of one every five seconds	Tell the first word that came to mind.	Frequency
Association Task 1	Wilson Test	A pair of words of which the connection between them is not apparent	Come up with the third word that connect two words.	Accuracy
Association Task 2	Wilson Test	A pair of words of which the connection between them is not apparent and five letters.	Choose one of the five letters in which one of them is the first letter of the correct connection.	Accuracy
Remote Associate Test/Compound remote associate test	Remote Associate Test	Three stimulus words draw from mutually remote associative cluster.	Come up with the fourth word that serve as a specific kind of associative connective link between the words.	Accuracy
	Compound Remote Associate Test	Three stimulus words draw from mutually remote associative cluster.	Come up with the fourth word that form compound word with each of the stimulus words.	Accuracy
	Visual Remote Associates Test	Three stimulus objects	Come up with the fourth object that generally co-occurs visually with the other shown three objects.	Accuracy

Importance of existing psychometric tools.

Divergent thinking. There are at least two reasons that divergent thinking has become the dominant concept of creative thinking. First, divergent thinking requires people to explore many mental categories, which often leads to novelty, which is the core property of creativity (Mark A. Runco & Selcuk Acar, 2019). A typical divergent thinking task provides people with one piece of information and asks them to come up with as many pieces of information as possible. Several aspects of the ideas (e.g., fluency, flexibility) indicate divergent thinking. Therefore, there is no correct answer in divergent thinking tasks, which constructs creative opportunities for participants (Finke et al., 1992).

Second, the divergent thinking tasks exhibit high reliability and discriminate validity. For example, the inter-rater reliability of divergent thinking tasks reached .90 (Meeker et al., 1985; Urban, 2005; Wallach & Kogan, 1965a). Also, existing research showed la ow correlation between creativity and intelligence scores (between 10% - 25%) (Andrews, 1930; Chassell, 1916; Colvin & Meyer, 1906; Dearborn, 1898; McCloy & Meier, 1939; Roe, 1953; Terman, 1940). More recent research proposed a threshold theory of intelligence (IQ) and creativity, saying that low levels of IQ correspond with low divergent thinking (Torrance, 1962). The others showed that, for high levels of IQ, divergent thinking is greater for individuals with an IQ below 120 than those above 120 (Guilford, 1981). These results imply that divergent thinking is a unique mind construct which deserves effortful investigation (Crockenberg, 1972; Kim, 2006; MacKinnon, 1965).

Convergent thinking. Creativity researchers are in praise of convergent thinking for at least two reasons. Initially, the typical process of convergent thinking requires and produces knowledge, while knowledge plays a crucial role in creative achievements (Cropley, 2006). For instance, convergent thinking tasks require individuals to employ existing knowledge to find a correct combination of several elements (prepare knowledge), develop a higher level of combination skills and achieve correct answers quicker afterwards (produce knowledge). Also, most of the descriptive creative process models start from “preparation”, which indicates that a creative process requires the accumulation of relevant knowledge (we discuss the argument in the Creative Process section of this chapter). Additionally, previous research made it clear that individual needs ten years of accumulation of knowledge and skills in a domain to generate creative outputs (Ericsson, 2006). Thus, convergent thinking is prominent in the creative process, supported by its close link with knowledge.

Second, the convergent thinking tasks belong to the traditional cognitive research approach that measures accuracy and speed (Anderson, 1991; Shepard & Cooper, 1986). They avoid the possible contaminations (e.g., inter-rater disagreement) of the subjective scoring method in divergent thinking tasks.

Creative cognition and neuroscience.

Existing psychometric tools support the development of creative cognition and creative neuroscience. In the early 1990s, cognitive psychologists officially developed a creative cognition approach that aims to explore the basic cognitive processes and structures of creative ideas production (Finke et al., 1992).

Cognition process framework. One creative cognition stream explores persons' cognitive process in AUT and (C)RAT. They found that creative thought replies to normal cognition, including attention, memory, and cognitive control and the creative outcomes come from the function of these basic cognitive processes (Abraham, 2018; Benedek & Fink, 2019; Sternberg & Lubart, 1996; Weisberg, 1986).

For example, researchers examined how different attentional states affect creative thoughts and achievements. The attentional states include focused attention, defocused attention (Benedek et al., 2014; Martindale, 1999), broad attention (Rowe et al., 2007), flexible attention (Vartanian, 2009), and leaky attention (Zabelina et al., 2015). They found that people who can focus and flexibly switch attention perform well in creative thought tasks, while people with broad and leaky attention report a higher level of creative achievement (Zabelina, 2018a; Zabelina & Ganis, 2018).

Existing research found that memory plays a role in altering the creative level of thoughts. For instance, existing research looked at memory organisation (Benedek et al., 2017; Kenett et al., 2014; Kenett et al., 2018), memory retrieval (Benedek & Neubauer, 2013; Merten & Fischer, 1999), memory construction process (directed vs. undirected) (Abraham & Windmann, 2007; Campbell, 1960; Gabora, 2018; Simonton, 2011; Ward et al., 1997), memory content (semantic memory vs. episodic memory) (Gilhooly et al., 2007; Leon et al., 2014; Nusbaum et al., 2017). The theories for memory organisation argue that associative hierarchies alter creative performance. However, the theories for memory retrieval suggest that AUT performance is altered by associative fluency (Benedek & Neubauer, 2013).

Additionally, empirical evidence showed that cognitive control and creative cognition are firmly related, but cognitive control is a double-edged sword. For example, the research examined how updating, inhibition, and shifting relates to divergent thinking. In detail, updating refers to monitoring incoming information and revising working memory by replacing absolute information with new and relevant information. Inhibition indicates the suppression of dominant but irrelevant responses tendencies and inhibition rather than shifting. Shifting refers to switching between tasks and mental states and disengaging irrelevant mental sets and tasks irrelevant to divergent thinking tasks (Friedman et al., 2006; Jonides & Smith, 1997; Miyake et al., 2000). They found that updating and inhibition rather than shifting are relevant to divergent thinking (Benedek et al., 2014). On the other hand, exposure to high inhibition demands leads to impaired inhibition ability and enhances fluency and originality in divergent thinking, but no such effect on RAT (Radel et al., 2015).

Some researchers reviewed the neuroscience evidence. The review suggests that cognitive processes such as goal-direct memory retrieval, domain response inhibition, and internally focused attention are related to creative performance. Among them, goal-direct memory retrieval indicates the ability to search episodic and semantic memory for task-relevant information. Domain response inhibition is the ability to suppress interference from dominant or salient response tendencies during divergent thinking. Finally, internally focused attention refers to self-generated thought processes and shielding internal processes from external interference (Beaty et al., 2019).

Cognitive neuroscience. The neuroscience research of divergent thinking is also fruitful (Roger E. Beaty et al., 2014; Beaty et al., 2019; Roger E Beaty, Paul J Silvia, et al., 2014; Benedek & Fink, 2019; Dietrich, 2004; Fink et al., 2009; Gabora, 2018). For example, researchers found that divergent thinking is associated with active alpha activity in parietal areas (Benedek et al., 2011), prefrontal cortex, and the right hemisphere (Fink et al., 2011; Jauk et al., 2012). In addition, existing research has been converging: divergent thinking requires the activation of both hemispheres, and striatum (Cools & D'Esposito, 2011; Mayseless et al., 2013; Zabelina et al., 2016) and prefrontal cortex, and with functional connectivity in the frontal, temporal, and parietal brain regions for semantic processing and combination of related information (Benedek et al., 2014; Bilder & Knudsen, 2014; De Dreu et al., 2012; Dietrich & Kanso, 2010).

The brain regions and networks in the prefrontal cortex and striatum interest genetic experts because both areas are closely linked to the neurotransmitter dopamine (DA) (Boot et al., 2017; Nijstad et al., 2010; Zabelina et al., 2016). Existing research found a U-shape relationship between DA[2] and divergent thinking. In detail, compared to low and high levels of DA, a medium level of DA is associated with a higher level of flexibility (Chermahini & Hommel, 2010), fluency (Akbari Chermahini & Hommel, 2012; Ueda et al., 2016), and originality in AUT (Agnoli et al., 2022).

Low predictive power of creative thinking tasks.

Accordingly, the reviewed evidence, creative cognition and neuroscience studies are overwhelmed by AUT and (C)RAT. However, existing creative thinking tasks disconnect with real-life creativity. For instance, divergent thinking performance may exhibit low predictive power toward real-life creativity (Hocevar, 1981; Zeng et al., 2011).

Divergent thinking tasks. For example, existing research found that self-reported creative activities show higher predictive power to real-life problem findings tasks than the divergent thinking task AUT (Okuda et al., 1991). The low predictive power is especially explicit for professional creativity. In detail, existing research found a low to medium correlation between divergent thinking and students' creativity (Marx = .48, Minor = .20, Mr = .30) (Drevdahl, 1956; Lowenfeld & Beittel, 1959; Merrifield et al., 1964) and occupational people's creativity (Max_r = .55, Min_r = .30, $Mean_r$ = .38) (Barron, 1955; Barron & Harrington, 1981). Also, individual divergent thinking could not predict the creative performance of architects (MacKinnon, 1965) and scientists (Gough, 1961) ($-1 < r < 0$) (for the review, see (Batey & Furnham, 2006)).

Creative personality research also indicates the disconnection between divergent thinking and creative achievements. For example, researchers linked divergent thinking with the big five personality traits: openness to experience, extraversion, agreeableness, neuroticism, and conscientiousness. The results showed that fluency and originality are positively related to social traits such as openness to experience (Walker & Jackson, 2014), extraversion (Batey et al., 2009),

---

[2] The DA was indicated by spontaneously Eye Blink Rate (sEBR) in the experiments, considering that sEBR is a reliable predictor of DA when it comes to creative behaviour (Boot et al., 2017)

or both (Furnham et al., 2009). However, a meta-analysis of creative genius personality showed that the people who obtain creative achievements have an anti-social tendency (Feist, 2010).

Convergent thinking tasks. Four studies conducted during 1963-1983 failed to find a significant relationship between a RAT and creative accomplishment (Andrews, 1967; Davis & Belcher, 1971; Worthen & Clark, 1971). Also, the accuracy in convergent thinking insight problems[3](DeYoung, 2020) shows no relation to self-report everyday creativity and creative achievement (Roger E Beaty, Emily C Nusbaum, et al., 2014). In addition, mathematical insight problems may be more predictive of real-life design problems compared to TTCT and RAT (Brougher & Rantanen, 2009).

Disconnections between creative thinking tasks and creative practice.

We propose that the low predictive power of creative thinking tasks may result from the disconnect between the creative thinking tasks and the creative practice. Creative practice refers to creative performance in a real-life context. There are two differences between creative thinking and creative practice. First, in terms of the outcome, creative thinking drives initial thoughts, while creative practice requires creative final products or achievements in a specific domain. Regarding the process that produces the outcome, creative thinking and creative practice refer to mental and descriptive behavioural processes. Here, we review the concepts for creative products (behaviour outcome) and creative process (behavioural process) and discuss their linkage with creative thinking tasks.

Creative products. The creativity definitions mainly focus on addressing the properties of creative ideas[4] and products. Back in the 1950s, the field forerunners proposed a two-criteria argument that creative ideas and products should be novel[5] and effective[6] (Barron, 1955; Guilford, 1950; Stein, 1953). In detail, an idea or product should be distinguishable from the established ones to be novel. Beyond

---

[3] An example of an insight problem is that: An American football team won 79-32, yet not one man scored as much as a single point. How is that possible? The correct answer is: It was a women's or coed football team.

[4] Both creative ideas and products in this thesis indicate behavioural outputs.

[5] Novel is sometimes labelled as original, unique, different, new, unusual, and uncommon.

[6] Effective is sometimes labelled as useful, appropriate, valuable, and logical.

that, a novel idea or product should exhibit utility to be creative (Runco & Jaeger, 2012; Simonton, 2011). For example, a creative product can be an original technology that solves a challenging problem, a scientific discovery that forwards our understanding of the universe or human beings, or an artistic painting that brings aesthetic pleasure with a unique expression. Since the 1960s, the two-criteria argument has become the standard definition of creativity (Bruner, 1962; Cattell & Butcher, 1968; Jackson & Messick, 1965; Kneller, 1965; Newell et al., 1962).

Subsequent systematic reviews also addressed the importance of novelty and effectiveness. For example, Besemer and her colleagues analysed 125 criteria for creativity in 90 pieces of literature. They found that creative products should be original and provide implications for future creative products (novelty). Also, creative products should fit the needs of a challenging situation (effectiveness) and combine seemly unrelated elements into a new entity (elaboration and synthesis) (Besemer & Treffinger, 1981). Recently, Henriksen reviewed 220 creativity assessments and found that creative ideas should have the quality of uniqueness to be attractive to the audience (novelty). In addition, they should bring (potential) value to the audience (effectiveness) and have a certain integrated aesthetic quality within the relevant task, domain, and context (wholeness) (Henriksen et al., 2015). Therefore, the essential role of novelty and effectiveness are full-bodied in creativity research.

Beyond academia, the industrial practice also admits the essential role of novelty and effectiveness in creative products. For example, U.S. Patent and Trademark Office has declared that an entity that "*invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent*". Specifically, the *new* (novelty) requires a patent to have no apparent connection with previous inventions and discoveries. Also, the *useful* (effectiveness) requires a patent to fulfil the intentioned purpose (Alexandria).

Instead, some researchers held up an implicit theory of creativity that disagrees with characterising creativity in specific dimensions. The implicit theory prefers evaluating ideas or products by a group of experts working in the corresponding domain, and the creativity of ideas or products is determined by the consensus among the experts (Amabile, 1982b; Hennessey, 1994; Hickey, 2001). Evaluating creative products via consensual assessment of experts is reasonable because domain-specific knowledge plays a crucial role in creative production (Baer et al., 2004; Kaufman et al., 2009). For example, many of the leading theories, such as the componential theory of creativity (Amabile, 1983), the evolving-system model of creativity (Gruber, 1988), the system model of creativity (Csikszentmihalyi, 2015), and the investment theory of creativity (Sternberg & Lubart, 1991), have addressed the importance of domain-specific knowledge in producing creative products.

However, the implicit theory does not undermine the essential role of novelty and effectiveness in creative ideas and products in everyday and domain-specific contexts. For everyday creativity (Kaufman & Beghetto, 2009; Richards et al., 1988), lay people associate it with unconventional thoughts (relevant to the novelty aspect of creativity) (Sternberg, 1985) or goodness and contributions to the society (relevant to the effectiveness aspect of creativity) (Niu & Sternberg, 2002; Rudowicz & Yue, 2000). For domain-specific creativity, the experts in engineering creativity (Cropley & Cropley, 2005; Shah et al., 2003; Thompson & Lordan, 1999), advertising creativity (Ang & Low, 2000; Haberland & Dacin, 1992; Kover et al., 1995; Rosengren et al., 2020), scientific creativity, and artistic creativity agree on the essential roles of novelty and effectiveness. Therefore, the standard definition of *creativity* such that the creative ideas and products require novelty and effectiveness.

Creative process. Creative process refers to the descriptive stages responsible for the generation of creative ideas and creative products (Lubart, 2001). It is crucial to differentiate between creative process, creative thinking, and creative cognition. In this thesis, the creative process refers to descriptive models that describe the behavioural actions that lead to creative ideas or products (Amabile, 1983; Campbell, 1960; Wallas, 1926). Creative thinking explores the cognitive structures or styles that enhance individuals' ability to produce creative ideas (Koestler, 1964; Mednick, 1962; R. C. Wilson et al., 1953) and construct corresponding psychometric tools. Creative cognition identifies how basic cognitive processes such as perception, attention, memory, and cognitive control contribute to the production of creative thought (Finke et al., 1992). We discussed creative thinking and creative cognition in previous sections. Here, we focus on the creative process.

Wallas constructed the earliest creative process model, including the stages of preparation, incubation, illumination, and verification (Wallas, 1926). The preparation stage has three tasks: identifying a problem or a challenge, gathering information, and putting effort into tackling the problem. This stage requires conscious work on the issue with problem-relevant knowledge and analytical skills. The incubation stage requires one to not think about the problem and work on other topics (or take a break from any thoughts and walk outside). The assumption is that, during the break, the person's mind would keep working on the problem and unconsciously associate (e.g., combine) information. This unconscious stage has the potential to bring us inspiration and "happy ideas". The illumination stage is when these "happy ideas" suddenly and immediately come to us in a clear image (known as the "aha moment"). In addition, the verification stage involves evaluating or further developing the "happy ideas". Some following descriptive models applied Wallas's model with minor adjustments. For example, McNally inserted a stage of immersion between the stages of preparation and incubation and replaced the stage of verification with the stages of explication and creative synthesis (McNally, 1982). Based on the models of McNally, Shaw specified the importance of validation with feedback loops at each stage (Shaw, 1989). Also, Carson applied the model to a family therapy domain (Carson, 1999).

Consistent with Wallas' model, Campell's two-stage model of blind-variation-and-selective-retention (BVSR) also supports the occurrence of the "aha moment", but with different interpretations or emphasis on the creative process (Campbell, 1960). In detail, Wallas's model assumes a unique construction of the creative process. In contrast, the BVSR model views knowledge expansion via creativity, learning, perception, and biological evolution as sharing the mechanism – BVSR. According to BVSR, the internal emission of ideational variations is blind. They are independent of the external context they occur, and the final solution and all potential ideas, no matter whether correct or incorrect, have the same probability of occurring. Once the internal emission blindly encounters a chance that one idea matches the selection criterion, people experience the "aha moment", and the creative process terminates. In other words, BVSR argues that a creative solution is always unexpected and surprising (Simonton, 2010). Subsequent studies use the model to describe the process of novel scientific discovery (Simonton, 2008, 2011).

Both Wallas's and BVSR models suppose that the creative process (e.g., knowledge expansion) differs from the non-creative process. However, a componential theory of creativity suggests no difference between creative and non-creative processes during problem-solving (Amabile, 1983). In detail, the theory suggests that all problem-solving starts with an assigned task or a problem (task presentation) and knowledge preparation from either memory retrieval or new knowledge learning (preparation). Then, one is likely to generate several possible outcomes with different information processing pathways (response generation) and test the appropriateness or usefulness of the possible outcomes (response validation). After the response validation, people can terminate the process or return to the task presentation depending on the progress toward the goal (outcome). Producing an outcome may require a long series of loops through the problem-solving process. The componential model of creativity proposes that the occurrence of a creative outcome is determined by task motivation, domain-specific skills, and creative-relevant skills. In detail, task motivation is high when the task matches one's existing preference and interests and is low when the task appears with external social and environmental constraints. Domain-relevant skills refer to the familiarity with a domain, the factual knowledge of the domain in question, and the technical skills required in the domain. Creative-relevant skills include creative cognitive styles, heuristics, creative working styles, personalities, and personal strategies for creative thinking.

The following creativity studies revised Amabile's model and applied it to domain-specific settings such as education (Treffinger, 1995), artistic production (Botella et al., 2013; Mace & Ward, 2002), organisational innovation (Cropley & Cropley, 2012), and engineering product design (Howard et al., 2007).

Consistent with descriptive models, the earliest creative cognition model – the Geneplore model - also proposes a two-stage creative process, including the generation and exploration phases (Finke et al., 1992). In the generation phase, people construct mental representations with various properties such as novelty, ambiguity, meaningfulness, emergence, incongruity, and divergence. These mental representations are called preventive structures, for which the examples are visual patterns, category exemplars, mental blends, and verbal combinations. After the generation of pre-inventive structures, people interpret them in meaningful ways via regeneration and modification in the exploration phase. If the initial preventive structure solves the problem within pre-determined constraints (e.g., specified product type, category, features, functions, components, or resource), it becomes a creative product. Otherwise, people return to the generation phase and generate another pre-inventive structure which might be more assuring. The likely cognitive process underlying the generation phase includes information retrieval, association, synthesis, transformation, analogical transfer, and categorical reduction. The likely cognitive process underlying the exploration phase includes attribute finding, conceptual interpretation, functional inference, contextual shifting, hypothesis testing, and searching for limitations. However, no one process is uniquely associated with one phase or another. Beyond that, the Geneplore model of creativity proposes an ordered, organised, and ordinary approach to studying creative cognition. For instance, it assumes that having the generation phase occur before the exploration phase may optimise the utility of the pre-inventive structure (ordered). Second, it assumes that the creative cognition process is constructed in an organised manner rather than random combinations (organised). Third, it assumes that everyone can think more creatively (ordinary).

We learn at least two things according to existing descriptive models of the creative process. First, all the models agree on a two-stage creative process, including generating and evaluating creative ideas and products. Second, researchers have different opinions regarding the relationship between creative and non-creative processes. For instance, Wallas assumed that the creative process is unique from the other knowledge processes. Campbell assumed that the creative process shares the BVSR mechanism with the knowledge expansion in perception and learning. On the other hand, Amabile and the Geneplore model assumed no difference between creative and non-creative processes and skills, while knowledge determines the creativity level of process outcome.

We now have grasped an overview of creative thinking and practice. Next, we illustrate the disconnections between the two and how the disconnections may result in the low predictive power of creative thinking tasks.

Disconnections between creative thinking and creative practice. There are two explicit disconnections between creative thinking and creative practice. First, the measures of previous creative thinking tasks could not grasp two dimensions of creative products. As we have introduced above, the standard definition of creativity explicitly addresses that creative ideas and products must be novel and effective. However, scoring systems of most creative thinking tasks grasp only one dimension of creative products. For instance, most divergent thinking tasks measured fluency, flexibility, originality, and frequency, which implies one's potential to produce novel ideas (see Figure 2) (Runco 2008). Additionally, most convergent thinking tasks grasp accuracy, which was associated with one's ability to produce effective and appropriate ideas (see Figure 3). None of them properly grasped both dimensions of creative products.

Second, divergent and convergent thinking tasks may not be able to simulate the information processing of creative processes. For example, all descriptive theories of the real-life creative process imply that the creative process starts from several pieces of information derived from prepared information. However, divergent thinking tasks always ask participants to give answers from one piece of information (see task input in Figure 2).

Also, all descriptive theories of the creative process indicated an open-ended format, such that the creative process is not to produce the right or wrong answer but to produce several outcomes (and offer a chance to select the best in a specific context). However, convergent thinking tasks were always in a close-ended format, which does not construct creative opportunities or opportunities to grasp creativity (see task output in Figure 3).

It is important to note that close-ended convergent thinking tasks themselves are not problematic because solving close-ended questions may require creative thinking. For instance, some researchers proposed that divergent thinking is the foundation of solving close-ended questions (Brophy, 1998). Also, the open-ended and close-ended questions may play distinctive roles in generating creative thoughts. For instance, a two-stage theory of creativity argues that the first stage of creative ideation is using divergent thinking to generate several novel ideas (open-ended stage). After that, people come to the second stage, where finding a correct and appropriate answer from the generated ideas is critical (close-ended stage) (Campbell, 1960; Cropley, 2016; Simonton, 1999). Some researchers, on the other side, proposed that the open-ended and close-ended stages emerge simultaneously (Perkins, 1998; Sternberg & Davidson, 2005) In this situation, the open-ended aspect involves an associative and effortless process, and the close-ended aspect involves a logical and rule-based process (Lin & Lien, 2013).

Although solving close-ended questions and creative thinking are mandatory for each other, the close-ended questions which measure one's creative thinking with one piece of information cannot grasp creative performance. There are three reasons. As we said before, real-life creative problem-solving in many domains involves open-ended problems without a single correct answer. For example, engineering creative problem solving has been defined as an open-ended process that allows multiple answers (Belski, 2017). Also, using a close-ended question with a single correct answer may stop the participants from coming up with original thoughts when originality is the key element that represents one's creative thinking (Finke et al., 1992). Consistently, existing research found that open-ended questions rather than close-ended questions benefit mathematical creativity (Livne et al., 2008; Wijaya, 2018) and extend the effort and time spent on the divergent investigation (Weinberger et al., 2016).

Admittedly, a close-ended question with a single correct answer can be open-ended when the answer can be obtained in multiple ways (Epstein et al., 1997). For example, some people may use two ways and obtain a single solution, while others may use ten ways and obtain a single solution. However, looking at the solution cannot detect how many ways they have tried. In other words, a creative thinking task should allow multiple observable solutions to make creative thinking measurable to grasp creative thinking from a behavioural level.

Integrative thinking.

When divergent and convergent thinking disconnect from creative practice, some researchers suggest further complementary processes that may help construct the connection (e.g., formulating problems, asking the right questions, extracting value from original ideas, and bringing ideas to reality) (Runco, 2008). Meanwhile, some researchers proposed that integrating divergent and convergent thinking may produce qualitatively new elements or entities and link to real-life creativity (Cropley, 2006; Tan, 2015; Zittoun et al., 2007). The latter argument is consistent with a widely mentioned but rarely validated thinking style – integrative thinking.

Integrative and convergent thinking look similar since both require integrating several pieces of information. The main difference is that integrative thinking is open-ended, which allows multiple answers (Belski, 2017) and constructs creative opportunities (Finke et al., 1992), while convergent thinking is close-ended that contains a single correct answer (Bowden & Jung-Beeman, 2003).

Importance.

If researchers pay extra attention during the literature review, they may find that creativity researchers have explicitly or implicitly emphasised the role of integrative thinking in the creative process. For example, some descriptive creative process models put synthesis as a mandatory stage in production (McNally, 1982).

Likewise, the definition of creativity adds synthesis as a mandatory dimension of creative ideas and products (Henriksen et al., 2015).

Here, we offer some relevant arguments that the leading scholars proposed. At an early stage of creativity research, some leading scholars started addressing the importance of integrative thinking. For example, Guilford regarded the ability to synthesise and reorganise information as an essential element of the ability to create (Guilford, 1950). In 1960s, Rhodes's description made this argument clearer. He proposed that, original ideas, as an output of creativity, are the by-products of a human mind thinking about the elements of a subject and their relationships and embody or articulate the elements into a new entity (Rhodes, 1961). At the same time, Mednick proposed an associative basis of creativity such that creative ideas result from combination of remotely associate and seemly unrelated elements (Mednick, 1962). For example, he cited the self-reflection of several creative genius (e.g., Albert Einstein) who shared a similar thought that a combinatory play of seemly unrelated elements is essential in their productive thought. Although Mednick's work focused on the distance of elements (remote associations) rather than the combination of elements (integration), his work is an important attempt in addressing the important role of integrative thinking in creativity.

Two years after the publication of Mednick's work, Koestler illustrated a possible mechanism of elements combined in human minds – bisociation (Dubitzky et al., 2012; Koestler, 1964). Koestler assumed that the human mind has patterns and fixed codes. The patterns are skills, habits, and abilities that shape our thoughts and behaviours, and we may have different patterns in different domains (e.g., philosophy and chess games). Fixed codes are rules we acquired from the external environments (e.g., moral arguments and chess game rules). As humans grow up, the patterns are more rigid, and the fixed codes are increasingly automated. When fixed codes in two different patterns are connected, people can see the associations between two unrelated and independent patterns. As a result, they may engage in further intellectual synthesis, which could result in creative ideas. Koestler claimed that bisociation might be the general mechanism of original ideas. It requires one to tolerate the chaos of unrelated or conflicting information and find new ways to mix them.

Consistent with the above theories, Rothenberg produced a series of case study analyses and found that creative genius or their works had shown the tendency to integrate contradicted ideas into one piece of novel work (Rothenberg, 1976). For example, Rothenberg introduced the story of Eugene O'Neill, a playwriter and Novel laureate in literature. In detail, O'Neill was inspired by the spontaneously opposite thoughts of his roommates and incorporated at least four logically opposite ideas in the central symbol of the play *The Iceman Cometh*. The symbol exhibit at least three different connotations. Rothenberg also talked about Frank Lloyds Wright - a well-known creative architect. Frank referred to his creative idea of constructing Organic Architecture as a simultaneous affirmative negation of three dimensions (i.e., building, furnishing, and surrounding) to achieve as many aspects of harmony between humans and nature as possible. More than that, he offered the creative stories of influential people in mathematics, science (Pinocare and Waston), music (Arnold Schoenberg), and poetry (Marianne Moore). He showed that these creative geniuses had experienced the spontaneously processing of opposite thoughts (i.e., Janusian thinking). Moreover, Rothenberg also introduced the concept of Homospatial thinking, which indicates that creative ideas are articulated by integrating two or more unrelated entities in the same space (i.e., Homospatial thinking). Overall, Rothenberg's work strengthens the theory of integrative thinking, especially for creative genius.

Another expert for creative genius research, Simonton, also tapped the importance and the potential mechanism of integrative thinking in the creative process (Simonton, 2011). In the extended model of BVSR, Simonton proposed that the creative process for complex and challenging problems requires more than a single process, but an integrative process based on *association richness*. In detail, when people create, they do not rely alone on the information associations in a divergent or convergent thinking manner. Instead, they may spontaneously engage in unconstrained associations via various thinking styles. The associative richness may result in several intersections in a semantic network. A successful solution to a problem may emerge at one of the intersections. Although the integration mechanism proposed by Simonton differs from Koestler, both experts agreed on the significant role of integration in the creative process.

Moreover, Carl Jung proposed that the opposition is essential in all works. Jung found validation for the "unitary world" in a symbol in every culture throughout history: the mandala or "magic circle," signifying both undifferentiated unity and integrated wholeness. In Jung's (CW volume 14) unus Mundus, in the "*potential* *world outside of time*", everything is interconnected, and there is no difference between psychological and physical facts and between past, present, or future. This borderline state where time, space, and eternity are "held together" by the magic circle of the mandala forms the backdrop for Jung's most basic formulation about the structure and dynamics of the psyche (Boeree, 2006; Jung, 2014). Here, the interconnected and unitary world also comes from creating an integration of oppositions.

Margaret Boden, the expert whose working stream is Creativity and Artificial Intelligence, also addressed the role of integrative thinking in her three-roads model of creativity. In detail, Boden proposes that creative ideas and products may be reached by combination, exploration, and transformation of mental elements in our conceptual spaces. The combination involves integrating familiar elements into unfamiliar entities without previously existing structures (this is about finding intersections). Exploration involves using previously existing rules (e.g., Chess game) to generate novel structures that may or may not have been realised before the exploration took place (this is about finding an alternative way of thinking). In addition, transformational creativity involves generating structures by altering previously existing conceptual spaces (this is about altering assumptions). Among the three, combination and exploration are the most common mental manipulation within conceptual spaces (Boden, 2009, 2010).

The neural complexity of a highly creative brain also implies the importance of integrative thinking: this is because the neural circuits of the highly creative brain are more highly interconnected and complex than the less creative one (Andreasen, 2005; Heilman et al., 2003). In other words, a highly creative brain not simply has strong activation in a particular area (e.g., right frontal or temporal area). It shows more robust connectivity between major associative regions, consistent with the idea that creative people can link remote associate elements. In detail, the more remote associates they generate, the more likely they find common ground for these remote associates (i.e., association integration). In other words, when one stimulus activates separate brain areas spontaneously, it activates the processing of more information which is more likely to result in information integration (Feist, 2010; Ramachandran & Hubbard, 2003).

Existing research also found that highly creative people can mentally synthesise visual parts. For example, Nicola Tesla, who contributed to the modern alternating current electricity supply system design, can mentally visualise and design every detail of new thought and make it a product (Pickover, 2015). Consistently, the cognitive psychologist Finke conducted a creativity task that measures one's mental ability to synthesise visual patterns. In the task, participants were asked to mentally assemble three parts to make a recognisable figure or easily named in 2 minutes. Then, they wrote down the figure's name and drew it. After that, judges rated the correspondence of the names and the drawing on a 5-point scale. The creative, highly correspondent output would be marked. It would be highly creative. Otherwise, it would be non-creative patterns, poor correspondence, wrong parts, or no pattern (Finke & Slayton, 1988). This task taps the integrative processing in the creative process. However, this task still constructs limited creative opportunities for participants, allowing people to create only one mental image. In addition, this task cannot simulate the real-life situation, such that the combination of formats does not construct the real-life products. In real-life, problem-solving is critical for products, and the function is the key.

Psychometric tools.

Though none of the existing psychometric tools claimed to be integrative thinking tasks, some of them start with several pieces of information and end with several pieces of information that may require the integration of information. Here, we name them "several-several tasks".

For example, a Figure Concept Task in Wilson Test provides people with twenty drawings of objects and individuals (e.g., drawing A: a child wearing a hat, drawing B: a woman wearing a hat, drawing C: young birds in a nest) and ask them to find the common features of two or more drawings (e.g., drawing A & B: wearing a hat, drawing A & C: young, drawing A & B: family). The performance is measured by frequency (Robert C Wilson et al., 1953; R. C. Wilson et al., 1953). A Hidden Shapes Task in GJTB asks people to find as many geometric figures as possible hidden in the given patterns and forms. The performance is measured by fluency, appropriateness, and originality (Getzels & Jackson, 1961). A Similarities Task in Wallach-Kogan Battery provides people with two objects (e.g., cat and mouse) and asks them to come up with many similarities between the two objects. The performance is measured by fluency and frequency (Wallach & Kogan, 1965a). An Adjusted Unusual Uses Task (Engineering) in Test of Creativity in Engineering provides people with two visualised objects and asks them to list many uses when the two objects are used together. The performance is measured by flexibility and originality (Harris, 1960a, 1960b). An Association Combination Task provides people with two seemly unrelated words (e.g., “summer” and “high”) and asks them to think of as many associations as possible that can relate the two words (e.g., “aeroplane”). The performance is measured by fluency (Benedek et al., 2012).

For some several-several tasks, the provided information is incorporated into a situational story. For example, a Just Suppose Task in TTCT provides people with an improbable situation (e.g., clouds had strings attached to them which hang down to earth) and asks them to list the possible outcomes of the situation. The performance is measured by flexibility and originality (Torrance, 1972). A Fable Ending Task in GJTB provides people with a fable in which the ending was missing (e.g., a mischievous dog bites people without warning) and asks them to write a moralistic ending, a humorous ending, and a sad ending. The performance is measured by fluency and originality (Getzels & Jackson, 1961).

Some several-several tasks incorporate domain-specific information and are used in student examinations. An Opportunity Identification Competence Assessment asks participants to generate many ideas that can lead to social, environmental, and economic gains for a new start-up in sustainable development. The performance is measured by fluency, flexibility, and elaboration (Baggen et al., 2018; Corbett, 2007). An Interaction Graph Task in C-SAT provides people with a graph of reversed changes in the amount of two variables and an effect that induces the changes and asks them to come up with many pairs of variables that can fit the graph. A Sugar Experiment Task in C-SAT provides people with a figure of an experiment and a hypothesis. The task asks them to consider many mandatory adjustments to the figure to prove the hypothesis. A String Experiment Task in C-SAT provides people with a Physics experiment and asks them to list many mandatory changes on the figure to achieve a goal (Sak & Ayas, 2013). The three tasks in C-SAT measure fluency, flexibility, and creativity in general (see Figure 4).

Figure 4

Origin, task input, task output, and measures of several-several tasks.

Task	Origin	Input	Output	Measures
Figure Concept	Wilson Test	20 pen and ink drawings of objects and individuals.	Find qualities or features that suggested by two or more drawings.	Frequency
Hidden Shapes	Getzel-Jackson Test Battery	Complex patterns and forms	Find a geometric figure hidden in the given information.	Fluency, appropriateness, and originality
Similarities	Wallach-Kogan Creativity Tests	Two objects	List as many similarities between two objects as you can.	Fluency and frequency
Unusual Uses 2	Test of Creativity in Engineering	Two visualised objects	List as many possible uses as possible when two objects used together.	Flexibility and originality
Association Combination	Association Tasks	A pair of unrelated stimulus-words	List as many associations that relate to both stimulus-words.	Fluency of adequate answers
Opportunity Identification Competence Assessment Test	Opportunity Identification Competence Assessment Test	The need to input for a business idea in the area of sustainable development, which lead to social, environmental or economic gains.	Generate ideas for this new start-up.	Fluency Innovativeness on a 6-point scale Fluency, elaboration, flexibility.
Interaction Graph	C-SAT	A graph of reverse changes in the amount of two variable and an effect that starts these changes	List as many pairs of variables as they can that fit the graph.	Fluency, flexibility, and creativity.
Sugar Experiment		A figure of an experiment and a hypothesis	List as many necessary changes on the figure to prove the hypothesis.	Fluency, flexibility, and creativity.
String Experiment		A figure of an experiment	List as many necessary changes on the figure to achieve a goal.	Fluency, flexibility, and creativity.

Weaknesses in grasping integrative thinking. However, the several-several tasks may not be able to represent integrative thinking. We propose two reasons here.

First, several several-several tasks require domain-specific knowledge, which may contaminate thinking skills assessment. For example, the Opportunity Identification Task requires business entrepreneurship knowledge in technology-based fields; The interaction graph task, sugar experiment task, and string experiment task in C-SAT require interdisciplinary science, chemistry, and physics knowledge, respectively. Therefore, we can attribute the task performance to neither pure domain-specific knowledge nor pure creative thinking.

The domain-specific knowledge is also frequently required in creative problem-solving tasks, which may produce concerns when experimenters consider employing them. For instance, an in-basket exercise asks people to respond to a packet of 22 problems as HR directors. This task requires the knowledge of organisational human resources (Shalley, 1991). A make-up problem in GJTB provides people with four paragraphs containing several numerical statements. Participants are asked to list as many mathematical problems as possible based on the given paragraphs. This task involves mathematics knowledge (Getzels & Jackson, 1961). A complex problem-solving task in a hypothetical World War 1 military context asks participants to generate a plan to handle the situation, which requires knowledge or experience in leadership (An et al., 2016). Besides, a Musical Expression Test requires participants to produce a one-minute improvisation based on step-by-step instruction that requires knowledge of music (Barbot & Lubart, 2012) (see Figure 5).

Figure 5

Origin, task input, task output, and measures of creative problem-solving tasks.

Task	Origin	Input	Output	Measures
In Basket Exercise	In Basket exercise	Brief description of the company and a packet of 22 problems presented to HR directors	Respond to the problems as the HR director	CAT
Make-up problems	Getzel-Jackson Test Battery	Four paragraphs, each containing a number of numerical statements (e.g., costs in building a house)	List as many mathematical problems that solved with the information given.	Fluency, appropriateness, and originality
Complex problem solving	Cued measures	Background information for a novel, ill-defined problem in a hypothetical military context, involving a hypothetical World War.	Problems relating to problem construction, information coding and category search, specification of the best-fitting categories, category combination and reorganization, idea evaluation, solution implementation, monitoring of implementation.	Overall effectiveness using a five-point scale.
	Un-cued measures		Read the problem and generate a plan to handle the situation.	Quality, originality, restriction awareness, units' coordination.
Alternative uses	Musical Expression Test	12 sound-production elements, which are playful and easy to use.	Identify the elements and create as many uses as possible of the identified elements.	Instrumental gesture
Composition			Create a 30s musical piece.	CAT
Improvisation		Self-identified 30s musical piece.	Record a 1m improvisation	CAT
Cause guessing	Real-life problem solving	A real-life problem: a first-year middle school student had enjoyed school life until a month ago and he does not want to go to school recently. Why would she/he act like this?	Guess and generate as many reasons as possible.	Fluency, flexibility, and originality
Creative problem solving			Choose one reason and discuss how to solve the problem by applying as many educational psychology theories as possible.	Appropriateness, originality, quality.

Furthermore, several-several tasks may not require integrative thinking. According to the Cambridge Dictionary, integrative thinking is the synonym of synthesis that refers to the acts or outputs of mixing or combining different ideas, influences, or things to make an entity that is new or different from the items considered separately[7]. However, the several-several tasks mostly ask people to find common qualities in given elements. For example, the Association Combination Task asks for common associations; the Unusual Uses Task 2 asks for common uses; The Figure Concept Task, Hidden Shapes Task, and Similarity Task ask for common shapes. The process of finding common qualities may not construct creative opportunities for creating qualitatively new entities.

Admittedly, there are figural tasks that ask people to come up with new visual elements. For example, there are various drawing tasks in TTCT and TCT-DP that ask people to draw pictures based on provided shapes. Similarly, a Divergent Pareidolias Task asks participants to draw pictures based on a natural landscape photograph. Also, a mental synthesis task asks people to mentally make a recognisable figure based on three selected shapes (see Figure 6). However, existing studies found specialised hubs and interactive systems in our brain for creativity in verbal and figural tasks (Zhu et al., 2017). For example, visual creativity in the figural task in TTCT is negatively related to functional connectivity in the precuneus and medial frontal cortex (MFC) of the posterior default mode network (pDMN). Nonetheless, verbal creativity in verbal tasks in TTCT is negatively related to functional connectivity in the medial prefrontal cortex (mPFC). Although visual and verbal creativity positively correlates with the functional connectivity between the default mode network (DMN) and frontoparietal network (FPN), the distinct aspects still require us to see visual and verbal creativity as different.

[7] https://dictionary.cambridge.org/dictionary/english/synthesis

Figure 6

Origin, task input, task output, and measures of figural creative thinking tasks.

Task	Origin	Input	Output	Measures
Drawing 1	Torrance Test of Creative Thinking	A shape (e.g., teardrop shape)	Draw a picture with the shape as an integral part and give titles for the drawing	Originality and elaboration
	The Test for Creative Thinking – Drawing Production	An incomplete drawing shape	Finish it without knowing what would come out of it in whatever way which.	Continuation, completion, new elements, connections, boundary breaking, perspective, humour and affectivity, unconventionality, Speed
Drawing 2	Torrance Test of Creative Thinking	An incomplete figure	Add lines to the figure and give titles for the drawings	Fluency and flexibility
		A pair of straight and parallel lines	Make pictures from the lines and give titles for the pictures	Fluency and flexibility
Divergent Pareidolias Task	Divergent Pareidolias Task	A photograph of natural landscapes	Let the imagination run free, and draw everything one can see, expect for the objects themselves, and say every time what it is	Fluency, flexibility, and originality
Mental Synthesis	Mental Synthesis Task	15 forms from which the parts would be selected, and experimenters indicated the name that would be used to designate each of the parts	Choose three forms. Close eyes and try to mentally assemble the parts to make a recognizable figure that could be easily named. If you come up with more than one recognizable figure, choose the best one.	Correspondence and creativity (i.e., appropriateness and originality)

Function Synthesis Task (FST).

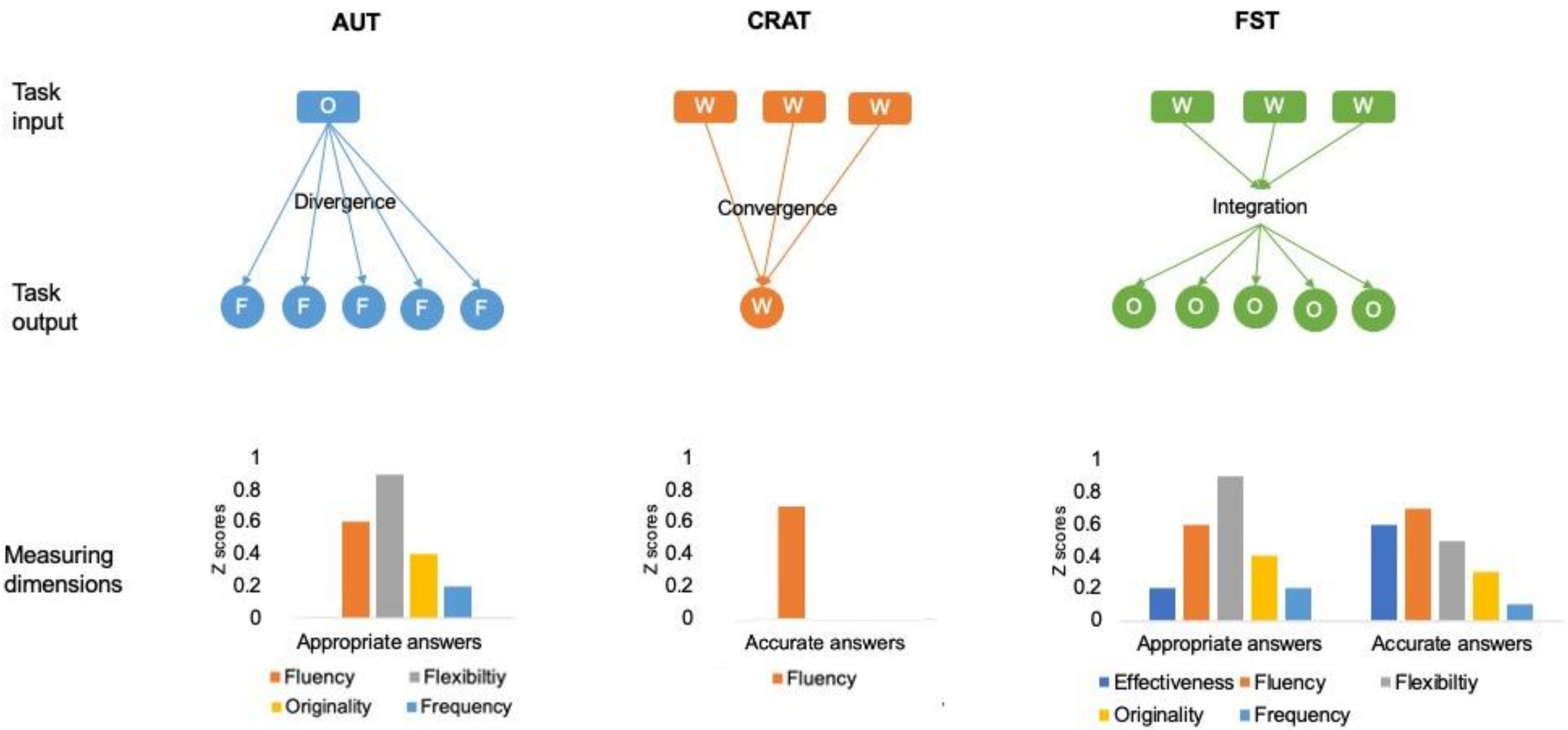
Divergent, convergent, and integrative thinking are essential elements in creative thinking. For the first two, the psychometric tools are widely employed but exhibit low predictive power due to the disconnects between creative thinking and creative products and creative thinking and creative processes.

Therefore, we proposed a new psychometric tool - FST – with the potential to simulate the integrative thinking process. In the FST, people see three functions and come up with objects that fulfil the three functions by naming as many objects as possible. Let’s take the functions “profit, advertise, decorate” as an example. A magazine is an appropriate answer for the given functions because we can use magazines to generate profits for publishers, advertise products, and decorate a bookshelf). The FST measures are effectiveness, fluency, flexibility, originality, and frequency of the responses at the appropriate and accurate levels.

FST has the potential to address some of the issues of divergent and convergent thinking tasks. In detail, FST simulates the starting and ending points of the creative process. It provides people with several pieces of information (i.e., three functions). It asks people to come up with as many pieces of information as possible (i.e., objects that fulfil all three functions). In addition, FST grasps two dimensions of creative products. In detail, appropriate, accurate effectiveness and fluency aim to indicate one’s ability to produce effective ideas. The other seven measures indicate one’s ability to create novel ideas. Therefore, FST has features of both AUT and CRAT but differs from them (see Figure 7).

Figure 7

Visualisation of information processing measures for AUT, CRAT, and FST.



Moreover, FST addresses the issues surrounding several-several tasks. For instance, the given functions (e.g., “entertain, immerse, recreate”) occurs in everyday contexts that do not require domain-specific knowledge. Also, thinking of the objects that can fulfil the three functions given requires people to integrate several pieces of elements into qualitatively new entities. This process may grasp the nature of integrative thinking. In addition, FST is a verbal task complementary to figural tasks.

Therefore, we conducted Study 1 to examine the discriminate validity of FST. In detail, the study assessed the statistical relationship among the measures of AUT, CRAT, and FST. We predicted that

Hypothesis 1: *FST grasps distinct and common aspects of creative thinking from AUT and CRAT.*

Shared measuring aspect (SMA).

To be more specific about the distinct and shared aspects in hypothesis 1, we introduce a new concept - the shared measuring aspect (SMA). There are three measuring aspects for each measure in AUT, CRAT, and FST: *task*, *accuracy*, and *method*. The *task* aspect has three categories: AUT, CRAT, and FST. Also, the *accuracy* aspect has two types: appropriateness and accuracy. Additionally, the *method* aspect has three categories: qualitative novelty, quantitative novelty, and effectiveness. Accordingly, each measure intersects with three measuring aspects (see Figure 8). For example, FST appropriate effectiveness is at the intersection of FST (*task*), appropriateness (*accuracy*), and effectiveness (*method*). AUT appropriate originality is at the intersection of AUT (*task*), appropriateness (*accuracy*), qualitative novelty (*method*). CRAT accurate fluency is at the intersection of CRAT (*task*), accuracy (*accuracy*), quantitative novelty (*method*).

Figure 8

Three measuring aspects of measures for AUT, CRAT, and FST.

Measures / Aspects	Task	Accuracy	Method
AUT appropriate fluency	AUT	Appropriateness	Quantitative novelty
AUT appropriate flexibility	AUT	Appropriateness	Quantitative novelty
AUT appropriate originality	AUT	Appropriateness	Qualitative novelty
AUT appropriate frequency	AUT	Appropriateness	Qualitative novelty
CRAT accurate fluency	CRAT	Accurateness	Quantitative novelty
FST appropriate fluency	FST	Appropriateness	Quantitative novelty
FST appropriate flexibility	FST	Appropriateness	Quantitative novelty
FST appropriate originality	FST	Appropriateness	Qualitative novelty
FST appropriate frequency	FST	Appropriateness	Qualitative novelty
FST appropriate effectiveness	FST	Appropriateness	Effectiveness
FST accurate fluency	FST	Accurateness	Quantitative novelty
FST accurate flexibility	FST	Accurateness	Quantitative novelty
FST accurate originality	FST	Accurateness	Qualitative novelty
FST accurate frequency	FST	Accurateness	Qualitative novelty
FST accurate effectiveness	FST	Accurateness	Effectiveness

SMA indicates the shared measuring aspects of any two of AUT, CRAT, and FST measures and there are at least two approaches to identify SMA values. The first approach assumes that the three shared measuring aspects have the same predictive power, so SMA is the number of shared measuring aspects (SMA^{N} = number of shared measuring aspects). For example, SMA between FST appropriate effectiveness and AUT appropriate originality is one because they only share *accuracy* aspect (i.e., $SMA^{N}_{\text{FST appropriate effectiveness \& AUT appropriate originality}} = 1$). For another example, SMA between FST appropriate effectiveness and CRAT accurate fluency is zero because they do not share any measuring aspect (i.e., $SMA^{N}_{\text{FST appropriate effectiveness \& AUT appropriate originality}} = 0$).

The second approach assumes that the three shared measuring aspects yield different predictive power toward shared variance of measures (here, we label SMA as SMA^{P}). Although $SMA^{N}_{\text{FST appropriate fluency \& FST accurate originality}} = SMA^{N}_{\text{FST appropriate fluency \& AUT appropriate originality}} = 1$, $SMA^{P}_{\text{FST appropriate fluency \& FST accurate originality}}$ could be larger than $SMA^{P}_{\text{FST appropriate fluency \& AUT appropriate originality}}$, if the *task* aspect has larger different predictive power than the *accuracy* aspect.

There are six possibilities of predictive power ranking of three measuring aspects (see Figure 9).

Figure 9

Scoring methods for SMA.

Scoring methods	Task	Accuracy	Method
$SMA^{P(TAM)}$	1	2	3
$SMA^{P(TMA)}$	1	3	2
$SMA^{P(ATM)}$	2	1	3
$SMA^{P(MTA)}$	2	3	1
$SMA^{P(MAT)}$	3	2	1
$SMA^{P(AMT)}$	3	1	2
*SMA^{N}	1	1	1

We assigned 1, 2, and 3 to the three measuring aspects based on six possibilities. The higher the assigned score, the higher the predictive power the measuring aspect has. Accordingly, SMA^{P} equals to the sum of assigned scores for shared measuring aspects and SMA^{P} alters when we employ different scoring methods. For example, when two measures share *task* and *method* aspects, $SMA^{P(T<A<M)} = 4$ but $SMA^{P(A<T<M)} = 5$. Here, we predicted that,

Hypothesis 2: *SMA can predict the shared variance (R) between AUT, CRAT, and FST measure in a positive direction.*

Study 1 Method

We conducted an online observational study which asked participants to finish three creative thinking tasks, including FST, AUT, and CRAT. We measured the performance in the three tasks. We also collected self-reported responses regarding self-efficacy in creative thinking and creative performance.

Participants.

We recruited 148 participants (62 female, 82 males, and four prefer not to say) on Amazon Mechanical Turk (MTurk). The average age of participants was 33.90-year-old (*SD* = 10.62). Most participants were from America (81.76%), 7.43 % were from India, and 10.81% were from Angola, Brazil, Britain, Canada, China, Columbia, and other areas of Asia and Europe. All participants identified themselves as English native-speaker. All participants gave consent and received £7.50 for 1 hour of their time. The UCL Ethics Committee approved the study.

Data screening. We excluded the responses from the participants who did an online search for the correct answers (*N* = 19): the participants who copied and pasted the online resources and who self-reported an online search behaviour. Furthermore, we excluded the responses from the participants who knew the tasks and the corresponding answers before our experiment (*N* = 42). For instance, the participants self-reported the experience and memory of the task and answers. In addition, we excluded the responses from the participants who did not put effort into our experiment (e.g., spent less than 30 seconds in each round; Wrote nothing or random words such as alphas; *N* =18)). After the data screening, 104 responses stayed in the analysis pool.

Post-hoc power analyses using GPower suggested that we had enough power (1.00) to detect the effect at the p = .05 level (*coefficient of determination* r^2 = .20, *effect size* $|r|$ = .44, *tail(s)* = 2, *err prob* = .05, *correlation H0* = 0)

Materials.

AUT. In each round of AUT, participants were asked to think of as many uses of an object as possible. Our study employed six objects: "brick", "hanger", "paperclip", "tire", "newspaper", and "mug".

To measure the performance in AUT, we created six answer pools of six AUT questions. For example, the answer pool of brick and paperclip were separate. For each AUT question, we put corresponding answers of all participants in the answer pool, allowing us to mark the answers without bias. The first marking criteria was appropriateness; We marked an answer as appropriate if it was an appropriate use of the object. The second marking criteria was unusualness; we gave one, two or three points for basic, alternative, and unusual uses, respectively. The third marking criterion was the category. For instance, "computers", "phones", and "VR" belong to one category – technology. We asked judges to determine the category and the categorisation logic based on their subjective experience and evaluation. The fourth marking criteria was the occurrence of an answer, equalling the number of answers divided by the number of all answers. For example, if the number of the answer "VR" and its similar answers "virtual reality" and "VR" was 10 in the answer pool, and the number of all answers was 232 in the answer pool, the occurrence of the answer "VR" was 0.043 (10/232). We gave one point to the answers that occurred more frequently than 0.05 and two points to answers that occurred more frequently than 0.01 and less than or equal to 0.05. We gave three points to answers that occurred less frequently than or equal to 0.01.

The marking criteria resulted in three measures of AUT, which were appropriate fluency (i.e., number of appropriate answers), appropriate flexibility (i.e., category of numbers of appropriate answers), and appropriate originality (i.e., originality score received of appropriate answers). The higher value of the AUT measures, the greater the divergent thinking that a participant displayed. Please see Appendix B for the full version: *Alternative Uses Task - Data Collection, Marking, and Cleaning*.

CRAT. Participants were asked to finish six novel rounds of CRAT randomly selected from eighteen rounds of CRAT. To prepare the questions for CRAT, we looked at the solving rate of 144 CRAT questions reported by a previous study (Bowden & Jung-Beeman, 2003). With the aim of averaging solving rate for each round was 50% - 70%, we randomly chose five questions for each round.

In each question, we provided participants with three stimulus words, and they were asked to produce the fourth word that made up a common compound word or phrase with each of the three stimulus words. For example, the correct answer to the question "cottage/Swiss/cake" is "cheese".

We measured the accurate fluency (i.e., the average number of correct answers in CRAT of each round). The higher the CRAT measures, the better participants' convergent thinking. Please see Appendix C for the full version: *Compound Remote Associate Test - Data Collection and Marking*.

FST. Participants were asked to finish six novel rounds of FST. Participants were given three functions and had to think of as many objects as possible that provided all three functions. The FST questions we employed were "interact/immerse/recreate", "profit/advertise/decorate", "customise/comfort/sanitise", "illuminate/alarm/contain", "protect/entertain/comfort", and "store/package/disseminate".

To measure the performance in FST, we put answers in answers pools (six FST questions, so we had six answer pools) as with the AUT. The first three marking criteria were effectiveness and purposefulness of an answer when fulfilling each of three functions. Specifically, if an answer fulfilled the first function given, the overall effectiveness score of this answer was 3. If an answer was not generated to fulfil the first function given but can provide the function naturally, the overall effectiveness score was 2. If an answer did not fit the first two situations but still fulfilled the first function, the overall effectiveness score was 1. If an answer could not fulfil the first function, the first effectiveness score was 0. Each answer received three effectiveness scores (there were three functions). The fourth marking criteria was appropriateness, and we marked an answer as appropriate if none of the effectiveness scores equalled zero. We marked an answer as wrong if one of the effectiveness scores equalled zero. The fifth marking criterion was accuracy. An accurate answer must satisfy three criteria. First, it must be an appropriate answer. Second, it must be a material object that could be touched and seen. Third, it must be able to fulfil the three functions given to the same users. For example, "an office building" is an appropriate answer for "profit, advertise, decorate" because it is not only able to make a profit or advertise information for companies but also able to decorate the urban landscape. However, "an office building" is not an accurate answer because it cannot provide functions to the same users. The sixth marking criterion was the originality of the answers. We gave one, two, or three points to easy-to-think-of objects, novel and expected objects, and objects that were novel and unexpected, respectively. The seventh marking criteria were flexibility, which followed AUT's criteria.

The seven marking criteria resulted in ten measures for FST. In an appropriate dimension, we had appropriate fluency, appropriate flexibility, appropriate originality (i.e., originality score of appropriate answers), appropriate effectiveness (i.e., an average of three effectiveness scores of appropriate answers), and appropriate frequency. In an accurate dimension, we had accurate fluency (i.e., number of accurate answers), accurate flexibility (i.e., category of numbers of accurate answers), accurate originality (i.e., originality score of accurate answers), and accurate effectiveness (i.e., an average of three effectiveness scores of accurate answers), and accurate frequency. We also measure RT. The higher the value of the FST measures, the greater the participant's integrative thinking. Please see Appendix D for the full version: *Function Synthesis Task - Data Collection, Marking, and Cleaning*.

Answer Pool Marking. We employed the Answer Pool Marking strategy in Study 1 and throughout the project. We provided marking instruction and training to two judges in this marking strategy. Following training, we asked the judges to mark all the answers in an answer pool. Note: In Answer Pool Marking, judges did not mark answers for each participant (Each Participant Marking) but marked each answer. Also, each answer appeared once. Compared to Each Participant Marking, Answer Pool Marking had three advantages. First, in Each Participant Marking, judges were likelier to correlate the performance in task A to task B. For example, they tended to score higher for participants' performance in task B if they performed well in Task A. However, Answer Pool Marking disassociated participants' performance in different tasks and avoided this bias. Second, in Each Participant Marking, judges were likelier to correlate fluency and originality. For instance, they gave higher originality scores to the answers from participants who came up with more ideas. However, Answer Pool Marking avoided this bias because judges could not associate individual fluency and originality. Third, Each Participant Marking required judges to mark the same answer several times, which may lead to inconsistent marks. For instance, participants marked the same answer differently for different participants simply because they were in a different moods. However, Answer Pool Marking avoided this error because judges were asked to mark one answer only once. Accordingly, we inferred that Answer Pool Marking also saved time for judges.

Creative self-efficacy inventory. We measured creative self-efficacy with an established creative self-efficacy inventory (Abbott, 2010). The inventory contained 28 statements describing activities that covered four dimensions of divergent thinking and creative performance. Divergent thinking includes fluency, flexibility, elaboration, and originality (Mark A. Runco & Selcuk Acar, 2019) Creative performance includes aptitude for the domain, impressing the field, and maintaining a creative personality (Csikszentmihalyi, 2014). An example of creative thinking statement was "come up with many possible solutions to a solution". An example of a creative performance statement was that "convince others that you have made a valuable contribution". Participants rated how confident they could do the activities on a 100-point scale. Please see Appendix E for the full version: *Creative Self-efficacy Inventory*.

We calculated the creative self-efficacy based on the ratings. Specifically, we summed the ratings for each dimension. The higher the summed ratings, the higher the confidence in the dimension. We also summed all ratings (*Cronbach's alpha* = .96). The higher the summed ratings, the higher the overall creative self-efficacy.

Procedure.

At the beginning of the experiment, we notified everyone interested in our experiment that they must be an English native speaker to join the experiment and asked them whether they were English native speakers. People who identified themselves as not English native speakers explained the importance of language fluency for our experiment and proceeded to the end of the experiment. People who identified as English native speakers were counted as participants and proceeded to the next step.

In the next step, participants read the participant information sheet and gave consent to participate. Next, we showed participants a brief introduction page which told them that existing research suggested that creativity is correlated with one's performance in AUT, CRAT, and FST. This stage is to increase their interest and attention to our experiment. Due to the high accessibility of AUT answers and CRAT answers from an online search, we replaced the names of AUT and CRAT as the Usage Task and the Relatedness Task, respectively. This introduction aimed to encourage the participants to take our experiment seriously. Name replacement aimed to increase the difficulty of finding the answers from the online resource.

Next, participants were asked to complete AUT, CRAT, and FST randomly. Participants encountered two sections for all tasks. The first section was a task instruction which showed participants the rules and an example of the task. Participants had to pass an instruction understanding test to proceed to the second page. The aim of the understanding test was to ensure that participants had paid enough attention and fully understood the instruction's key messages (e.g., rules). The second section was the six rounds of the main task. It was important to note that there was a self-evaluation page after each round of AUT and FST. On the self-evaluation page, participants were asked to drop their answers into one of three categories regarding unusualness.

After the creative thinking tasks, participants finished a manipulation check which asked them to report where they got the answers from and a questionnaire regarding creative self-efficacy. In the end, we collected demographic information.

Please see Appendix F for the full version: *Study 1 Script*.

Study 1 Results

Inter-rater reliability. We provided marking instruction and training to two judges from psychology and engineering backgrounds. The inter-rater reliability for all measures was acceptable (*Cronbach's α* > .60). Therefore, none of the measures was excluded. Please see Table 1 for the details.

Table 1

Intraclass Correlation Coefficients (ICC) with Two-way Mixed Effects Model for Measures in FST and AUT (N of Items = 2).

		95% *CI*	
Measures	*ICC*	Lower Bound	Upper Bound
AUT appropriate fluency	.91	.89	.92
AUT appropriate flexibility	.89	.84	.92
AUT appropriate originality	.90	.83	.93
FST appropriate fluency	.97	.96	.98
FST appropriate flexibility	.94	.90	.97
FST appropriate originality	.95	.87	.98
FST appropriate effectiveness	.92	.89	.94
FST accurate fluency	.83	.45	.92
FST accurate flexibility	.79	.37	.91
FST accurate originality	.81	.67	.89
FST accurate effectiveness	.84	.78	.89

a. Cronbach's *α* is the average measures of intraclass correlation coefficients.

b. Cronbach's *α* using an absolute agreement definition.

Statistical assumptions. To prepare an appropriate dataset for statistical analysis of Maximum Likelihood Factor Analysis, Pearson Correlation, and Linear Regression, we did a series of assumption tests on the variables. All the variables met the assumption of non-zero variances, no outlier, related pairs, linearity, and no autocorrection. The skewness (between ± 2) & kurtosis (between ±5), and Shapiro–Wilk test indicated that the data contained approximately normally distributed error. Please see Table 2 for Descriptive Statistics of all variables in this study and Table 3 for normality test results. Further assumption tests could be found in the following running texts.

Table 2

Descriptive Statistics for Measures of FST, AUT, CRAT, and Creative Self Efficacy.

Measures	*M*	*SD*
AUT appropriate fluency	5.47	3.08
AUT appropriate flexibility	4.49	1.41
AUT appropriate originality	1.87	0.28
AUT appropriate frequency	1.17	0.09
CRAT accurate fluency	1.88	1.38
FST appropriate fluency	2.86	1.94
FST appropriate flexibility	2.88	0.99
FST appropriate originality	1.48	0.21
FST appropriate frequency	1.26	0.15
FST appropriate effectiveness	2.10	0.19
FST accurate fluency	1.35	1.00
FST accurate originality	1.39	0.52
FST accurate flexibility	1.31	0.56
FST accurate frequency	2.15	0.64
FST accurate effectiveness	0.01	0.01
Creative self-efficacy	56.31	20.70

Table 3

Normality Tests for Measures of AUT, CRAT, FST, and Creative Self-efficacy.

Measures	Skewness	Kurtosis	Shapiro-Wilk	
			Statistics	*p* value
AUT appropriate fluency	1.50	3.93	.90**	<.001
AUT appropriate flexibility	0.44	0.45	.98	.120
AUT appropriate originality	-1.18	1.52	.91**	<.001
AUT appropriate frequency	0.07	-0.61	.98	.070
CRAT accurate fluency	0.72	-0.25	.93**	<.001
FST appropriate fluency	1.36	2.21	.90**	<.001
FST appropriate flexibility	0.85	0.70	.95**	.001
FST appropriate originality	0.27	-0.02	.99	.323
FST appropriate frequency	0.02	-0.59	.98	.072
FST appropriate effectiveness	-0.19	-0.17	.99	.430
FST accurate fluency	1.30	1.61	.89**	<.001
FST accurate originality	-0.45	1.78	.89**	<.001
FST accurate flexibility	0.30	1.84	.93**	<.001
FST accurate frequency	-2.15	5.06	.76**	<.001
FST accurate effectiveness	0.25	-0.40	.98	.062
Creative self-efficacy	-0.31	-0.03	.99	.334

*$p < .05$ (2-tailed). **$p < .01$ (2-tailed).

To examine the construct validity, we examined how AUT, CRAT, and FST measures relate to each other. In detail, we conducted an exploratory factor analysis and Pearson correlation analysis.

Exploratory factor analysis.

Choosing a factor analysis method. For the investigation of the factor structure from a bunch of variables, there were various common-used factor analysis methods such as confirmatory factor analysis (CFA), principal component analysis (PCA), and exploratory factor analysis (EFA). We employed EFA rather than CFA because FST was a new task which had not been examined before. Specifically, CFA was inappropriate for our study since CFA required a well-established hypothesis or theoretical foundations about the latent variables (i.e., factors) for the manifest variables. On the other hand, EFA was appropriate as it allowed the dataset to find the underlying factor structures (Cronbach & Meehl, 1955).

We employed EFA rather than PCA since the results of EFA were more stable than that of PCA. In detail, the key process of EFA dissociated the shared variance, the unique variance, and the error variance of a manifest variable. The objective of EFA was to form an underlying factor structure via analysing how the shared variance of manifest variables was covaried. However, the process of PCA did not dissociate shared variance and unique variance of a manifest variable. The objective of PCA was to reduce the number of manifest variables via analysing how all the variance of manifest variables was covaried. Therefore, PCA would yield different variances when factors were correlated and uncorrelated. However, EFA would yield the same and stable results in these two situations.

Among the factor extraction methods of EFA, we chose Maximum Likelihood Factoring Analysis (MLFA) since our variables were relatively normally distributed. Statistical theorists suggested that MLFA should be the best choice for EFA because it computes the model's goodness of fit and should be employed when data are normally distributed (Jöreskog & Lawley, 1968; Lawley & Maxwell, 1973).

MLFA with all variables. Initially, we examined the factorability of our dataset. Kaiser-Meyer-Olkin measure (KMO = .83) and Bartlett's test of sphericity (X^2 (105) = 2037.86, $p < .001$) indicated that our sample was adequate for factor analysis. However, we excluded the variable CRAT accurate fluency due to its low communality (0.10) and the low primary factor loading (0.15). The two low values indicated that CRAT accurate fluency did not share a common variance with the other variables.

After that, we conducted an MLFA with an oblique method (i.e., oblimin rotation) for the remaining 14 variables in FST and AUT. The Kaiser-Meyer-Olkin measure (KMO = .83), Bartlett's test of sphericity (X^2 (91) = 2018.13, $p < .001$), and goodness-of-fit test, X^2 (52, N = 104) = 394.05, $p < .001$, indicated that our sample was adequate for factor analysis. The primary factor loadings of all variables were above .50 (see Table 4) the correlations between the factors were above .32 or above (see Table 5), indicating that our dataset was adequate for oblimin rotations.

Table 4

Pattern Matrixa and Communalities from a Maximum Likelihood Factor Analysis of 14 Items in AUT and FST.

Items	Factor loading			Communalities
	1	2	3	
Component 1: AUT				
AUT appropriate flexibility	**1.01**	-0.03	0.01	1.00
AUT appropriate fluency	**0.94**	0.04	-0.03	0.91
AUT appropriate frequency	**0.69**	0.05	-0.05	0.49
AUT appropriate originality	**0.53**	-0.04	0.04	0.28
Component 2: FST appropriate quantity and accurate all				
FST accurate fluency	-0.01	**1.05**	-0.14	0.94
FST accurate flexibility	0.01	**1.03**	-0.07	0.98
FST appropriate fluency	0.04	**0.88**	-0.03	0.78
FST appropriate flexibility	0.06	**0.80**	0.05	0.75
FST accurate effectiveness	0.01	**0.74**	0.19	0.77
FST accurate frequency	-0.02	**0.69**	0.23	0.71
FST accurate originality	0.06	**0.61**	0.36	0.84
Component 3: FST appropriate quality				
FST appropriate originality	0.01	-0.04	**0.98**	0.93
FST appropriate frequency	0.02	0.02	**0.94**	0.92
FST appropriate effectiveness	0.05	0.19	**0.65**	0.65

Note. N = 104. The rotation method was oblimin rotation with Kaiser Normalization. Rotation converged in 4 iterations. The primary factor loadings were in bold.

Table 5

Internal Consistency (Cronbach's α, cf. α), Composite Reliability (CR), Average Variance Extracted (AVE), and Component Correlation Matrix from a Maximum Likelihood Factor Analysis of 14 Items in AUT and FST.

Component	*α*	*CR*	*AVE*	1	2	3
1: AUT	.69	.88	.66	-		
2: FST appropriate quantity and accurate all	.91	.94	.71	.54	-	
3: FST appropriate quality	.92	.9	.76	.36	.64	-

Note. N = 104. The rotation method was oblimin rotation with Kaiser Normalization. The square root of *AVE* of component 1 was .82. The square root of *AVE* of component 2 was .84. The square root of *AVE* of component 3 was .87.

Based on the criteria that the eigenvalue was larger than 1, The oblimin rotations extracted a three-factor model, which explained 78.08% of the variance. Factor 1 consisted of all the variables in AUT, so we labelled factor 1 as AUT. Factor 2 consisted of two quantitative variables (i.e., fluency and flexibility) in FST appropriate dimension and all the variables in FST accurate dimension, so we labelled the second factor 2 as FST appropriate quantity and accurate all. Finally, factor 3 consisted of three qualitative variables (i.e., originality, frequency, and effectiveness) in FST appropriate dimension, so we labelled factor 3 as FST appropriate quality.

For items in factor 1 (AUT), the internal consistency was less than but close to .70 (*Cronbach's α* = .69). There was an increase in internal consistency by eliminating the item AUT appropriate originality (*Cronbach's α* = .72)[8]. The

---

[8] We excluded the AUT appropriate originality that shared the least variance with the other AUT variables and replicated the analysis. The results were consistent with the reported results, except for two differences. First, in the reported results, the first factor was AUT, and the second was

composite reliability score was .88, and AVE was .66, indicating acceptable reliability and convergent validity of the AUT factor. The square root of AVE was larger than the correlations among three factors, indicating a discriminant validity of the factor.

For items in factor 2 (FST appropriate quantity and accurate all), the internal consistency was above .70 (*Cronbach's α* = .91). No substantial increase in internal consistency was achieved by eliminating more items. The composite reliability score was .94, and AVE was .71, indicating acceptable reliability and convergent validity of the factor - FST appropriate quantity and accurate all. The square root of AVE was larger than the correlations among three factors, indicating a discriminant validity of the factor.

For items in factor 3 (FST appropriate quality), the internal consistency was above .70 (*Cronbach's α* = .92). No substantial increase in internal consistency was achieved by eliminating more items. The composite reliability score was .90, and AVE was .76, indicating acceptable reliability and convergent validity of the factor - FST appropriate quality. The square root of AVE was larger than the correlations among three factors, indicating a discriminant validity of the factor.

MLFA with appropriate - dimension variables. We replicated the above analysis with the variables in AUT, CRAT, and the FST appropriate dimension. We reasoned that: FST (10 variables) had more variables than AUT (4 variables) and CRAT (1 variable), and the variables in FST were correlated with each other. This situation may strengthen the latent variable related to FST and bias the three tasks' underlying factor structure.

Initially, we examined the factorability of our dataset. Kaiser-Meyer-Olkin measure (*KMO* = .78) and Bartlett's test of sphericity (X^2 (45) = 961.36, p < .001) indicated that our sample was adequate for factor analysis. However, we excluded variable CRAT accurate fluency due to low communality (.09) and the low primary factor loading (.18). The two low values indicated that CRAT accurate fluency did not share a common variance with the other variables.

FST Appropriate Quantity and Accurate All. However, the order of these two factors was reversed when we excluded AUT appropriate originality. Second, in the reported results, the valence of the factor AUT was positive. However, the valence of the factor AUT was reversed when we excluded AUT appropriate originality.

After that, we conducted an MLFA with an oblique method (i.e., oblimin rotation) for the remaining nine variables in FST and AUT. The Kaiser-Meyer-Olkin measure (*KMO* = .77), Bartlett's test of sphericity (X^2 (36) = 945.35, p < .001), and goodness-of-fit test, X^2 (19, N = 104) = 199.39, p < .001, indicated that our sample was adequate for factor analysis. The primary factor loadings of all variables were above .40[9] (see Table 6). The correlations between the factors were above .32 or above (see Table 7), indicating that our dataset was adequate for oblimin rotations.

Table 6

Pattern Matrixa and Communalities from a Maximum Likelihood Factor Analysis of 9 Items in AUT and FST Appropriate Dimension.

Items	Factor loading		Communalities
	1	2	
Factor 1: FST appropriate dimension			
FST appropriate frequency	**1.04**	-0.11	0.99
FST appropriate originality	**0.97**	-0.1	0.86
FST appropriate effectiveness	**0.8**	0.02	0.65
FST appropriate flexibility	**0.51**	0.28	0.47
FST appropriate fluency	**0.46**	0.3	0.43
Factor 2: AUT			
AUT appropriate flexibility	0.01	**0.98**	0.97
AUT appropriate fluency	0	**0.96**	0.93
AUT appropriate frequency	-0.04	**0.73**	0.51
AUT appropriate originality	0.02	**0.51**	0.27

Note. N = 104. The rotation method was oblimin rotation with Kaiser Normalization. Rotation converged in 4 iterations. The primary factor loadings were in bold.

[9] We excluded the FST appropriate fluency whose primary factor loading was lower than .500 and replicated the analysis. The results showed no local minimum was found in 25 iterations, and extraction was terminated.

Table 7

Internal Consistency (Cronbach's α, cf. α), Composite Reliability (CR), Average Variance Extracted (AVE), and Component Correlation Matrix from a Maximum Likelihood Factor Analysis of 9 Items in AUT and FST Appropriate Dimension.

Component	*α*	*CR*	*AVE*	1	2
1: FST appropriate dimension	.78	.88	.62	-	
2: AUT	.69	.88	.67	.46	-

Note. N = 104. The rotation method was oblimin rotation with Kaiser Normalization. The square root of *AVE* of component 1 was .79. The square root of *AVE* of component 2 was .81.

Based on the criteria that the eigenvalue was larger than 1[10], the oblimin rotations extracted a two-factor model, which explained 67.40% of the variance. Factor 1 consisted of the variables in FST appropriate dimension, so we labelled it as FST appropriate. Factor 2 consisted of the variables in AUT, so we labelled it as AUT. The internal consistency values[11], composite reliability, and AVE indicated both factors' acceptable reliability and convergent validity. In addition, the square roots of the AVE of both factors were larger than the correlations between them, indicating a discriminant validity of the factor.

MLFA with accurate - dimension variables. Following the logic of the last analysis**,** we replicated the above analysis for the variables in AUT and CRAT and the variables in the FST accurate dimension. The extraction was terminated until we excluded the CRAT accurate fluency (communality value = .23), which shared the least common variance with the other variables.

[10] Considering that the MLFA with all variables extracted three variables, we tried extracting three factors here. However, no local minimum was found in 25 iterations, and the extraction was terminated.

[11] We excluded the AUT appropriate originality, which lowered the internal consistency of AUT measures. The results showed no local minimum was found in 25 iterations, and extraction was terminated.

The results of the factorability examination were like the above analysis. Kaiser-Meyer-Olkin measure (*KMO* = .82) and Bartlett's test of sphericity (X^2 (45) = 1125.57, p < .001) indicated that our sample was adequate for factor analysis. However, we excluded variable CRAT accurate fluency due to low communality (.09) and the low primary factor loading (.21).

We conducted an MLFA with an oblique method (i.e., oblimin rotation) for the remaining nine variables in FST and AUT. The Kaiser-Meyer-Olkin measure (*KMO* = .83), Bartlett's test of sphericity (X^2 (36) = 1104.26, p < .001), and goodness-of-fit test, X^2 (19, N = 104) = 129.82, p < .001, indicated that our sample was adequate for a piece of factor analysis. The primary factor loadings of all variables were above .50 (see Table 8) the correlations between the factors were above .32 or above (see Table 9), indicating that our dataset was adequate for oblimin rotations.

Table 8

Pattern Matrixa and Communalities from a Maximum Likelihood Factor Analysis of 9 Items in AUT and FST Accurate Dimension.

Items	Factor loading		Communalities
	1	2	
Factor 1: AUT			
AUT appropriate flexibility	**1.01**	-0.02	1.00
AUT appropriate fluency	**0.94**	0.03	0.91
AUT appropriate frequency	**0.69**	0.02	0.49
AUT appropriate originality	**0.53**	-0.02	0.27
Factor 2: FST accurate dimension			
FST accurate flexibility	-0.01	**1**	0.98
FST accurate fluency	-0.02	**0.97**	0.92
FST accurate effectiveness	0	**0.87**	0.76
FST accurate frequency	-0.02	**0.84**	0.68
FST accurate originality	0.07	**0.83**	0.75

Note. N = 104. The rotation method was oblimin rotation with Kaiser Normalization. Rotation converged in 3 iterations. The primary factor loadings were in bold.

Table 9

Internal Consistency (Cronbach's α, cf. α), Composite Reliability (CR), Average Variance Extracted (AVE), and Component Correlation Matrix from a Maximum Likelihood Factor Analysis of 9 Items in AUT and FST Accurate Dimension

Component	*α*	*CR*	*AVE*	*1*	*2*
1: AUT	.94	.88	.66	-	
2: FST accurate dimension	.69	.96	.82	.54	-

Note. N = 104. The rotation method was oblimin rotation with Kaiser Normalization. The square root of *AVE* of component 1 was .81. The square root of *AVE* of component 2 was .90.

Based on the criteria that the eigenvalue was larger than 1[12], the oblimin rotations extracted a two-factor model, which explained 75.15% of the variance. Factor 1 consisted of the variables in AUT, so we labelled it as AUT. Factor 2 consisted of the variables in FST, so we labelled it as FST accurate dimension. The internal consistency values[13], composite reliability, and AVE indicated both factors' acceptable reliability and convergent validity. In addition, the square roots of the AVE of both factors were larger than the correlations between them, indicating a discriminant validity of the factor.

MLFA with FST variables. We also explored the latent variables of the ten measured variables for FST. The Kaiser-Meyer-Olkin measure (KMO = .82), Bartlett's test of sphericity (X^2 (45) = 1643.02, p < .001), and goodness-of-fit test, X^2 (26, N = 104) = 358.37, p < .001, indicated that our sample was adequate for a piece

[12] Considering that the MLFA with all variables extracted three variables, we tried extracting three factors here. However, the three-factor model that was extracted was not significant, X^2 (7, N = 104) = 10.32, p = .17.

[13] We excluded the AUT appropriate originality and replicated the analysis. The results were consistent with the reported results, except for two differences. First, in the reported results, the first factor was AUT, and the second factor was FST appropriate quantity and accurate all. However, the order of these two factors was reversed when we excluded AUT appropriate originality. Second, in the reported results, the valence of the factor AUT was positive. However, the valence of the factor AUT was reversed when we excluded AUT appropriate originality.

of factor analysis. The primary factor loadings of all variables were above .60 (see Table 10). The correlations of between the factors were above .32 or above (see Table 11), indicating that our dataset was adequate for oblimin rotations.

Table 10

Pattern Matrixa and Communalities from a Maximum Likelihood Factor Analysis of 10 Items in FST.

Items	Factor loading		Communalities
	1	2	
Factor 1: FST appropriate quality and accurate all			
FST accurate fluency	**1.05**	-0.14	0.94
FST accurate flexibility	**1.04**	-0.08	0.98
FST appropriate fluency	**0.90**	-0.03	0.78
FST appropriate flexibility	**0.83**	0.05	0.75
FST accurate effectiveness	**0.75**	0.18	0.63
FST accurate frequency	**0.68**	0.23	0.71
FST accurate originality	**0.63**	0.37	0.83
Factor 2: FST appropriate quantity			
FST appropriate originality	-0.05	**1**	0.93
FST appropriate frequency	0.01	**0.95**	0.92
FST appropriate effectiveness	0.21	**0.64**	0.63

Note. N = 104. The rotation method was oblimin rotation with Kaiser Normalization. Rotation converged in 4 iterations. The primary factor loadings were in bold.

Table 11

Internal Consistency (Cronbach's α, cf. α), Composite Reliability (CR), Average Variance Extracted (AVE), and Component Correlation Matrix from a Maximum Likelihood Factor Analysis of 10 Items in FST.

Component	*α*	*CR*	*AVE*	1	2
1: FST appropriate quantity and accurate all	.91	.95	.73	-	
2: FST appropriate quality	.92	.91	.77	.54	-

Note. N = 104. The rotation method was oblimin rotation with Kaiser Normalization. The square root of *AVE* of component 1 was .86. The square root of *AVE* of component 2 was .88.

Based on the criteria that the eigenvalue was larger than 1, the oblimin rotations extracted a two-factor model, which explained 82.28 % of the variance. Factor 1 consisted of appropriate fluency, appropriate flexibility, and all variables in accurate dimension, so we labelled it as FST appropriate quantitative novelty and accurate all. Factor 2 consisted of appropriate originality, fluency, and effectiveness, so we labelled it as FST appropriate qualitative novelty and qualitative effectiveness. The internal consistency values, composite reliability, and AVE indicated both factors' acceptable reliability and convergent validity. In addition, the square roots of the AVE of both factors were larger than the correlations between them, indicating a discriminant validity of the factor.

MLFA with AUT variables. We also explored the latent variables of the four manifest variables for AUT. The Kaiser-Meyer-Olkin measure (*KMO* = .69), Bartlett's test of sphericity (X^2 (6) = 351.67, p < .001), and goodness-of-fit test, X^2 (2, N = 104) = 18.28, p < .001, indicated that our sample was adequate for a piece of factor analysis. Based on the criteria that the eigenvalue was larger than 1, MLFA extracted only one factor with all four variables, and the solution could not be rotated.

Pearson correlation analysis.

In this section, we present the Pearson correlation results in tables (see Table 12 – 13) and demonstrate them in text after the tables.

Table 12

Correlations Cross Creative Self-efficacy and Creative Thinking (All Participants, N = 104).

	Creative Self-efficacy
AUT appropriate fluency	.00
AUT appropriate flexibility	.02
AUT appropriate originality	-.02
AUT appropriate frequency	.17
CRAT accurate fluency	.23*
FST appropriate fluency	.17
FST appropriate flexibility	.18
FST appropriate originality	.20*
FST appropriate frequency	.17
FST appropriate effectiveness	-.28**
FST accurate fluency	.01
FST accurate flexibility	-.04
FST accurate originality	-.02
FST accurate frequency	-.09
FST accurate effectiveness	-.19

*$p < .05$ (2-tailed). **$p < .01$ (2-tailed).

Table 13

Correlations for AUT, CRAT, and FST Measures (All Participants, N = 104).

Pearson Correlations	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. AUT appropriate fluency	-														
2. AUT appropriate flexibility	.94**	-													
3. AUT appropriate originality	.53**	.63**	-												
4. AUT appropriate frequency	.58**	.58**	.54**	-											
5. CRAT accurate fluency	.20*	.23*	.22*	.16	-										
6. FST appropriate fluency	.46**	.48**	.30**	.26**	.24*	-									
7. FST appropriate flexibility	.46**	.48**	.29**	.25*	.26**	.93**	-								
8. FST appropriate originality	.11	.12	.01	.11	.01	.27**	.26**	-							
9. FST appropriate frequency	.13	.11	.00	.12	.04	.35**	.33**	.57**	-						
10. FST appropriate effectiveness	.24*	.24*	.20*	-.02	.14	.03	-.07	-.26**	-.06	-					

11. FST accurate fluency	.45**	.47**	.35**	.19	.24*	.92**	.82**	.24*	.29**	.17	-				
12. FST accurate flexibility	.36**	.37**	.31**	.09	.12	.69**	.70**	.27**	.38**	.16	.71**	-			
13. FST accurate originality	.30**	.31**	.09	.11	.02	.30**	.35**	.47**	.36**	-.08	.27**	.45**	-		
14. FST accurate frequency	.16	.18	.22*	-.06	-.01	.20*	.23*	.03	.12	.39**	.22*	.48**	.36**	-	
15. FST accurate effectiveness	-.06	-.03	.05	-.17	-.03	.11	.16	-.27**	-.18	.11	.17	.11	-.01	.47**	-

*p < .05 (2-tailed). **p < .01 (2-tailed).

Correlations of AUT. The results showed significant positive associations among the measures in AUT. Specifically, there was a strong association ($r >= .90$) between appropriate fluency and flexibility. All the other variables were associated at moderate level ($.40 <= r < .70$).

Correlations of appropriate dimension in FST. The results showed significant positive associations among the appropriate-dimension measures of FST. Specifically, there were strong associations between appropriate fluency and flexibility. There were moderate associations between appropriate originality and frequency. There were weak associations ($0 < r < .40$) between appropriate fluency and originality, between appropriate fluency and frequency, and between appropriate flexibility and frequency. The results also showed a significant negative association between appropriate effectiveness and originality, while effectiveness was not associated with other measures.

Correlations of accurate dimension in FST. The results showed significant positive associations among the accurate-dimension measures of FST. Specifically, there was a high association ($.70 <= r < .90$) between accurate fluency and flexibility. There were moderately associations between accurate flexibility and originality, accurate flexibility and frequency, and accurate frequency and effectiveness. However, accurate effectiveness was not associated with other measures.

Correlations of cross dimensions in FST. The results showed significant positive associations among all measures of FST. For the measures that were partially in the same dimension (e.g., appropriate fluency and accurate fluency), there was a strong association between appropriate and accurate fluency. There were moderate associations between appropriate and accurate flexibility and between appropriate and accurate originality. However, there was no significant association between appropriate and accurate frequency and between appropriate and accurate effectiveness.

There was a high association between appropriate flexibility and accurate flexibility for the measures in a completely different dimension (e.g., appropriate fluency and accurate flexibility). There was a moderate association between appropriate flexibility and accurate fluency. Notably, there was a significant negative association between appropriate originality and accurate effectiveness. Except for the above, all the other variables were positively associated at a weak level or not associated.

Correlations among AUT, CRAT, and FST. The results showed significant positive associations among measures in AUT, CRAT, and FST. In detail, both AUT appropriate fluency and flexibility were moderately associated with FST appropriate fluency, appropriate flexibility, and accurate fluency, and weakly associated with FST appropriate effectiveness, accurate flexibility, and accurate originality. AUT appropriate originality was weakly associated with FST fluency, flexibility, effectiveness in the appropriate dimension and FST fluency, flexibility, and frequency in the accurate dimension. AUT appropriate frequency was weakly associated with FST appropriate fluency and flexibility. CRAT accurate fluency was weakly associated with appropriate fluency, flexibility, and originality in AUT and appropriate and accurate fluency in FST. None of the other associations was significant.

Correlations among creative self-efficacy and others. The results showed significant positive associations between creative self-efficacy and creative thinking. For instance, creative self-efficacy was weakly associated with CRAT accurate fluency and FST appropriate originality. Also, there was a significant negative association between creative self-efficacy and FST appropriate effectiveness.

Predictive power of SMA towards AUT, CRAT & FST correlations.

Statistical assumptions. We firstly examined statistical assumptions for dependent variables in linear regressions (i.e., correlations among AUT, CRAT, and FST measures[14]). The data met the assumption of independent errors as the Durbin-Watson value was larger than 1 and less than 3. The normal P-P plot of standardised residuals did not show completely points on the line. The scatterplot of standardised predicted values showed that the variables met the assumptions of homogeneity of variance. There is not an issue of heteroscedasticity.

Correlations between SMA and correlations. The results showed that correlations among AUT, CRAT, and FST measures were significantly positively associated with SMA (see Table 14).

---

[14] To clean the data, we excluded the correlations that were not significant. However, the results remained consistent when we included the correlations that were not significant.

Table 14

Correlations Between SMA and Shared Variance Among AUT, CRAT, and FST (N = 77).

	1	2	3	4	5	6	7	8
1. SMA^{N}	-							
2. $SMA^{P(T<A<M)}$	.93**	-						
3. $SMA^{P(T<M<A)}$	.95**	.94**	-					
4. $SMA^{P(A<T<M)}$	.93**	.95**	.84**	-				
5. $SMA^{P(M<T<A)}$	.94**	.80**	.94**	.76**	-			
6. $SMA^{P(M<A<T)}$	.94**	.75**	.84**	.80**	.95**	-		
7. $SMA^{P(A<M<T)}$	.95**	.84**	.81**	.93**	.85**	.94**	-	
8. R	.76**	.76**	.70**	.78**	.63**	.65**	.74**	-

**p* < .05 (2-tailed). ***p* < .01 (2-tailed).

Regression of SMA towards correlations. Following the correlation results, we conducted single linear regressions to examine the predictive power of different SMA on the correlations. The results showed that SMA^{N} explained a significant amount of the variance in correlations among AUT, CRAT, and FST, F (1, 75) = 99.48, p <.001, R^2 = .57, $R^2_{Adjusted}$ = .56. In detail, SMA^{N} was a significant predictor of correlations among creative thinking measures, B = .26, *Beta* = .76, t (75) = 9.97, p < .001, *95% CI* [.21, .31] (*Durbin-Watson* = 1.975, *std. residual (min)* = -3.91, *std. residual (max)* = 1.76).

Regression of SMA towards correlations. Following the correlation results, we conducted single linear regressions to examine the predictive power of different SMA on the correlations. The results showed that SMA^{N} explained a significant amount of the variance in correlations among AUT, CRAT, and FST, F (1, 75) = 99.49, p <.001, R^2 = .57, $R^2_{Adjusted}$ = .56. In detail, SMA^{N} was a significant predictor of correlations among creative thinking measures, B = .26, *Beta* = .76, t (75) = 9.97, p < .001, *95% CI* [.21, .31] (*Durbin-Watson* = 1.98, *std. residual (min)* = -3.91, *std. residual (max)* = 1.76).

The results showed that $SMA^{T<A<M}$ explained a significant amount of the variance in correlations among AUT, CRAT, and FST, F (1, 75) = 103.97, p <.001, R^2 = .58, $R^2_{Adjusted}$ = .58. In detail, $SMA^{T<A<M}$ was a significant predictor of

correlations among creative thinking measures, *B* = .12, *Beta* = .76, *t* (75) = 10.20, *p* < .001, *95% CI* [.10, .15] (*Durbin-Watson* =1.89, *std. residual (min)* = -3.43, *std. residual (max)* =1.72).

The results showed that $SMA^{T<M<A}$ explained a significant amount of the variance in correlations among AUT, CRAT, and FST, F (1, 75) = 70.96, p <.001, R^2 = .49, $R^2_{Adjusted}$ = .48. In detail, $SMA^{T<M<A}$ was a significant predictor of correlations among creative thinking measures, *B* = .11, *Beta* = .70, *t* (75) = 8.42, *p* < .001, *95% CI* [.09, .14] (*Durbin-Watson* =1.94, *std. residual (min)* = -3.57, *std. residual (max)* = 2.11).

The results showed that $SMA^{A<T<M}$ explained a significant amount of the variance in correlations among AUT, CRAT, and FST, F (1, 75) = 120.44, p <.001, R^2 = .62, $R^2_{Adjusted}$ = .61. In detail, $SMA^{A<T<M}$ was a significant predictor of correlations among creative thinking measures, *B* = .13, *Beta* = .79, *t* (75) = 10.97, *p* < .001, *95% CI* [.11, .15] (*Durbin-Watson* = 1.93, *std. residual (min)* = -3.53, *std. residual (max)* = 1.52).

The results showed that $SMA^{M<T<A}$ explained a significant amount of the variance in correlations among AUT, CRAT, and FST, F (1, 75) = 50.38, p < .001, R^2 = .40, $R^2_{Adjusted}$ = .39. In detail, $SMA^{M<T<A}$ was a significant predictor of correlations among creative thinking measures, *B* = .10, *Beta* = .63, *t* (75) = 7.10, *p* < .001, *95% CI* [.07, .13] (*Durbin-Watson* = 2.10, *std. residual (min)* = -3.64, *std. residual (max)* = 1.98).

The results showed that $SMA^{M<A<T}$ explained a significant amount of the variance in correlations among AUT, CRAT, and FST, F (1, 75) = 55.16, p < .001, R^2 = .42, $R^2_{Adjusted}$ = .42. In detail, $SMA^{M<A<T}$ was a significant predictor of correlations among creative thinking measures, *B* = .10, *Beta* = .65, *t* (75) = 7.43, *p* < .001, *95% CI* [.08, .13] (*Durbin-Watson* = 2.06, *std. residual (min)* = -3.71, *std. residual (max)* = 1.62).

The results showed that $SMA^{A<M<T}$ explained a significant amount of the variance in correlations among AUT, CRAT, and FST, F (1, 75) = 89.69, p < .001, R^2 = .55, $R^2_{Adjusted}$ = .54. In detail, $SMA^{A<M<T}$ was a significant predictor of correlations among creative thinking measures, *B* = .12, *Beta* = .74, *t* (75) = 9.47, *p* < .001, *95% CI* [.10, .15] (*Durbin-Watson* = 1.88, *std. residual (min)* = -3.76, *std. residual (max)* = 1.36).

Exploratory analysis on gender difference.

The results of the independent samples test showed that, compared to male participants, female participants received significantly higher scores for appropriate fluency, flexibility, and originality in AUT and FST. Female participants also received a significantly higher score for FST accurate originality. None of the other measures revealed a significant difference between female and male participants (see Table 15).

Table 15

Results of Independent Sample Tests Examining the Gender Differences for Creative thinking.

Measures	Female (*N*=45)		Male (*N*=59)		*t* (102)	*p* value	95% *CI*	
	M	*SD*	*M*	*SD*			Low	Up
AUT appropriate fluency	6.59	3.26	4.62	2.67	3.38**	.001	0.81	3.12
AUT appropriate flexibility	4.96	1.27	4.13	1.41	3.11**	.002	0.30	1.36
AUT appropriate originality	1.95	0.18	1.81	0.32	2.65**	.009	0.04	0.25
AUT appropriate frequency	1.18	0.09	1.16	0.09	1.19	.237	-0.01	0.06
CRAT accurate fluency	2.06	1.27	1.74	1.45	1.17	.245	-0.22	0.86
FST appropriate fluency	3.23	1.92	2.58	1.93	1.71	.090	-0.10	1.41
FST appropriate flexibility	3.10	1.04	2.71	0.93	2.03*	.045	0.01	0.78
FST appropriate originality	1.54	0.21	1.43	0.20	2.80**	.006	0.03	0.19
FST appropriate frequency	1.26	0.15	1.26	0.15	0.29	.772	-0.05	0.07
FST appropriate effectiveness	2.08	0.20	2.11	0.18	-0.67	.503	-0.10	0.05
FST accurate fluency	1.53	0.91	1.21	1.05	1.65	.103	-0.07	0.71
FST accurate flexibility	1.36	0.56	1.28	0.55	0.80	.427	-0.13	0.31
FST accurate originality	1.53	0.50	1.29	0.52	2.30**	.023	0.03	0.44

FST accurate frequency	2.15	0.58	2.15	0.69	0.02	.983	-0.25	0.26
FST accurate effectiveness	0.01	0.01	0.01	0.01	-1.45	.151	0.00	0.00
Creative self-efficacy	58.76	23.23	54.43	18.53	1.06	.293	-3.79	12.45

*p < .05 (2-tailed). **p < .01 (2-tailed).

Study 1 Discussion

Study 1 supported hypothesis 1, that FST grasps distinct and common aspects of creative thinking from AUT and CRAT, and hypothesis 2, that SMA can predict the shared variance between AUT, CRAT, and FST.

MLFA. According to MLFA results, the measures that shared the *task* aspect shared the most variance (cluster 1). In detail, AUT measures shared the most variance. FST appropriate fluency and flexibility and all accurate measures shared variance (cluster 2). Also, FST measures appropriate originality, frequency, and effectiveness shared variance (cluster 3). However, CRAT accurate fluency did not share variance with AUT and FST measures.

There are two takeaways from the results. First, we can infer that AUT, CRAT, and FST may grasp different aspects of creative thinking because AUT, CRAT, and FST measures were grouped in different clusters. The finding implies that three thinking tasks represent qualitatively different thinking styles. In detail, AUT represents divergent thinking, CRAT represents convergent thinking, and FST may represent integrative thinking. We can also infer that divergent, convergent, and integrative thinking represent different cognitive processes. Second, FST may grasp more aspects of creative thinking than AUT and CRAT because FST measures were divided into more clusters than AUT and CRAT measures. The finding is consistent with previous arguments that integrative thinking combines divergent and convergent thinking (Cropley, 2006; Tan, 2015; Zittoun et al., 2007).

However, MLFA results cannot tell a comprehensive story to us. For example, we cannot infer the predictive power of the *accuracy* aspect and *method* aspects' predictive power toward the shared variance. Therefore, we conducted a Pearson correlation analysis which associated measures in the three tasks.

Pearson correlations. The Pearson correlation results offer us five pieces of information.

1. Many significant associations were positive (60 of 62).
2. There were strong positive correlations ($r > .92$) between the measures that shared the *method* aspect in quantitative novelty, the *accuracy* aspect in appropriateness, and the *task* aspect in AUT or FST.
3. There were high positive correlations between the measures that shared the m*ethod* aspect in quantitative novelty and the *task* aspect in AUT or FST ($.69 < r =< .92$).
4. There were moderate positive correlations between the measures that shared the *accuracy* aspect in appropriateness and the *task* aspect in AUT or FST ($.52 < r < .64$).
5. There was a weak negative correlation between FST appropriate originality and FST appropriate and accurate effectiveness ($-.27 =< r < -.25$).

Pearson correlation results supported MLFA results and offered us further information. First, the results showed that AUT, CRAT, and FST grasped different aspects of creative thinking, which again support our argument that AUT, CRAT, and FST represent different thinking styles.

Second, cross-task measures may grasp the similar aspect of creative thinking. For instance, fluency and flexibility of appropriate answers grasped the most similar aspects as they shared the highest level of variance. Also, appropriate answers in the same tasks and fluency and flexibility of appropriate and accurate answers shared a high level of variance. In other words, the number, and the category number of (appropriate) answer, though in different creative-thinking tasks, may grasp a similar aspect of creative thinking.

We can see that all the cross-task measures that shared variance were in AUT and FST rather than CRAT. What can we learn from it? On the one hand, the overlap between AUT and FST measures indicates the overlap between divergent and integrative thinking. On the other hand, the lack of overlap between CRAT measures and the others may be attributed to two potential explanations. For instance, convergent thinking may overlap with neither divergent nor integrative thinking. Another explanation is that CRAT is a language-based task while AUT and FST are functions/uses-based. The differences in the knowledge base may induce the differences in the measures. Therefore, it is important to note that, though the CRAT measure did not share variance with AUT and FST measures, CRAT may share some cognitive processes with AUT and FST.

Third, originality and effectiveness in FST were negatively correlated with each other. It implies that originality and effectiveness may grasp opposite aspects of creative thinking. Also, the novelty and effectiveness of the produced ideas FST changed in the opposite ways. Therefore, we can infer that the thinking style grasped by FST may not be able to benefit the originality and effectiveness of idea generation at the same time.

However, the patterns of the significant correlations between .19 to .50 were unclear. Therefore, we conducted a regression analysis, measuring SMA's predictive power towards the correlations.

Regression analysis. The linear regression results showed that all the scoring methods for SMA significantly predicted the shared variance between creative thinking measures. Again, regression analysis results supported the previous argument that AUT, CRAT, and FST grasped different aspects of creative thinking. They also showed that the appropriate and accurate answers represented different aspects of creative thinking, and quantitative novelty measures, qualitative novelty measures, and effectiveness measures also grasped different aspects of creative thinking. More than that, the results showed that $SMA^{A>T>M} > SMA^{T>A>M} > SMA^{N} > SMA^{A>M>T} > SMA^{T>M>A} > SMA^{M>A>T} > SMA^{M>T>A}$.

The results supported hypothesis 1 and informed us how different aspects of SMA weighted in predicting shared variances. In detail, compared to the difference between the measures in different *task* aspects, the differences between the measures in different *accuracy* aspects may be larger. In addition, the difference between the measures in different *method* aspects may be the least among the three aspects. However, it is important to note that: the difference between the measures in different *accuracy* aspects was the largest maybe because it already entails the difference in the *task* aspect, such that AUT only has appropriate answers and CRAT only has accurate answers. Therefore, the difference in the *accuracy* aspect can still be attributed to the *task* aspect.

Strength.

Advance understanding of creative thinking with FST. The generation of FST advanced our understanding of creative thinking by showing that creative thinking is not limited to the widely accepted and employed concepts such as divergent and convergent thinking. Integrative thinking may be crucial in real-life creative performance in social contexts. Accordingly, our research highlights the importance of understanding integrative thinking in the creative process, and FST has the potential to change the situation that creative thinking research is overwhelmed by AUT and CRAT.

Limitations and suggestions.

Can FST represent integrative thinking? Although we generated FST based on integrative thinking, it is still early to define FST as an integrative thinking task. One of the reasons is that the creativity research field does not have a widely accepted definition of integrative thinking. Therefore, we do not define integrative thinking but differentiate it from divergent and convergent thinking. In detail, the three thinking skills' starting and ending points of information processing are different. The generation of FST was also based on the difference. Therefore, the link between FST and integrative thinking may disappear if their information process does not match.

Suggestions. Therefore, we suggest that future studies further explore the relationship between integrative thinking and FST. Here are the directions:

1. Work on the definition of integrative creativity.
2. Produce more empirical evidence regarding the physiology, neuroscience, cognitive, affective, and behavioural mechanisms underlying FST and compare it with AUT and CRAT.
3. Actively linking the outcome of the first two steps would help further advance the understanding of integrative thinking.

This chapter has introduced FST and links it with AUT and CRAT. The next chapter links creative thinking with social comparison and explore how comparative social feedback affects people's performance in different creative thinking tasks.

Chapter 3: Creative Thinking and Social Comparison

As an important quality of human beings, creative thinking is either affected by the social context or a solution to social issues. This chapter focuses on the former view. Here, we review the theories that specify the role of social context on creativity and talk about the implicit assumptions of these theories. Next, we highlight why studying the link between creative thinking and social comparison is important. In addition, we introduce the concept of social comparison and review the empirical evidence that investigates how social comparison affects creative thinking. Moreover, we discuss the inconsistency of existing evidence and propose a new experimental paradigm (Repeated Rounds of Creativity Tasks with Social Feedback; RRCTSF) to study the topic. After that, we demonstrate three studies (with one pilot study) that examined the effect of different kinds of comparative social feedback (i.e., competition and star rating) on CRAT, AUT, and FST.

Literature Review

Creativity is affected by social context: social psychology of creativity.

An educational creativity psychologist Plucker defined creativity as "*the interaction of aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context*" (Plucker et al., 2004).

Several theories specify the role of social context on creativity. For example, the system model of creativity proposes that creative accomplishment results from the interaction of three forces. They are a field ("*selects from the variations produced by individuals those that are worth preserving*"), a cultural domain ("*selects new ideas or forms that are worth preserved and transmitted to the following generations*"), and the individual ("*brings about some change in the domain*") (Csikszentmihalyi, 2014). Also, Boden distinguished personal and historical creativity and proposed that personal creativity is an intrapersonal process (Boden, 2004).

Some researchers suggest that social and environmental factors interact with personality characteristics and cognitive skills to affect the creative process and output. For example, Amabile proposed that social and environmental factors that induce extrinsic constraints alter intrinsic task motivations and influence creative performance in the componential theory of creativity. Also, Amabile and her colleagues conducted a series of experiments studying the effect of the social context (e.g., evaluation, competition, surveillance, coaction) on kids' creativity

(Amabile et al., 1990) and found inconsistent results. After that, they applied the theory in an organisational setting and found that social context is a double-edged sword for creativity.

Amabile and her colleagues also found that social context is a double-edged sword for creativity because social context induces extrinsic motivations. On the one hand, extrinsic motivations may make people feel that they are dictated, undermining creativity. For example, creative performance in an organisation would be undermined by criticising new ideas, generating political problems, emphasising the status quo, being conservative in top management, and under excessive pressure (Amabile, 2011). On the other hand, extrinsic motivations may confirm people's competence and induce their excitement about the work. For instance, creative performance in an organisation would be enhanced by inducing a sense of positivity, creating collaborative team spirit, allowing autonomy, encouraging new ideas, and having an innovation-encouraging vision in the top management (Amabile, 1983, 1993; Amabile & Mueller, 2008).

The investment theory of creativity proposes consistent arguments for extrinsic motivations. In detail, creative people are willing to pursue unknown areas with growth potential, persistent when encountering resistance, and eventually receive admiration (buy low and sell high). The intelligence, knowledge, thinking styles, personality, motivations, and environments are six distinct but related resources for the creative production of these people. Furthermore, the environmental resources indicate that an individual's creativity would not display without a supportive and rewarding environment (Sternberg & Lubart, 1996).

More recently, a dual pathway model of creativity incorporates situational and dispositional variables on creative performance. The dual pathway model suggests that novel and effective ideas are obtained via two qualitatively different cognitive processes – cognitive flexibility and persistence. In the flexibility pathway, people follow the two-stage generation and evaluation process. In detail, they have broad attentional focus and switch flexibly between approaches to the task. It is related to increased dopamine and reduced inhibition and allows more distant associations, which lead to more original thoughts. However, this one may reduce cognitive control and increase distractibility. Therefore, we may need cognitive inhibition later to evoke our internal goals and intentions to take out poor solutions (evaluation). On the other hand, people systematically and effortfully explore the possibilities in a few categories or perspectives in the persistence pathway. Exploring a limited number of categories may result in obvious and readily available solutions at the beginning. However, in-depth exploration may lead to an original response within the category of people persisted. Furthermore, the process requires a high level of cognitive control to remove irrelevant thoughts from working memory and fully focus on the task. Therefore, they are more focused and less flexible in the persistence pathway (Nijstad et al., 2010).

It remains unknown whether the two pathways are negatively related. Some researchers argued that cognitive processes must play different roles in different pathways, so two pathways are negatively related. For instance, attention must focus on a task and ignore distraction in persistence. However, this also reduces the breadth of attentional focus and prevents flexibility. Therefore, the two pathways might require attentional states, but we can switch from more flexible to more systematic processing modes. For instance, we can be flexible and persistent when stuck requires in-depth exploration (Geneplore model). However, this means that the executive function is important in creative thoughts.

Two assumptions.

Bottom-up process vs. top-down process. According to the above theories that address the importance of social context on creativity, social context affecting the creative process is a bottom-up process (Simonton, 2010), which refers to a stimuli-detection mechanism. In detail, bottom-up processing of social stimuli is driven by people's ability to detect social stimuli. Based on this assumption, the creative process is an intrapersonal process by default. The effect of social context is determined by the salience of the social stimuli and individual ability to trigger the social effect on creativity. However, the social psychology of creativity can also be a top-down process which is a knowledge-driven mechanism. For instance, top-down processing of social stimuli is driven by people's acquired predictions about others' thoughts and behaviour. Based on this assumption, the creative process is an interpersonal process by default, and the effect of social factors is determined by the people's social expectations and acquired response rules (Sarter et al., 2001).

Early cognitive psychologists refer to higher-order cognition as internalising the social world. For example, Mead (1934) proposed that people systematically process the rule-based social world into every aspect of mind operation. This process determined cognitive processing and behaviour (Mead & Schubert, 1934).

Vygotsky emphasised that people constantly internalise others' actions around them, so social interaction forms the fundamental of higher-order cognition such as intelligence (Sternberg, 2003; Vygotsky, 1997), p.24. Therefore, an individual creative potential may start with the internalisation of social interactions by not just copying but rather a transformation or reorganisation of incoming information and mental structures based on existing knowledge.

Automatic vs. controlled social cognition. The existing theories of social psychology of creativity also assume social cognition in creativity as a controlled processing, such that people detect social stimuli when experimenters deliver them deliberately. However, the human mind and social environment are inseparable and mutually constitutive concepts. Therefore, the processing of social information may be automatic. For being creative, self-evaluation can be an interactive referential process in which people compare their ideas against other people's ideas, relevant culture domains, and experts' and audiences' acceptability in relevant domains (Sawyer, 2006; Silvia et al., 2008; Sternberg, 2006). The automatic social cognition in the creative process also explains the research that found creative people have anti-social tendencies (Feist, 1998). It takes effort to not compare with others and not think about what others think. Therefore, isolating oneself during the creative process can save cognitive resources.

Accordingly, we can infer that the novelty-oriented creative process is more likely to induce automatic social cognition than the accuracy-oriented process. For example, the novelty measure of AUT and FST (not CRAT) may spontaneously induce social comparison because novel ideas are those that others can rarely think of, so the others' performance is the reference point. Without awareness, AUT and FST may automatically activate social signals within one's existing knowledge and change their behaviour. Therefore, people doing AUT and FST and receiving feedback about their performance may process the given information, their ideas, and the potential ideas of the other people at the same time. On the other hand, CRAT that has correct answers may not evoke social comparison automatically because the reference point is a pre-determined logically correct answer. Because social comparison has been automatically induced in AUT and FST rather than CRAT, the deliberately given social comparison stimuli may have a different effect on performance in CRAT and AUT & FST. Thankfully, existing research examined how social comparison affects creative thinking. Before we review the empirical evidence, please allow us to introduce social comparison.

Social comparison.

Human beings seek reference points to evaluate their opinions, values, and abilities (i.e., self-evaluation). In some contexts, existing objective standards serve as a reference point. When the objective standard is not available or insufficient, people are likely to consider how other people think or do and compare with them. More than the motive of self-evaluation, people tend to do upward social comparison to improve themselves and downward comparison to enhance a positive sense of self (Festinger, 1954; Helgeson & Mickelson, 1995).

Take the case of a creative artist, Vincent Van Gogh, as an example. Vincent used his teacher’s feedback as a reference point to gauge his learning progress. Also, he intentionally compared with superior others (i.e., upward comparison) to estimate how much he had learned and what he still needed to learn (Johnson & Stapel, 2007; Wheeler, 1966). In addition, Vincent compared with inferior others (downward comparison). When he needed to make decisions about work selling and idea construction, he discerned bad ideas by referring to the people whose works he deemed worse than his own (Bandura & Jourden, 1991). When Vincent felt threatened by other gifted artists’ achievements, he also do downward comparisons to maintain a positive sense of self (Hakmiller, 1966).

Social comparison affects creativity.

Previous creativity research generated empirical evidence regarding the effect of social comparison on creativity. In detail, existing studies examined how the presence (Michinov & Primois, 2005; Redifer et al., 2021; Shalley & Oldham, 1997; Strong & Gray, 1972; Van Knippenberg et al., 1981), the expectations (Clark & Goldsmith, 2006; De Vet & De Dreu, 2007; Van de Ven et al., 2011) and the direction (Van de Ven et al., 2011) of comparative social feedback in ranking competition (Amabile, 1982a; Balietti et al., 2016; Conti et al., 2001; Eisenberg & Thompson, 2011; Landers et al., 2019; Raina, 1968; Van Leeuwen & Baas, 2017) or zero-sum competition (Bittner & Heidemeier, 2013; Erat & Gneezy, 2016) affect divergent thinking, convergent thinking, and creative design or side products of creativity at individual and group levels.

The positive effect of expecting and receiving comparative social feedback seems relatively robust on divergent and convergent thinking tasks at the individual level. For instance, when performing in a “Ask and Guess Test” and an “Imagination Test”, secondary school students who were told that the top three performers were written on the school bulletin board generated more ideas with higher flexibility than the students who were told nothing about the competition (Raina, 1968). When the experimenter asked a group of participants to attack or defend a castle by completing the UUT, the defenders generated more ideas than the attackers and the participants (they knew nothing about the castle game) (Van Leeuwen & Baas, 2017). When controlling the intrinsic motivation, the originality of ideas generated by psychological undergraduates who were told that they were competing against a top ranking was higher than those who were told nothing about the competition (Landers et al., 2019). Business undergraduates who were told their performance would be compared with the other participants generated more ideas with higher flexibility than those who were told nothing about the comparison (Shalley & Oldham, 1997).

In addition, Canadian University’s students and staff who were told that the top three performers would receive cash prizes and that the best performer would be acknowledged generated more creative music than those told nothing about the competition (Eisenberg & Thompson, 2011). When undertaking an interactive drawing task on a computer, the participants whose monetary rewards were based on their performance generated more innovative drawings than those whose monetary rewards were fixed (Balietti et al., 2016). Furthermore, the positive effect of comparative social feedback remains for convergent thinking – CRAT. A correlational study supported this positive association by showing a positive association between benign envy (i.e., a comparative social emotion), the intention to study, and the performance in CRAT (Van de Ven et al., 2011).

More than that, comparative social information also benefits group creativity. For example, university instructors and training consultants were allocated into 3-4 person groups online. They were given five days to express as many ideas as possible about the definitions and examples of groups or teams’ constructions. Compared to the groups that received no information about other group members’ activities, the groups that received information about group members’ contributions produced more ideas with better quality (Michinov & Primois, 2005). The information included 1) the number of ideas each member produced, 2) the number of times they logged in to the platform, and 3) the time they spent on the task.

However, the positive effect of expecting comparative social feedback still has a chance to flip for gender differences. For instance, a group of 6-10-year-old boys were told to design paper collages. Some of them were told that the top three performers would receive a prize (experimental group), and some were told that the prize winner would be raffled off (control group). The results showed that the experimental group created more creative paper collages than the control group (Conti et al., 2001). However, the positive effects of competition on creative product design disappear for girls. In Conti's and Amabile's experiment, the 6-10-year-old girls in the competition group performed less creative than those in the no-competition group (Amabile, 1982a; Conti et al., 2001).

Also, the positive effect of expecting comparative social feedback may disappear when we replace a creative thinking task with a creative problem-solving task - an "in-basket exercise". For example, there was no difference between the performance of undergraduate students from Organisational Behaviour who were told that their performance would be compared with the other participants and those who knew nothing about comparison (Shalley, 1995).

More than that, existing evidence shows that social comparison undermines creativity. For instance, Bittner and Heidemeier recruited German Master's students and allocated them to three groups. A promotion focus group started the experiment by writing the ideal self and developmental goals. A prevention focus group wrote about their ought and responsibilities, and a control group did not write anything. After that, they finished a mindset questionnaire and an instance task. The instance task consisted of five words, and participants indicated the words that had similarities in that they had categories in common (e.g., "compass", "clock", "roadmap", "polar star", "course"). The results showed that the promotion focus group exhibited a less competitive mindset and pointed out more solutions in an instance task than the prevention group. Also, the control group did not differ in mindset but pointed out more solutions in instance task than the prevention group. The following experiment also allocated people into different groups and measured performance on AUT for 10 minutes. The competition group was told that their performance was compared to others, and only the winner received rewards. The cooperation group was told that their performance was added to others' performance, and their summed scores determined the final rewards. The results showed that competition significantly indirectly affected fluency in AUT, mediated by the prevention focus level. In detail, the competition group exhibited a significantly higher prevention focus than the cooperation and control groups and a lower prevention focus associated with higher AUT fluency (Bittner & Heidemeier, 2013).

Consistently, Erat and colleagues found the negative effect of competition on insight problems. In detail, they asked participants to develop a hidden and non-obvious solution for a word/picture puzzle, and they were told that their performance was evaluated. In the control condition, participants were told about the evaluation thing. In the piece-rate condition, one's evaluation outcome determined their received monetary rewards. In the competition condition, one's evaluation was compared with another participants, and only the winner received monetary rewards. The results showed that participants in piece-rate and competition groups put more effort into rebus puzzle problems than control groups but did not perform better. Also, competition groups performed worse than control groups (Erat & Gneezy, 2016).

The above studies examined how the presence and expectation of comparative social feedback affect creativity. Other studies examined how the direction of comparative social feedback (e.g., upward and downward comparison) affects creativity. For example, Redifer and colleagues asked participants to do UUT and metaphor tasks. The researchers provided participants with three kinds of artificial comparative social information in which people saw the average and their own scores. Participants' scores were higher and lower than the average in positive and negative conditions. In the control condition, participants did not receive a score. The results showed that, compared to negative and no feedback, positive feedback for creative thinking tasks led to significantly lower perceived difficulties during the subsequent tasks.

Additionally, although the experiment did not find the effect of feedback on creative performance, they found that lower perceived difficulty was significantly associated with higher fluency scores (rather than originality scores) during subsequent UUT and metaphor tasks (Redifer et al., 2021). Therefore, the results showed that downward comparison lowered the perceived difficulties of the divergent thinking task, which may benefit fluency. Beyond that, Van de Van found that upward comparison benefited performance in the CRAT only when participants feel a possibility to improve and experience benign envy (Van de Ven et al., 2011).

Further studies found that engaging in comparative social information undermines side-products of creativity. For example, receiving a low-ranking position for performance on AUT could reduce one's creative self-efficacy (Strong & Gray, 1972). It also reduce the likability and perceived value of the AUT (Van Knippenberg et al., 1981). For first-year undergraduate students who were sensitive to social comparison, verbalising ideas when completing the UUT undermined the ideas' originality (De Vet & De Dreu, 2007). For undergraduates in the southeast US, the attention to social comparison information negatively affected their creativity in purchasing decisions (Clark & Goldsmith, 2006).

Interpretations for inconsistent findings.

Existing studies yield inconsistent effects of social comparison on creativity. There are at least five potential causes, including inconsistent social comparison manipulation, contaminated monetary rewards, contaminated expected evaluation, inconsistent creativity measures, and other moderators.

First, the studies induced social comparison in different ways. As we can see in the evidence review, some studies induced social comparison with competition, while the others induced ranking. Some studies induced social comparison with expecting feedback while others with receiving feedback. The studies that manipulated social comparison in the same way, may induce different stress levels affected by task instructions or experiment contexts, while stress had a complex effect on creativity (Byron et al., 2010).

Second, the studies employed different creativity tasks. As we have mentioned, different thinking tasks may induce different levels of automatic social cognition. The unknown and complex social cognition may alter the effect of social comparison.

Third, most studies did not dissociate monetary and social rewards while monetary rewards have a separate effect on creativity. In detail, studies that induced participants with expected comparative feedback normally linked the competition outcome or ranking outcome with monetary rewards such as cash (Balietti et al., 2016; Eisenberg & Thompson, 2011; Van Leeuwen & Baas, 2017) and art material rewards (Amabile, 1982a; Conti et al., 2001). However, considering that monetary rewards have a large and complex impact on creativity (Eisenberger et al., 1998; Eisenberger & Rhoades, 2001; Wang & Holahan, 2017), we could not attribute the effect of the stimulus to either social feedback or monetary reward. For example, compared with a person in the lower-ranking position in a contest, a participant who gains a higher-ranking position would receive greater monetary rewards. Therefore, participants may be creative for higher ranking positions or more money.

Fourth, existing studies did not dissociate different social rewards, such as non-comparative evaluation (e.g., an objective score) and comparative evaluation, while both have complex impacts on creativity. Our research focused on the effect of comparative social feedback on creativity. However, there is another research stream showing that expected evaluation affects creativity inconsistently (Amabile et al., 1990; Baer, 1997, 1998b; Hennessey, 1989; King & Gurland, 2007; Lei et al., 2020; LOKŠA & LOKŠOVÁ, 2000; Shalley, 1995; Shalley & Perry-Smith, 2001; Slijkhuis et al., 2013; Wang et al., 2017; Yuan & Zhou, 2008; Zhou & Oldham, 2001). The overall inconsistent results could result from evaluation, comparison, or both.

Even though existing studies induced participants with the pure ranking position without monetary rewards and with a strictly controlling effect of evaluation, the results may still alter in different studies as comparative social feedback would evoke a multitude of psychological mechanisms. Different people may perceive different levels of difficulties in creativity tasks at a cognition level. At a motivation level, expecting or receiving comparative social feedback evokes social comparison spontaneously (Festinger, 1954) In social comparison, people use the information to evaluate their own opinions, beliefs, abilities, or decisions (Bandura & Jourden, 1991), to maintain a good sense of themselves (Hakmiller, 1966), or to motivate themselves to improve (Johnson & Stapel, 2007; Wallace & Gruber, 1989; Wheeler, 1966). At the affection level, engaging in comparative social feedback would elicit admiration, envy (Nabi & Keblusek, 2014; Van de Ven, 2017), the joy of winning, and the fear of losing (Astor et al., 2013). Also, individual creativity would be affected by social comparison sensitivity (Clark & Goldsmith, 2006; De Vet & De Dreu, 2007), the fear of losing out (Cheng & Hong, 2017), and the type of envy (i.e., benign envy and malicious envy) that was recalling (Van de Ven et al., 2011). Also, the regulatory focus (Bittner & Heidemeier, 2013) and the directions of comparison (Van de Ven et al., 2011) may moderate the relationship. Therefore, the evoked psychological mechanisms not only mediate but also moderate the relationship between social comparison and creativity.

Repeated rounds of creativity tasks with social feedback (RRCTSF).

To address the experimental issues in the previous studies, we produced a paradigm (RRCTSF) to examine the effect of zero-sum competition and evaluation on AUT, CRAT, and FST. In RRCTSF, we took out monetary rewards and dissociated evaluation and competition. Also, the manipulation of comparative social feedback remained consistent throughout different creative thinking tasks. Considering that existing research found inconsistent effects of comparative social feedback on individual creative thinking, and this is the first behavioural study that rigorously compares the effect of comparative social feedback on different creative thinking tasks our studies answer two exploratory questions:

First: *how does comparative social feedback affect individual performance in AUT, CRAT, and FST?*

Second: *does comparative social feedback have different effects on individual performance in AUT, CRAT, and FST?*

Study 2 (Pilot) Method

We employed a within-subject design with social feedback, either competition, star rating, or control, as within subject conditions. The dependent variables were speed and accurate fluency in Chinese Compound Remote Associate Problems (CCRAP) (Wu & Chen, 2017). The study was generated on the Qualtrics platform.

Participants.

One hundred and one Mandarin Native speakers (64 women, M_{age} = 29.16 years) were recruited using a convenience sampling strategy. They were all given an information sheet and signed on informed consent.

Materials.

CCRAP. CCRAP integrated CRAT's design, considered the different linguistic attributes of Chinese and English and fitted the language habit of Mandarin speakers. CCRAP consisted of 120 items. Participants saw three stimulus words in each item. They came up with the fourth word that made up an actual two-character word with each of the three stimulus words. For instance, the correct answer of the item: "火(fire)/落(fall)/ 指 (indicate)" was "点 (dot)", since "点 (dot)" made up "点(ignition)/落点(drop-point)/指点 (direction)". We measured accurate fluency (Thissen, 1983) in CCRAP. The accurate fluency in CCRAP indicated participants' convergent thinking. The higher the accurate fluency, the greater the convergent creativity.

We also measured the speed in CCRAP. This dependent variable fulfilled an exploratory purpose because speed could have different indications on the task performance. In detail, the time spent in a task (speed) may be associated with the creative thinking ability, the effort that a participant put into creative thinking, the fatigue that participants experienced, or the typing proficiency in either positive or negative directions.

Procedure.

Participants were told to play a word puzzle game (the word puzzles were CCRAP items). They read instructions, which explained the rules of the experiment. To enter the main experiment, participants had to answer seven questions to ensure they fully understood the instruction. Considering the game's complexity, each participant had three opportunities to answer the questions. The game consisted of 18 rounds, each of which consisted of five CCRAP items. The rounds were shown in random order. The difficulties of items were calculated based on normative data for

CCRAP, and the accumulated difficulties for each round were the same. The questions in each round were random and non-repeated. These 18 rounds were divided into three blocks, each representing one condition (i.e., six rounds per block/condition).

In the control condition, participants were told that they were playing independently; they received the number of correct answers after each round. In the competition condition, participants were told that they were competing with another player randomly selected from our participant pool. The competitors differed in each round; they received feedback informing the number of correct answers and the competition winner after each round. In the star rating conditions, participants were told that a star rating system would evaluate their performance. The system considered participants' results, other players' results, and the difficulties of questions; they received feedback informing the number of correct answers and the star rating at the end of each round. The information about the number of correct answers reflected participants' true performance, while the competition's winner was selected randomly, and the star ratings were given randomly.

Participants encountered three conditions in random order. They had 2 minutes to complete each round, but they did not have to run out of time. At the end of the experiment, we asked for gender and age. They left their email address if they wanted us to update the correct answer for all CCRAP items.

Study 2 (Pilot) Results

Most participants did not finish the experiment in the pilot study. As a result, we did not have sufficient power in the pilot, and none of the dependent variables showed a significant result. However, participants provided valuable comments on the RRCTSF paradigm in terms of length and difficulties of the word puzzle items. For example, they reported the paradigm's limitation, which helped us to generate an efficient experiment in the next study. Specifically, participants' feedback centred around five things. First, five items per round were too lengthy and too difficult. Second, two minutes gave them too much time pressure, and they could not finish the test. Third, the star- rating system in the star rating condition was not trustworthy. Fourth, winner selection did not make sense since we did not provide them with the competitor's number of correct answers. Finally, five participants stopped the experiment in the middle of the game and continued several hours later. The pause would easily make our manipulation invalid. To overcome these limitations, we adjusted the procedure in the following studies.

Study 2 Method

We replicated the experimental design in the pilot study but changed the materials (i.e., replaced CCRAP with CRAT) and adjusted procedures.

Participants.

One hundred and one English Native speakers (43 women, M_{age} = 37.43 years) were recruited on Mturk using a simple random sampling strategy. All participants were workers on Mturk and were recruited based on the principle - first come, first serve. Participants were all given an information sheet and signed on informed consent. They received $4.09 as a participant fee.

Post-hoc power analyses using GPower suggested that we had enough power (.98) to detect the effect at the p = .05 level (*partial* η^2 = .05, *effect size f* = .23, *number of groups* = 1, *number of measurements* = 3, *corr among rep measures* = .5, *nonsphericity correction* ε = 1).

Materials.

CRAT. In CRAT, participants saw three stimulus words and came up with the fourth word that made up a common compound word or phrase with each of the three stimulus words. For instance, if they saw “cottage/Swiss/cake”, they should come up with the fourth word “cheese” since it is made up of “cottage cheese/Swiss cheese/cheesecake”. We replicated the measurement in the pilot study.

Motives of social comparison. We measured social comparison motives using a well-known questionnaire (Helgeson & Mickelson, 1995). This inventory included six sub-scales which focused on the motives for self-improvement, motives for a common bond, motives for altruism, self-enhancement, self-destruction, and self-evaluation, respectively (see Appendix G).

Individualism and collectivism. We were interested in the relationship between ideologies on creative thinking, so we measured individualism and collectivism using a well-validated questionnaire (Triandis & Gelfand, 1998). This inventory included four sub-scales; one focused on horizontal individualism— in which people were seeking a unique position in a group —and included items such as "*I would rather depend on myself than others*". Second, one focus on vertical individualism — in which people wanted high status and were aggressive in competitions—included items such as "*Winning is everything*". Third, one focused on horizontal collectivism —in which people viewed themselves as being similar to others — and included items such as "*I feel good when I cooperate with others*". Finally, one focused on vertical collectivism — in which people were willing to sacrifice their benefit for the sake of in-group objectives — and included items such as "*Parents and children must stay together as much as possible*" (see Appendix H). The reliabilities for both scales were acceptable (see Table 16).

We measured individualism and collectivism in and only in study 2 since the author was interested in the link between culture and creativity at the time of experiment conduction (i.e., personal exploratory interests). The relationship between cultural background and creative thinking was not the focus of this research project and does not fit well with research logic holistically. Therefore, we reported relevant results but would not discuss them in further detail.

Table 16

Scale Reliability for Motives of Social Comparison and Individualism & Collectivism.

	Cronbach's α	N of Items
Self-improvement	.81	5
Self enhancement	.87	5
Self-evaluation	.78	3
Common bond	.79	4
Horizontal individualism	.64	4
Vertical individualism	.74	4
Individualism	.72	8
Horizontal collectivism	.73	4
Vertical collectivism	.72	4
Collectivism	.79	8

Procedure.

We replicated the procedure in the pilot study. However, considering the limitations of the pilot study, we made some changes to improve the efficiency of this paradigm (please see Appendix I for *Study 2 Script*) In detail:

1. There were three items in one round.

2. Time limits for each round were 3 minutes.

3. The star rating was given based on the performance (please see Appendix J for *Scoring and Star Rating Algorithm for CRAT (Study 2)*).

4. We added a competitor allocation page before each round of the competition block. The allocation page showed “we are allocating you a competitor”, and the page lasted 3 seconds.

5. The feedback of the competition comprised both the winner and the number of correct answers of two players (the correct number of correct answers of competitors was set randomly).

6. We emphasised that participants could not stop in the middle of each block.

7. We added two questionnaires at the end of the experiment. The questionnaires measured cultural background and motivations of social comparison.

8. In the debriefing script, which was used to reveal the study’s true purpose, we asked participants if they suspected that we were not being completely honest with them (57 participants did not suspect our study, and 44 of them suspected our honesty).

Study 2 Results

Statistical assumptions. To prepare an appropriate dataset for within subject repeated measures ANOVA, we did a series of assumption tests on the dependent variables. All the variables met the assumption of independence of variables, and sphericity. There was no outlier. The skewness (between ± 2) & kurtosis (between ±3), and Shapiro–Wilk test indicated that the data contained approximately normally distributed error. Please see Table 17 for descriptive statistics and Table 18 for normality tests of all measures.

Table 17

Descriptive Statistics for Measures of Speed, Accuracy, Individualism, Collectivism, and Social Comparison Motivations.

	M	*SD*
Speed _ Competition	47.01	30.38
Speed _ Star rating	48.85	30.17
Speed _ Control	43.40	25.44
Accurate fluency _ Competition	12.78	3.83
Accurate fluency _ Star rating	12.80	3.85
Accurate fluency _ Control	12.36	3.91
Self-improvement	3.98	0.80
Self enhancement	2.96	1.06
Self-evaluation	3.94	0.88
Common bond	3.57	0.93
Horizontal individualism	4.141	0.648
Vertical individualism	2.965	0.952
Individualism	3.553	0.644
Collectivism _ Horizontal	3.688	0.770
Collectivism _ Vertical	3.507	0.820
Collectivism	3.598	0.679

Table 18

Normality Tests for Measures of Speed, Accuracy, Individualism, Collectivism, and Social Comparison Motivations.

			Shapiro-Wilk	
	Skewness	Kurtosis	Statistic	*p* value
Speed _ Competition	1.62	2.91	.86**	<.001
Speed _ Star rating	1.07	0.62	.90**	<.001
Speed _ Control	1.33	2.30	.90**	<.001
Accurate fluency _ Competition	-0.45	-0.52	.95**	.001
Accurate fluency _ Star rating	-0.43	-0.60	.95**	<.001
Accurate fluency _ Control	-0.49	-0.31	.96**	.004
Self-improvement	-1.40	2.46	.89**	<.001
Self enhancement	-0.47	-0.54	.94**	<.001
Self-evaluation	-1.66	3.47	.82**	<.001
Common bond	-0.55	-0.03	.96**	.004
Horizontal individualism	-0.72	0.50	.94**	<.001
Vertical individualism	0.06	-0.32	.98	.155
Individualism	-0.05	-0.44	.99	.548
Collectivism _ Horizontal	-0.72	0.82	.94**	<.001
Collectivism _ Vertical	-0.37	0.32	.97*	.032
Collectivism	-0.82	1.67	.96**	.005

*p < .05 (2-tailed). **p < .01 (2-tailed).

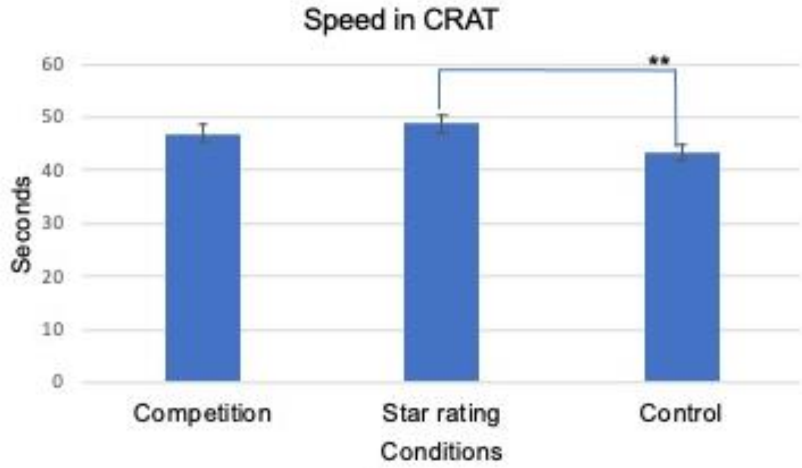
CRAT ANOVA. We performed three-way repeated measures ANOVA on the speed to determine whether participants spent more or less time in different conditions. There was a significant main effect of *speed*, F (1, 100) = 5.32, p = .006, η^2= .05, Ω = .83. The following-up t-test[15] showed that participants spent more time, t (100) = 3.06, p = .009, *95% CI* [1.11, 9.80], in the star rating condition (M = 48.85, SD = 3.00) than control condition (M = 43.40, SD = 2.53). However, the speed did not reveal significant differences between the competition (M = 47.01, SD = 3.02) and the star rating conditions, t (100) = -1.04, p = .904, *95% CI* [-6.16, 2.48], and

[15] *p* values in this thesis were Bonferroni corrected.

between the competition and control conditions, t (100) = 2.35, p = .062, *95% CI* [-0.12, 7.35] (see Figure 10). Also, the main effect did not reveal a significant difference for CRAT accurate fluency, F (1, 100) = 1.54, p = .218, η^2= .02, Ω = .32.

Figure 10

CRAT speed in competition, star rating, and control conditions.



Correlations of CRAT measures. We correlated speed and CRAT accurate fluency. The results showed significant positive correlations between speed and accurate fluency in the star rating and control conditions. In detail, speed and accurate fluency was weakly associated in star rating condition, r = .26, p = .009, and in control condition, r = .20 p = .045. However, speed and accurate fluency were not significantly associated in the competition condition.

Correlations with self-reported measures. The results showed that there were significant positive associations between motives of self-improvement and CRAT accurate fluency. In detail, the motive of self-improvement was weakly associated with accurate fluency in the competition and star rating conditions (see Table 19).

Table 19

Correlations Cross CRAT Speed, Accurate Fluency, and Motives of Social Comparison.

	Self-improvement	Self enhancement	Self-evaluation	Common bond
Speed _ Competition	.15	.10	.04	.11
Speed _ Star rating	.14	.02	-.06	.06
Speed _ Control	.13	.11	.00	.12
Accurate fluency _ Competition	.26**	-.05	.20	.06
Accurate fluency _ Star rating	.21*	-.07	.16	.08
Accurate fluency _ Control	.12	-.02	.13	.05

**p* < .05 (2-tailed). ***p* < .01 (2-tailed).

The results also showed significant positive associations between individualism and accurate fluency in the star rating condition. In detail, the vertical individualism and the average of horizontal and vertical individualism were weakly associated with accurate fluency in the star rating condition (see Table 20).

Table 20

Correlations Cross CRAT Speed, Accuracy, and Individualism.

	Horizontal individualism	Vertical individualism	Individualism
Speed _ Competition	.08	-.05	.00
Speed _ Star rating	.17	-.05	.05
Speed _ Control	.07	-.07	-.02
Accurate fluency _ Competition	.07	.16	.15
Accurate fluency _ Star rating	.14	.25*	.26**
Accurate fluency _ Control	.08	.11	.12

**p* < .05 (2-tailed). ***p* < .01 (2-tailed).

The results did not show a significant association between collectivism and CRAT accurate fluency (see Table 21).

Table 21

Correlations Cross CRAT Speed, Accuracy, and Collectivism.

	Horizontal collectivism	Vertical collectivism	Collectivism
Speed _ Competition	.07	.08	.09
Speed _ Star rating	.09	.05	.08
Speed _ Control	.12	.18	.18
Accurate fluency _ Competition	-.01	-.05	-.04
Accurate fluency _ Star rating	-.04	.02	-.01
Accurate fluency _ Control	.11	.10	.12

**p* < .05 (2-tailed). ***p* < .01 (2-tailed).

Exploratory analysis on gender difference. The results of the independent samples test showed that speed and accurate fluency in CRAT did not reveal a significant difference between female and male participants (see Table 22).

Table 22

Gender Difference in CRAT Measure (N_{female} = 43, N_{male} = 58, df = 99).

Measures	Gender	*M*	*SD*	*t*	*p* values	95% CI Lower	95% CI Upper
Speed _ Competition	Female	49.29	33.41	0.65	0.518	-8.19	16.14
	Male	45.32	28.09				
Speed _ Star rating	Female	50.88	32.23	0.58	0.564	-8.56	15.61
	Male	47.35	28.74				
Speed _ Control	Female	41.97	23.47	-0.49	0.629	-12.69	7.71
	Male	44.46	26.96				
Accurate fluency _ Competition	Female	12.70	4.14	-0.19	0.85	-1.68	1.39
	Male	12.84	3.62				
Accurate fluency _ Star rating	Female	13.07	3.78	0.60	0.55	-1.07	2.01
	Male	12.60	3.91				
Accurate fluency _ Control	Female	12.44	4.01	0.19	0.85	-1.42	1.72
	Male	12.29	3.87				

Study 3 Methods

We replicated the experimental design in Study 2 but changed the materials (i.e., replaced CRAT with AUT) and adjusted the procedure.

Participants.

We recruited 64 English Native speakers (25 females, M_{age} = 36.61, SD_{age} = 10.62) from Mturk using a simple random sampling strategy and a first come, first serve principle. Participants were all given an information sheet and signed on informed consent. They received $4.09 as a participant fee.

Using GPower, we conducted a priori power analyses based on the results in Study 1. The results suggested that a sample size of 55 could offer enough power (.84) to detect the effect at the p = .05 level (*partial* η^2 = .05, *effect size f* = .23, *number of groups* = 1, *number of measurements* = 3, *α err prob* = .05, *power* = .83, *corr among rep measures* = .20, *nonsphericity correction* ε = 1).

Materials.

AUT. As we did in Study 1, we asked participants to think of as many uses of household objects as possible in six rounds. We measured appropriate fluency, appropriate originality, and appropriate frequency. We did not measure appropriate flexibility since previous studies showed that it was positively associated with appropriate fluency. For the details of the measuring strategy, please refer to AUT in Materials in Study 1.

Procedure.

The procedure was replicated in Study 2 (see Appendix K for *Study 3 script*). Participants read AUT instructions and had to pass the instruction test to enter the main experiment. The experiment consisted of 6 rounds, each of which consisted of one round of AUT. The six rounds were divided into three blocks, representing one condition (i.e., two rounds per block/condition). The manipulations of the competition, star-rating, and control conditions remained almost the same as in Study 2, with reasonable changes. First, we added a competitor allocation page to make the competition more trustworthy. Also, we offered participants 5 minutes to finish each round as AUT may require longer typing time than CRAT. In addition, participants received subject scores rather than the number of correct answers in the feedback screens (please see Appendix L for *Scoring and Star Rating Algorithm for AUT (Study 3)*). Participants encountered three conditions randomly, and the six different

AUT objects were randomly matched with conditions. For instance, the probabilities for brick to appear in competition, star-rating, and control conditions were the same. After the main experiment and the self-report questionnaires for motives of social comparison and individualism & collectivism (see Materials in Study 2), participants reported demographic information and received a debriefing statement. Please see Table 23 for the sale reliability of self-report questions.

Table 23

Scale Reliability for Motives of Social Comparison and Individualism and Collectivism.

	Cronbach's α	N of Items
Self-improvement	.82	5
Self enhancement	.86	5
Self-evaluation	.66	3
Common bond	.78	4
Horizontal individualism	.74	4
Vertical individualism	.80	4
Individualism	.74	8
Horizontal collectivism	.73	4
Vertical collectivism	.82	4
Collectivism	.83	8

Study 3 Results

Inter-rater reliability. We provided marking instruction and training to two judges: one each from psychology and medical science backgrounds. Following training, we asked judges to ask mark the originality an answer pool. Inter-rater reliability was found between two judges for all markings (*Cronbach's α* = .89)**.**

Statistical assumptions. To prepare an appropriate dataset for within subject repeated measures ANOVA, we did a series of assumption tests on the dependent variables. All the variables met the assumption of independence of variables, and sphericity. There was no outlier. The skewness (between ± 2) & kurtosis (between ±10) indicated that the data contained approximately normally distributed error. Please see Table 24 for descriptive statistics and Table 25 for normality tests of all measures.

Table 24

Descriptive Statistics for AUT Measures of Speed, Accuracy, Individualism, Collectivism, and Motives of Social Comparison.

	M	*SD*
Speed _ Competition	119.11	68.77
Speed _ Star rating	121.13	72.88
Speed _ Control	115.53	62.43
Appropriate fluency _ Competition	8.58	3.33
Appropriate fluency _ Star rating	8.53	3.28
Appropriate fluency _ Control	8.42	3.45
Appropriate originality _ Competition	1.93	0.21
Appropriate originality _ Star rating	1.91	0.16
Appropriate originality _ Control	1.90	0.19
Appropriate frequency _ Competition	0.99	0.24
Appropriate frequency _ Star rating	0.96	0.25
Appropriate frequency _ Control	0.96	0.24
Self-improvement	4.16	0.67
Self enhancement	3.23	1.00
Common bound	3.82	0.82
Self-evaluation	3.97	0.74
Horizontal individualism	3.82	0.71
Vertical individualism	3.04	0.99
Individualism	3.43	0.67
Horizontal collectivism	3.93	0.71
Vertical collectivism	3.71	0.82
Collectivism	3.82	0.67

Table 25

Normality Tests for AUT Measures of Speed, Accuracy, Individualism, Collectivism, and Motives of Social Comparison.

			Shapiro-Wilk	
	Skewness	Kurtosis	Statistic	*p* value
Speed _ Competition	0.86	-0.05	.92**	<.001
Speed _ Star rating	0.77	-0.45	.92**	<.001
Speed _ Control	0.76	0.04	.94**	.005
Appropriate fluency _ Competition	0.27	-0.91	.96*	.039
Appropriate fluency _ Star rating	0.39	-0.89	.95*	.011
Appropriate fluency _ Control	0.37	-0.91	.95*	.015
Appropriate originality _ Competition	-0.28	-0.11	.99	.854
Appropriate originality _ Star rating	-0.08	0.54	.99	.896
Appropriate originality _ Control	-0.38	-0.10	.99	.671
Appropriate frequency _ Competition	0.05	0.12	.99	.911
Appropriate frequency _ Star rating	0.42	2.97	.95*	.011
Appropriate frequency _ Control	0.31	0.11	.99	.748
Self-improvement	-1.82	6.83	.85**	<.001
Self enhancement	-0.95	0.22	.90**	<.001
Common bound	-1.04	1.07	.91**	<.001
Self-evaluation	-1.12	1.30	.89**	<.001
Horizontal individualism	-0.84	1.89	.95**	.006
Vertical individualism	-0.12	-0.89	.97	.098
Individualism	-0.75	1.58	.96*	.046
Horizontal collectivism	-1.55	2.98	.85**	<.001
Vertical collectivism	-1.18	1.30	.89**	<.001
Collectivism	-1.56	3.54	.88**	<.001

*p < .05 (2-tailed). **p < .01 (2-tailed).

AUT ANOVA. We performed a three-way (competition vs star-rating vs control) repeated measures ANOVA measures. None of AUT measures - speed F (1, 63) = 0.46, p = .633, η^2= .01, Ω = .12, appropriate fluency F (1, 63) = 0.25, p = .783, η^2= .00, Ω = .09, appropriate originality, F (1, 63) = 1.13, p = .325, η^2= .02, Ω = .25, and appropriate frequency, F (1, 63) = 0.85, p = .431, η^2= .01, Ω = .19, revealed a significant difference in the three conditions.

Correlations of AUT measures. The results showed that there were significant positive associations among measures for AUT. In the competition condition, appropriate fluency was moderately correlated with speed, and appropriate frequency was weakly associated with speed, appropriate fluency, and appropriate originality (see Table 26). In the star rating condition, appropriate fluency was weakly associated with speed. Appropriate originality was weakly associated with speed and moderately associated with appropriate originality. Also, the appropriate frequency was moderately associated with appropriate fluency (see Table 27). In the control condition, there were weak or moderate correlations between the two AUT measures (see Table 28).

Table 26

Correlations for AUT Measures in Competition Condition.

	1	2	3	4
Speed	-			
Appropriate fluency	.53**	-		
Appropriate originality	.24	.21	-	
Appropriate frequency	.27**	.26**	.25**	-

*p < .05 (2-tailed). **p < .01 (2-tailed).

Table 27

Correlations for AUT Measures in Star Rating Condition.

	1	2	3	4
Speed	-			
Appropriate fluency	.40**	-		
Appropriate originality	.33**	.44**	-	
Appropriate frequency	.17	.44**	.23	-

*p < .05 (2-tailed). **p < .01 (2-tailed).

Table 28

Correlations for AUT Measures in Control Condition.

	1	2	3	4
Speed	-			
Appropriate fluency	.45**	-		
Appropriate originality	.46**	.45**	-	
Appropriate frequency	33**	.31**	.41**	-

*p < .05 (2-tailed). **p < .01 (2-tailed).

Correlations with self-reported measures. The results showed a significant negative association between motives of self-improvement and AUT measures. In detail, the motive of self-enhancement was weakly associated with appropriate frequency in the competition condition (see Table 29).

Table 29

Correlations Cross AUT Measures and Motives of Social Comparison.

	Self-improvement	Self enhancement	Common bound	Self-evaluation
Speed _ Competition	.22	-.03	.16	.02
Speed _ Star rating	.11	.00	.01	.00
Speed _ Control	.05	.05	-.05	.02
Appropriate fluency _ Competition	.02	-.16	.08	.03
Appropriate fluency _ Star rating	-.05	-.18	.06	-.05
Appropriate fluency _ Control	.03	-.11	.12	-.05
Appropriate originality _ Competition	-.13	-.12	-.15	-.03
Appropriate originality _ Star rating	-.09	-.04	-.19	-.08
Appropriate originality _ Control	-.13	-.16	-.01	.02
Appropriate frequency _ Competition	.07	-.28*	.03	.01
Appropriate frequency _ Star rating	-.20	-.08	.07	-.20
Appropriate frequency _ Control	.06	-.18	.15	.01

**p* < .05 (2-tailed). ***p* < .01 (2-tailed).

The results also showed significant negative associations between individualism and AUT measures. In detail, horizontal individualism was weakly associated with appropriate originality in the control condition, and vertical individualism was weakly associated with appropriate frequency in the star rating condition. Also, average individualism was weakly associated with appropriate frequency in both star rating and control conditions (see Table 30).

Table 30

Correlations Cross AUT Measures and Individualism.

	Horizontal individualism	Vertical individualism	Individualism
Speed _ Competition	.10	-.02	.04
Speed _ Star rating	.04	.10	.09
Speed _ Control	.06	.18	.17
Appropriate fluency _ Competition	-.02	-.21	-.17
Appropriate fluency _ Star rating	-.12	-.17	-.19
Appropriate fluency _ Control	-.09	-.15	-.16
Appropriate originality _ Competition	-.02	-.17	-.13
Appropriate originality _ Star rating	.01	.04	.04
Appropriate originality _ Control	-.30*	-.08	-.22
Appropriate frequency _ Competition	-.04	-.24	-.20
Appropriate frequency _ Star rating	-.10	-.28*	-.26*
Appropriate frequency _ Control	-.23	-.18	-.25*

*p < .05 (2-tailed). **p < .01 (2-tailed).

The results also showed significant negative associations between collectivism and AUT measures. In detail, vertical collectivism was weakly associated with appropriate originality in the competition condition, and average collectivism was weakly associated with appropriate originality in the competition condition and appropriate frequency in the competition condition (see Table 31).

Table 31

Correlations Cross AUT Measures and Collectivism.

	Horizontal collectivism	Vertical collectivism	Collectivism
Speed _ Competition	.10	.05	.08
Speed _ Star rating	-.02	.07	.03
Speed _ Control	-.10	-.06	-.09
Appropriate fluency _ Competition	.08	.02	.05
Appropriate fluency _ Star rating	-.09	-.06	-.08
Appropriate fluency _ Control	-.04	-.09	-.07
Appropriate originality _ Competition	-.17	-.29*	-.27*
Appropriate originality _ Star rating	-.02	-.02	-.02
Appropriate originality _ Control	-.10	-.10	-.11
Appropriate frequency _ Competition	-.23	-.22	-.25*
Appropriate frequency _ Star rating	-.17	-.05	-.12
Appropriate frequency _ Control	-.17	-.23	-.23

*p < .05 (2-tailed). **p < .01 (2-tailed).

Exploratory analysis on gender difference. The results of the independent samples test showed that none of the AUT measures revealed a significant difference between female and male participants (see Table 32).

Table 32

Gender Difference in AUT Measure (N_{female} = 25, N_{male} = 39, df = 62).

Measures	Gender	*M*	*SD*	*t*	*p* values	95% CI Lower	95% CI Upper
Speed _ Competition	Female	134.35	77.25	1.43	0.157	-9.92	59.94
	Male	109.34	61.80				
Speed _ Star rating	Female	120.76	78.29	-0.03	0.974	-38.23	37.02
	Male	121.37	70.25				
Speed _ Control	Female	108.80	59.04	-0.69	0.494	-43.16	21.06
	Male	119.85	64.89				
Appropriate fluency _ Competition	Female	9.54	3.45	1.89	0.064	-0.10	3.25
	Male	7.96	3.15				
Appropriate fluency _ Star rating	Female	9.32	3.31	1.56	0.125	-0.37	2.96
	Male	8.03	3.21				
Appropriate fluency _ Control	Female	8.99	3.33	1.06	0.292	-0.83	2.71
	Male	8.05	3.52				
Appropriate frequency _ Competition	Female	1.01	0.24	0.36	0.722	-0.10	0.15
	Male	0.99	0.25				
Appropriate frequency _ Star rating	Female	0.94	0.30	-0.42	0.679	-0.16	0.10
	Male	0.97	0.22				
Appropriate frequency _ Control	Female	0.94	0.24	-0.55	0.582	-0.16	0.09
	Male	0.97	0.24				
Appropriate originality _ Competition	Female	1.87	0.20	-1.69	0.096	-0.20	0.02
	Male	1.96	0.21				
Appropriate originality _ Star rating	Female	1.95	0.17	1.55	0.126	-0.02	0.14
	Male	1.89	0.14				
Appropriate originality _ Control	Female	1.89	0.18	-0.33	0.742	-0.12	0.08
	Male	1.90	0.20				

Study 4 Method

We replicated the experiment design in Study 2 but changed the materials (i.e., replaced CRAT with FST) and adjusted the procedure.

Participants.

We recruited 52 participants (21 females, M_{age} = 35.54, SD_{age} = 11.17) from Mturk using a simple random sampling strategy and a first come, first serve principle. Participants were all given an information sheet and signed on informed consent. They received $4.09 as a participant fee.

According to a priori power analysis in Study 3, we needed 55 participants to achieve enough power (.84). We also did post hoc power analyses based on the significant results in this study. The results suggested that 52 participants offered enough power (*partial* η^2 = .09, *effect size f* = .31, *number of groups* = 1, *number of measurements* = 3, *corr among rep measures* = .50, *nonsphericity correction* ε = 1).

Materials.

FST. As we did in Study 1, we provided participants with three functions and asked them to think of as many objects as possible that fulfilled all three functions. We measured fluency, flexibility, originality, frequency, and effectiveness in appropriate and accurate dimensions. For the details of the measuring strategy, please refer to FST in Study 1 Materials. Also, we were interested in the number of answers produced (all fluency). The more answers one came up with, the more effort a participant invested into the task and the higher potential they had for original ideas (Runco, 2010).

Creative self-efficacy. We employed the 28-statement creative self-efficacy inventory. Please refer to the creative self-efficacy inventory in Materials in Study 1 (*Cronbach's* α = .95).

Procedure.

We replicated the procedure in Study 2 and Study 3 (see Appendix M for *Study 4 Script*). Participants read the FST instruction and had to pass the instruction test to enter the main experiment. Like Study 3, the experiment consisted of six rounds, each of which consisted of one round of FST. We divided them into three blocks, each representing one condition. The manipulations of the competition, star-rating, and control condition remained almost the same as in Study 3, with reasonable changes (see Appendix N for *Scoring and Star Rating Algorithm for FST*

(Study 4)). For example, we offered participants 7 minutes to finish each round as FST may require longer typing time than AUT and CRAT. Also, our study aim was not to examine the effect of speed but to allow participants to think as creatively as possible - the creativity-encouraged mindset may require a longer time. Participants encountered three conditions in a random order in the main experiment. Afterwards, they did questionnaires for creative self-efficacy, reported demographic information, and received a debriefing statement.

Study 4 Results

Inter-rater reliability. We provided marking instruction and training to two judges from psychology and engineering backgrounds. The inter-rater reliability for all measures was acceptable (see Table 33) (*Cronbach's α* > .60). Therefore, none of the measures was excluded.

Table 33

Intraclass Correlation Coefficients (ICC) with Two-way Mixed Effects Model for Measures in FST (N of Items = 2).

		95% *CI*	
Measures	*ICC*	Lower Bound	Upper Bound
FST appropriate fluency	.94	.95	.98
FST appropriate flexibility	.92	.88	.99
FST appropriate originality	.91	.84	.99
FST appropriate effectiveness	.91	.86	.98
FST accurate fluency	.78	.37	.93
FST accurate flexibility	.76	.27	.92
FST accurate originality	.79	.60	.97
FST accurate effectiveness	.82	.72	.96

a. Cronbach's *α* is the average measures of intraclass correlation coefficients.

b. Cronbach's *α* using an absolute agreement definition.

Statistical assumptions. To prepare an appropriate dataset for within subject repeated measures ANOVA, we did a series of assumption tests on the dependent variables. All the variables met the assumption of independence of variables, and sphericity. There was no outlier. The skewness (between ± 3) & kurtosis (between ±7) indicated that the data contained approximately normally distributed error. Please see Table 34 for descriptive statistics and Table 35 for normality tests of all measures.

Table 34

Descriptive Statistics for FST Measures, Individualism, Collectivism, and Motives of Social Comparison.

	M	*SD*
Speed _ Competition	167.43	83.81
Speed _ Evaluation	169.44	84.80
Speed _ Control	158.57	87.40
All fluency _ Competition	11.33	5.67
All fluency _ Star rating	10.06	5.11
All fluency _ Control	10.34	5.00
Appropriate fluency _ Competition	4.17	2.55
Appropriate fluency _ Star rating	3.74	2.08
Appropriate fluency _ Control	4.35	2.49
Appropriate flexibility _ Competition	2.57	1.56
Appropriate flexibility _ Star rating	2.26	1.12
Appropriate flexibility _ Control	2.56	1.39
Appropriate originality _ Competition	1.33	0.54
Appropriate originality _ Star rating	1.27	0.45
Appropriate originality _ Control	1.36	0.38
Appropriate frequency _ Competition	2.21	0.76
Appropriate frequency _ Star rating	2.17	0.69
Appropriate frequency _ Control	2.30	0.54
Appropriate effectiveness _ Competition	1.32	0.47
Appropriate effectiveness _ Star rating	1.35	0.48
Appropriate effectiveness _ Control	1.45	0.41

Accurate fluency _ Competition	1.62	1.61
Accurate fluency _ Star rating	1.39	1.38
Accurate fluency _ Control	1.48	1.61
Accurate flexibility _ Competition	1.04	0.89
Accurate flexibility _ Star rating	0.85	0.70
Accurate flexibility _ Control	0.81	0.78
Accurate originality _ Competition	0.81	0.60
Accurate originality _ Star rating	0.69	0.47
Accurate originality _ Control	0.66	0.56
Accurate frequency _ Competition	1.31	0.87
Accurate frequency _ Star rating	1.18	0.77
Accurate frequency _ Control	1.17	0.90
Accurate effectiveness _ Competition	0.85	0.59
Accurate effectiveness _ Star rating	0.85	0.55
Accurate effectiveness _ Control	0.77	0.57
Creative self-efficacy	62.72	14.79

Table 35

Normality Tests for FST Measures, Individualism, Collectivism, and Motives of Social Comparison.

			Shapiro-Wilk	
	Skewness	Kurtosis	Statistic	*p* value
Speed _ Competition	0.26	-1.26	.93**	.003
Speed _ Evaluation	0.15	-1.31	.93**	.004
Speed _ Control	0.34	-1.22	.92**	.001
All fluency _ Competition	0.48	-1.25	.87**	<.001
All fluency _ Star rating	0.96	-0.25	.85**	<.001
All fluency _ Control	0.67	-0.48	.97	.170
Appropriate fluency _ Competition	0.63	0.37	.92**	.002
Appropriate fluency _ Star rating	0.62	0.86	.96	.097
Appropriate fluency _ Control	0.51	-0.44	.97	.228
Appropriate flexibility _ Competition	0.67	0.92	.92**	.003
Appropriate flexibility _ Star rating	0.28	0.35	.95*	.022
Appropriate flexibility _ Control	1.04	1.65	.96	.064
Appropriate originality _ Competition	-0.69	-0.07	.94*	.010
Appropriate originality _ Star rating	-0.62	-0.14	.88**	<.001
Appropriate originality _ Control	-1.07	2.20	.91**	.001
Appropriate frequency _ Competition	-1.26	0.76	.90**	<.001
Appropriate frequency _ Star rating	-1.11	0.95	.90**	<.001
Appropriate frequency _ Control	-1.86	5.27	.92**	.002
Appropriate effectiveness _ Competition	-1.09	0.55	.95*	.039
Appropriate effectiveness _ Star rating	-0.87	0.16	.88**	<.001
Appropriate effectiveness _ Control	-0.92	1.58	.89**	<.001
Accurate fluency _ Competition	2.01	6.34	.89**	<.001
Accurate fluency _ Star rating	1.51	2.39	.89**	<.001
Accurate fluency _ Control	1.38	1.33	.91**	.001
Accurate flexibility _ Competition	0.88	-0.27	.88**	<.001
Accurate flexibility _ Star rating	0.86	0.11	.92**	.002
Accurate flexibility _ Control	1.11	0.94	.96*	.047

Accurate originality _ Competition	0.46	-1.04	.92**	.002
Accurate originality _ Star rating	0.28	-0.64	.94**	.007
Accurate originality _ Control	0.66	-0.32	.95*	.037
Accurate frequency _ Competition	0.35	-1.01	.91**	.001
Accurate frequency _ Star rating	0.08	-0.77	.94*	.013
Accurate frequency _ Control	0.27	-1.02	.96	.099
Accurate effectiveness _ Competition	0.61	-0.28	.93**	.007
Accurate effectiveness _ Star rating	0.11	-0.66	.97	.169
Accurate effectiveness _ Control	0.17	-0.98	.96	.091
Creative self-efficacy	-0.03	-0.51	.99	.918

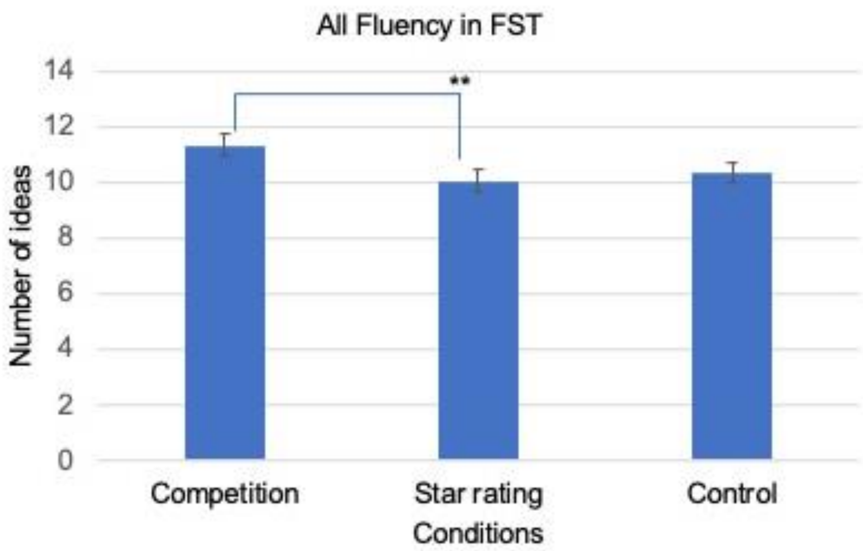
$^{*}p < .05$ (2-tailed). $^{**}p < .01$ (2-tailed).

FST ANOVA. We performed a three-way (competition vs star-rating vs control) repeated measures ANOVA on speed, fluency, flexibility, originality, and frequency in appropriate and accurate dimensions. There was a significant main effect of *all fluency*, F (1, 51) = 5.05, p = .008, η^2 = .09, Ω = .81. The following-up t-test[16] showed that participants produced more ideas, t (51) = 2.93, p = .015, *95% CI* [0.20, 2.34], in competition condition (M = 11.33, SD = 5.67) than star rating condition (M = 10.06, SD = 5.11). All fluency in control condition (M = 10.34, SD = .69) did not significantly differ from competition t (51) = 2.30, p = .077, *95% CI* [-0.08, 2.06], and star rating conditions t (51) = -0.71, p = 1.000, *95% CI* [-1.26, 0.08] (see Figure 11). None of the other measures – speed, F (1, 51) = 0.98, p = .381, η^2 = .02, Ω = .22, appropriate fluency, F (1, 51) = 1.95, p = .148, η^2 = .04, Ω = .40, appropriate flexibility, F (1, 51) = 1.45, p = .238, η^2 = .03, Ω = .31, appropriate originality, F (1, 51) = 0.73, p = .486, η^2 = .01, Ω = .17, appropriate frequency, F (1, 51) = 0.76, p = .47, η^2 = .02, Ω = .18, appropriate effectiveness, F (1, 51) = 1.97, p = .145, η^2 = .04, Ω = .40, revealed a significant difference in the three conditions.

[16] *p* values were Bonferroni corrected.

Figure 11

FST fluency in competition, star rating, and control conditions.



Correlations of FST measures. The results showed that, in the competition condition, there were significant positive associations between each FST measure in appropriate and accurate dimensions. Also, the time spent doing FST (speed) was weakly associated with all fluency. However, all fluency and speed were not associated with other FST measures (see Table 36).

In the star rating condition, there were significant positive associations between each FST measure in appropriate and accurate dimensions. However, accurate fluency was not associated with appropriate flexibility, originality, frequency, and effectiveness, and accurate flexibility was not associated with appropriate originality. In addition, all fluency was negatively associated with appropriate effectiveness but not the other FST measures. Finally, speed was not associated with any of the FST measures.

In the control condition, there were significant positive associations between each two FST measures in appropriate and accurate dimensions, except that accurate fluency was not associated with appropriate originality and effectiveness. Also, all fluency was positively associated with the other FST measures, such that there was a moderate correlation between all fluency and appropriate fluency. All fluency was also weakly correlated with accurate fluency, frequency, and effectiveness. Speed was not associated with any of the FST measures.

Table 36

Correlations for FST Measures in Competition Condition (N = 52).

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. All fluency	-												
2. Appropriate fluency	.17	-											
3. Appropriate flexibility	.15	.84**	-										
4. Appropriate originality	-.02	.63**	.68**	-									
5. Appropriate frequency	-.06	.66**	.71**	.94**	-								
6. Appropriate effectiveness	-.17	.62**	.63**	.86**	.91**	-							
7. Accurate fluency	-.02	.69**	.53**	.38**	.38**	.40**	-						
8. Accurate flexibility	-.12	.61**	.67**	.46**	.48**	.50**	.87**	-					
9. Accurate originality	-.19	.59**	.58**	.58**	.56**	.55**	.65**	.75**	-				
10. Accurate frequency	-.19	.56**	.57**	.52**	.55**	.54**	.67**	.77**	.95**	-			
11. Accurate effectiveness	-.22	.52**	.53**	.47**	.49**	.53**	.63**	.74**	.93**	.95**	-		
12. Creative self-efficacy	-.01	-.10	-.13	-.16	-.11	-.14	-.03	-.05	-.05	-.03	-.01	-	
13. Speed	.30*	.20	.19	.06	-.05	-.05	.25	.20	.08	.02	-.03	.08	-

*p < .05 (2-tailed). **p < .01 (2-tailed).

The results showed a significant negative association between creative self-efficacy and an FST measure. In detail, creative self-efficacy was moderately associated with accurate fluency in the control condition. Creative self-efficacy was not significantly associated with other FST measures.

Exploratory analysis on gender difference. The results of independent samples test showed a significant difference in accurate fluency in star rating condition, t (50) = -2.05, p = .045, *95% CI* [-.85, -.01], such that female participants (M = 0.92, SD = 0.76) produced fewer accurate answer for FST than male participants (M = 1.35, SD = 0.73) when participants received star ratings. None of the other measures revealed significant difference between female and male participants (see Tables 37 – 39).

Table 37

Gender Difference in FST Measure in Competition Condition (N_{female} = 21, N_{male} = 31, df = 50).

Measures	Gender	*M*	*SD*	*t*	*p* values	95% CI Lower	Upper
Speed	Female	2.15	0.31	0.04	0.968	-0.14	0.15
	Male	2.15	0.22				
All fluency	Female	10.83	5.83	-0.51	0.610	-4.07	2.41
	Male	11.66	5.63				
Appropriate fluency	Female	3.56	2.44	-1.44	0.156	-2.46	0.41
	Male	4.59	2.58				
Appropriate flexibility	Female	2.35	1.55	-0.86	0.394	-1.27	0.51
	Male	2.73	1.57				
Appropriate originality	Female	1.19	0.59	-1.53	0.133	-0.53	0.07
	Male	1.42	0.49				
Appropriate frequency	Female	2.08	0.92	-1.09	0.279	-0.66	0.20
	Male	2.31	0.63				
Appropriate effectiveness	Female	1.22	0.57	-1.25	0.218	-0.43	0.10
	Male	1.39	0.38				
Accurate fluency	Female	1.25	1.01	-1.36	0.180	-1.52	0.29
	Male	1.86	1.89				
Accurate flexibility	Female	0.87	0.74	-1.14	0.261	-0.79	0.22
	Male	1.15	0.97				
Accurate originality	Female	0.66	0.49	-1.47	0.148	-0.58	0.09
	Male	0.91	0.65				
Accurate frequency	Female	1.15	0.79	-1.07	0.291	-0.75	0.23
	Male	1.41	0.92				
Accurate effectiveness	Female	0.71	0.46	-1.42	0.162	-0.57	0.10
	Male	0.95	0.66				

Table 38

Gender Difference in FST Measure in Star Rating Condition (N_{female} = 21, N_{male} = 31, df = 50).

						95% CI	
Measures	Gender	*M*	*SD*	*t*	*p* values	Lower	Upper
Speed	Female	2.19	0.28	0.93	0.359	-0.08	0.21
	Male	2.13	0.24				
All fluency	Female	9.71	4.56	-0.40	0.694	-3.50	2.35
	Male	10.29	5.51				
Appropriate fluency	Female	3.69	2.13	-0.13	0.899	-1.27	1.12
	Male	3.77	2.09				
Appropriate flexibility	Female	2.35	1.35	0.42	0.674	-0.51	0.78
	Male	2.21	0.96				
Appropriate originality	Female	1.34	0.51	0.90	0.372	-0.14	0.37
	Male	1.22	0.41				
Appropriate frequency	Female	2.09	0.74	-0.62	0.538	-0.51	0.27
	Male	2.22	0.65				
Appropriate effectiveness	Female	1.35	0.50	0.07	0.947	-0.27	0.28
	Male	1.34	0.47				
Accurate fluency	Female	1.43	1.56	0.17	0.868	-0.72	0.85
	Male	1.36	1.26				
Accurate flexibility	Female	0.88	0.87	0.25	0.801	-0.35	0.45
	Male	0.83	0.56				
Accurate originality	Female	0.65	0.56	-0.55	0.586	-0.34	0.20
	Male	0.72	0.41				
Accurate frequency	Female	0.92	0.76	-2.05*	0.045	-0.85	-0.01
	Male	1.35	0.73				
Accurate effectiveness	Female	0.68	0.55	-1.88	0.066	-0.59	0.02
	Male	0.96	0.53				

Table 39

Gender Difference in FST Measure in Control Condition (N_{female} = 21, N_{male} = 31, df = 50).

Measures	Gender	*M*	*SD*	*t*	*p* values	95% CI Lower	Upper
Speed	Female	2.12	0.32	0.16	0.878	-0.15	0.17
	Male	2.11	0.26				
All fluency	Female	10.07	5.02	-0.31	0.756	-3.31	2.42
	Male	10.51	5.05				
Appropriate fluency	Female	3.89	2.16	-1.08	0.284	-2.17	0.65
	Male	4.65	2.68				
Appropriate flexibility	Female	2.31	1.09	-1.06	0.295	-1.21	0.37
	Male	2.73	1.56				
Appropriate originality	Female	1.35	0.47	-0.03	0.973	-0.22	0.21
	Male	1.36	0.31				
Appropriate frequency	Female	2.28	0.67	-0.14	0.890	-0.33	0.29
	Male	2.31	0.44				
Appropriate effectiveness	Female	1.40	0.47	-0.65	0.522	-0.31	0.16
	Male	1.48	0.37				
Accurate fluency	Female	1.33	1.44	-0.52	0.605	-1.16	0.68
	Male	1.57	1.74				
Accurate flexibility	Female	0.89	0.82	0.61	0.546	-0.31	0.58
	Male	0.76	0.76				
Accurate originality	Female	0.79	0.65	1.34	0.187	-0.11	0.53
	Male	0.58	0.48				
Accurate frequency	Female	1.34	1.01	1.09	0.283	-0.23	0.79
	Male	1.06	0.82				
Accurate effectiveness	Female	0.84	0.63	0.66	0.515	-0.22	0.43
	Male	0.73	0.53				

Study 2 – 4 Discussion

Social feedback and creative thinking. The results showed that social feedback affected the speed in CRAT and the individual performance in FST but affected neither performance nor speed in AUT. In detail, participants spent more time in CRAT when they received star rating outcomes than the controlled outcomes. Participants produced more ideas in FST when they received competition outcomes than star rating outcomes. In other words, star rating feedback made participants to spend longer time into CRAT but did not change their performance. Compared to the star rating feedback, competition feedback encouraged participants to put more effort into Study 4 and did benefit the production of ideas. Therefore, social feedback did show an impact on the effort on creative thinking but did not show an impact on the creative thinking outcome.

The findings were inconsistent with the studies that found a positive effect of comparative social feedback on divergent and convergent thinking tasks (Balietti et al., 2016; Eisenberg & Thompson, 2011; Landers et al., 2019; Michinov & Primois, 2005; Raina, 1968; Shalley & Oldham, 1997; Van de Ven et al., 2011; Van Leeuwen & Baas, 2017). They were also inconsistent with the studies that found negative effects (Bittner & Heidemeier, 2013; Erat & Gneezy, 2016). However, the result was consistent with the study, which found that expecting comparative social feedback had no impact on creative problem solving in "in-basket exercise" (Shalley, 1995). The inconsistency between our findings and existing evidence was reasonable because we employed a new experimental paradigm that included novel social comparison manipulations and removed the contaminated monetary rewards from social rewards. To fully understand the impact of social comparison on creative thinking, we need more research in the field.

Competition vs star rating for FST all fluency. In addition, the competition feedback made participants produce more ideas than the star rating condition. However, none of the existing empirical evidence showed the difference between competition and star rating. We suggest two potential mechanisms here. First, the competition may induce more stress than star-rating because of the fear of losing or the joy of winning. However, the star rating system always shows participants a view of earning stars, which may not induce as much stress and motivation as competition (Akinola et al., 2019; Talbot et al., 1992; Yeh et al., 2015).

AUT, CRAT vs FST. Considering that the same manipulation induced a different impact on AUT, CRAT, and FST, the results for studies 2- 4 also supported hypothesis 1 (see Chapter 2). In detail, FST may grasp different elements from AUT and CRAT. However, we should be aware of the confounding variables such that studies 2 – 4 employed different participants. Therefore, the different impacts of social feedback in different studies may result from the differences in the sample.

Objective vs subjective feedback. Another confounding variable is the way that we represented the performance. In detail, participants received the number of correct/accurate answers in CRAT and FST but received the score of answers in AUT. Here, the number of correct/accurate answers was objective. In contrast, the answers' scores were subjective, plus subjective feedback may induce mentalising (i.e., the inference of others' minds (Frith & Frith, 2006; V. K. Lee & L. T. Harris, 2013)) and affect performance. The control condition in AUT may be a social condition. In other words, we propose that the control condition in AUT induced mentalising of the scoring system. In contrast, the control condition in CRAT and FST did not induce mentalising, which may lead to no impact on AUT and the impact on CRAT and FST.

Automatic social cognition in AUT. The different results in different studies may be explained by our argument that novelty-orientated tasks rather than accuracy-orientated tasks induced automatic social comparison (please see literature review in this chapter). That is, the feedback format in AUT (e.g., creativity score) was more novelty-orientated and CRAT and FST (e.g., the number of correct/accurate answers) were more accuracy-orientated. This explanation is consistent with our above argument that AUT's control condition was probably still a social condition.

Strength.

Link social comparison and creative thinking with RRCTSF. Studies 2 – 4 offered empirical evidence regarding how comparative social feedback affects different kinds of creative thinking with a novel experimental paradigm – RRCTSF. RRCTSF is the first paradigm that allows online experiments in the field of social psychology of creativity. Considering the difficulties of manipulating social context in an online experiment and the difficulties of giving dynamic and trustworthy feedback on individuals' creative potential or performance, RRCTSF has mainly addressed three issues that existed in the research stream of social comparison and creativity (see Chapter 3).

First, RRCTSF dissociates monetary and social rewards from a competitive context by taking out the material rewards. Our result supports existing research findings, which dissociated competition and material rewards (Shalley & Oldham, 1997). Hence, we suggest that social researchers pay more attention to social stimulus dissociation. We could learn from existing studies that we normally treat material rewards as an incentive for competition. However, material rewards seem to contaminate the positive effect of social competition.

The second issue has been addressed to an extent: we argued that different anticipation of evaluation outcome would change the effect of expected evaluation, so we controlled the anticipation by providing participants with the evaluation outcome after each round. Our result was inconsistent with most existing research, which found negative effects of expected evaluation. We attributed the inconsistency to the differences between the natures of expected evaluation and evaluation outcome – evoking expectation does not provide information while providing outcome provides social information. To fully address the second issue, we suggest two further studies - one could control the anticipation of expected evaluation. For instance, give participants examples of evaluation. Another could compare the effect of expected evaluation and received an evaluation.

The third issue, which is about inconsistent measures of creativity, has been addressed. We employed the most widely used divergent thinking task (AUT), convergent thinking task (CRAT), and our new thinking task (FST) and made the social comparison effect comparable.

Limitations and suggestions.

External validity. One of the limitations of this paradigm is the external validity (generalisation) of the results, which is a common limitation in most laboratory studies (Davis, 2008). For instance, participants may not need to do six rounds of AUT, CRAT, and FST in real-life contexts. This intense workflow could not simulate real-life situations, so the effect of social comparison may not remain in real-life situations.

Mixed effect of expectations. The paradigm’s second limitation relates to the “repeated rounds of tasks”, which may induce mixed effects of receiving the feedback of the previous round(s) and expecting the feedback of the current round, which is a common limitation within subject experiments (i.e., carryover effect). Here is an example. When a participant received competition feedback in round 2 and was going to receive star rating feedback in round 3, the participant’s performance in round 3 could be affected by 1) the win-or-lose feedback in round 2, 2) the expectation of the star rating in round 3, or 3) both. Therefore, the effect of different stimuli was interacting with each other, which is a confounding variable.

Subjective feedback. The other limitation relates to “feedback” when the paradigm is used for the tasks with non-objective feedback. The uniqueness of feedback for creative tasks (especially divergent thinking tasks) is that the reference point of the feedback is normally subjective. Even though we told participants that we had a pre-established scoring system, which was used to increase the level of objectivity, it was still hard to guarantee that this scoring system would be perceived as an objective reference point for participants. The problem is that if an experiment requires participants to trust the feedback, a lack of objectivity may induce the lack of trust. Again, it is important to note that this limitation is special for the tasks that allow creative opportunities (e.g., AUT) but not for the cognitive tasks (e.g., CRAT), which have objective answers.

Limited comparability. Another limitation of this sub-project is that, even though we can predictively compare the effect of comparative social feedback on AUT, CRAT, and FST, we could not quantify the differences statistically.

Suggestions. Based on the limitations, we suggest future studies replicating the paradigm using a between-subject design. In other words, participants would participate either competition, star rating, or control condition. Accordingly, they would not be able to guess the study’s intention and would have more cognitive resources to finish the tasks. In addition, we suggest comparing the effect of non-objective and objective feedback to the tasks by dissociating the non-objective and objective aspects of creativity outcomes. For example, future studies could compare the effect of fluency-only feedback, originality-only feedback, frequency-only feedback, mixed feedback, and no feedback on AUT and FST performance. We predict that originality- and frequency-only feedback would induce mentalising or social comparison mindset and have different impacts on creative thinking compared to the fluency-only feedback. It would also be worthwhile to study the effect of feedback directions. For example, we can compare the effect of win/loss in competition conditions. We can also compare the effect of different levels of star rating in the star rating condition. Moreover, future studies can work on the trustworthiness of the feedback in RRCTSF. Moreover, future studies can make the results comparable.

Overall, this chapter has explored how social feedback affects creative thinking and found that comparative social feedback may change participants task performance in CRAT and FST. The next chapter investigates how creative thinking affects social problem-solving.

Chapter 4: Divergent Thinking and Stereotypes in Advertising

When social context affects creative thinking, creative thinking solves societal problems in return. For example, evolutionary anthropologists claimed that creativity might be an important booster for society establishment (Diamond, 2014). Also, the stories of creative genius in science, architecture, and arts (see Chapter 2) have taught us that creativity benefits the development of (at least) science and arts, playing a crucial role in human evolution and sustainability. Recently, increasing evidence shows a negative association between divergent thinking and stereotypical thinking, especially for people who exhibits low personal need for structure (Gocłowska et al., 2014). In other words, the creativity could be positively related to stereotype avoidance (Sassenberg et al., 2017) and promote diversity, equality, and inclusion. Therefore, we linked advertising stereotypes and creativity and examined how stereotype avoidance affects marketers' creative thinking and audiences' perceived creativity.

Literature Review

Advertising stereotypes.

Researchers and the advertising industry are paying increased attention to advertising stereotypes because advertisers typically use stereotypes in their creative products as a communication tactic. However, such stereotypes may impact both brands and society and the social responsibility of media-based advertisements (Baker, 2014; Cho et al., 2017). For example, stereotypical advertisements tend to overrepresent ideal and attractive imagery, such as skinny women (Eisend & Langner, 2010; Elliott & Elliott, 2005). They also underrepresent imagery of minority groups, such as homosexual couples (Bowen & Schmid, 1997; Oakenfull et al., 2008). In addition, stereotypical advertisements represent some social groups with generalised biases. For example, older people are depicted with poor health or stay-at-home spouses as undereducated(Neuhaus, 2011).

Advertising stereotypes undermine human well-being.

The above stereotypical depictions in advertising are double-edged swords. In detail, they help to communicate with the target audience efficiently (Eisend et al., 2014; Lysonski & Pollay, 1990; Pick & Eisend, 2014) because stereotypical depictions are congruent with widespread beliefs about social reality (Johnson & Grier, 2012; Lee & Schumann, 2004). However, on the other hand, such stereotyped depictions significantly undermine the audience's cognitive processes and well-being

(Behm-Morawitz & Mastro, 2009; McKenzie et al., 2018; Rosengren et al., 2013). For example, over-exposure to attractive bodies in advertisements induces upward comparison (Gulas & McKeage, 2000; Martin & Kennedy, 1993; Mastro et al., 2009; Richins, 1991). The upward comparison can increase audience body anxiety(Halliwell & Dittmar, 2004) and self-objectification (Fredrickson & Roberts, 1997; Zimmerman & Dahlberg, 2008). It also reduces self-satisfaction (Bissell & Rask, 2010; Wan et al., 2013) and performance on cognitive tasks (Davies et al., 2002; Steele & Aronson, 1995). Also, under-representation or biased representations of some social groups even foster prejudice and discrimination (Fiske, 1998; Fiske et al., 2004; Harris & Fiske, 2006, 2007) and, in turn, unfairly limit their career opportunities (Knoll et al., 2011; Taylor & Stern, 1997). The situation suggests that avoiding stereotypes in mass communications is socially responsible behaviour.

Advertising stereotypes result in negative audience reaction.

Robust evidence demonstrates that stereotype avoidance has a positive effect while stereotype use produces harmful effects on audience reactions. For example, schema incongruity processing theory—representing minority groups in advertisements induces a mismatch between the audience's expectations and the perceived information (Lee & Schumann, 2004), which motivates in-depth information processing of advertising products (Grier & Brumbaugh, 1999). Existing research found that in-depth information processing induces self-categorisation change for people who are open-minded towards minority groups (Brumbaugh & Grier, 2006; Grier et al., 2006). The processing also increases social connectedness and empathy and evokes positive reactions towards the corresponding social groups and advertising brands (Åkestam et al., 2017b; Grier & Brumbaugh, 1999; Rosengren et al., 2013). They also found that people who hold negative impressions of the minority groups perceive the advertisement as irrelevant to themselves and react neutrally (Aaker et al., 2000).

Stereotypical depictions of minority groups also negatively affected the behaviour of minority groups. A recent meta-analysis demonstrated that when perceiving advertising imagery of majority groups, members from minority groups evoke negative attitudes towards the advertisements and the advertising brands and lower purchase intention towards the advertising products (Eisend & Hermann, 2019). Moreover, stereotypical information in advertisements generates reactance in target audiences: arousal and defensive responses; audiences detect a persuasive intention of communication (Brehm, 1966; Hammock & Brehm, 1966; Henderson-King et al., 2001; Petty & Cacioppo, 1979). For example, advertising imagery that delivers stereotypical information about females (e.g., physical characteristics and occupation roles) may induce a high level of advertising reactance and lead to more negative attitudes towards the advertisement and the advertising brand (Åkestam et al., 2017a).

There is a large gap between oneself and others when perceiving stereotypical information, and this gap also explains reactions to stereotyped advertisements (McLeod et al., 2001). Target audiences tend to overestimate the advertisements' negative impact on other people (Dahlén et al., 2013; Scharrer, 2002; Youn et al., 2000). Therefore, the audience is likely to react negatively to stereotyped advertisements since they may be concerned about the negative impacts on members of the social groups represented with bias (Eisend, 2015).

Stereotype avoidance elicits positive audience reaction.

However, advertisements that avoid stereotypes elicit positive audience responses. For example, consider schema incongruity processing theory—representing minority groups in advertisements induce a mismatch between the audience's expectations and the perceived information (Lee & Schumann, 2004), which motivates in-depth information processing of advertising products (Grier & Brumbaugh, 1999). For people who are open-minded toward minority groups, in-depth information processing induces self-categorisation change (Brumbaugh & Grier, 2006; Grier et al., 2006). It also increases social connectedness and empathy and evokes positive reactions toward the corresponding social groups and advertising brands (Åkestam et al., 2017b; Grier & Brumbaugh, 1999; Rosengren et al., 2013). Conversely, people with negative impressions of the minority groups perceive the advertisement as irrelevant and react neutrally (Aaker et al., 2000).

Link stereotypes with divergent thinking.

Stereotype.

According to flexible social cognition theory (Harris & Harris, 2017), people hold multiple probable attributions when considering another person. Some of those attributions are relevant to social category-based stereotypes, which result from a mental societal hierarchical structure that categorises other people at different levels (Ellemers et al., 1999). The other attributions are stereotype-irrelevant or -inconsistent information, such as the trait information inferred from the person's behaviour in a context. This information does not correspond to existing stereotypes (Bai et al., 2020; Fiske et al., 1999; Fiske & Neuberg, 1989). People can access either set of information if motivated to do so (Harris, 2021).

Stereotype activation is an oversimplified process in which people link another person with the attributes relevant to the most accessible or salient social category, such as gender (Clement & Schiereck, 1973; Harris & Fiske, 2009; Nelson & Klutas, 2000). In other words, stereotype activation is a process of close associations in which people link an element (i.e., the other person) with its closely associated elements from a salient category (i.e., stereotypes-consistent information).

On the other hand, stereotype avoidance is when people employ stereotype-irrelevant or -inconsistent information, which helps them perceive the person as unique (Park et al., 1991). The process indicates remote associations in which people connect two elements (i.e., the other person and stereotype-inconsistent information) that seems irrelevant. There are various stereotype avoidance approaches. For instance, researchers provided counter-stereotypical exemplars (Blair et al., 2001) and formed counter-stereotypic intentions (Blair & Banaji, 1996) to decrease stereotypes. They also trained people to reject stereotypic targets after a social category prime (Kawakami et al., 2000) and taught the low stereotype consensus among people (Sechrist & Stangor, 2001). In addition, forming concrete implementation intentions for egalitarianism is also a stereotype avoidance approach (Mendoza et al., 2010).

We can see that stereotype avoidance is likely to share similar cognitive mechanisms with divergent thinking, in which people explore different mental categories and make associations between remote elements (Joy P Guilford, 1967; Runco & Acar, 2012; Mark A Runco & Selcuk Acar, 2019).

Divergent thinking benefits stereotype avoidance.

The close link between stereotype avoidance and creative thinking is supported by the work of Sassenberg and his colleagues. They conducted a series of studies showing that priming a creative mindset exerts efficient control over stereotype activation. For example, in a lexical decision task, people were asked to respond to stereotypical targets after African American primes. The results showed that, after writing down three situations/activities/ ideas in which participants were creative, participants responded slower after writing down situations/activities/ ideas in which they were thoughtful and after they wrote nothing (Sassenberg &

Moskowitz, 2005). In other words, creative thoughts may decrease stereotypes. The creative priming also facilitates remote associations when the experimenters employed primes and targets irrelevant to human beings (“sugar”, “tea”, “sweet”) in the lexical decision task (Sassenberg et al., 2017).

Stereotype avoidance benefits divergent thinking.

Further empirical evidence supports the similar mechanisms of stereotype avoidance and creative thinking. For example, avoiding stereotypes by priming stereotype-inconsistent information boosts creativity. In one study, cognitive flexibility in a pasta-naming task was more significant for people who were asked to generate words to describe a female mechanic (stereotype-inconsistent) than for those who were asked to label a male mechanic (stereotype-consistent) (Gaither et al., 2015). When developing a novel idea and sketching a poster for a nightclub, people perform better when required to generate stereotype-inconsistent category combinations rather than stereotype-consistent ones (Gocłowska & Crisp, 2013; Wen et al., 2019; Zuo et al., 2019). When replacing the poster sketching task with a Chinese idiom riddle test, stereotype-inconsistent information is relevant to a better performance than stereotype-consistent information (Zuo et al., 2019). Overall, existing studies arrive at the same conclusion: stereotype-inconsistent information induces more individual creative thinking compared to stereotype-consistent information.

Further evidence supporting a link between stereotype avoidance and creative thinking comes from studies of racial identity and creative thinking. Priming multi-racial identities benefit multi-racial people's convergent thinking and mono-racial people's divergent thinking, while priming a mono-racial mindset does not affect mono-racial people's convergent thinking (Gaither et al., 2015). There are two possible reasons for multi-racial people being more creative than mono-racial people. First, multi-racial people may have more knowledge accessibility than mono-racial people (Gocłowska & Crisp, 2014). Second, multi-racial people may need greater cognitive flexibility than mono-racial people (Kharkhurin, 2011; Prior & MacWhinney, 2010) due to the need to identify (Leong & Ward, 2000; Phinney & Devich-Navarro, 1997) and integrate cultural inconsistency (Cheng et al., 2008; Tadmor et al., 2009). Cognitive flexibility helps multi-racial people switch between different cultural frameworks (Hong et al., 2000). In a word, the multi-identity broadens the breadth of self-identity (Der-Karabetian & Balian, 1992; King & Ruiz-

Gelices, 2003) and extends the scope of people's ideas (Hadis, 2005; Zhai & Scheer, 2004).

The discussion thus far demonstrates a positive relationship between stereotype avoidance and creative thinking. However, can this positive relationship maintain when we talk about stereotype avoidance and perceived creativity? As far as we know, existing research has not investigated this link, so we compare audiences' responses to advertising stereotypes and advertising creativity.

Advertising creativity.

Unlike advertising stereotypes, advertising creativity receives undoubted appreciation in the marketing field. Creativity is a long-standing mandatory criterion of a good advertisement (Bernstein, 1974; Edwards, 1956). Marketing education commonly includes one or two creativity-focused chapter(s) in the advertising student textbook (Smith & Yang, 2004). The advertising industry has set several impactful advertisements awards, such as Clio's, which advertising agencies admire (Smith & Yang, 2004; Till & Baack, 2005).

Previous studies conceptualised advertising creativity in multiple dimensions (e.g., divergence, relevance, effectiveness, and connectedness), with divergence and relevance as two necessary dimensions (Ang et al., 2007; Ang et al., 2014; Ang & Low, 2000; Kover et al., 1995; Kover et al., 1997; Tellis, 1997). In the field of advertising creativity, *divergence* is defined as the extent to which advertising messages are original (Haberland & Dacin, 1992), unexpected (Ang & Low, 2000; Haberland & Dacin, 1992), novel (Ang & Low, 2000; Kover et al., 1995), exciting, new (Kover et al., 1995), imaginative (Duke, 2001), different, unusual (Smith & Yang, 2004), or unique (Smith et al., 2008). In addition, *relevance* (Till & Baack, 2005) indicates that creative advertisements should be meaningful, appropriate, or valuable to the audience (Ang & Low, 2000; Haberland & Dacin, 1992; Smith & Yang, 2004).

Audiences have similar positive reactions to advertising creativity. Compared to non-creative advertisements, creative advertisements induce more positive feelings (Ang et al., 2007; Baack et al., 2008). Creative advertisements also elicit greater information processing (Smith et al., 2008; Smith et al., 2007; Yang & Smith, 2009) and a higher level of perceived product quality (Modig & Rosengren, 2013). More than that, audiences exhibit greater attention (Pieters et al., 2002), likings (Kover et al., 1995; Lehnert et al., 2013; Mannix & Neale, 2005), memory recalls

(Pick et al., 1991; Sheinin et al., 2011), and purchase intentions toward creative advertisements and corresponding brands and products (Till & Baack, 2005).

Considering the meaningful role that media plays in our society (Blowfield, 2005; Herbst et al., 2013; Pollay, 1986) and the impactful role that stereotypes and creativity play in developing a positive brand reputation and human well-being (Cravens & Piercy, 2006; Kotler et al., 1990), we use a stereotype intervention to disrupt stereotype activation, encourage stereotype avoidance, and highlight the stereotypes-creativity link for real-life marketers. We predicted that:

Hypothesis 3: *The stereotype intervention reduces marketers' stereotypicality and benefits their divergent thinking,*

especially when motivated to develop and maintain good relationships with potential and existing customers (Hawkins & Mothersbaugh, 2010). Here, stereotypicality represents the extent to which a marketer's dependence on stereotypical imagery.

We also measure whether stereotype avoidance relates to perceived creativity when watching the advertisements. Based on similar audience responses toward advertisements that avoid stereotypes and highlight creativity, we predicted that:

Hypothesis 4: *There is a negative relationship between the stereotypicality of advertisements and the audiences' perceived creativity and purchase intention.*

Here, stereotypicality represents the extent that an advertisement exhibits stereotypical imagery.

Study 5 (Pilot) Method

We began with a pilot study to develop and benchmark our dependent variable. Then, we developed a *consumer labelling task* in collaboration with marketers to match a crucial decision in developing media communications and marketing campaigns as closely as possible.

Participants.

We recruited an independent advertising non-professional general population sample (N = 152, 69 women, M_{age} = 33.3 years) who were residents where our marketers worked using Amazon Mechanical Turk (MTurk). Most generic participants were residents of the USA (68%), 29% were residents of the UK, and 3% were from the Netherlands and Western countries (i.e., Canada and Italy).

However, we did not exclude these latter participants since they are geographically and culturally close to the populations where our marketers worked and were likely to share similar stereotypes with our marketers.

Materials.

Consumer labelling task. We developed consumer labelling task to reflect a crucial decision in developing media communications and marketing campaigns. Participants read real-world consumer segmentation data and selecting many, few, or none amongst seventeen consumer labels that characterized the consumer segmentation data.

For example, a piece of consumer segmentation data included consumers' attitudes to life (e.g., "life is too boring, need more adventure and excitement"), brand preference (e.g., buyers index for Cable TV Networks), hobbies (e.g., sports, read), and gender split (e.g., male (42%) and female (58%)). Also, the seventeen customer labels were *creative*, *eco-friendly*, *risk seeker*, *active*, *loyal*, *unhealthy*, *price driven*, *anxious*, *adventurous*, *lazy*, *security preferred*, *convenience preferred*, *self-conscious*, *socialising*, *status-driven*, *fickle*, and *introverted*. (See Appendix O for the further details of the *Consumer Labelling Task*).

We assigned stereotypicality scores to the consumer labels based on the selection frequency. The more frequently the participants selected a consumer label, the more stereotypical the consumer label. We summed the selection frequency of each consumer label and ranked them. The higher ranking of a consumer label, the higher the stereotypicality of that label to the consumer segmentation data.

Procedure.

Participants completed two consumer labelling tasks with different consumer segmentation data. Consumer labels were the same for two tasks and were presented in random order. Participants also described the consumer segmentation in an open-ended question (we did this for exploratory and did not discuss it further).

Study 5 (Pilot) Results

Consumer labels. We conducted a chi-square analysis to determine whether the choice of consumer labels significantly differed. The results showed that, in Task 1, the proportion of participants who selected the most stereotypical label *convenience preferred* (1[17]) significantly differed from those who selected *unhealthy* (2), *anxious* (4), *lazy* (5), *self-conscious* (6), *price driven* (8), *status driven* (12), *risk seeker* (13), *security preferred* (14), *fickle* (16), and *introverted* (17), but did not differ by those who selected *active* (3), *creative* (7), *adventurous* (9), *loyal* (10), *eco-friendly* (11), and *socialising* (15) (see Tables 40 – 41 for descriptive statistics of the consumer labelling tasks).

Table 40

Descriptive statistics of Consumer Labels in Consumer Labelling Tasks 1.

	Task 1		
Consumer Labels	Occurrence	Stereotypicality	Ranking
Creative	6.09%	5.24	7
Eco-friendly	4.79%	3.82	11
Risk seeker	4.46%	3.44	13
Active	7.40%	6.33	3
Loyal	5.22%	4.27	10
Unhealthy	9.03%	6.75	2
Price driven	6.53%	5.14	8
Anxious	7.29%	5.78	4
Adventurous	5.99%	4.63	9
Lazy	6.86%	5.55	5
Security preferred	4.46%	3.35	14
Convenience preferred	8.92%	7.54	1
Self-conscious	6.86%	5.44	6
Socialising	4.46%	3.31	15
Status driven	4.68%	3.52	12
Fickle	3.81%	2.79	16
Introverted	3.16%	2.12	17

Note. [a] the ranking number is based on the stereotypicality value.

[17] Ranking of the label appears in parenthesis.

Table 41

Descriptive statistics of Consumer Labels in Consumer Labelling Tasks 2.

	Task 2		
Consumer Labels	Occurrence	Stereotypicality	Ranking
Creative	9.64%	7.15	2
Eco-friendly	5.38%	3.74	9
Risk seeker	5.88%	3.97	8
Active	13.02%	10.00	1
Loyal	6.88%	5.27	7
Unhealthy	3.88%	2.16	13
Price driven	4.38%	2.67	10
Anxious	3.00%	1.68	14
Adventurous	9.26%	6.72	3
Lazy	2.75%	1.58	15
Security preferred	3.88%	2.60	11
Convenience preferred	3.50%	2.21	12
Self-conscious	7.13%	5.37	6
Socialising	7.89%	5.75	5
Status driven	8.51%	6.37	4
Fickle	2.50%	1.36	17
Introverted	2.50%	1.54	16

Note. [a] the ranking number is based on the stereotypicality value.

Stereotypicality. We conducted a paired sample *t*-test to determine whether the stereotypicality of the labels at higher stereotypical ranking positions was significantly higher than those at lower stereotypical ranking positions. The results showed that, in Task 1, the stereotypicality of the most stereotypical label *convenience preferred* (1) was not significantly higher than the second and the third stereotypical labels (i.e., *unhealthy* (2) and *active* (3)) but was significantly higher than the other thirteen labels.

Study 5 Methods

Armed with the consumer labelling task, we considered the link between stereotyping and creative thinking in the generation of advertisements. Specifically, we used a novel stereotype intervention that

- Provided marketers with multi-racial identity information about the self,
- Disrupted the employment of stereotype-consistent information,
- Encouraged the use of stereotype-inconsistent information, and
- Made salient a link between stereotyping and creativity.

We predicted that the stereotype intervention reduces marketers' stereotypicality and benefit their creative thinking.

We employed a two *condition* (intervention vs no intervention) X 2 *time* (pre- vs post-intervention) mixed design, with the former as a between-subjects factor. We measured whether stereotype intervention impacts marketers' stereotypicality, divergent thinking, and creative outputs derived from real consumer segmentation data.

Participants.

We recruited 136 marketers (93 women, M_{age} = 36.10 years) using convenient samples and snowballing techniques from a stable of brands across a multi-national company. Most responses were from the UK participants (55%), 26% were from the USA, 13% were from the Netherlands, and 6% were from other European countries. Sixty-nine participants participated in the experimental condition subject to our intervention, while the remainder comprised the control condition and did not receive the intervention. We did not analyse any data before data completion, determined by a specific cut-off date.

We did a pattern analysis of missing data and found that 52.9% of participants did not complete the main measures. Therefore, we report results only for those participants who completed all measures (N = 53; 34 in the experimental group, 19 in the control group) (Jakobsen et al., 2017).

Post-hoc power analyses using GPower suggested that we had enough power (1.00) to detect the effect at the p = .05 level (*partial* $\eta 2$ = .09, *effect size f* = .31, *number of groups* = 2, *number of measurements* = 4, *corr among rep measures* = .50, *nonsphericity correction* ε = 1).

Materials.

Consumer labelling task. To measure stereotypicality in advertisement creation, we employed the two consumer labelling tasks from Study 5 (Pilot).

AUT. To measure divergent thinking (i.e., creativity), participants completed the AUT, where they thought of as many uses as possible for a brick and a mug. We assigned fluency scores for each participant, equalling the number of appropriate answers. The higher the fluency, the greater the participant's divergent thinking.

We assigned an appropriate frequency score to each answer based on its frequency[18] in a generic sample answer pool. If the frequency of an answer was 5% or above, the frequency score of the answer was 0. If the frequency of an answer was between 1% to 5%, the frequency of the answer was 1. If the frequency was below 1%, the frequency score of the answer was 2. The higher the frequency a participant performed, the greater the participant's divergent thinking.

Procedure.

All participants provided informed consent and completed an online test consisting of one round of the AUT, a consumer labelling task, the gender bias scale, and demographic information. The experimental group was exposed to the stereotype intervention the following day, while the control group did not receive any deliberate intervention. In addition, participants in the experimental group were asked to complete a post-test two weeks after the intervention. Meanwhile, the control group completed a post-test three days after the pre-test. The post-test replicated all tasks in the pre-test with two consumer labelling tasks (one task was repeated and the other novel) (see Appendix P for *Study 5 Script*).

Stereotype intervention. The stereotype intervention was promoted to marketeers as a workshop to boost creativity, not as a stereotype reduction workshop. There were two parts to the intervention. In the first part, we primed participants with a multi-identity mindset. Participants completed a DNA swab and received ancestral DNA information while amongst their team members who were also participants. They then joined a brief lecture that provided a primer on DNA and information on how to interpret their results. We thought that the DNA results, either for themselves or their team members, would be surprising enough for them to

---

[18] Frequency of an answer = (Occurrence of the answer across all participants / The number of all answers) X 100

consider that each person is more than simply the single generic ethnic category assigned to them societally.

In the second part, usually the following day, we focused on stereotype reduction. This approach made salient the marketeers' professional selves and the responsibility they have to those roles to be creative and to avoid perpetuating negative stereotypes. Building on their DNA experience, we discussed psychological illusions that suggested perception could differ from reality. We linked stereotypes and creativity, explaining that moving away from reliance on stereotypes could boost the creativity of advertisements. We discussed how stereotypes are acquired and processed in the brain. We gave them opportunities to discuss within their brand teams what processes may better avoid reliance on contextual events that promote the use of stereotypes as heuristics. We provided them with a toolkit to combat the employment of stereotype-consistent information. Also, we asked participants to reflect on their creative thinking process and identify the pressure points where stereotype activation was likely to occur, for instance, deadlines that promote heuristic thinking.

Study 5 Results

Statistical assumptions. To prepare an appropriate dataset for mixed repeated measures ANOVA, we did a series of assumption tests on the variables. All the variables met the assumption of non-zero variances, independence of variables, homogeneity of variance, and sphericity. There was no outlier. The skewness (between ± 2) & kurtosis (between ±2) indicated that the data contained approximately normally distributed error. Please see Table 42 for descriptive statistics and Table 43 for normality tests of all measures.

Table 42

Descriptive Statistics for Measures in Study 5 (N = 34).

	M	*SD*
Labelling stereotype - pre - exp	775.91	308.79
Labelling stereotype - post - exp	531.13	315.35
AUT appropriate fluency - pre - exp	2.94	2.47
AUT appropriate fluency - post - exp	3.40	2.91
AUT appropriate frequency - pre - exp	2.34	2.56
AUT appropriate frequency - post - exp	3.12	3.53
Labelling stereotype - pre - ctrl	827.00	298.01
Labelling stereotype - post - ctrl	698.13	206.59
AUT appropriate fluency - pre - ctrl	3.11	2.67
AUT appropriate fluency - post - ctrl	3.03	2.25
AUT appropriate frequency - pre - ctrl	2.47	3.13
AUT appropriate frequency - post - ctrl	2.21	2.33

Table 43

Normality test for Measures in Study 5.

	Skewness	Kurtosis	Shapiro-Wilk Statistic	Shapiro-Wilk *p* value
Labelling stereotype – pre[1] – exp[2]	-0.90	0.47	.93*	.027
Labelling stereotype – post[3] - exp	0.22	-0.90	.93*	.038
AUT appropriate fluency - pre - exp	0.64	-1.21	.79**	<.001
AUT appropriate fluency - post - exp	0.70	-1.07	.78**	<.001
AUT appropriate frequency - pre - exp	0.97	-0.52	.79**	<.001
AUT appropriate frequency - post - exp	1.16	0.67	.82**	<.001
Labelling stereotype - pre – ctrl[4]	-0.42	-0.42	.94	.271
Labelling stereotype - post - ctrl	0.40	0.40	.92	.137
AUT appropriate fluency - pre - ctrl	0.75	0.75	.75**	<.001
AUT appropriate fluency - post - ctrl	0.63	0.63	.83**	.003
AUT appropriate frequency - pre - ctrl	1.89	1.89	.76**	<.001
AUT appropriate frequency - post - ctrl	1.00	1.00	.86*	.011

*p < .05 (2-tailed). **p < .01 (2-tailed).

Notes. [1]pre indicates pre-intervention condition. [2]exp indicates experimental condition. [3]post indicates post-intervention condition. [4]ctrl indicates control condition.

Stereotypicality of consumer labels[19]. We performed a mixed measures ANOVA to examine the effect of the stereotype intervention on consumer label scores. There was a significant main effect of time, F (1, 51) = 23.15, p < .001, Ω = 0.67, such that stereotypicality scores pre-intervention (M = 794.23, SD = 303.09) was higher than post-intervention (M = 591.00, SD = 290.55)[20].

The *condition* X *time* interaction did not reveal a significant effect. Nonetheless, we conducted simple effect tests to probe our hypotheses. There was

[19] We found no significant differences on this measure for all participants, including those with missing data.

[20] Participants did the label task 1 in the pre-intervention task and did label tasks 1 & 2 in the post-intervention. Therefore, the stereotypicality score pre-intervention was the score of label task 1, and the stereotypicality score post-intervention was the average score of label task 1 and label task 2.

a significant difference between the experimental and control condition after the stereotype intervention, *t* (51) = -2.07, *p* = .044, *95% CI* [-167.00, -80.72], such that the stereotypicality scores were lower in experimental condition (*M* = 531.13, *SD* = 315.35) than the control condition (*M* = 698.13, *SD* = 206.59). The stereotypicality before the intervention did not reveal significant differences *t* (51) = -0.59, *p* = .561, *95% CI* [-226.49, 124.32], between the experimental (*M* = 775.91, *SD* = 308.79) and control conditions (*M* = 827.00, *SD* = 298.01).

There was also a significant difference between the stereotypicality scores in the experimental condition, *t* (33) = 4.96, *p* < .001, *95% CI* [144.30, 345.26], such that the stereotypicality scores were higher in pre-intervention condition (*M* = 775.91, *SD* = 308.79) than post-intervention (*M* = 531.13, *SD* = 315.35). There was also a significant difference between the stereotypicality scores in the control condition, *t* (18) = 2.37, *p* = .029, *95% CI* [14.59, 243.14], such that the stereotypicality scores were higher in pre-intervention condition (*M* = 827.00, *SD* = 298.01) than post-intervention (*M* = 698.13, *SD* = 206.59).

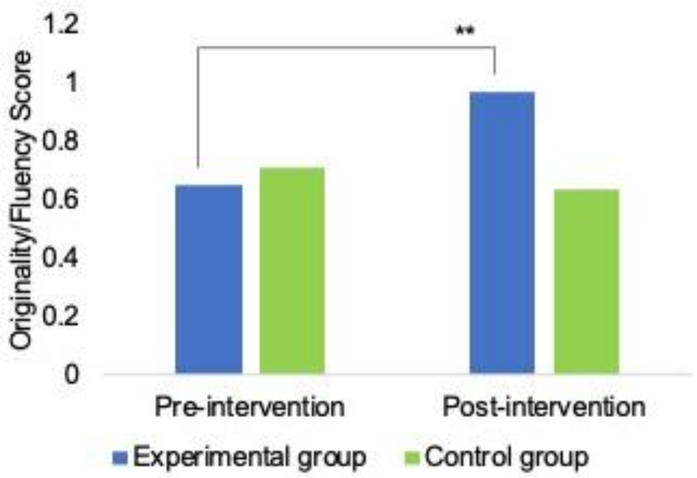
Following the significant differences pre- to post-intervention in both experimental and control conditions, we compared stereotypicality scores separately in pre-intervention (task 1) with stereotypicality scores in the repeated task (task 1) and the non-repeated task (task 2). In the experimental condition, there was a significant reduction of stereotypicality scores in repeated tasks, *t* (33) =4.19, *p* < .001, *95% CI* [134.13, 386.81], and non-repeated tasks, *t* (33) = 3.64, *p* < .001, *95% CI* [101.20, 356.98], such that the stereotypicality scores was higher in pre-intervention condition than the repeated task (*M* = 515.44, *SD* = 375.73) and the non-repeated task (*M* = 546.82, *SD* = 396.65) in post-intervention.

In the control condition, there was a significant reduction of stereotypicality scores in repeated tasks, *t* (33) = 3.40, *p* = .003, *95% CI* [64.61, 274.13], such that the stereotypicality scores was higher in pre-intervention condition than task 1 in post-intervention (*M* = 657.63, *SD* = 182.99). However, in task 2, the stereotypicality scores in the control condition did not reveal significant differences, *t* (18) = 1.28, *p* = .216, *95% CI* [-56.49, 233.22], pre- (*M* = 827.00, *SD* = 298.01) to post-intervention (*M* = 738.63, *SD* = 277.52).

Divergent thinking in the AUT[21]. To examine the effect of stereotype intervention on divergent thinking, we performed a mixed ANOVA on frequency in the AUT. There was a significant condition X time interaction, *F* (1, 51) = 5.22, *p* = .027, *Ω* = 0.31 (see Figure 12). However, neither condition, *F* (1, 51) = 0.27, *p* = .607, *Ω* = .08, nor time, *F* (1, 51) = 0.39, *p* = .536, *Ω* = .09, revealed significant differences in the main effects.

Figure 12

AUT frequency in experimental and control groups.



We followed up this interaction with simple effect tests. There was a significant difference between the frequency in the experimental condition, *t* (33) = -2.82, *p* = .008, *95% CI* [-.54, -.09], such that frequency was lower pre-intervention (*M* = 0.65, *SD* = 0.32) than post-intervention (*M* = 0.97, *SD* = 0.72). The originality in the control condition did not reveal significant differences pre- to post-intervention. The frequency before the intervention did not reveal significant difference, *t* (51) = 1.76, *p* = 0.085, *95% CI* [-.05, .70].

[21] For all participants, including those with missing data, there was a significant main effect of *condition* on divergent thinking after the intervention, *F* (1, 62) = 5.113, *p* = .027, *CI* [-0.00, 0.22], such that frequency post-intervention in experimental condition (*M* = 1.00, *SD* = 0.76) was higher than control condition (*M* = 0.61, *SD* = 0.52). There were no other significant main effects or interactions.

We performed a mixed ANOVA on the fluency in the AUT to determine whether the marketers produced more or fewer ideas after the intervention. Neither condition, F (1, 51) = 0.02, p = .882, Ω =.05, time, F (1, 51) = 0.42, p = .520, Ω = .10, nor condition X time interaction, F (1, 51) = 0.84, p = .363, Ω = .15, revealed significant differences.

Exploratory analysis. The experimental manipulation affected the stereotypicality in the labelling task and the frequency in the AUT and showed the same pattern. Therefore, we conducted mediation analyses to examine whether the significant effects on the stereotypicality in the labelling task were mediated by the frequency in the AUT. In addition, we examined whether the significant effect of *the intervention* on post-intervention stereotypicality mediated post-intervention frequency scores. The indirect effects showed that the relationship between post-intervention frequency and post-intervention stereotypicality was insignificant.

Also, we examined whether the significant effect of *time* on stereotypicality was mediated by the frequency in the experimental group. The indirect effects showed that the relationship between the frequency in the experimental group and the stereotypicality in the experimental group was not significant. Both results implied that the AUT frequency did not mediate the effect of the intervention on the stereotypicality in the labelling task.

Study 6 (Pilot) Method

Study 5 demonstrated a link between stereotype avoidance and creative thinking from the perspective of advertising generation. We next look at the link from the perspective of advertising consumption. To create appropriate stimuli for the main experiment, we conducted a pilot observational study that measured the perceived stereotypicality of protected characteristic groups (i.e., race/ethnicity, gender, sexual orientation, and disability) in a series of video advertisements. To predict the effects of stereotypicality on perceived creativity, we also measured the perceived creativity and emotion valence towards the video advertisements.

Participants.

We recruited 61 participants (women = 39, M_{age} = 31.43 years) from the online participant subject pool Prolific Academic. We decided on the sample size based on a rule of thumb to estimate a normal distribution (Simmons et al., 2011). All participants had been residents in the UK for more than five years. More than half of the participants had a bachelor's degree (66%). Most participants did not report a long-standing disability (92%). Most participants were White British (85%), 7% were Asian British, 5% were Black British, and 3% were mixed ethnicities. Also, most participants reported themselves as heterosexuals (75%), 21% reported themselves as LGBTQ, and 4% preferred not to report. Participants gave informed consent and received £7.50 for 1 hour of their time.

Post-hoc power analyses using GPower suggested that we had enough power (1.00) to detect the effect at the p = .05 level (r^2 = .83, *effect size* |r| = .911, *correlation* ρ *H0* = 0).

Materials.

Advertisements. We acquired video advertisements from an online database https://adsoftheworld.com for advertisement stimuli. We chose the website since it contained advertisements aired in the UK. We set up three criteria when selecting video advertisements. First, the advertisements were disseminated in 2017, three years before data collection, so they were neither too old to be irrelevant nor too recent to be highly memorable for UK residents. Secondly, the length of each advertisement was 30 seconds to facilitate sufficient stimuli to boost statistical power without placing an undue burden on participants. Finally, they were all live-action videos featuring human actors instead of animation or animated objects or animals. At the time point of video acquirement, 66 video advertisements satisfied the three criteria, and we employed them as experiment stimuli.

The advertising brand or product covered various industries such as gambling industries (e.g., Betway), food products (e.g., Bisto), financial service comparison website (e.g., Go compare), personal care product (e.g., Remington air plates), and car (e.g., Mazda). (Please see Appendix Q for List of Advertising Brands and Products)

Familiarity. We measured the familiarity of each advertisement. Participants respond to "*how familiar are you with the advertisement?"* on a 7-point Likert scale. The higher the rating, the more familiarity with the advertisement.

Perceived stereotypicality scales. We measured the perceived stereotypicality towards each advertisement with four question sets. Each question set asked participants to evaluate the degree to which an advertisement represented one protected characteristic group of people. The groups included women, LGBTQIA+, ethnic minorities (BAME), and the disabled. Within each question set, the participants started with a yes-no question to evaluate whether an advertisement represented a specific group of people. If participants answered yes, we asked them to assess the degree of representativeness, significance, stereotypicality, and social interactions for the chosen category/categories on a 7-point Likert scale. The statements were "*how explicit is the character to its social category (e.g., women)*?" "*How significant a role did the character play in the advertisement?*" "*How stereotypical is the role being performed by the character?*" and "*How positive/negative would you describe the interaction?*" If participants answered no, we proceeded participants to the next question set. After answering four question sets about four social groups, participants rated the advertisement's overall stereotypicality on a 7-point Likert scale.

We calculated the perceived stereotypicality of each advertisement based on participants' ratings on the 7-point Likert scale. Specifically, we summed the ratings on representativeness, significance, interaction, and overall inclusivity. Then, we subtracted the stereotypicality ratings (scale reliability: *Cronbach's α* = .60). The higher the summed ratings, the less stereotypical the advertisement.

Perceived creativity scales. We asked participants to evaluate the divergence, relevance, and overall creativity on 7-point Likert scales to measure the perceived creativity towards each advertisement. First, participants rated how much did agree/disagreed with five statements: "*the advertisement was different*", "*the advertisement was uncommon*", "*the advertisement was relevant to you*", "*the advertisement was meaningful to you*", and "*how creative do you think the advertisement was*" (Smith et al., 2008). We summed the ratings on the five statements (scale reliability: *Cronbach's α* = .94). The higher the summed ratings, the more creativity the participants perceived from the advertisement.

Emotion valence scales. To measure the level of positive and negative emotions elicited by each advertisement, we asked participants to evaluate emotion valence on a 7-point Likert scale. Participants rated emotional valence with two statements. They were "*how positive do the advertisement make you feel?*" and "*how negatively does the advertisement make you feel?*" (Ang et al., 2014).

Procedure.

At the beginning of this study, all participants gave consent and answered a question about the year they began UK residence. We proceeded only with the participants living in the UK for five years or above to the main task to ensure our sample was aware of British societal stereotypes. Participants watched an advertisement in each round of the main task and rated it regarding its familiarity, stereotypicality, creativity, and emotional valence. Participants rated 22 advertisements in the main task, with a time limit of 5 minutes per advertisement. The advertisements did not repeat and were randomly selected from the larger set of 66 advertisements. Participants then provided demographic information.

Study 6 (Pilot) Results

Statistical assumptions. To prepare an appropriate dataset for Pearson correlations, we did a series of assumption tests on the variables. All the variables met the assumption of related pairs, linearity, and homogeneity of variance. There was no outlier. The skewness (between ± 3) & kurtosis (between ±5) indicated that the data contained approximately normally distributed error. Please see Table 44 for descriptive statistics and Table 45 for normality tests of all measures.

Table 44

Descriptive Statistics for Measures in Study 6 (Pilot), N = 61.

	M	*SD*
Familiarity	1.02	0.96
Perceived stereotypicality	11.08	4.67
Positive emotion	3.09	1.37
Negative emotion	1.18	0.88
Perceived creativity	2.72	1.09
Perceived divergence	3.15	1.24
Perceived convergence	2.00	1.17
Perceived difference	3.34	1.31
Perceived uncommonness	2.96	1.22
Perceived relevance	2.14	1.17
Perceived meaningfulness	1.86	1.25
Perceived overall creativity	3.29	1.17
Perceived LGBQIA+	0.33	1.18
Perceived BAME	2.99	1.82
Perceived disability	0.11	0.26
Perceived female	5.05	2.31
Perceived overall inclusivity	2.60	1.15

Table 45

Normality test for Measures in Study 6 (Pilot).

			Shapiro-Wilk	
	Skewness	Kurtosis	Statistic	*p* value
Familiarity	1.29	0.99	.85**	<.001
Perceived stereotypicality	0.53	0.62	.98	.35
Positive emotion	0.07	-0.41	.99	.87
Negative emotion	0.88	0.83	.93**	.002
Perceived creativity	0.21	0.31	.99	.747
Perceived divergence	-0.12	-0.39	.99	.909
Perceived convergence	0.85	0.67	.95*	.013
Perceived difference	-0.02	-0.31	.99	.990
Perceived uncommonness	-0.16	-0.53	.98	.583
Perceived relevance	0.79	0.54	.95*	.023
Perceived meaningfulness	0.83	0.51	.94**	.006
Perceived overall creativity	-0.12	-0.60	.99	.709
Perceived LGBQIA+	7.29	55.31	.24**	<.001
Perceived BAME	0.47	-0.13	.97	.188
Perceived disability	2.21	4.31	.57**	<.001
Perceived female	0.05	-0.44	.98	.409
Perceived overall inclusivity	0.51	-0.10	.98	.261

*$p < .05$ (2-tailed). **$p < .01$ (2-tailed).

Correlations. We correlated familiarity, perceived stereotypicality, creativity, and positive and negative emotions of each participant. Results indicated that familiarity, positive emotion, and perceived creativity yield significant positive associations (see Table 46). The correlation between positive emotion and perceived creativity was strong (see Figure 13a – 13c). Perceived stereotypicality and negative emotion were associated with neither of the other variables.

Table 46

Correlations Between Familiarity, Perceived Stereotypicality, Perceived Creativity, Positive Emotion, and Negative Emotion.

Variable	1	2	3	4	5
1. Familiarity	-				
2. Positive stereotypicality	-.12	-			
3. Perceived creativity	.50**	.13	-		
4. Positive emotion	.47**	.20	.83**	-	
5. Negative emotion	.035	-.07	.12	-.02	-

*p < .05 (2-tailed). **p < .01 (2-tailed).

Figure 13a

A positive correlation between familiarity and positive emotion.

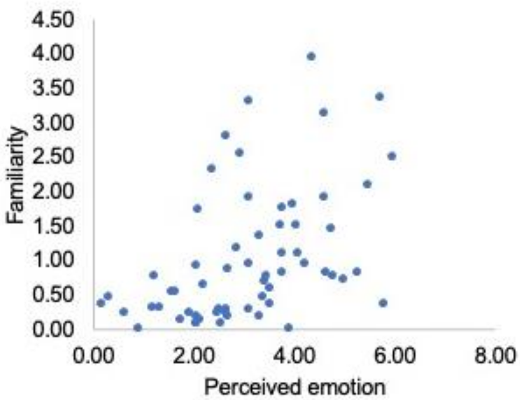


Figure 13b

A positive correlation between familiarity and perceived creativity.

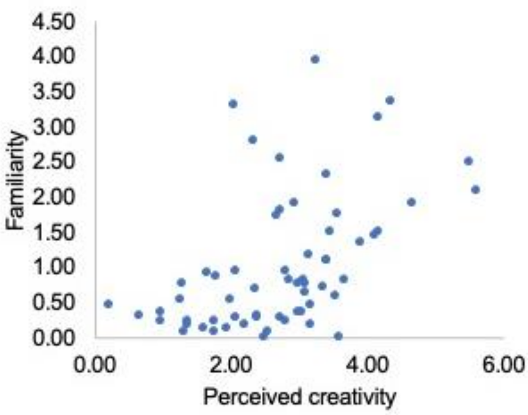
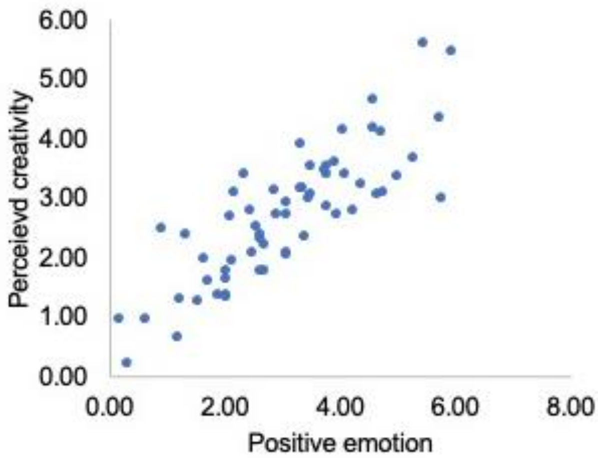


Figure 13c

A positive correlation between perceived creativity and positive emotion.



Study 6 Method

To examine the effect of stereotypicality of advertising on audience responses, we conducted an online experiment that manipulated stereotypicality across three levels (high, medium, low) in a within-subject design. In addition, we measured the perceived creativity of the advertisements. Finally, we also measured purchase intention toward the advertising product as a proxy for the effectiveness of the advertisements as an exploratory variable.

Participants.

We recruited 102 participants (women = 67, M_{age} = 30.46 years) on Prolific Academic. Half of the participants had a bachelor's degree or above. Most participants did not report a long-standing disability (91%). Most participants were White British (86%), 7% were Asian British, 4% were Black British, and 3% were mixed ethnicities. Most participants reported themselves as heterosexuals (84%), 15% reported themselves as LGBTQ, and 1% preferred not to report. Participants gave informed consent and received £7.50 for 1 hour of their time.

Post-hoc power analyses using GPower suggested that we had enough power (1.00) to detect the effect at the p = .05 level (*partial* $\eta 2$ = .63, *effect size f* = 1.31, *number of groups* = 3, *number of measurements* = 2, *corr among rep measures* = .50).

Materials.

Advertisements. To manipulate the stereotypicality levels of advertising imagery, we ranked 66 selected advertisements[22] based on perceived stereotypicality scores from the pilot study. We selected seven advertisements per stereotypicality level: most stereotypicality (*M* = 2.20, *SD* = 0.84), medium stereotypicality (*M* = 11.23, *SD* = 1.29), and least stereotypicality (*M* = 26.11, *SD* = 5.62). Advertisements in the same ranking range (i.e., most, medium, least). (Please see Appendix Q for the details of selected advertisements)

Perceived creativity scales*.* We asked participants to evaluate the unexpectedness, uniqueness, and overall creativity on a 0 – 7 points Likert scale. The scale measured the perceived creativity toward each advertisement. In detail, participants rated how much they agreed/disagreed with three statements. The statements were, "*the advertisement was typical of the kind of advertisements I see.*" "*The advertisement was unique*." and "*how creative was the advertisement?*" The first statement was reversed coded. The scale reliability showed low internal inconsistency across the three statements (*Cronbach's α* = .63), which increased substantially (*Cronbach's α* = .93) if we excluded the responses from the first statement. Therefore, we separated the first statement from the remaining two statements. The higher the ratings for the first statement, the greater the perceived unexpectedness of the advertising product. We summed the rating for the second and the third statements: the higher the summed rating, the greater the perceived creativity of the advertising product.

Purchase intention scales. To measure the purchase intention towards the advertising product, participants rated how much they agreed/disagreed with two statements: "*I am likely to purchase the product advertised*." and "*I would not recommend this product to a friend*." on a 0 – 7 points Likert scale. The second statement was reverse coded. The results of scale reliability showed a low internal inconsistency of two statements (*Cronbach's α* = .57), so we treated the two statements as separate variables. The higher the ratings for either statement, the more willingness to purchase the advertising product for the self (purchase intention) or recommend the advertising product to a friend (recommend intention).

---

[22] Inter-rater reliabilities of three advertisements were .95, .50, and .68.

Procedure.

We designed the study on the platform Qualtrics Survey Software. At the beginning of this study, all participants gave consent and reported their year of residence in the UK. Therefore, we only proceeded with participants living in the UK for five years or above to the main task. Participants evaluated all 21 advertisements in the main task. In detail, participants watched a 30-second advertisement rated either high, medium, or low on stereotypicality. Next, they were given 5 minutes to rate on the Likert scales for purchase intention and perceived creativity. We also measured memory recall of the advertisements as an exploratory variable; we do not discuss these results further. Participants then provided demographic information (see Appendix R for *Study 6 Script*).

Study 6 Results

Statistical assumptions. To prepare an appropriate dataset for between subject ANOVA, we did a series of assumption tests on the variables. All the variables met the assumption of independence of variables and homogeneity of variance. There was no outlier. The skewness (between ± 1) & kurtosis (between ±2) indicated that the data contained approximately normally distributed error. Please see Table 47 for descriptive statistics and Table 48 for normality tests of all measures.

Table 47

Descriptive Statistics for Measures in Study 6, N = 102.

	M	*SD*
Purchase intention _ low[1]	1.62	0.56
Purchase intention _ medium[2]	1.50	0.55
Purchase intention _ high[3]	1.71	0.61
Recommend intention _ low	5.42	1.24
Recommend intention _ medium	5.30	1.20
Recommend intention _high	5.62	1.21
Unexpectedness _ low	3.89	1.02
Unexpectedness _ medium	4.20	0.90
Unexpectedness _ high	4.93	0.93
Uniqueness _ low	3.13	1.05
Uniqueness _ medium	3.41	1.06
Uniqueness _ high	4.49	0.95
Creativity _ low	3.26	1.06
Creativity _ medium	3.62	1.13
Creativity _ high	4.66	1.06
Uniqueness & Overall _ low	3.19	1.02
Uniqueness & Overall _ medium	3.51	1.06
Uniqueness & Overall _ high	4.57	0.96
Average of perceived creativity _ low	3.43	0.83
Average of perceived creativity _ medium	3.74	0.81
Average of perceived creativity _ high	4.65	0.77

Notes. [1]low indicates low stereotypicality condition. [2]medium indicates medium stereotypicality condition. [3]high indicates high stereotypicality condition.

Table 48

Normality test for Measures in Study 6.

			Shapiro-Wilk	
	Skewness	Kurtosis	Statistic	*p* value
Purchase intention _ low	-0.98	0.55	.93**	<.001
Purchase intention _ medium	-0.12	-0.43	.99	.555
Purchase intention _ high	-0.38	-0.04	.99	.346
Recommend intention _ low	-0.65	1.32	.97*	.011
Recommend intention _ medium	-0.19	-0.16	.99	.707
Recommend intention _high	-0.56	0.70	.98	.090
Unexpectedness _ low	0.20	1.87	.97*	.013
Unexpectedness _ medium	0.02	0.36	.99	.616
Unexpectedness _ high	0.01	-0.66	.99	.320
Uniqueness _ low	-0.33	-0.03	.98	.222
Uniqueness _ medium	-0.09	-0.23	.99	.727
Uniqueness _ high	-0.33	0.25	.99	.358
Creativity _ low	-0.46	0.17	.98	.072
Creativity _ medium	-0.29	-0.05	.99	.584
Creativity _ high	-0.78	1.09	.96**	.006
Uniqueness & Overall _ low	-0.41	0.28	.99	.349
Uniqueness & Overall _ medium	-0.19	-0.11	.99	.640
Uniqueness & Overall _ high	-0.53	0.86	.98	.141
Average of perceived creativity _ low	-0.41	0.73	.98	.218
Average of perceived creativity _ medium	-0.17	0.40	.99	.492
Average of perceived creativity _ high	-0.04	-0.17	.99	.875

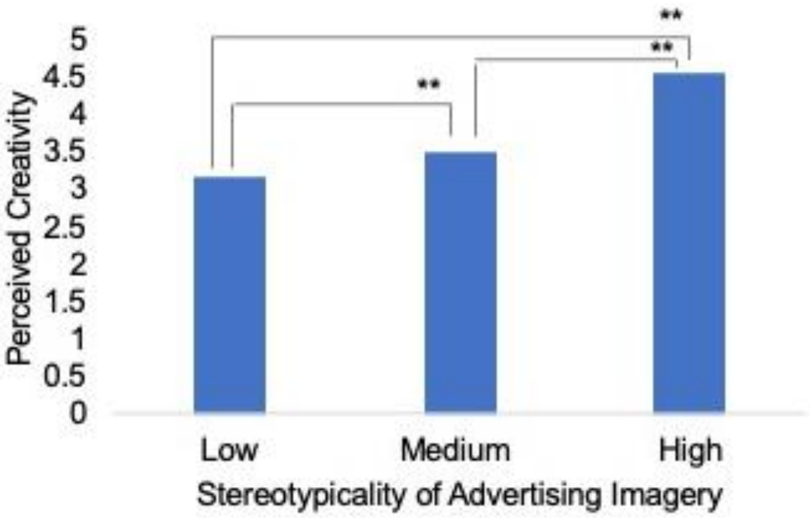
**p* < .05 (2-tailed). ***p* < .01 (2-tailed).

Unexpectedness. We conducted a repeated-measures ANOVA to examine the effect of stereotypicality of advertising imagery on the perceived unexpectedness towards the advertisements. There was a significant main effect of stereotypicality on perceived unexpectedness, F (2, 100) = 71.97, p < .001, Ω = .85. We followed up on this main effect with a post hoc paired sample t-test. There was a significant difference between high stereotypicality and medium stereotypicality levels, t (101) = .73, p < .001, *95% CI* [.52, .93], such that the participants perceived advertising imagery with a high level of stereotypicality (M = 4.93, SD = 0.93) as more unexpected than advertising imagery with a medium level (M = 4.20, SD = .90). There was a significant difference between high and low stereotypicality levels, t (101) = 1.04, p < .001, *95% CI* [.82, 1.26], such that participants perceived advertising imagery with a high level of stereotypicality as more unexpected than advertising imagery with a low level (M =3.89, SD = 1.02). There was a significant difference between the medium stereotypicality condition and the low stereotypicality condition, t (101) = .31, p = .003, *95% CI* [.90, 1.12], such that participants perceived the advertising imagery with a medium level of stereotypicality as more unexpected than the advertising imagery with a low level. The results suggested that participants expect low stereotypicality in advertising.

Perceived creativity. We conducted a repeated-measures ANOVA to examine the effect of stereotypicality of advertising imagery on the perceived uniqueness and creativity towards the advertisements. There was a significant main effect of stereotypicality on perceived uniqueness and creativity, F (2, 100) = 168.66, p < .001, Ω = 1.29. We followed up on this main effect with a post hoc paired sample t-test. There was a significant difference between high stereotypicality and medium stereotypicality levels, t (101) = 1.06, p < .001, *95% CI* [.88, 1.24], such that the participants perceived advertising imagery with a high level of stereotypicality (M = 4.57, SD = .96) as more unique and creative than advertising imagery with a medium level (M = 3.51, SD = 1.06). There was a significant difference between high and low stereotypicality levels, t (101) = 1.38, p < .001, *95% CI* [1.18, 1.58], such that participants perceived advertising imagery with a high level of stereotypicality as more unique and creative than advertising imagery with a low level (M =3.19, SD = 1.02). There was a significant difference between the medium stereotypicality condition and low stereotypicality condition, t (101) = .32, p < .001, *95% CI* [.13, .51], such that participants perceived the advertising imagery with a medium level of stereotypicality as more unique and creative than the advertising imagery with a low level. This finding is unexpected and suggests that highly stereotypical advertisements are perceived as not only more unexpected, but more creative as well (see Figure 14).

Figure 14

Perceived creativity in low, medium, and high stereotypicality conditions.



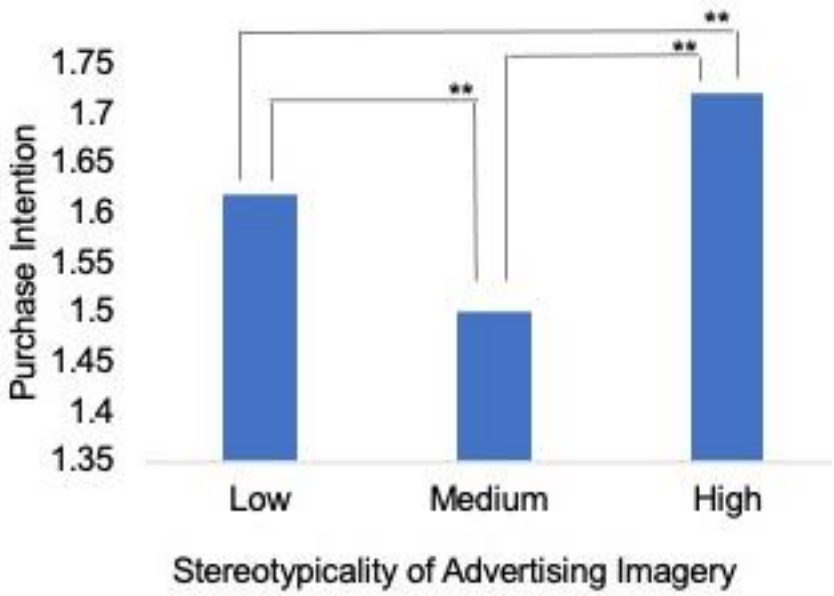
Purchase intention. We conducted a repeated-measures ANOVA to test the effects of stereotypicality of advertising imagery on the purchase intention towards the advertising products. There was a significant main effect of stereotypicality on purchase intention, F (2, 100) = 10.69, p < .001, Ω = .33. We followed up the main effect with post hoc pairwise comparisons. There was a significant difference between the high and medium stereotypicality conditions, t (101) = .21, p < .001[23], *95% CI* [.10, .32], such that the advertising imagery with a high level of stereotypicality (M = 1.72, SD = .61) induced more purchase intention than the medium level (M = 1.50, SD = .55). There was a significant difference between low stereotypicality condition and medium stereotypicality condition, t (101) = .11, p = .002, *95% CI* [.01, .22], such that the advertising imagery with a low level of stereotypicality induced more purchase intention than the medium level (M = 1.62, SD = .56). There was no significant difference, t (101) = -1.93, p = .057, *95% CI*

[23] The p values in the t-tests were already adjusted for multiple comparisons: Bonferroni.

[-.19, .00], such between the high stereotypicality condition and the low stereotypicality condition. This result demonstrates a quadratic relationship between stereotypicality and purchase intention, such that people are more willing to buy products that are marketed using high and low stereotypical imagery (see Figure 15).

Figure 15

Purchase intention in low, medium, and high stereotypicality conditions.



Recommendation intention. We conducted repeated-measures ANOVA to test the effects of stereotypicality of advertising imagery on the recommend intention towards the advertising products. There was a significant main effect of stereotypicality on recommend intention, *F* (2, 100) = 7.17, *p* = .001, *Ω* = .27. We followed up the main effect with post hoc pairwise comparisons. There was a significant difference between the high and medium stereotypicality conditions, *t* (101) = .31, *p* < .001, *95% CI* [.11, .52], such that the advertising imagery with a high level of stereotypicality (*M* = 5.62, *SD* = 1.21) induced more recommendation intention than the medium level (*M* = 5.30, *SD* = 1.20). There was a significant difference between the high stereotypicality and low stereotypicality conditions, *t* (101) = 1.20, *p* = .035, *95% CI* [.01, .39], such that the advertising imagery with a high level of stereotypicality induced more recommendation intention than the low level (*M* = 5.42, *SD* = 1.24). There was no significant difference *t* (101) = 1.28, *p* = .203, *95% CI* [-.06, .30], between the medium stereotypicality and low stereotypicality conditions. These results differed from the purchase intention results and suggested that the relationship between purchase recommendations and stereotypicality was stepwise.

Gender difference. We conducted an independent samples test to examine the gender differences. The results showed a significant gender difference on purchase intention in low stereotypicality condition, *t* (101) = 2.38, *p* = .019, *95% CI* [.05, .50], such that female participants (*M* = 1.71, *SD* = 051) reported more purchase intention towards the advertising imagery with low level of stereotypicality than male participants (*M* = 1.44, *SD* = .62). Female and male participants did not show significant differences in other measures (see Table 49). Other demographic variables did not reveal significant differences in our variables.

Table 49

Gender Difference (N $_{female}$ = 67, N $_{male}$ = 35, df = 100).

Measures	Gender	*M*	*SD*	*t*	*p* values	95% CI Lower	95% CI Upper
Purchase intention	Female	1.64	0.46	0.67	0.504	-0.14	0.28
	Male	1.57	0.60				
Recommend intention	Female	5.50	1.13	0.63	0.533	-0.32	0.61
	Male	5.35	1.09				
Unexpectedness	Female	4.30	0.82	-0.72	0.474	-0.45	0.21
	Male	4.42	0.77				
Perceived creativity	Female	3.50	0.77	1.34	0.184	-0.11	0.57
	Male	3.27	0.93				

Study 5 – 6 Discussion

Study 5 results supported hypothesis 3 and showed that the stereotype intervention, which encouraged stereotype avoidance, efficiently decreased marketers' stereotypical inferences, and improved the marketers' ability to produce original ideas. The finding is consistent with existing research that claimed positive effects of multi-identity information (Cheng et al., 2008; Gaither et al., 2015) and stereotype-inconsistent information (Gocłowska & Crisp, 2013; Wen et al., 2019; Zuo et al., 2019) on originality. Therefore, Study 5 successfully validated the laboratory link between stereotype avoidance and divergent thinking in the advertising field.

Study 6 exhibited a positive relationship between advertising stereotypicality and perceived creativity. It also showed a U-shaped function of advertising stereotypicality on purchase intention, indicating that the audience reacted most positively (high perceived creativity and high purchase intention) towards the advertisements with high stereotypicality and reacted least positively towards the medium stereotypical advertising. In addition, the purchase intention toward low stereotypical advertisements was lower than high stereotypical ones and higher than medium stereotypical ones.

Unexpectedness. The unexpectedness of the advertisement may drive the U-Shaped function. Let us consider two facts. First, the effect of the unexpectedness of advertisements followed a U-Shaped function. Second, the unexpectedness item was the first question in the scales which was more likely to drive participants' understanding of creativity and intention to purchase and less likely to be affected by the evaluation of the other items. Therefore, it is reasonable to infer that our participants probably had watched medium and low stereotypical advertising imagery more often than high stereotypical imagery, which drives the differences in perceived creativity. Accordingly, answering perceived creativity questions before purchase intention may produce a priming effect, which makes the results contradict the existing research that found negative audience reactions towards advertising stereotypes (Åkestam et al., 2017a). However, the explanation for unexpectedness does not have strong support from empirical evidence, so it warrants further exploration.

Cognitive load theory. Cognitive load theory is a possible alternative explanation for the U-shaped function for audience reaction. In our study, participants were asked to perceive two pieces of information: minority group imagery and product information. After the first round, when they knew the task content, they were likely to add and process another piece of information – advertisement evaluation. Therefore, we infer that evaluating the medium may be more difficult than the high and low stereotypical advertisements. Therefore, the hesitation in making an evaluation costs cognitive load from processing product information.

For the high and low stereotypical advertisements, processing stereotype-consistent information (stereotype activation) carried less cognitive load than processing stereotype-inconsistent information (stereotype avoidance) (Sherman & Frost, 2000; Sherman et al., 1998; Wigboldus et al., 2004). Therefore, high stereotypical advertisements may save cognitive load from processing product information. Overall, not being able to process enough product information may decrease perceived creativity and purchase intention towards medium stereotypical advertisements. On the other hand, processing product information may increase perceived creativity and purchase intentions of highly stereotypical advertisements (Åkestam, 2017; Leonard, 2014). Suppose cognitive load theory is the mechanism here, we highlight that the generalisation of the second study is limited because people are not required to stereotype scoring when perceiving advertisements in real life.

Inconsistency with Study 6 (Pilot) and Study 6. The main study finding is inconsistent with the pilot study, which found no relationship between perceived creativity and stereotypicality. The inconsistency may result from either of three reasons—first, the difference in design between an association and a causal study. Second, the different understandings of perceived creativity. In the pilot study, the perceived creativity scale implicitly links “creativity” to “difference”, “uncommonness”, “relevance”, and “meaningful”. In Study 6 (pilot), the scale implicitly links “creativity” to “unexpectedness” and “uniqueness”. Considering that one’s concept of creativity is adapted to change according to the provided information (Sternberg, 2020), the difference in measurement scale may trigger different definitions of creativity for participants in two studies. Third, the difference in the sample, such that the finding is not generalisable across types of protected characteristic groups. For instance, the perceived stereotypicality of the pilot-study participants might differ from that in Study 6. Unfortunately, we did not measure perceived stereotypicality in Study 6, so we cannot compare, and we suggest future studies be aware of this issue.

Overall, Study 6 supports the positive link between stereotype avoidance and creativity by showing a positive effect of stereotype intervention on creative thinking. The second study against the link exhibited a U-shaped function of advertising stereotypicality on perceived creativity.

Study 5 – Study 6 comparisons. The inconsistent effects of the two studies are reasonable because they employed participants with different motivations for being more creative. In detail, the participants in Study 5 (marketers) are more likely to process stereotype-inconsistent and creative information because they have strong motivations to do so (Bernstein, 1974; Edwards, 1956; Hawkins & Mothersbaugh, 2010). However, the participants in Study 6 (the public) may not have such motivations. Therefore, they may selectively perceive, process, and recall stereotype-consistent information to reinforce prior expectations toward outgroup members (Fiske, 1998; Nickerson, 1998) and save cognitive resources (Sherman & Frost, 2000; Sherman et al., 1998; Wigboldus et al., 2004; Zuberi & Bonilla-Silva, 2008).

Moreover, two studies operationalised stereotypicality and creativity in different ways. Study 5 manipulates participants' stereotypicality and wants to know how manipulation affects creative thinking. Study 6 manipulates advertisements' stereotypicality and is interested in the manipulation impacts participants' perceived creativity. Therefore, the positive link between stereotype avoidance and creativity worked in advertisement generation. However, the positive link does not work in advertisement perception. We suggest that future studies be aware of the operationalisations of stereotype avoidance and creativity when illustrating the link. We also suggest that the advertising industry be aware of the gap between generated and perceived advertising content.

Strength.

Link stereotype and creativity with transdisciplinary collaborations. Previous studies showed that stereotypical depictions still occupy mainstream media (Furnham & Paltzer, 2010; Grau & Zotos, 2016; Mastro & Stern, 2003; Plakoyiannaki & Zotos, 2009). The situation may result from a lack of validated tools that help marketers think less stereotypically and more creatively. In other words, there was a knowledge gap such that the field lacked a field study that validates the link between stereotype avoidance and creativity. Furthermore, the disciplinary fragmentation of creativity research and practice may be responsible for the knowledge gap. For example, creativity researchers often find it challenging to recruit sufficient marketers to join an experiment due to the lack of communication between academia and industry (Amabile, 1996; West et al., 2019). Therefore, we collaborated with the industry, and our transdisciplinary research projects validated the psychological link between stereotype avoidance and creativity in advertising practice and connected advertising generation and advertising perception.

Limitations and suggestions.

The lack of advertising creativity measures. However, the creativity measurement of study 5 limits its power in a specific domain. In detail, we measured marketeers' divergent thinking skills rather than the novelty or effectiveness of the product design in the advertising domain. Thus, we can infer that our workshop increased marketeers' creative potential rather than their creative performance in advertising.

Suggestions. Accordingly, we suggest that future field studies consider both creative potential based on thinking skill tests and creative performance based on novelty judgment of a product.

Furthermore, we suggest that future studies pay more attention to laypeople's perceived creativity to detect the gap between creativity generation and perception. For instance, we can include perceived creativity as a dependent variable when investigating advertising creativity (Rosengren et al., 2020; Smith et al., 2008). As we can see in our study 6, the unexpectedness or creativity scale may drive audiences' purchase intention. For example, further studies can explore whether participants' intention of stereotype avoidance affects their perceived advertisement creativity. Knowing audience perception of creativity and linking it with advertising strategy can be crucial for brand development. Besides, future studies should be aware of how they form the creativity scale when measuring perceived creativity. For instance, in our second study, linking creativity with unexpectedness and relevance may have yielded different results.

Chapter 5: Creative Thinking and Product Design in Engineering Education

The last chapter validated the association between divergent thinking and stereotype avoidance in real-life advertising. This chapter investigated the association between creative thinking and real-life engineering problem-solving. First, we discuss the importance and knowledge gaps in engineering creativity research. Then, we discuss integrative thinking's role in creative engineering design. Afterwards, we illustrate two studies (one pilot study) examining the predictive power of AUT, CRAT, and FST towards engineering students' creative design for COVID-19 prevention. Moreover, we examined the effect of comparative social feedback on product design because teamwork is important in engineering education (Lingard & Barkataki, 2011), and the presence of others is inevitable in engineering students' real-life contexts.

Literature Review

The importance of engineering creativity.

In contemporary engineering education, educators and researchers are putting more effort into developing engineering students' creativity. For engineering students, creativity captures their ability to generate original, novel, effective, and potentially valuable ideas when dealing with open-ended questions or problematic situations (Belski, 2017; Lawshe & Harris, 1960; Pereira, 1999)**.** This ability requires divergent and convergent thinking to become aware of, observe, imagine, conceptualise, and rearrange existing elements to generate a new idea that may not be apparent to the professionals in a specific engineering discipline (Belski, 2017; Farid et al., 1993; Hirshfield & Koretsky, 2020). Additionally, engineering educators point out that the ability to discover nature with imagination is as essential as domain-specific knowledge when generating sustainable solutions (Schexnayder & Anderson, 2011). Also, existing research demonstrated that incorporating creativity development in engineering classes could establish an interactive atmosphere (Stouffer et al., 2004). Moreover, creativity provides a competitive advantage for engineering students in their career development (Richards, 1998).

Creativity development in engineering education.

Since mid-90s, educators have incorporated *design* and various creativity training programmes and techniques (e.g., TRIZ; (Al'tshuller, 1999)) into engineering education curricula (Clapham, 1997; Clapham & Schuster, 1992). However, these activities did show notable positive effects in the last three decades. For example, the public still rarely links “invent” and “creative” with engineering (Wulf, 1998). Furthermore, industrial employers are still not satisfied with the creativity of engineering graduates (Cropley, 2012; Richards, 1998). Also, educators do not appreciate the creativity traits of engineering students (Cropley & Cropley, 2010), and for those educators who would like to incorporate creativity into their classes, there exists insufficient knowledge to facilitate incorporation (Amoussou et al., 2011; Darling-Hammond et al., 2012; Haertel et al., 2012). In addition, engineering students comment that their programmes of study provide little encouragement and support for creative thinking and design skills (Carpenter, 2016). Overall, creativity development does not proceed well in engineering education.

Knowledge gaps.

One of the reasons for the restricted development of creativity in engineering education may be the lack of appropriate theoretical support. For instance, existing creativity theories (e.g., divergent thinking) are disconnected from engineering practice and do not produce high-quality engineering solutions (Hirshfield & Koretsky, 2020). Additionally, the lack of empirical evidence may impede creativity development in engineering education. For instance, the impact of social factors on creativity is a widely investigated research theme (Amabile, 2018; Hennessey, 2003), exploring how competition (Shalley & Oldham, 1997), expected evaluation (Shalley & Perry-Smith, 2001; Wang et al., 2017), time pressures (Moreau & Dahl, 2005), surveillance (Amabile et al., 1990), and stereotypes (Gocłowska et al., 2013; Zuo et al., 2019) affect individual creativity. However, there are no such experiments focused on engineering students. Therefore, our research aimed to address these twin knowledge gaps. Specifically, we conducted an experiment to explore the effect of a social factor—social comparison—and the ability of other factors theoretically linked to creativity to predict engineering students’ ideation and product design creativity.

Creative engineering design – an integrative thinking process.

Previous studies employed divergent thinking and corresponding tasks to measure and develop engineering creativity (Kudrowitz & Dippo, 2013). However, divergent thinking is disconnected from real-life problem solving (Hirshfield & Koretsky, 2020). The disconnection may result from the difficulty in simulating complex engineering problem-solving[24]. Specifically, AUT asks people to produce many novel ideas that stem from one piece of information. However, engineering product design requires engineering students to be creative with several pieces of information (e.g., customer preferences, market needs, financial budgets, and technological constraints) (Howard et al., 2007; Onarheim, 2012; Scopelliti et al., 2014).

For example, the process models of engineering product design illustrate the complexity of information processing in the engineering field. One model and subsequent task—the Knowledge-Driven Design Process (KDDP) — requires engineering students to create several cognitive or knowledge spaces and fill as much information into these spaces in a seemly random order. The design process ends when there is sufficient information in the spaces to proceed with the design (Hatchuel & Weil, 2003). A second model—the Linear Type Design Process (LTDP)—simplified the process in the sequence of market needs, task analyses, conceptual designs, embodiment designs, and detailed designs. Even though LTDP prioritises market needs, engineering product design still struggles to separate such needs from the task requirements and technology (Pahl & Beitz, 1984). Both models imply that engineering product design starts from several pieces of information. For instance, the KDDP describes several pieces of information abstractly (i.e., required information in cognitive or knowledge space), while LTDP names several pieces of information specifically as market needs, tasks, and technologies.

Admittedly, some product design models imply the possibility of starting the product design from one piece or limited pieces of information and then engaging divergent thinking on each piece of information. For example, the Divergent-Convergent Style Process (DCSP) proposes that engineering students gain and evaluate collected information first, generate ideas, and solve problems based on each piece of information (Howard et al., 2007).

However, we propose that initiating a design from evaluated information may limit engineering creativity. Here is the logic. Divergent thinking in engineering is not limited to producing new ideas or solutions but includes generalising existing ideas or solutions in different contexts (David H. Cropley, 2015b). For example, an

[24] We discussed the disconnections between creative thinking and real-life creativity in Chapter 2. The disconnection remains when it comes to the engineering creativity field.

engineering solution may not be correct in one context but may be correct in others. Also, the solution may not be effective now but could be helpful at a future time point (Niiniluoto, 2016). Therefore, the process may not be completely creative if one evaluates information initially and processes the information deemed valid or correct. It would be best for engineering students to keep alternatives or "incorrect" solutions in mind while aiming to design a genuinely creative product. Therefore, we propose that starting product design from several pieces of information and continuing to integrate multiple pieces of information throughout the design process is crucial in creative engineering product design.

Our integrative thinking perspective on engineering product design is consistent with definitions and theories of engineering and technology. For instance, engineering educators highlight that engineering education encourages students to develop solutions that work with nature in a holistic and integrated manner (Schexnayder & Anderson, 2011). Engineering development requires "Ingenium", which includes the ability to integrate various bits of information quickly (Verene, 1981). Engineering aims to analyse and synthesise learned information (Felder, 1987; Isaksen & Parnes, 1985). The Integrated Creative-Design process model (ICDP) argues that the creative thinking process should be integrated into the product design process, including analysing task, concept, and embodiment (Howard et al., 2007).

The integration in engineering product design may look like convergent thinking since both require integrating several pieces of information. In detail, the convergent thinking task CRAT provides participants with three words and asks them to produce one correct answer that could form a compound word with each of the given words. The difference between engineering integration and convergent thinking is at their endpoint. Specifically, engineering integration is an open-ended question that allows multiple answers (Belski, 2017), while CRAT is a close-ended question with a single correct answer (Bowden & Jung-Beeman, 2003). Therefore, CRAT as a convergent thinking task may not be able to represent engineering creative product design. From another perspective, convergent thinking is conceptualised as evaluative thinking in some engineering design models. Part of the creative process leads to one or several appropriate and valuable solutions (Cropley, 2006). Even though we admit that evaluative thinking starts from several pieces of information and allows several correct answers, it is far from creative thinking. *Evaluation* is a process that uses numerous criteria to select from existing options. Novelty, a fundamental element of creativity, is not mandatory in the evaluative process.

FST simulates the integrative process, grasps the creative process in engineering, and captures two dimensions of creative ideas (i.e., novelty and effectiveness). Therefore, we predicted that:

Hypothesis 5: *FST predicts creativity in engineering product design. The better the FST performance, the higher the creativity in engineering product design.*

Empirical evidence showed that comparative social feedback could impact individual creativity (see Chapter 3 for literature review). Therefore, we ask an exploratory question here: *Does social comparison feedback alter creativity in engineering product design?*

To test the hypotheses and answer the exploratory research question, we conducted two online experiments. The pilot study aimed to validate two engineering creative design tasks. In addition, the main study aimed to 1) examine the predictive power of AUT, CRAT, and FST toward engineering creative product design and 2) examine the effect of social comparison on engineering creativity.

Study 7 (Pilot) Method

In a pilot study, we generated and validated two engineering design tasks (i.e., an ideation task (IT) and a product design task (PDT)). In detail, we conducted an online observational study asking engineering students to finish FST, AUT, CRAT, and then IT and PDT. We measured the performance in the five tasks. We also collected self-reported responses regarding creative self-efficacy, creative personality identity, and grit.

Participants.

To collect data, we asked the administrators of all engineering departments at University College London (UCL) to circulate a piece of research invitation to all students in their departments. Accordingly, the administrators in the department of civil, environmental, and geomatic engineering and the department of biochemical engineering circulated the research invitation.

We received responses from 14 students (women = 6) from UCL. Most participants were master's students (68.75%), and 31.25 % were undergraduate students. All participants gave consent and opted to receive individual creativity feedback via email and enter a random draw to receive a £30 Amazon voucher. The UCL Ethics Committee approved the study.

Materials.

AUT. We measure participants' AUT performance as we did in Study 1, but with two adjustments. First, participants finished six rounds of AUT in Study 1, but they finished two rounds of AUT in this study. In detail, we employed "tire" and "newspapers" as AUT questions, and the questions were randomly selected. We decided to avoid participants' fatigue because the current study has more tasks than Study 1. Second, we excluded one measure - appropriate frequency since Study 1 showed that originality and frequency strongly correlated with each other. Also, the current study did not have enough sample size to create an informative and objective answer pool.

CRAT. We measured CRAT performance as we did in Study 1, but with two adjustments. First, we curtailed the round number from six to two for the same reason we discussed in AUT. In detail, participants were asked to finish the non-repeated two rounds of CRAT. Each round was made of five questions. Second, we prepared seven language versions of CRAT rather than pre-screened English native speakers. For instance, participants chose from Chinese (Wu & Chen, 2017), Dutch (Chermahini et al., 2012), English (Bowden & Jung-Beeman, 2003), Japanese (Seki et al., 2010), Romania (Olteţeanu et al., 2019), Russian (Toivainen et al., 2019), and Spanish (Peláez-Alfonso et al., 2020) based on their first language. If none of them was their first language, participants selected the one with whom they were most familiar. We did this adjustment to enlarge the sample size because UCL was an international school, and many students did not identify themselves as English native speakers. Please see Appendix S for *Compound Remote Associate Test in Different Languages*.

FST. We measured FST performance as we did in Study 1, but with two adjustments. The adjustments followed the logic of AUT adjustments. In detail, we again reduced six rounds to two rounds. We also excluded appropriate frequency and accurate frequency as measures. In this study, we employed "interact/immerse/recreate" and "protect/entertain/comfort" as questions.

IT. We generated IT that measured creativity in idea generation. At the beginning of the IT, we introduced participants to two ways that COVID-19 was transmitted. Specifically, they were told:

"*Direct Transmission: COVID-19 is transmitted by respiratory particles of someone infected with COVID-19 that reach healthy people directly. The particles are emitted through sneezing, coughing, or even talking. These droplets from an infected person are packed with millions of viral particles on whom fall in close range and infect whatever they land.*

Fomite Transmission: COVID-19 is transmitted by touching an infective surface (e.g., keys, telephones, power button) and touching your face."

Participants were asked to think of novel and valuable COVID-19 prevention products for a cosy restaurant according to the background information. We asked them not to criticise any of their ideas and write down as many ideas as possible.

We employed the marking system and the measures of FST to measure the performance in IT. The only difference was the criteria of appropriate answers and accurate answers. In detail, we marked answers that prevented at least one way of COVID-19 transmission as appropriate answers. We also marked answers that prevented both ways of COVID-19 transmission in a cosy restaurant as the most accurate answers. Like FST, the higher scores the IT measures, the greater the creative ideation displayed by the participant.

PDT. The PDT was a follow-up task of IT. We provided participants with the background information of COVID-19 transmission and the ideas they came up with in IT. Based on this information, participants were asked to design a novel and effective COVID-19 prevention product for a cosy restaurant by answering four long-text format questions - "*What is your product?*" "*What are the functions of this product?*" "*What are the materials needed?*" and "*How can we generalize this product to other contexts?*"

We employed the Creative Solution Diagnosis Scale (CSDS) (Cropley & Kaufman, 2012) to measure participants' product design creativity. The scale measured five dimensions of a product: relevance and effectiveness, problematization, propulsion, elegance, and genesis. The relevance and effectiveness dimension evaluated to what extent a product was valuable and appropriate in a specific context. The problematization dimension evaluated to what extent a product could detect and solve the issues with the existing solutions. The propulsion dimension measured how much a product added novel contributions to existing solutions. Both problematization and propulsion were in the novelty dimension. The elegance dimension expected a product to be qualified from various perspectives (e.g., safe to use, consistent, and sustainable). Finally, the genesis dimension assessed to what extent a product was helpful in the current situation for which they were generated and could apply in other unrelated situations. The measure of PDT was the scores in each dimension and totals. We also measured RT. The higher the score of measures in PDT, the greater the participant's product design creativity. Please see Appendix T for *Engineering Creative Design – Data Collection, Marking, and Cleaning*.

Short Scale of Creative Self. We measured creative self-concept variables such as creative self-efficacy (CSE) and creative personal identity (CPI) using a well-established short scale of creative self (SSCS) (Karwowski, 2014; Karwowski et al., 2012). CSE was about the self-description of creative abilities (Karwowski, 2011) and CPI was about the importance of creativity in one's self-description (Jaussi et al., 2007). We provided eleven statements that were used to describe oneself. An example of CSE was "*I trust my creative abilities*". An example of CPI was "*my creativity is important for who I am*". Participants decided how each statement described themselves on a 5-point Likert scale.

We calculated the creative self-concept variables based on the 5-point Likert scale. Considering the internal consistency of 6 CSE items was low (*Cronbach's α* = .69) and achieved an acceptable level excluding item 6 (*Cronbach's α* = .79), we summed the ratings on items 3, 4, 5, 8, and 9 for the level of CSE. The higher the summed ratings, the higher the self-reported creative abilities. We summed the rating of items 1, 2, 7, 10, and 11 for the level of CPI (*Cronbach's α* = .91). The higher the summed ratings, the importance of creativity in self-description. Please see Appendix U.

Scale of Grit. Considering the important role that the inconsistency of short-term interests and long-term interests and the perseverance of effort (i.e., grit) played in engineering students' retention and academic achievements (see a review (Direito et al., 2021)), we measured the grit of our participants who were engineering students. We gave participants twelve statements to describe themselves on a well-established grit scale (Duckworth & Quinn, 2009), An example of inconsistent interests was "*I often set a goal but later choose to pursue a different one*". An example of perseverance was that "*I finish whatever I begin*". Participants decided how each statement described themselves on a 5-point Likert scale.

We calculated the grit based on participants' ratings on the 5-point Likert scale. Specifically, we summed the ratings on the first six statements for the level of interest's inconsistency (*Cronbach's α* = .89). The higher the summed ratings, the higher the inconsistency of short-term interests and long-term goals. Considering the internal consistency of six CSE items was low (*Cronbach's α* = .63) and achieved an acceptable level when excluded item 8 and item 10 (*Cronbach's α* = .74), we summed the ratings on items 7, 9, 11, and 12 for perseverance. The higher the summed ratings, the greater the perseverance of effort. Please see Appendix V.

Procedure.

In the beginning, all participants read an overview and an information sheet of the experiment and consented to participate. Next, participants were asked to complete the AUT, CRAT, and FST randomly. As we did in Study 1, we called AUT and CRAT "Usages Task" and "Word Puzzle Task", respectively. Considering that we have disseminated FST in Study 1, which may have built publicity, we also changed the name of FST to "Object Task". Each task consisted of two pages. Participants read a task instruction on the first page, and they had to pass an instruction test to proceed to the second page. On the second page, participants were given 5 minutes to finish one round of the main task.

Following AUT, CRAT, and FST, all participants answered three questions about their experience thus far. The first question asked the participants to select the answer sources of creative thinking tasks, either original ideation, memory extraction, or online resource. The second question asked them to select the effort they put into the creative-thinking tasks: all, some, or none. Finally, the third question asked whether participants expected to receive a performance evaluation of the three creative thinking tasks.

After the experience report, all participants completed the IT and PDT sequence, with 5 and 25 minutes allocated to the two tasks, respectively. Once participants completed the tasks, they reencountered the game experience questions. They had to report the answer sources, employ effort, and evaluate expectations for IT and PDT.

Next, participants finished the grit scale and the SSCS. We then collected demographic information. In the end, participants decided whether they requested performance feedback from us and whether they entered a random draw to receive a

£30 Amazon voucher. We also provided a debrief form in which we disclosed and explained the areas of deception. Please see Appendix W for *Study 7 (Pilot) Script*.

Study 7 (Pilot) Results

Dataset preparation.

Inter-rater reliability. A high degree of inter-rater reliability was found between the two judges for most measures (*Cronbach's α* > .80). Although the accurate originality and effectiveness yielded relatively low inter-rater reliability, we did not exclude them because they were acceptable (> .60). However, we suggest the following studies be aware of these two measures' low and acceptable inter-rater reliability (see Table 50).

Table 50

Intraclass Correlation Coefficients (ICC) with Two-way Mixed Effects Model for Measures in AUT, FST, IT, and PDT (N of items = 2).

Measures	*ICC*	*95% CI*	
		Lower Bound	Upper Bound
AUT appropriate fluency	1.00	.99	1.00
AUT appropriate originality	.94	.82	.98
AUT appropriate flexibility	.93	-.07	.99
FST appropriate fluency	.98	.89	.99
FST appropriate originality	.86	.61	.95
FST appropriate flexibility	.91	.73	.97
FST appropriate effectiveness	.86	.52	.96
FST accurate fluency	.90	.70	.97
FST accurate originality	.81	.46	.93
FST accurate flexibility	.86	.62	.95
FST accurate effectiveness	.61	-.18	.86
IT appropriate fluency	.99	.96	1.00
IT appropriate originality	.94	.81	.98
IT appropriate flexibility	.95	.85	.98
IT appropriate effectiveness	.97	.92	.99
IT accurate fluency	.95	.84	.98
IT accurate originality	.71	.15	.90
IT accurate flexibility	.88	.65	.96
IT accurate effectiveness	.94	.83	.98
PDT relevance & effectiveness	.94	.84	.98
PDT problematisation	.93	.79	.98
PDT propulsion	.85	.54	.95
PDT elegance	.85	.47	.95
PDT generalisation	.92	.75	.97

a. Cronbach's α is the average measures of intraclass correlation coefficients.

b. Cronbach's α using an absolute agreement definition.

Data screening. We followed the data screening rules for Study 1. We excluded the response from one participant who did not finish the experiment (put effort into four tasks and wrote nothing in one task). There were 15 responses from participants who stayed in the analysis pool after the data exclusion (N = 15).

Statistical assumptions. To prepare an appropriate dataset for statistical analysis, we did a series of assumption tests on the variables in all five tasks and questionnaires. All the variables met the assumption of non-zero variances. There was no outlier in the variables. The skewness & kurtosis and Shapiro–Wilk test indicated that the data contained approximately normally distributed error (see Table 51 – 52)..

Table 51

Descriptive Statistic for FST, AUT, CRAT, and Creative Self Efficacy (N = 15).

	M	*SD*
AUT appropriate fluency	7.70	4.02
AUT appropriate flexibility	6.38	2.41
AUT appropriate originality	1.99	0.03
CRAT accurate fluency	3.20	7.89
FST appropriate fluency	3.13	2.41
FST appropriate flexibility	2.42	1.14
FST appropriate originality	1.81	0.17
FST appropriate effectiveness	1.74	0.12
FST accurate fluency	1.92	1.24
FST accurate flexibility	1.53	0.64
FST accurate originality	1.95	0.24
FST accurate effectiveness	1.67	0.14
IT appropriate fluency	6.10	6.15
IT appropriate flexibility	3.40	1.72
IT appropriate originality	1.09	0.12
IT appropriate effectiveness	2.31	0.49
IT accurate fluency	1.13	0.84
IT accurate flexibility	0.90	0.36
IT accurate originality	1.18	0.68
IT accurate effectiveness	1.78	1.27
PDT total	65.23	202.57
PDT relevance & effectiveness	10.20	6.74
PDT problematisation	6.73	4.92
PDT propulsion	12.93	11.07
PDT elegance	22.47	29.70
PDT generalisation	12.90	4.19
CSE	3.41	0.47
CPI	3.57	0.77
Consistency for goals	3.38	0.93
Perseverance	3.80	0.65

Table 52

Normality Tests for Measures of AUT, CRAT, FST, and Creative Self-efficacy.

			Shapiro-Wilk	
	Skewness	Kurtosis	Statistics	*p* value
AUT appropriate fluency	0.65	-0.34	.93	.252
AUT appropriate flexibility	0.25	-0.82	.98	.920
AUT appropriate originality	-0.37	-0.76	.94	.415
CRAT accurate fluency	0.12	-1.59	.86	.023
FST appropriate fluency	0.46	-0.88	.92	.213
FST appropriate flexibility	0.16	-0.64	.97	.777
FST appropriate originality	0.17	0.02	.96	.740
FST appropriate effectiveness	0.32	2.39	.93	.267
FST accurate fluency	0.54	-0.80	.93	.265
FST accurate flexibility	1.41	1.95	.86	.024
FST accurate originality	0.28	0.57	.96	.756
FST accurate effectiveness	0.48	0.56	.95	.450
IT appropriate fluency	-1.05	1.68	.92	.162
IT appropriate flexibility	-1.06	2.26	.92	.187
IT appropriate originality	-2.06	7.12	.75	.001
IT appropriate effectiveness	-2.86	9.35	.66	.000
IT accurate fluency	0.27	-0.54	.92	.179
IT accurate flexibility	-0.12	-0.56	.91	.140
IT accurate originality	0.42	0.68	.91	.135
IT accurate effectiveness	-1.08	-0.79	.69	<.001
PDT total	-0.50	2.72	.93	.226
PDT relevance & effectiveness	-1.38	1.09	.82	.006
PDT problematisation	1.52	4.30	.81	.005
PDT propulsion	0.12	0.94	.96	.613
PDT elegance	-0.65	1.83	.95	.509
PDT generalisation	-0.71	2.20	.88	.052
CSE	0.39	-0.92	.94	.414
CPI	-0.39	-0.11	.96	.604
Consistency for goals	-0.23	-1.06	.96	.638

Perseverance	-0.52	0.22	.95	.445

*p < .05 (2-tailed). **p < .01 (2-tailed).

Pearson correlation analysis.

In this section, we present the results of the Pearson correlation analysis. We first presented the correlations among the measures in creative thinking tasks (i.e., AUT, CRAT and FST) (see Table 53). We then presented the correlations between creative thinking and engineering design.

Table 53

Correlations among AUT, CRAT, and FST Measures (All Participants, N = 15)

	1	2	3	4	5	6	7	8	9	10	11	12
1. AUT appropriate fluency	-											
2. AUT appropriate flexibility	.90**	-										
3. AUT appropriate originality	.45	.62*	-									
4. CRAT accurate fluency	.36	.39	.27	-								
5. FST appropriate fluency	.36	.40	.17	.32	-							
6. FST appropriate flexibility	.40	.43	.10	.36	.91**	-						
7. FST appropriate originality	-.05	-.02	.28	-.11	.09	-.16	-					
8. FST appropriate effectiveness	.12	.18	.20	.36	-.02	.07	-.36	-				
9. FST accurate fluency	.42	.53*	.35	.26	.91**	.79**	.28	.07	-			
10. FST accurate flexibility	.43	.51	.10	.41	.66**	.85**	-.19	.21	.69**	-		
11. FST accurate originality	-.05	-.11	.08	-.19	-.05	-.24	.85**	-.37	.07	-.23	-	
12. FST accurate effectiveness	.00	-.13	-.09	.47	-.19	-.13	-.38	.77**	-.26	-.09	-.31	-

*p < .05 (2-tailed). **p < .01 (2-tailed).

Correlations of AUT. The results showed significant positive associations among the measure of AUT. Specifically, there was a high association between appropriate fluency and appropriate flexibility. There was a moderate association between appropriate originality and appropriate flexibility. There was no association between appropriate fluency and appropriate originality.

Correlations of FST. The results showed significant positive associations among the quantitative measures of FST. Specifically, appropriate fluency, appropriate flexibility, accurate fluency, and accurate flexibility were moderately, highly, or strongly associated. Also, there were significant positive associations among the qualitative measure of FST. Specifically, there were high associations between appropriate and accurate originality and between appropriate and accurate effectiveness. Also, originality measures and effectiveness measures were not associated with each other.

Correlations between FST, AUT, and CRAT. The results showed a significant positive association between the FST and AUT measures. Specifically, AUT appropriate flexibility was moderately associated with FST accurate fluency. Besides, none of the others was significant.

Correlations of IT. The results showed significant positive associations among the appropriate-dimension measures of IT (see Table 54). All four measures were either moderately or highly associated with each other. Also, there were significant positive associations among the accurate-dimension measures of IT. Most measures were either moderately or highly associated with each other, except for that between accurate flexibility and originality. Besides, there were significant positive associations between the appropriate-dimension measures and the accurate-dimension measures. Specifically, appropriate fluency was moderately associated with accurate fluency, originality, and effectiveness but not with accurate flexibility. Appropriate flexibility was moderately associated with accurate originality but not with accurate fluency, flexibility, and effectiveness. Appropriate originality was moderately associated with accurate fluency, accurate originality, and accurate flexibility but not associated with accurate effectiveness. Finally, appropriate effectiveness was not associated with any measures in the accurate dimension.

Table 54

Correlations among IT Measures (All Participants, N = 15)

	1	2	3	4	5	6	7	8
1. Appropriate fluency	-							
2. Appropriate flexibility	.67**	-						
3. Appropriate originality	.69**	.87**	-					
4. Appropriate effectiveness	.64*	.56*	.77**	-				
5. Accurate fluency	.56*	.40	.57*	0.35	-			
6. Accurate flexibility	.51	.48	.53*	0.48	.80**	-		
7. Accurate originality	.57*	.54*	.70**	0.37	.70**	.49	-	
8. Accurate effectiveness	.53*	.27	.51	0.36	.77**	.58*	.68**	-

*p < .05 (2-tailed). **p < .01 (2-tailed).

Correlations of PDT. The results showed that there were significant positive associations among the PDT measures. Specifically, the total score in PDT was strongly associated with propulsion, elegance, and generalisation and was highly associated with relevance & effectiveness and problematisation. In addition, the subcategories were highly associated (see Table 55)..

Table 55

Correlations for PDT Measures (All Participants, N = 15).

PDT measures	1	2	3	4	5	6
1. Total	-					
2. Relevance & effectiveness	.83**	-				
3. Problematisation	.83**	.58*	-			
4. Propulsion	.93**	.63*	.78**	-		
5. Elegance	.95**	.75**	.68**	.89**	-	
6. Generalisation	.95**	.88**	.85**	.84**	.84**	-

$^*p < .05$ (2-tailed). $^{**}p < .01$ (2-tailed).

Correlations between IT and PDT. The results showed that there were significant negative associations between the IT measure and PDT measures. Specifically, IT appropriate flexibility was moderately associated with PDT propulsion, $r = -.53$, $p = .044$, and PDT elegance $r = -.52$, $p = .046$. None of the other associations between IT and PDT was significant (see Table 56).

Table 56

Correlations Cross IT and PDT Measures (All Participants, N = 15).

	PDT total	PDT effectiveness	PDT problematisation	PDT propulsion	PDT elegance	PDT generalisation
IT appropriate fluency	-.28	-.28	-.14	-.31	-.27	-.20
IT appropriate flexibility	-.46	-.24	-.27	-.53*	-.52*	-.36
IT appropriate originality	-.17	-.10	-.09	-.18	-.19	-.16
IT appropriate effectiveness	-.19	-.17	-.15	-.20	-.14	-.20
IT accurate fluency	.12	.18	-.05	.11	.16	.08
IT accurate flexibility	-.11	.13	-.18	-.20	-.12	-.11
IT accurate originality	.07	-.04	.02	.11	.13	-.04
IT accurate effectiveness	.18	.04	.14	.26	.20	.11

*p < .05 (2-tailed). **p < .01 (2-tailed).

Correlations between creative thinking and design. The results showed a significant negative association between the measures in creative thinking tasks and engineering design tasks. Specifically, CRAT accurate fluency, r = -.59, p = .02, was moderately associated with PDT relevance and effectiveness. FST appropriate effectiveness was moderately associated with IT accurate flexibility, r = -.52, p = .047. FST accurate effectiveness, r = -.65, p = .008, and none of the other associations between creative thinking and engineering design measures was significant. AUT measures were not associated with IT and PDT measures (see Table 57 – 60).

Table 57

Correlations Cross AUT and IT, and AUT and PDT Measures (All Participants, N = 15).

	AUT appropriate fluency	AUT appropriate flexibility	AUT appropriate originality
IT appropriate fluency	0.11	0.14	0.01
IT appropriate flexibility	-0.20	-0.12	0.23
IT appropriate originality	-0.02	0.06	0.20
IT appropriate effectiveness	-0.20	-0.14	-0.16
IT accurate fluency	0.24	0.11	-0.07
IT accurate flexibility	-0.07	-0.28	-0.21
IT accurate originality	0.04	0.04	0.17
IT accurate effectiveness	0.28	0.14	-0.02
PDT total	0.37	0.44	0.32
PDT effectiveness	0.20	0.27	0.38
PDT problematisation	0.30	0.36	0.36
PDT propulsion	0.50	0.51	0.30
PDT elegance	0.29	0.37	0.15
PDT generalisation	0.41	0.50	0.45

*p < .05 (2-tailed). **p < .01 (2-tailed).

Table 58

Correlations Cross CRAT and IT, and AUT and PDT Measures (All Participants, N = 15).

	CRAT accurate fluency
IT appropriate fluency	.40
IT appropriate flexibility	.19
IT appropriate originality	.22
IT appropriate effectiveness	.15
IT accurate fluency	-.05
IT accurate flexibility	-.16
IT accurate originality	.13
IT accurate effectiveness	.10
PDT total	-.44
PDT relevance & effectiveness	-.59*
PDT problematisation	-.30
PDT propulsion	-.23
PDT elegance	-.45
PDT generalisation	-.43

*p < .05 (2-tailed). **p < .01 (2-tailed).

Table 59

Correlations Cross FST appropriate dimension, IT and PDT (All Participants, N = 15).

	FST appropriate fluency	FST appropriate flexibility	FST appropriate originality	FST appropriate effectiveness
IT appropriate fluency	.01	.17	-.14	-.22
IT appropriate flexibility	-.09	-.13	.18	-.17
IT appropriate originality	-.08	-.09	.21	-.03
IT appropriate effectiveness	.09	.16	.17	-.05
IT accurate fluency	-.04	.07	.04	-.26
IT accurate flexibility	.02	.01	.19	-.52*
IT accurate originality	-.48	-.42	.24	-.01
IT accurate effectiveness	-.17	-.05	-.21	.16
PDT total	-.10	-.04	-.02	.08
PDT effectiveness	.04	.01	.17	-.28
PDT problematisation	-.18	-.09	-.28	.24
PDT propulsion	-.13	-.04	-.08	.33
PDT elegance	-.13	-.08	.06	.05
PDT generalisation	.00	.08	-.09	.02

*$p < .05$ (2-tailed). **$p < .01$ (2-tailed).

Table 60

Correlations Cross FST accurate dimension and IT and PDT (All Participants, N = 15).

	FST accurate fluency	FST accurate flexibility	FST accurate originality	FST accurate effectiveness
IT appropriate fluency	-.13	.18	-.29	-.01
IT appropriate flexibility	-.08	-.19	-.17	.03
IT appropriate originality	.01	-.02	-.22	.06
IT appropriate effectiveness	.10	.25	-.10	.04
IT accurate fluency	-.09	-.02	-.22	-.19
IT accurate flexibility	-.13	-.18	-.03	-.22
IT accurate originality	-.40	-.30	.02	.06
IT accurate effectiveness	-.28	-.13	-.36	.30
PDT total	.07	.12	-.01	-.37
PDT effectiveness	.17	-.03	.07	-.65**
PDT problematisation	-.06	.15	-.22	-.10
PDT propulsion	.07	.18	-.03	-.08
PDT elegance	.04	.09	.09	-.40
PDT generalisation	.14	.19	-.08	-.43

*p < .05 (2-tailed). **p < .01 (2-tailed).

Correlations of self-reported measures and the others. The results showed no association between self-reported measures (i.e., CSE, CPI, Inconsistency of goals, and perseverance) and the measures in the five creative tasks (see Table 61).

Table 61

Correlations Cross Self-Report Measures and AUT, CRAT, FST, IT, and PDT (All Participants, N = 15).

	CSE	CPI	Consistency for goals	Perseverance
AUT appropriate fluency	-.19	-.28	.10	.03
AUT appropriate flexibility	-.06	-.24	.01	.03
AUT appropriate originality	.20	-.02	.24	.07
CRAT accurate fluency	-.09	.07	-.14	.10
FST appropriate fluency	-.19	-.11	.08	.22
FST appropriate flexibility	.10	.23	-.03	.14
FST appropriate originality	-.21	-.01	-.09	-.13
FST appropriate effectiveness	.00	-.33	.25	.18
FST accurate fluency	-.09	.03	-.28	.07
FST accurate flexibility	.26	.32	-.19	.20
FST accurate originality	-.13	.09	.16	-.17
FST accurate effectiveness	-.14	-.31	.30	.06
IT appropriate fluency	.12	.17	-.09	-.26
IT appropriate flexibility	.03	.08	-.22	-.44
IT appropriate originality	.00	-.03	-.39	-.26
IT appropriate effectiveness	.26	.33	-.44	-.15
IT accurate fluency	-.31	-.21	-.11	-.08
IT accurate flexibility	-.15	.09	-.21	-.04
IT accurate originality	-.21	-.30	.04	-.08
IT accurate effectiveness	-.19	-.37	.14	.03
PDT total	.26	-.06	-.03	.32
PDT effectiveness	.29	.15	-.08	.22
PDT problematisation	.49	.03	-.02	.27
PDT propulsion	.10	-.19	.01	.36
PDT elegance	.12	-.17	-.05	.33
PDT generalisation	.41	.10	.02	.23

**p* < .05 (2-tailed). ***p* < .01 (2-tailed).

Single linear regressions.

In this section, we present the results of linear regressions. We firstly conducted and presented the results of linear regressions based on significant associations between measures in creative thinking tasks and engineering design tasks. For instance, since the two variables were significantly associated, we employed the FST's accurate effectiveness to predict PDT relevance and effectiveness. After that, we conducted and presented the results of multiple linear regressions. We employed selected FST measures, AUT measures, and CRAT accurate fluency to predict each measure of IT and PDT. We talked about the details of predictors selections below when applicable.

Statistical assumptions. We further examined statistical assumptions for dependent variables in linear regressions (i.e., IT and PDT measures). The data met the assumption of independent errors as the Durbin-Watson value was larger than one and less than three. The normal P-P plot of standardised residuals showed not completely points on the line but close. The scatterplot of standardised predicted values showed that the variables met the assumptions of homogeneity of variance without heteroscedasticity issues. Please see Table 62.

Table 62

Durbin-Watson and Standardized Residuals for IT Measures and PDT Measures.

		Std. Residual	
Measures	Durbin-Watson	Minimum	Maximum
IT appropriate fluency	1.79	-2.46	1.49
IT appropriate originality	1.66	-3.11	1.43
IT appropriate flexibility	1.46	-2.58	1.59
IT appropriate effectiveness	1.56	-3.32	0.76
IT accurate fluency	1.64	-1.39	1.86
IT accurate originality	1.59	-1.45	2.26
IT accurate flexibility	1.61	-1.28	2.00
IT accurate effectiveness	1.31	-1.62	0.87
PDT relevance and effectiveness	1.96	-2.06	0.95
PDT problematisation	2.57	-1.28	2.53
PDT propulsion	1.98	-1.63	1.78
PDT elegance	2.26	-2.09	1.41
PDT generalisation	2.22	-1.98	1.65

Predictive power of FST effectiveness. The correlation results showed that FST appropriate effectiveness was significantly associated with IT accurate flexibility. Therefore, we conducted a single linear regression to examine the predictive power of FST appropriate effectiveness on IT accurate flexibility. The results showed that FST appropriate effectiveness explained a significant amount of the variance in IT accurate flexibility, F (1, 13) = 4.80, p = .047, R^2 = .27, $R^2_{Adjusted}$ = .21. In detail, FST appropriate effectiveness was a significant predictor of IT accurate flexibility, B = -0.89, *Beta* = -.52, t (13) = -2.13, p = .047, *95% CI* [-.77, -.01].

The correlation results also showed that FST accurate effectiveness was significantly associated with PDT relevance and effectiveness. Therefore, we conducted a single linear regression to examine the predictive power of FST accurate effectiveness on PDT relevance and effectiveness. The results showed that FST accurate effectiveness explained a significant amount of the variance in PDT relevance and effectiveness, F (1,13) = 9.61, p = .008, R^2 = .43, $R^2_{Adjusted}$ = .38. In detail, FST accurate effectiveness was a significant predictor of PDT relevance and effectiveness, B = -4.49, *Beta* = -.65, t (13) = -3.10, p = .008, *95% CI* [-7.61, -1.36].

Predictive power of CRAT. The correlation results also showed that CRAT accurate fluency was significantly associated with PDT relevance and effectiveness. Therefore, we conducted a single linear regression to examine the predictive power of CRAT accurate fluency on PDT relevance and effectiveness. The results showed that CRAT accurate fluency explained a significant amount of the variance in PDT relevance and effectiveness, F (1,13) = 7.07, p = .020, R^2 = .35, $R^2_{Adjusted}$ = .30. In detail, CRAT accurate fluency was a significant predictor of PDT relevance and effectiveness, B = -.55, *Beta* = -.59, t (13) = -2.66, p = .02, *95% CI* [-.10, -.13].

Predictive power of FST. To prepare appropriate predictors for multiple linear regression, we conducted collinearity diagnostics for FST measures. First, we excluded the measures with the highest VIF until the VIF of all measures were below .10. Then, we excluded the measures with the lowest tolerance value until the tolerance value of all measures was above .20. We took out appropriate fluency, flexibility, and originality based on the above logic. The results showed that multicollinearity was not a concern for the rest of the five measures of FST (see Table 63).

Table 63

Collinearity Statistics for The FST Measures in The Multiple Linear Regression.

FST measures	Tolerance	VIF
Appropriate effectiveness	0.44	2.28
Accurate fluency	0.43	2.34
Accurate originality	0.77	1.31
Accurate flexibility	0.31	3.26
Accurate effectiveness	0.31	3.27

We conducted a series of multiple linear regression to examine the predictive power of FST measures on each of the IT and PDT measures. The results showed that the five measures in FST did not explain significant variance in any of the IT and PDT measures. Considering the small sample size of this study, we reported the F models with relatively high R^2 values (> .40) and contained significant predictors or suppressors. The results showed that the five FST measures explained an amount of the variance in PDT relevance and effectiveness, F (5,9) = 3.09, p = .068, R^2 = .63, $R^2_{Adjusted}$ = .43, PDT propulsion, F (5,9) = 1.58, p = .261, R^2 = .47, $R^2_{Adjusted}$ = .17, PDT elegance, F (5,9) = 2.64, p = .098, R^2 = .59, $R^2_{Adjusted}$ = .37, and PDT generalisation, F (5,9) = 2.17, p = .148, R^2 = .55, $R^2_{Adjusted}$ = .29.

In detail, FST appropriate effectiveness was a significant predictor of PDT relevance and effectiveness, *B* = -8.68, *Beta* = -1.26, *t* (9) = -3.45, *p* = .007, *95% CI* [-14.38, -23.00], PDT propulsion, *B* = -9.06, *Beta* = -1.03, *t* (9) = -2.34, *p* = .044, *95% CI* [-17.84, -.84], PDT elegance, *B* = -19.63, *Beta* = -1.36, *t* (9) = -3.54, *p* = .006, *95% CI* [-32.17, -7.1], PDT generalisation, *B* = -6.98, *Beta* = -1.29, *t* (9) = -3.17, *p* = .011, *95% CI* [-11.96, -2.00]. Also, FST accurate effectiveness significantly strength the predictive power of the models of PDT propulsion, *B* = 11.04, *Beta* = 1.17, *t* (9) = 2.66, *p* = .026, *95% CI* [1.66, 20.43], PDT elegance, *B* = 17.98, *Beta* = 1.16, *t* (9) = 3.03, *p* = .014, *95% CI* [4.57, 31.39], PDT generalisation, *B* = 5.72, *Beta* = .99, *t* (9) = 2.43, *p* = .038, *95% CI* [.40, 11.05].

Predictive power of AUT. We conducted collinearity diagnostics for AUT measures to prepare appropriate predictors for multiple linear regression. Based on the measure-exclusion logic illustrated above, we took out appropriate flexibility. The results showed that multicollinearity was not a concern for the other five measures of AUT. We conducted a series of multiple linear regression to examine the predictive power of AUT measures on each of the IT and PDT measures. The results showed that the two measures in AUT did not explain significant variance in any of the IT and PDT measures. None of the F models contained a significant predictor (see Table 64).

Table 64

Collinearity Statistics for The AUT Measures in The Multiple Linear Regression.

AUT measures	Tolerance	VIF
Appropriate fluency	0.99	1.01
Appropriate originality	0.99	1.01

FST vs. AUT. In the above analysis of FST and AUT, FST had five predictors while AUT had two. Therefore, we conducted further analysis on FST's appropriate fluency and originality, making the analysis of FST and AUT more comparable. Multi-collinearity was not a concern for the two measures. We conducted a series of multiple linear regressions to examine the predictive power of the two FST measures on each IT and PDT measure. The results showed that the two measures in FST did not explain significant variance in any of the IT and PDT measures. None of the F models contained a significant predictor.

Study 7 Method

We employed a between-subjects design where we manipulated comparative social feedback. Participants either received top-ranking feedback (downward comparison group), bottom-ranking feedback (upward comparison group), or no feedback (control group). We measured creativity thinking using the AUT, CRAT, and FST, and creative product design using the IT and PDT.

Participants.

Based on a pre-test power analysis, we intended to recruit 162 participants to obtain an effect size *f* of .31 at the *p* = .05 level. Using the online participant subject pool Prolific Academic, we recruited 176 participants (women = 51; M_{age} = 23.77), with 124 engineering students and 52 engineers who had a bachelor's degree or above. According to the pre-screen self-report, 66 participants were English Native speakers, and 110 were at proficiency level. The participant gave consent and received £6.50 for 1 hour of their time. The University Ethics Committee approved the study. Based on the experimenters' one-by-one check and the outlier detection, we found that 18 participants copied and pasted online answers, which left us 144 participants (women = 38, M_{age} = 28.34) in sample pool.

Post-hoc power analyses using GPower suggested that we had enough power (.85) to detect the effect at the p = .05 level (*partial* $\eta 2$ = .055, *effect size f* = .24, *number of groups* = 3, *number of measurements* = 2, *corr among rep measures* = 0.5).

Materials.

AUT. We employed one round of AUT to measure divergent thinking. In this study, participants were asked to think of as many uses of "tire" as possible. We replicated the measures and the marking process in Study 1.

CRAT. We employed ten questions from the CRAT to measure convergent thinking. The ten questions were selected based on their difficulty ratings from normative data reported by a previous study (Bowden & Jung-Beeman, 2003). The selection requirements were: 1) the questions were neither too difficult that no participants answered them within 2 seconds nor too easy that 80% of participants answered them within 30 seconds, and 2) the difficulties of the selected questions varied since the percentage of participants solving questions varied. According to the normed data, 1%-24% of their participants solved the ten questions we selected within 2 seconds, and 10%-74% of their participants solved the questions within 30 seconds. We measure CRAT performance as we did in Study 1. Please see Appendix X for *Compound Remote Associate Test Questions in Study 7*.

FST. We employed one round of the FST to measure participants' integrative thinking. The three functions we provided were "interact, immerse, recreate". We measure FST performance as we did in Study 1.

IT & PDT. We replicated IT and PDT in the Study 7 (Pilot) (Please see Materials in Study 7 (Pilot)).

Self-improvement motivation scale. We implemented a self-improvement motivation questionnaire adapted from Kurman (2006) to fit our tasks. Participants rated items on a scale of 1 (not at all) to 7 (very much), e.g. "*In my opinion, I should have done more for these tasks*". The internal consistency was α = .40 (see Appendix Y).

Sensitivity to social comparison scale. We also implemented the Attention to Social Comparison Scale, e.g. "*My behavior often depends on how I feel others wish me to behave.*", and Short Creative Self-Efficacy Scale, e.g. "*I trust my creative abilities.*" (Karwowski et al., 2018; Lennox & Wolfe, 1984). The internal consistency was α = .815 (see Appendix Z).

Perceived threat scale. We developed a perceived threat questionnaire with internal consistency α = .913. Participants rated items from 1 (not at all) to 7 (very much), e.g. "*I feel nervous or distressed thinking about my rank compared to others*" (see Appendix AA).

Procedure.

We replicated most steps in Study 7 (Pilot) for the procedure. However, we added a manipulation before participants encountered IT and PDT. In detail, participants first did AUT, CRAT, and FST and answered three questions about their experience thus far.

Then, we differentiated experimental groups (i.e., upward and downward groups) from the control group by asking an additional question to the experimental groups. In this additional question, we told participants that we administered the creative thinking tasks to 1500 undergraduates, and there was a ranking pool for the undergraduates' performance. The participants needed to select a performance range they thought they achieved compared to the 1500 undergraduates. After the selection, we proceeded the experimental groups to a waiting page. On this page, we told participants there was a creativity scoring system that evaluated their performance in the creative thinking tasks. Participants were also told that the evaluation was based on their answers' novelty, originality, usefulness, and accuracy. Participants waited for up to 1 minute to get the results. The creativity scoring system was a deception, and the role of the waiting page was to make the pre-determined ranking more authentic. Participants in the downward comparison group received pre-determined ranking feedback saying that "compared to the performance of the other 1500 university undergraduates, your performance in three creative thinking tasks is ranked in Best Performance Range". To make the evaluation more specific, we visualised the ranking, indicating that they were one of the top-100s compared to the 1500 undergraduates.

On the other hand, participants in the upward comparison group received pre-determined ranking feedback saying that their performance was ranked in the Worst Performance Range. The visualisation indicated that they were one of the bottom-200s compared to the 1500 undergraduates. Participants in the control group were not exposed to information about the ranking feedback, and we proceeded to the next step right after they reported their game experience.

After that, all participants completed the IT and PDT and reencountered the game experience questions. We asked the experimental groups whether they remembered and trusted their ranking in the creative thinking tasks. Also, we asked participants to report their self-improvement motivation, sensitivity to social comparison, perceived threat, and demographic information. We also provided a debriefing form in which we disclosed and explained the areas of deception (please see Appendix AB for *Study 7 Script*).

Study 7 Results

Inter-rater reliability. We provided marking instruction and training to two judges: one each from psychology and engineering backgrounds. Inter-rater reliability was obtained for all measures (*Cronbach's* $\alpha > .60$) (see Table 65).

Table 65

Intraclass Correlation Coefficients (ICC) with Two-way Mixed Effects Model for Measures in AUT, FST, IT, and PDT (N of items = 2).

Measures	*ICC*	*95% CI*	
		Lower Bound	Upper Bound
AUT appropriate fluency	.99	.96	1.00
AUT appropriate originality	.92	.66	.98
AUT appropriate flexibility	.94	.05	.95
FST appropriate fluency	.97	.83	.93
FST appropriate originality	.83	.33	.95
FST appropriate flexibility	.92	.89	.92
FST appropriate effectiveness	.85	.28	.93
FST accurate fluency	.87	.48	.97
FST accurate originality	.82	.76	.92
FST accurate flexibility	.85	.38	.95
FST accurate effectiveness	.63	.16	.86
IT appropriate fluency	.98	.92	1.00
IT appropriate originality	.89	.63	.98
IT appropriate flexibility	.96	.93	.98
IT appropriate effectiveness	.95	.80	.97
IT accurate fluency	.94	.72	.98
IT accurate originality	.73	.47	.90
IT accurate flexibility	.87	.43	.96
IT accurate effectiveness	.96	.75	.96
PDT relevance & effectiveness	.95	.74	.98
PDT problematisation	.91	.62	.95
PDT propulsion	.84	.28	.95
PDT elegance	.88	.67	.96
PDT generalisation	.93	.89	.97

a. Cronbach's *α* is the average measures of intraclass correlation coefficients.

b. Cronbach's *α* using an absolute agreement definition.

Statistical assumptions. To prepare an appropriate dataset, we did a series of assumption tests. The results showed that our dataset met the assumption of non-zero variances and independent errors (3 > *Durbin-Watson value* > 1). Furthermore, according to Skewness and Kurtosis, most variables were close to normal distribution. However, the results of the Shapiro-Wilk test showed that most creativity measures were not normally distributed. Therefore, we conducted both parametric tests and non-parametric tests for the dataset. Here, we report the results of parametric tests (e.g., Pearson correlation, linear regression, t-test, and ANOVA) and specify when non-parametric tests could not detect the same results. Please see Table 66 - 71 for the descriptive statistics and normality tests.

Table 66

Descriptive Statistics for All Measures (Control Group, N = 42).

	M	*SD*
AUT appropriate fluency	6.86	2.74
AUT appropriate flexibility	5.50	1.87
AUT appropriate originality	1.99	0.17
AUT appropriate frequency	2.24	0.33
CRAT accurate fluency	3.05	3.21
FST appropriate fluency	5.05	2.66
FST appropriate flexibility	3.24	1.41
FST appropriate originality	1.59	0.30
FST appropriate frequency	2.34	0.39
FST appropriate effectiveness	1.83	0.32
FST accurate fluency	3.42	2.10
FST accurate flexibility	2.16	0.98
FST accurate originality	1.57	0.12
FST accurate frequency	2.31	0.48
FST accurate effectiveness	1.84	0.41
IT appropriate fluency	4.38	1.75
IT appropriate flexibility	3.81	1.56
IT appropriate originality	1.55	0.42
IT appropriate frequency	2.71	0.24
IT appropriate effectiveness	2.01	0.26
IT accurate fluency	2.55	1.25
IT accurate flexibility	2.06	1.04
IT accurate originality	1.52	0.46
IT accurate frequency	2.48	0.46
IT accurate effectiveness	2.16	0.41
PDT total	67.75	10.15
PDT relevance & effectiveness	10.48	1.38
PDT problematisation	8.16	2.29
PDT propulsion	13.69	2.45

PDT elegance	22.38	4.05
PDT genesis	15.05	2.51
Self-improvement	37.57	4.26
Perceived threat	7.79	2.93
Social Comparison Motivation	36.83	8.08
Creative self	3.77	0.91

Table 67

Normality Tests for All Measures (Control Group).

			Shapiro-Wilk	
	Skewness	Kurtosis	Statistics	*p* value
AUT appropriate fluency	0.89	0.21	.92**	.005
AUT appropriate flexibility	0.53	0.05	.95	.057
AUT appropriate originality	-0.12	-0.26	.98	.478
AUT appropriate frequency	-0.26	-0.65	.97	.336
CRAT accurate fluency	0.89	-0.11	.85**	<.001
FST appropriate fluency	1.43	1.85	.86**	<.001
FST appropriate flexibility	0.77	0.43	.95*	.042
FST appropriate originality	0.15	-0.35	.98	.598
FST appropriate frequency	-0.96	1.93	.94*	.040
FST appropriate effectiveness	0.39	-0.24	.98	.589
FST accurate fluency	1.41	2.31	.87**	<.001
FST accurate flexibility	0.79	0.35	.91**	.003
FST accurate originality	-0.04	-0.55	.95	.092
FST accurate frequency	-0.45	0.18	.95	.070
FST accurate effectiveness	0.42	-0.27	.97	.367
IT appropriate fluency	0.18	-0.67	.96	.202
IT appropriate flexibility	0.45	0.27	.97	.350
IT appropriate originality	0.43	-0.83	.94*	.020
IT appropriate frequency	-0.30	-0.73	.91**	.003
IT appropriate effectiveness	0.31	-0.01	.99	.831
IT accurate fluency	0.39	-0.91	.89**	.001
IT accurate flexibility	0.56	-0.34	.94*	.029
IT accurate originality	0.40	-0.12	.96	.102
IT accurate frequency	-1.07	1.52	.89**	.001
IT accurate effectiveness	1.62	4.02	.81**	<.001
PDT total	-0.55	1.60	.95	.061
PDT relevance & effectiveness	-0.58	1.11	.96	.096
PDT problematisation	0.37	0.76	.96	.108

PDT propulsion	-0.73	1.31	.96	.111
PDT elegance	-0.70	0.45	.96	.122
PDT genesis	-0.24	1.16	.97	.263
Self-improvement	-0.09	-0.32	.98	.727
Perceived threat	0.54	-0.29	.95	.074
Social comparison motivation	-0.19	-0.58	.98	.495
Creative self	-1.01	0.68	.91**	.004

*p < .05 (2-tailed). **p < .01 (2-tailed).

Table 68

Descriptive Statistics for All Measures (Upward Group, N = 52).

	M	*SD*
AUT appropriate fluency	6.39	2.27
AUT appropriate flexibility	4.93	1.76
AUT appropriate originality	1.99	0.22
AUT appropriate frequency	2.24	0.25
CRAT accurate fluency	2.71	2.86
FST appropriate fluency	3.91	2.74
FST appropriate flexibility	2.81	1.62
FST appropriate originality	1.48	0.39
FST appropriate frequency	2.26	0.49
FST appropriate effectiveness	1.74	0.41
FST accurate fluency	2.50	1.81
FST accurate flexibility	1.79	1.13
FST accurate originality	1.38	0.54
FST accurate frequency	1.97	0.68
FST accurate effectiveness	1.62	0.64
IT appropriate fluency	5.64	2.55
IT appropriate flexibility	4.63	1.92
IT appropriate originality	1.48	0.31
IT appropriate frequency	2.80	0.20
IT appropriate effectiveness	1.93	0.18
IT accurate fluency	2.96	1.57
IT accurate flexibility	2.29	1.06
IT accurate originality	1.45	0.38
IT accurate frequency	2.53	0.45
IT accurate effectiveness	2.13	0.34
PDT total	73.86	10.26
PDT relevance & effectiveness	10.85	1.26
PDT problematisation	8.86	2.24
PDT propulsion	14.44	3.09

PDT elegance	24.07	3.02
PDT genesis	15.45	2.86
Self-improvement	38.90	4.33
Perceived threat	8.29	3.11
Social comparison motivation	36.62	7.93
Creative self	3.69	0.68

Table 69

Normality Tests for All Measures (Upward Group).

	Skewness	Kurtosis	Shapiro-Wilk	
			Statistics	*P* value
AUT appropriate fluency	1.72	5.25	.86**	<.001
AUT appropriate flexibility	1.67	3.83	.84**	<.001
AUT appropriate originality	-0.83	2.46	.94*	.016
AUT appropriate frequency	-0.23	-0.48	.96	.120
CRAT accurate fluency	1.06	0.29	.85**	<.001
FST appropriate fluency	1.76	4.38	.85**	<.001
FST appropriate flexibility	1.42	2.59	.88**	<.001
FST appropriate originality	0.30	0.22	.98	.391
FST appropriate frequency	-0.54	0.13	.96	.055
FST appropriate effectiveness	0.51	0.58	.94*	.014
FST accurate fluency	0.97	0.43	.91**	.001
FST accurate flexibility	0.77	0.18	.93**	.005
FST accurate originality	-0.23	0.70	.96	.119
FST accurate frequency	-0.83	1.13	.93**	.006
FST accurate effectiveness	-0.34	0.40	.96	.084
IT appropriate fluency	1.10	1.42	.91**	.001
IT appropriate flexibility	0.69	0.79	.95*	.020
IT appropriate originality	0.41	-0.68	.96	.075
IT appropriate frequency	-1.52	4.21	.85**	<.001
IT appropriate effectiveness	0.21	0.10	.99	.969
IT accurate fluency	0.95	1.28	.90**	<.001
IT accurate flexibility	0.35	-0.26	.97	.132
IT accurate originality	0.27	0.40	.97	.169
IT accurate frequency	-0.81	-0.15	.88**	<.001
IT accurate effectiveness	-2.39	6.63	.75**	<.001
PDT total	-1.08	2.09	.92**	.002
PDT relevance & effectiveness	-0.70	1.49	.94*	.017
PDT problematisation	-0.99	1.46	.91**	.001

PDT propulsion	-1.16	2.28	.89**	<.001
PDT elegance	-0.75	0.66	.96	.088
PDT genesis	-0.97	1.86	.94*	.010
Self-improvement	-0.40	1.08	.98	.354
Perceived threat	0.41	-0.16	.96*	.046
Social Comparison Motivation	-1.21	1.50	.90**	<.001
Creative self	-0.40	-0.43	.97	.184

*p < .05 (2-tailed). **p < .01 (2-tailed).

Table 70

Descriptive Statistics for All Measures (Downward Group, N = 50).

	M	*SD*
AUT appropriate fluency	7.78	3.57
AUT appropriate flexibility	5.83	1.80
AUT appropriate originality	1.93	0.16
AUT appropriate frequency	2.25	0.34
CRAT accurate fluency	2.54	2.52
FST appropriate fluency	4.63	2.54
FST appropriate flexibility	3.13	1.59
FST appropriate originality	1.45	0.36
FST appropriate frequency	2.27	0.48
FST appropriate effectiveness	1.71	0.39
FST accurate fluency	3.11	2.34
FST accurate flexibility	1.97	1.09
FST accurate originality	1.36	0.49
FST accurate frequency	2.06	0.70
FST accurate effectiveness	1.68	0.60
IT appropriate fluency	4.71	2.22
IT appropriate flexibility	4.05	1.74
IT appropriate originality	1.63	0.40
IT appropriate frequency	2.74	0.27
IT appropriate effectiveness	2.01	0.23
IT accurate fluency	2.84	1.39
IT accurate flexibility	2.37	1.13
IT accurate originality	1.61	0.58
IT accurate frequency	2.50	0.59
IT accurate effectiveness	2.15	0.48
PDT total	71.86	10.38
PDT relevance & effectiveness	11.20	1.09
PDT problematisation	7.95	2.24
PDT propulsion	13.41	3.10

PDT elegance	24.36	2.71
PDT genesis	14.94	3.38
Self-improvement	38.06	4.38
Perceived threat	8.20	2.84
Social comparison motivation	36.38	6.64
Creative self	3.88	0.65

Table 71

Normality Tests for All Measures (Downward Group).

			Shapiro-Wilk	
	Skewness	Kurtosis	Statistics	*p* value
AUT appropriate fluency	1.06	1.68	.93**	.007
AUT appropriate flexibility	0.22	0.79	.97	.164
AUT appropriate originality	-0.05	-0.74	.96	.061
AUT appropriate frequency	-0.47	-0.37	.96	.112
CRAT accurate fluency	1.07	0.79	.87**	<.001
FST appropriate fluency	1.31	3.00	.91**	.001
FST appropriate flexibility	1.34	4.04	.91**	.001
FST appropriate originality	-0.61	5.08	.91**	.001
FST appropriate frequency	-2.23	9.62	.83**	<.001
FST appropriate effectiveness	-1.51	6.48	.89**	<.001
FST accurate fluency	1.23	1.19	.88**	<.001
FST accurate flexibility	0.74	1.24	.94*	.015
FST accurate originality	-0.95	2.40	.90**	<.001
FST accurate frequency	-1.58	2.58	.84**	<.001
FST accurate effectiveness	-1.08	2.32	.88**	<.001
IT appropriate fluency	1.51	3.01	.87**	<.001
IT appropriate flexibility	1.21	2.17	.91**	.001
IT appropriate originality	0.10	-0.96	.96	.080
IT appropriate frequency	-0.81	-0.12	.86**	<.001
IT appropriate effectiveness	-0.11	-0.69	.98	.495
IT accurate fluency	0.49	0.37	.94*	.014
IT accurate flexibility	0.00	-0.70	.97	.173
IT accurate originality	0.08	0.53	.98	.354
IT accurate frequency	-2.03	5.96	.79**	<.001
IT accurate effectiveness	-2.76	8.81	.69**	<.001
PDT total	-0.46	1.00	.97	.245
PDT relevance & effectiveness	-0.31	0.02	.96	.081
PDT problematisation	-0.34	0.36	.97	.143

PDT propulsion	-0.74	0.66	.95*	.039
PDT elegance	-0.46	0.02	.97	.330
PDT genesis	-0.87	1.04	.91**	.001
Self-improvement	-0.27	-0.01	.96	.130
Perceived threat	0.24	-0.31	.97	.156
Social comparison motivation	-0.16	0.04	.97	.305
Creative self	-1.17	1.99	.92**	.002

*p < .05 (2-tailed). **p < .01 (2-tailed).

Next, we conducted three clusters of analysis with different aims. The first one was for the predictive power of creative thinking on engineering creativity. Then the second was for social comparison and engineering creativity. Finally, the third part was for correlations among creative thinking tasks.

Predictive power of creative thinking toward engineering creativity.

To examine the predictive power of AUT, CRAT, and FST towards IT and PDT, we started by conducting the Pearson correlation analysis to figure out the relationship among all the measures. After we grasped a broad picture, we conducted a linear regression analysis accordingly. In addition, we employed the control condition for the analysis to control confounding variables.

Correlations of creative thinking and engineering creativity. To determine the relationship between creative thinking and engineering creativity, we correlated the performance in AUT, CRAT, or FST with the performance in IT and PDT in the control group (see Table 72 – 75).

Table 72

Correlations AUT and IT, and AUT and PDT Measures (All Participants, N = 42).

	AUT appropriate fluency	AUT appropriate flexibility	AUT appropriate originality	AUT appropriate frequency
IT appropriate fluency	.42**	.34*	.19	.45**
IT appropriate flexibility	.31*	.19	.06	.34*
IT appropriate originality	.08	-.01	.19	-.05
IT appropriate frequency	.02	.05	.04	-.17
IT appropriate effectiveness	.04	-.02	.15	-.12
IT accurate fluency	.35*	.29	.12	.19
IT accurate flexibility	.27	.18	.03	.17
IT accurate originality	.14	.05	.10	.00
IT accurate frequency	.15	.15	.13	.08
IT accurate effectiveness	.15	.10	.27	.18
PDT total	.22	.15	-.06	-.03
PDT relevance & effectiveness	.13	.00	-.24	.03
PDT problematisation	.10	.08	.04	-.10
PDT propulsion	.21	.23	.08	-.06
PDT elegance	.24	.14	-.15	.09
PDT genesis	.12	.08	.01	-.13

*p < .05 (2-tailed). **p < .01 (2-tailed).

Table 73

Correlations Cross CRAT and IT, and AUT and PDT Measures (All Participants, N = 42).

	CRAT accurate fluency
IT appropriate fluency	-.05
IT appropriate flexibility	.07
IT appropriate originality	-.16
IT appropriate frequency	-.44**
IT appropriate effectiveness	-.21
IT accurate fluency	.01
IT accurate flexibility	.11
IT accurate originality	-.09
IT accurate frequency	-.19
IT accurate effectiveness	-.18
PDT total	-.17
PDT effectiveness	.01
PDT problematisation	-.18
PDT propulsion	-.28
PDT elegance	-.11
PDT genesis	-.07

*p < .05 (2-tailed). **p < .01 (2-tailed).

Table 74

Correlations Cross FST appropriate dimension and IT and PDT (All Participants, N = 42).

	FST appropriate fluency	FST appropriate flexibility	FST appropriate originality	FST appropriate frequency	FST appropriate effectiveness
IT appropriate fluency	.39**	.17	.17	.30	.14
IT appropriate flexibility	.309*	.08	.15	.25	.16
IT appropriate originality	.15	.22	.12	-.22	.07
IT appropriate frequency	.15	.11	-.04	-.03	-.09
IT appropriate effectiveness	.21	.18	.03	-.19	.08
IT accurate fluency	.28	.09	.18	.01	.17
IT accurate flexibility	.21	.07	.24	.00	.24
IT accurate originality	.22	.23	.20	-.14	.12
IT accurate frequency	.20	-.03	.24	-.10	.25
IT accurate effectiveness	.28	.13	.10	-.14	.23
PDT total	.10	.31*	.06	-.04	-.01
PDT relevance & effectiveness	.13	.17	.05	.05	-.03
PDT problematisation	.13	.21	-.03	-.05	-.08
PDT propulsion	.06	.04	.05	-.21	.07
PDT elegance	.12	.37*	.13	.04	.04

PDT genesis	-.05	.34*	-.03	-.02	-.10

*p < .05 (2-tailed). **p < .01 (2-tailed).

Table 75

Correlations Cross FST accurate dimension and IT and PDT (All Participants, N = 42).

	FST accurate fluency	FST accurate flexibility	FST accurate originality	FST accurate frequency	FST accurate effectiveness
IT appropriate fluency	.29	.04	.18	.13	.19
IT appropriate flexibility	.23	-.02	.18	.08	.23
IT appropriate originality	.19	.37*	.06	-.23	.01
IT appropriate frequency	.18	.17	-.09	.08	-.17
IT appropriate effectiveness	.19	.18	.02	-.21	.06
IT accurate fluency	.20	.04	.25	-.18	.26
IT accurate flexibility	.14	.02	.27	-.18	.31*
IT accurate originality	.21	.32*	.09	-.22	.06
IT accurate frequency	.19	-.01	.09	-.23	.18
IT accurate effectiveness	.25	.10	.03	-.27	.23
PDT total	.01	.28	.06	.11	-.09
PDT relevance & effectiveness	.09	.12	.03	.10	-.08
PDT problematisation	.12	.28	.04	.17	-.10
PDT propulsion	.06	.13	.02	-.06	-.03
PDT elegance	-.04	.22	.13	.09	-.01

PDT genesis	-.10	.32*	-.05	.16	-.20

*p < .05 (2-tailed). **p < .01 (2-tailed).

The results showed significant positive associations between IT and AUT. Specially, appropriate fluency, appropriate frequency, and appropriate flexibility in AUT were moderately (.40 <= r < .70) or weakly (r < .40) associated with IT appropriate fluency. Appropriate fluency and frequency in AUT were weakly associated with IT appropriate flexibility. There was also a weak association between AUT appropriate fluency and IT accurate fluency. Also, there was a moderate negative association between CRAT accurate fluency and IT appropriate frequency.

For FST appropriate-dimension measures, appropriate fluency was weakly associated with IT appropriate fluency and flexibility. FST appropriate flexibility was weakly associated with total score, elegance, and genesis in PDT. For FST accurate-dimension measures, accurate flexibility was weakly associated with IT appropriate originality, IT accurate originality, and PDT genesis. FST accurate relevance was weakly associated with IT accurate flexibility

Considering that several measures in AUT, CRAT, or FST significantly correlated with IT and PDT, we conducted multiple linear regressions to examine the predictive power of AUT, CRAT, and FST towards engineering creativity in IT and PDT.

Predictive power of AUT. To prepare appropriate predictors for multiple linear regression, we referred to the correlations among AUT measures in the control condition (see Table 76). If the Pearson correlation between two measures within one task was larger than .70, we excluded one of the two to avoid multicollinearity issues. For example, there was a positive correlation between appropriate fluency and appropriate flexibility of AUT, r = .89, p < .001. Considering that appropriate fluency, r = .51, p < .001, was yielded a larger correlation with other measure(s) (i.e., appropriate frequency) than appropriate flexibility, r = .50, p = .001, we excluded the AUT appropriate fluency from this piece of analysis. The Tolerance was larger than 0.1, and the VIF was lower than 10. Therefore, multicollinearity was not a concern for AUT measures (see Table 77).

Table 76

Correlations for AUT Measures in Control Condition (N = 42).

AUT measures	1	2	3	4
1. Appropriate fluency	-			
2. Appropriate flexibility	.89**	-		
3. Appropriate originality	.24	.25	-	
4. Appropriate frequency	.51**	.50**	.38*	-

*p < .05 (2-tailed). ** p < .01 (2-tailed).

Table 77

Collinearity Statistics for The AUT Measures That Were in The Multiple Linear Regression.

AUT measures	Tolerance	VIF
Appropriate flexibility	.75	1.34
Appropriate originality	.85	1.17
Appropriate frequency	.68	1.46

We conducted a series of multiple linear regression to examine the predictive power of AUT creativity on IT and PDT creativity. The results showed that, appropriate fluency, appropriate frequency, and appropriate originality in AUT explained a significant amount of the variance in IT appropriate fluency, F (3,38) = 3.58, p = .022, R^2 = .22, $R^2_{Adjusted}$ = .16. In detail, AUT appropriate frequency was a significant predictor, B = 1.96, *Beta* = .37, t (40) = 2.12, p = .040, *95% CI* [.09, 3.83]. However, appropriate fluency, frequency, and originality in AUT did not explain a significant amount of the variance in other IT and PDT measures.

Predictive power of CRAT. We conducted a series of single linear regressions to examine the predictive power of CRAT creativity on IT and PDT creativity. The results showed that, CRAT accurate fluency explained a significant amount of the variance in IT appropriate frequency, F (1,40) = 9.47, p = .004, R^2 = .19, $R^2_{Adjusted}$ = .17. In detail, CRAT accurate fluency was a significant predictor, B = -.03, *Beta* = .44, t (40) = -3.08, p = .004, *95% CI* [-.05, -.01]. However, CRAT accurate fluency did not explain a significant amount of the variance in any other measures in IT and PDT. Multi-collinearity was not a concern because we had only one measure for CRAT.

Predictive power of FST. To prepare appropriate predictors for multiple linear regression, we referred to the correlations of FST measures in the control condition (see Table 56 - 58). Because some measures in appropriate and accurate dimensions were highly correlated with each other ($r >= .70$), we did regression analysis for appropriate-dimension measures and accurate-dimension measures separately. Also, the effectiveness measures were significantly correlated with the originality measures in both dimensions. We excluded the latter because originality measures showed a more normally distributed pattern than effectiveness measures in the data set. According to collinearity diagnosis, multicollinearity was not a concern for FST measures (see Table 78).

Table 78

Collinearity Statistics for FST Measures in Multiple Linear Regression.

FST measures	Tolerance	VIF
Appropriate fluency	.59	1.70
Appropriate flexibility	.95	1.05
Appropriate originality	.63	1.60
Appropriate frequency	.85	1.17
Accurate fluency	.60	1.68
Accurate flexibility	.89	1.12
Accurate originality	.55	1.83
Accurate frequency	.89	1.13

We conducted a series of multiple linear regressions to examine the predictive power of FST creativity on IT and PDT creativity. For the appropriate dimension in FST, the results showed that fluency, flexibility, originality, and frequency explained a significant amount of the variance in PDT genesis, F (4,37) = 2.67, p = .047, R^2 = .22, $R^2_{Adjusted}$ = .14, In this case, appropriate flexibility was a significant predictor, B = 1.06, *Beta* = .59, t (40) = 3.25, p = .002, *95% CI* [.40, 1.72] and appropriate fluency was a significant suppressor, B = -.38, *Beta* = -.40, t (40) = 2.12, p = .041, *95% CI* [-.74, -.02].

For the accurate dimension in FST, the results showed that fluency, flexibility, originality, and frequency explained a significant amount of the variance in IT appropriate originality, F (4,37) = 3.20, p = .024, R^2 = .26, $R^2_{Adjusted}$ = .18, and PDT genesis, F (4,37) = 3.67, p = .013, R^2 = .28, $R^2_{Adjusted}$ = .21. In detail, accurate flexibility was a significant predictor for IT appropriate originality prediction, B = .22, *Beta* = .51, t (40) = 2.68, p = .011, *95% CI* [.05, .38] where accurate frequency was a significant suppressor, B = -.32, *Beta* = -.37, t (40) = -2.44, p = .020, *95% CI* [-.59, -.05]. Also, accurate flexibility was a significant predictor for PDT genesis, B = 1.70, *Beta* = .09, t (40) = 3.52, p = .001, *95% CI* [.72, 2.68] where accurate fluency was a significant suppressor, B = -.62, *Beta* = -.52, t (40) = -2.89, p < .001, *95% CI* [-1.06, -.19]. However, appropriate measures and accurate measures in FST did not explain much of the variance in other IT and PDT measures.

Further analysis for FST predicative power. To increase the comparability of FST predictive power with AUT and CRAT, we conducted multiple linear regression on FST towards IT and PDT. First, we simulated the analysis for AUT and examined the predictive power of flexibility, originality, and frequency in appropriate and accurate dimensions in FST. The collinearity diagnosis showed that multicollinearity was not a concern here (see Table 79). Then, we simulated and analysed for CRAT and examined the predictive power of FST accurate fluency.

It is important to note that, although we employed the same measures to enhance comparability, we did not conclude that the flexibility, originality, and frequency in FST and AUT refer to the same thing. The relationship among these measures requires further investigation, and our analysis served an exploratory goal.

Table 79

Collinearity Statistics for FST (Exclude Fluency) in Multiple Linear Regression.

FST measures	Tolerance	VIF
Appropriate flexibility	.90	1.11
Appropriate originality	.96	1.04
Appropriate frequency	.93	1.08
Accurate flexibility	.89	1.12
Accurate originality	.86	1.17
Accurate frequency	.89	1.13

The multiple linear regression showed that accurate flexibility, accurate originality, accurate frequency in FST explained a significant amount of the variance in IT appropriate originality, *F* (3,38) = 4.35, *p* = .010, R^2 = .26, $R^2_{Adjusted}$ = .20, and IT accurate originality, *F* (3,38) = 3.29, *p* = .031, R^2 = .21, $R^2_{Adjusted}$ = .14. In detail, accurate flexibility was a significant predictor for IT appropriate originality prediction, *B* = .21, *Beta* = .48, *t* (40) = 3.21, *p* = .003, *95% CI* [.08, .08], where accurate frequency was a significant suppressor, *B* = -.32, *Beta* = -.37, *t* (40) = -2.49, *p* = .017, *95% CI* [-.59, -.06]. Also, accurate flexibility was a significant predictor for PDT genesis, *B* = .20, *Beta* = .42, t (40) = 2.71, *p* = .010, *95% CI* [.05, .35], where accurate frequency was a significant suppressor, *B* = -.33, *Beta* = -.34, *t* (40) = -2.20, *p* = .034, *95% CI* [-.63, -.03]. Except for that, accurate flexibility, originality, and frequency did not explain a significant amount of the variance in other IT and PDT measures.

Furthermore, the results showed that appropriate flexibility, originality, and frequency in FST did not explain a significant amount of the variance in any of the measures in IT and PDT. The results also showed that FST accurate fluency did not explain significant variance in any of the measures in IT and PDT.

Social comparison influences engineering creativity.

Effect of social comparison on engineering creativity. We conducted a one-way between-subject ANOVA to compare the effect of social comparison on creative performance in IT, PDT, and self-reported measures in the upward comparison group, downward comparison group, and control group[25]. We followed up significant main effects with post hoc independent sample *t*-tests with Bonferroni corrections and considered confidence intervals (*CI*) that do not include zero as significant.

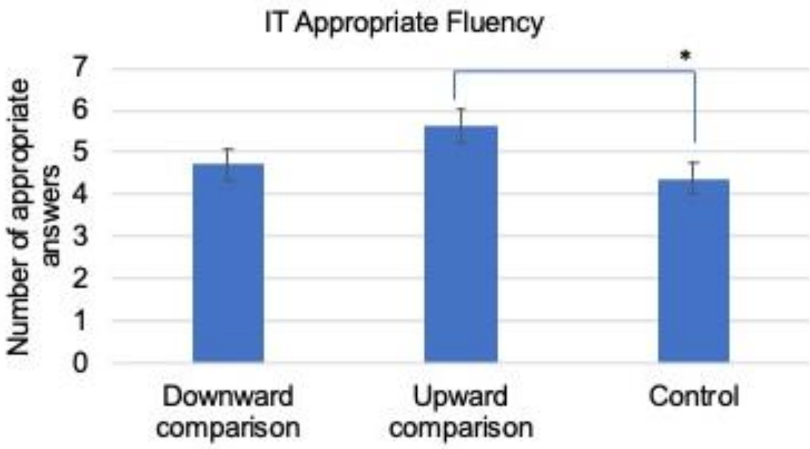
There was a significant main effect of social comparison on appropriate fluency of ideation creativity, *F* (2, 141) = 4.11, *p* = .02, *Ω* = .73. Follow-up tests revealed a significant difference between appropriate fluency between the upward comparison group and the control group, *t* (92) = 2.71, *p* = .023, *95% CI* [.13, 2.37], such that appropriate fluency was higher in the upward comparison group (*M* = 5.63, *SD* = 2.55) than the control group (*M* = 4.38, *SD* = 1.75). The appropriate fluency did

[25] The results of the nonparametric test Kruskal-Wallis Test supported the findings in ANOVA.

not reveal a significant difference, *t* (100) = -1.95, *p* = .114, *95% CI* [-.15, 1.99], between the upward comparison and downward comparison groups (*M* = 4.71, *SD* = 2.22), and *t* (90) = 2.71, *p* = .778, *95% CI* [-.1.46, 0.80], between the downward comparison and control groups (see Figure 16).

Figure 16

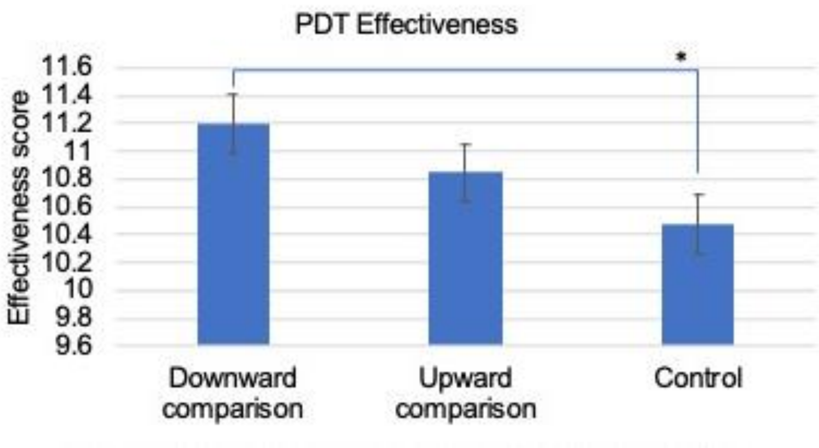
IT appropriate fluency in downward, upward, and control conditions.



There was a significant main effect of social comparison on product design creativity for the dimensions of relevance & effectiveness, *F* (2, 141) = 5.99, *p* = .023, *Ω* = .70. Follow-up tests revealed a significant difference, *t* (90) = 2.81, *p* = .018, *95% CI* [.09, 1.35], such that relevance & effectiveness was higher in the downward comparison group (*M* = 11.20, *SD* = 1.09) than the control group (*M* = 10.48, *SD* = 1.38). The relevance & effectiveness did not reveal a significant difference, *t* (100) = 0.25, *p* = .458, *95% CI* [-.95, 0.24], between the upward (*M* = 10.85, *SD* = 1.26) and downward groups, and *t* (92) = 0.26, *p* = .460, *95% CI* [-.25, .99], between upward and control groups (see Figure 17).

Figure 17

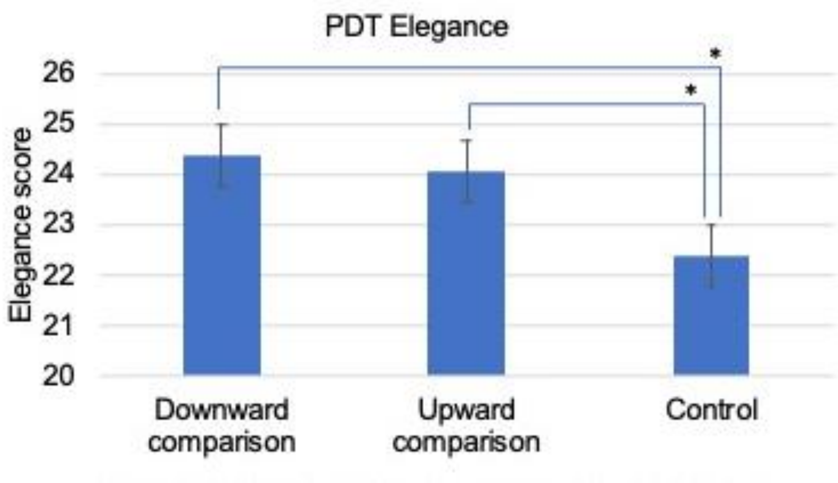
PDT effectiveness in downward, upward, and control conditions.



There was a significant main effect of elegance, *F* (2, 141) = 4.79, *p* = .010, *Ω* = .80. Follow-up tests revealed a significant difference, *t* (92) = 2.31, *p* = .041, *95% CI* [.05, 3.32], such that elegance was higher in the upward comparison group (*M* = 24.07, *SD* = 3.02) than the control group (*M* = 22.38, *SD* = 4.05. There was also a significant difference, *t* (90) = 2.87, *p* = .013, *95% CI* [.33, 3.63], such that elegance was higher in the downward comparison group (*M* = 24.36, *SD* = 2.71) than the control group. There was no significant difference *t* (100) = 0.65, *p* = 1.000, *95% CI* [-1.86, 1.27], between the upward and downward comparison groups. None of the other measures revealed a significant difference (see Figure 18).

Figure 18

PDT elegance in downward, upward, and control conditions.



Effect of feedback expectation on engineering creativity. After doing IT and PDT, participants were asked to report whether they expected a creativity evaluation. Here, we compared the creativity performance and self-reported scores (i.e., self-improvement motivation, perceived threat, comparison motivation, and creative self-efficacy) of participants who expected creativity evaluations and those who did not. The independent sample test results showed that, in downward comparison group, participants who did not expect evaluations (M = 36.48, SD = 3.21) reported a lower level of self-improvement motivation, t (47) = -2.28, p = .027, *95% CI* [-4.96, -.31], than those who expected evaluation (M = 39.39, SD = 4.50). In upward comparison group, participants who did not expect evaluations (M = 6.77, SD = 2.58) reported a lower level of perceived threat, t (51) = -4.05, p < .001, *95% CI* [-4.50, -1.52], than those who expected evaluation (M = 9.78, SD = 2.82). In control group, participants who did not expect evaluations (M = 3.41, SD = .22) reported a lower level of creative self-efficacy, t (40) = -2.66, p = .011, *95% CI* [-1.23, -.17], than those who expected evaluation (M = 4.11, SD = .68). None of the other creativity performance and self-report dimensions revealed significant differences for feedback expectations.

Correlations among AUT, CRAT, and FST. Considering that participants did AUT, CRAT, and FST before social comparison manipulation, we conducted a Pearson correlation analysis for the measures of three tasks, which aims to see how they correlate with each other (see Table 80).

Table 80

Correlations among AUT, CRAT, and FST measures (All participants, N = 144)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. AUT appropriate fluency	-														
2. AUT appropriate flexibility	.78**	-													
3. AUT appropriate originality	.16	.09	-												
4. AUT appropriate frequency	.52**	.43**	.26**	-											
5. CRAT accurate fluency	.11	-.03	.11	.03	-										
6. FST appropriate fluency	.29**	.28**	.06	.06	-.04	-									
7. FST appropriate flexibility	.28**	.29**	.09	.08	-.03	.76**	-								
8. FST appropriate originality	.12	.18*	.11	-.03	.06	.18*	.19*	-							
9. FST appropriate frequency	.07	.05	.13	.07	.00	.31**	.34**	.24**	-						

10. FST appropriate effectiveness	.02	.08	.04	-.09	.13	.10	.03	.79**	.02	-					
11. FST accurate fluency	.12	.15	.11	-.05	-.09	.80**	.44**	.06	.21*	.05	-				
12. FST accurate flexibility	.10	.13	.26**	-.05	-.08	.62**	.70**	.13	.29**	.01	.72**	-			
13. FST accurate originality	.07	.11	.17*	-.04	.06	.25**	.27**	.72**	.09	.68**	.24**	.36**	-		
14. FST accurate frequency	-.02	.06	.14	.02	-.06	.35**	.35**	.06	.61**	.00	.45**	.57**	.35**	-	
15. FST accurate effectiveness	.02	.03	.12	-.07	.12	.17*	.08	.56**	-.03	.79**	.17*	.18*	.85**	.19*	-

*p < .05 (2-tailed). **p < .01 (2-tailed).

Correlations of AUT. The results showed significant positive associations among the measures in AUT. Specifically, there was a high association between appropriate fluency and appropriate flexibility. Also, the appropriate frequency was moderately associated with appropriate fluency and flexibility and weakly associated with appropriate originality. However, appropriate originality was not associated with appropriate fluency and appropriate flexibility.

Correlations of appropriate dimension in FST. There were significant positive associations among the measures in FST appropriate dimension. Specifically, there was a high association between appropriate fluency and appropriate flexibility. On the other hand, there was a weak association between appropriate originality and appropriate frequency. Moreover, appropriate fluency and flexibility were weakly associated with appropriate originality and frequency. Other than that, appropriate effectiveness was highly correlated with appropriate originality but none of the other measures.

Correlations of accurate dimension in FST. There were significant positive associations among the measures in FST accurate dimension. Specifically, there were high associations between accurate fluency and accurate flexibility, and between accurate originality and accurate effectiveness. Moreover, accurate frequency was moderately associated with accurate fluency and accurate flexibility. In addition, the other associations among FST accurate-dimension measures were significant and weak (below .40).

Correlations of cross dimensions in FST. The results showed significant positive associations among all measures of FST. For the measures that were partially in the same dimension (e.g., appropriate fluency and accurate fluency), there were high correlations between appropriate fluency and accurate fluency, appropriate flexibility and accurate flexibility, appropriate originality and accurate originality, and appropriate effectiveness and accurate effectiveness. There were moderate correlations between appropriate frequency and accurate frequency.

For the measures in a completely different dimension (e.g., appropriate fluency and accurate flexibility), there were moderate associations between appropriate fluency and accurate flexibility, appropriate flexibility and accurate fluency, appropriate originality and accurate effectiveness, and appropriate effectiveness and accurate originality. The rest of the associations were either weak or not significant.

Correlations among AUT, CRAT, and FST. The results showed significant correlations among the measures in FST and AUT. For example, AUT appropriate fluency was weakly associated with FST appropriate fluency and flexibility but not the other FST measures. AUT appropriate flexibility was weakly associated with FST appropriate fluency, flexibility, and originality. However, it did not associate with the other FST measures. In addition, AUT appropriate originality was weakly associated with FST accurate flexibility and originality. It did not associate with other FST measures. Also, AUT frequency was not associated with any of the FST measures.

Study 7 Discussion

Hypothesis 5: FST predicts creativity in engineering product design. The better the FST performance, the greater creativity in engineering product design.

Predictive power on ideation creativity. The results supported hypothesis 5. In detail, the selected six measures in FST (i.e., appropriate flexibility, appropriate frequency, appropriate originality, accurate fluency, accurate frequency, and accurate originality) significantly explained the variance in four of eight measures in IT (i.e., fluency and flexibility of appropriate and accurate answers). The highest R^2, lowest R^2, and average R^2 were 46.9%, 33.5%, and 39.4%, respectively. Meanwhile, the selected three measures in AUT (i.e., appropriate fluency, appropriate frequency, and appropriate originality) significantly explained the variance in five measures in IT (i.e., fluency and flexibility of appropriate and accurate answers and originality of accurate answers). The highest R^2, lowest R^2, and average R^2 were 43.5%, 23.4%, and 35.54%. The CRAT measures (i.e., accurate fluency) did not significantly explain the variance in IT. Therefore, compared to the performance in AUT and CRAT, the performance in FST explained the most considerable variance in the ideation creativity of engineering students.

However, hypothesis 5 was accepted with at least two conditions. First, FST may explain the greatest variance in ideation creativity only if we measure creativity with two or more levels of accuracy. For instance, when we had three measures of appropriate answers in FST, which was the same as AUT measures, the significant predictive power of FST to IT disappeared. Similarly, if we left only one measure – accurate fluency in FST, which was the same as the CRAT measure, the significant predictive power of FST on IT also disappeared. The disappearance of significant predictive power implied that the best predictive power in FST – IT regression models might result from the high quantity rather than the high quality of FST predictors. The second condition was that the best predictive power of FST existed if we evaluated the predictive power of the task based on their significant R^2 values. For example, if we evaluated the predictive power of AUT and FST according to the largest, lowest, and average values of R^2, FST performed better since its R^2 were higher than AUT. However, if we evaluated the predictive power according to the number of dimensions a task predicted, AUT performed better than FST since AUT predicated three dimensions of IT, but FST predicted only two.

The shared cognitive styles between AUT and IT may explain the better predictive power of AUT than FST when controlled by the number of predictors and compared to the number of dimensions. In detail, AUT provides participants with one piece of information (i.e., an object) and asks them to generate many ideas that stem from the information. FST provides participants with three pieces of information (i.e., three functions) and asks participants to generate many ideas that must satisfy all three pieces of information. The IT provides participants with three pieces of information (i.e., two ways that COVID-19 was transmitted and the user context of the product). It asks participants to generate many ideas that ideally but not mandatorily satisfy all the given information. The IT also encourages participants to follow their thinking flows, not criticize any idea's appropriateness and accuracy, and write down all the pop-up ideas. Considering that human beings tend to save cognitive resources when perceiving the social context (Macrae, Milne, et al., 1994), participants may be less likely to stick with satisfying all three pieces of the given information in IT when they are allowed to hatch ideas by processing one piece of information. Therefore, the information processing between AUT and IT may be more similar than between FST and IT.

Within all significant predictive models, we found that the appropriate fluency in AUT and appropriate flexibility in FST, rather than the other measures, were the two significant predictors of ideation creativity. The pattern may be because, in both AUT and FST predictor lists, at least two measures for the quality aspect of answers and the predictive power of single quality measures were weakened (Nathans et al., 2012; Nimon & Oswald, 2013). For example, appropriate originality and frequency in AUT and FST represented the quality of answers because they measured the novelty and unexpectedness of appropriate answers. The scores of accurate answers in FST also represented the quality of answers because accuracy is a quality criterion. However, AUT appropriate fluency and FST appropriate flexibility were the only quantity measures in their respective predictor lists, so their predictive power was relatively high.

One more important finding was that CRAT did not predict the ideation creativity of engineering students at all, which reflects the criticisms of traditional creativity tasks. For example, CRAT may create a language barrier for non-native speakers, disconnecting the task from domain-specific or general creativity (Cortes et al., 2019). We admit this possible explanation because accurate fluency in CRAT was not associated with any IT measures. However, we propose that the issue at root may be that it cannot grasp various dimensions of creativity. In detail, CRAT only measures accurate fluency does not mean that participants do not process novel and appropriate ideas during the task. People who process more accurate answers may not be those who process more novel and appropriate answers. As we know, if we employed only accurate fluency in FST in regression, the predictive power of FST also disappeared to an extent. Therefore, the absence of CRAT's predictive power may result from language barriers, inflexible question format (i.e., one answer box per question), and an incomprehensive scoring system (i.e., one dimension).

Predictive power on product design creativity. The results supported hypothesis 5. In detail, the selected six measures in FST significantly explained 30.0%, 37.4%, 41.1%, 43.0%, and 43.8% of the variance in problematization, propulsion, elegance, genesis, and total creativity scores of PDT. When we excluded the measures of accurate answers, the selected three measures of appropriate answers in FST still significantly explained 24.0%, 35.5%, 35.4%, 38.2%, 40.9% of the variance in problematization, propulsion, elegance, genesis, and total creativity scores of PDT. Even when we left only accurate fluency in FST in the predictor list, this one predictor still significantly explained 9.7%, 31.1%, and 12.4% of the relevance & effectiveness, elegance, and total creativity scores of PDT. FST appropriate frequency predicted problematization, FST appropriate originality predicted propulsion, and accurate fluency predicted elegance. However, the measures in AUT and CRAT did not significantly explain the variance in PDT performance.

There were two main takeaways from our results. First, compared to AUT and CRAT, FST is the creative thinking task that shows the best predictive power to PDT creativity for engineering students. Our previous argument could explain this finding that FST simulated a complex thinking process of creative product design in the engineering field, which could not be simulated by divergent or convergent thinking tasks.

Second, the quality measures such as appropriate frequency, appropriate originality, and accurate fluency in FST predicted the product design creativity of engineering students. In contrast, none of the quantity measures predicted the product design creativity. The results may be because that evaluative thinking is an essential thinking skill in engineering product design, and the quality measures rather than quantitative measures in FST could grasp this thinking skill. In detail, the quantity measures indicate one’s ability to generate many ideas in different categories. Accurate fluency indicates one’s ability to generate many ideas with high value and usefulness. Appropriate frequency and originality indicate one’s ability to generate original ideas. The three kinds of idea generation require different thinking skills. To produce many ideas, people have to think broadly and flexibly and may need to leave the category of the produced idea deliberately (Baer, 2014). However, generating many useful and valuable ideas requires people to think broadly and evaluate the usefulness or value of the ideas based on their existing knowledge (Groborz & Necka, 2003).

Similarly, generating original ideas requires flexible thinking. It requires evaluating the originality of ideas based on the comparison with existing ideas or their subjective criteria about what is new (Mayseless et al., 2015). In other words, FST quantity measures required divergent thinking, while FST quality measures required divergent and evaluative thinking (David H. Cropley, 2015a, 2015c) Therefore, the quality measures for originality and accuracy, rather than quantity, predicted the creative performance in product design in engineering students.

Effect of social comparison on engineering creativity. Participants in the social comparison groups performed better or spent longer time on IT and PDT than those in the control group. For IT, participants in the upward comparison group produced more appropriate ideas, and the answers covered more categories than those in the control group. Participants in the downward comparison group spent longer on IT than those in the control group. For PDT, the product design of the upward comparison group was more propulsive and elegant than those of the control group. The product design of the downward comparison group was more relevant & effective, and elegant than those in the control group.

Although our study employs a between-subject design, the creativity baselines of the three groups were at the same level for the creative thinking tasks (i.e., pre-manipulation). Therefore, we could attribute the positive effect to our social comparison stimuli rather than baseline differences between different groups (Charness et al., 2012). The only exception is that FST accurate fluency is higher for the control group than for the upward comparison group. However, the higher baseline of the control group strengthens the positive effect of social comparison because this advantage of the control group disappeared after the social comparison manipulation.

Shared neural network. The positive effect of social comparison on creativity may result from a shared neural network of social comparison and creativity. Existing research found that perceiving visual creativity is associated with prefrontal cortex regions such as the MPFC and the dorsolateral prefrontal cortex (DLPFC) (Aziz-Zadeh et al., 2013). Doing creative tasks is also strongly related to resting functional connectivity between MPFC and other brain regions such as PCC (Wei et al., 2014) and the middle temporal gyrus (Shamay-Tsoory et al., 2011). Meanwhile, MPFC is strongly associated with “mentalising”, an automatic and spontaneous cognitive process in a social context (V. Lee & L. Harris, 2013). For instance, people in the context of social comparison automatically recruit a brain network of mentalising (Swencionis & Fiske, 2014). Therefore, the prefrontal cortex regions as a shared neurocognitive process of creativity and social comparison may explain higher creativity in the social comparison groups than in the control group. Besides, the activation of MPFC and DMPFC for the upward comparison may be more robust than the downward comparison and the control group, considering that people of low status may be more likely to make inferences about others (Muscatell et al., 2012). The difference also explains the highest creativity in the upward comparison group in our case.

Perceptions and motivations. Except for a shared neurocognitive mechanism, perceived threats and self-improvement motivations may explain the positive effect of social comparison. On the one side, people making downward comparisons tend to maintain a positive self-evaluation (Crocker et al., 1987) and avoid the risk of loss (Abdellaoui et al., 2007). Thus, they may perceive high threats from others and put increasing efforts into the allocated task (Buunk & Gibbons, 2007; Taylor & Lobel, 1989). In our study, this may be why the downward comparison group spent longer on IT than the control group. Also, people making upward comparisons tend to improve themselves (Wood, 1989, 1996) or repair the destroyed self-evaluation (Johnson & Stapel, 2007). Therefore, they are motivated to perform a better job. Therefore, self-improvement motivation may explain in our study that the upward comparison group performed better than the control group in IT and PDT.

Inconsistency. Although most of our findings supported existing theories and results, but there is an exception. In detail, existing findings suggested that social comparison would improve the novelty of ideas in brainstorming (Glăveanu et al., 2019; Michinov et al., 2015; Michinov & Primois, 2005; Shepherd et al., 1995), but we did not find an improvement in originality or frequency in IT.

The inconsistency may result from the difference between individual and group creativity, such that we examined individual ideation creativity while previous studies examined group brainstorming creativity. Another explanation of the inconsistency may be the delayed improvement of novelty in the creative thinking process (Beaty & Silvia, 2012; George & Wiley, 2020). In detail, previous studies found that producing many ideas was a perquisition of producing original ideas (Milgram et al., 1978). In our study, the 5 minutes allocated to IT were too short for participants to digest the background information and come up with original thoughts. The explanation was possible because our study showed a positive effect of social comparison on the novelty scores (e.g., propulsion and elegance) in the following 25-minute creativity task (i.e., PDT).

Strength.

Investigating engineering creativity with a novel task. First, our study offers more empirical evidence regarding engineering creativity, which is considered a field that deserves more research. Also, we employed a hot topic, COVID-19, which is effective in the current situation and offers some inspiration for future studies. For example, we generated and used a COVID-19 prevention task to grasp engineering students' creativity. The value of this task has two aspects. First, the task itself is original and effective. Second, our reflection on the task generation may offer future studies some inspiration. In detail, we used COVID-19 transmission as the background information and asked participants to generate products to stop transmission. The logic behind it is that, in late 2020, COVID-19 transmission was a severe disease that everyone had some understanding of it, most people's life was affected by it, and there was no single confirmed effective solution or product to stop it. Accordingly, this hot topic allows participants to be creative (Finke et al., 1992), motivates participants' problem-solving spirit, and allows solutions from various disciplines.

Limitations and suggestions.

The task may not be replicable. However, the COVID-19 task is very context-sensitive and may not be replicable when COVID-19 is neither common knowledge nor wholly solved.

Order effect. Considering that PDT comes after IT, the effect of social comparison on product design creativity may be contaminated by the cognitive process (e.g., self-evaluation) in ideation creativity. In other words, the significant effect on PDT may not come from the pure effect of social comparison stimuli but results from the interaction effect of social comparison and the cognitive process in ideation.

The lack of external validity. Moreover, the tasks and measures of engineering creativity in our study could not cover the whole picture of practical product design for several reasons: 1. In our experiment, the IT and PDT focused on the brainstorming stage and conceptual design stage, which was at the very beginning of practical product design (Howard et al., 2008), and the effect of social comparison on conceptual design may not be extended to later stages. 2. IT and PDT are individual creativity tasks, while the engineering product design is teamwork in practice. 3. IT and PDT set a tight time limit, but practical engineering product design may allow longer. 4. The evaluation of IT and PDT was based on subjective markings, and the results may be changed if the judges changed.

Suggestions. First, we suggest future studies to critically evaluate whether IT and PDT create a common baseline of knowledge for all participants. For example, if a replication study is conducted when COVID-19 is almost over or there is a single correct answer, it is better to find other content for IT and PDT. The criteria for the content can be found in the Novelty and Effectiveness paragraphs we have discussed above.

Second, we suggest future studies put PDT right after the social comparison stimulus (instead of the IT). In this way, we can see the effect of social comparison on PDT without the confounding variable.

The replications of our experiment can also strengthen the validity of the predictive power analysis aspect. For example, our predictive power analysis of creative thinking tasks employed the control group that limited data points and reduced external validity. Therefore, future studies may replicate our findings by eliminating the social comparison stimulus.

Moreover, future studies may replace IT and PDT with other engineering creativity tasks or measures (e.g., number of patents). They may replace an individual work context with a group work context. Besides, future studies may conduct a field study that can simulate a practical engineering product design process.

6. General Discussion

Overall, we conducted four pilot studies and seven main studies to investigate our central purpose – to advance the understanding of creative thinking and its role in real-life circumstances. This chapter starts with the integration of individual main findings. Then, we talk about the limitations that existed in all studies. After that, we integrate the implications of all studies and demonstrate them from the views of theory, methodology, education, organisational management, and policymaking. In the end, we offer a conclusion.

Integrated Main Findings

The individual main findings could be integrated into at least three creativity research streams.

Function synthesis task.

We generate and validate a new psychometric tool, FST, based on the concept of integrative thinking. We found that FST grasped different aspects of creative thinking compared to AUT and CRAT. In detail, Study 1 and Study 7 (Pilot) showed that the AUT, CRAT, and FST measures shared the most variance within each task. Study 1 and Study 7 (pilot) also showed that FST measures grasp more aspects of creative thinking when compared with AUT and CRAT. Study 2 – 4 showed that the comparative social feedback induced different impacts on AUT, CRAT, and FST. Study 7 showed that AUT, CRAT, and FST exhibited different levels of predictive power toward engineering creative product design. Overall, the evidence discloses that AUT, CRAT, and FST grasp different aspects of creative thinking, and FST may grasp more aspects than the others.

Although AUT, CRAT, and FST grasped different things, we can still detect overlap between them. For example, Study 1 and Study 7 (Pilot) showed that appropriate fluency and flexibility in AUT and FST shared strong variance ($r > .80$). However, FST did not share variance with CRAT. In Study 2, comparative social feedback affected CRAT speed and FST all fluency. However, the feedback did not affect AUT performance. More than that, Study 7 showed that AUT and FST exhibited predictive power toward engineering creative ideation tasks. However, CRAT did not exhibit predictive power. Therefore, FST shares similarities with AUT and CRAT.

As a new psychometric tool, FST also relates closely to real-life situations. For example, Study 2 – 4 showed that star rating feedback affected the ideas generated in FST, while it did not affect the ideas in AUT and CRAT [26]. In Study 7, AUT and FST explained the considerable variance in creative ideation for engineering students while CRAT did not. Also, FST predicted the creative product design of engineering students while AUT and CRAT did not. Therefore, FST may have a stronger connection with real-life creativity in the engineering field.

The gender difference in FST performance remained unclear, but our findings may provide clues for future studies. For example, Study 1 showed that females received higher scores than males for FST appropriate fluency, flexibility, originality, and accurate originality. Meanwhile, Study 2 showed that male participants performed better than female participants for FST accurate fluency. Future studies may explore the factors that drive the performance gap.

Social comparison and creativity.

In addition to investigating FST, we also explore the effect of comparative social feedback on creativity. For example, study 7 showed that upward and downward comparison ranking feedback could benefit engineering students' performance in a COVID-19 prevention design task. However, Study 2 - 4 showed that the competition and star rating feedback did not affect creative thinking.

There are several explanations for the inconsistent results—first, the type of feedback matters. In detail, the ranking would benefit creative thinking but rather than competition and star rating would not. Second, the task matters. For instance, comparative social feedback may affect the thinking in creative product design tasks. However, it does not alter performance in AUT, CRAT, and FST. Third, confounding variables such as length of the study, sample, and the experimental paradigm may also result in inconsistent results.

Creativity in advertising and engineering.

Beyond the above, we explored the role of creativity in advertising and engineering. For example, Study 5 showed that an un-stereotype intervention could benefit divergent thinking. Additionally, Study 7 showed that AUT and FST predicted engineering creative product design, which implies that divergent thinking and the

---

[26] Comparative social feedback affected CRAT speed but not task performance.

creative thinking skill required for FST [27] could benefit engineering creativity. The findings inform us that creative thinking could solve real-life problems practically.

Last but not the least, Study 6 found that perceived creativity would react to social stimuli in the real-life advertisement. For example, study 4 showed a positive effect of advertising stereotypes on perceived creativity and a reversed U-shaped effect of advertising stereotypes on purchase intention. In other words, creative thinking and perceived creativity play a role in real-life situations.

General Limitations and Suggestions.

The study-specific limitations are discussed in corresponding chapters. Here, we proposed four general limitations that share by all studies.

Creative product definition. First, we employed the two-criteria standard definition to create novel and effective ideas and products (Runco & Jaeger, 2012). Therefore, we suggest that future studies employ other definitions of creativity (e.g., implicit theory of creativity) when replicating our studies.

Two judges for subjective evaluations. Second, we employed two judges and trained them for the AUT and CRAT markings. Although their marks exhibited accepted consistency, the validity is still limited. In detail, the change of either the training process or judges may change the results of subjective creativity evaluations in our studies (e.g., originality and effectiveness marks). Therefore, we suggest that future studies hire more judges to examine the replicability of our results.

Online studies. We collected responses online. Therefore, we have confounding variables that people may be multi-tasking or in different social contexts. For instance, some participants may do the task at home independently, while some may do it in a coffee shop. In addition, the participants who took the study in a public space may receive more examples for ideation creativity tasks. Therefore, we suggest future studies replicating our studies in an in-person context and comparing the results.

Are FST-related findings replicable? One of the main contributions of our project is FST. However, we have limited empirical evidence. Therefore, we could not guarantee that the relationship between FST measures, the relationship between

[27] Although we generated FST based on the concept of integrative thinking, it would be better NOT to link them directly for now. We may need a widely accepted single definition for integrative thinking and more empirical evidence for FST.

AUT, CRAT, and FST measures, the effect of comparative social feedback, and the predictive power toward engineering creative product design are replicable in the future. Therefore, we suggest future studies to examine the replicability of our studies.

Integrated Implications

We integrate the implications from individual chapters and offer suggestions for developing theory, methodology, education, organisational management, and policy making.

Theoretical implications.

Link creative thinking with integrative thinking. Existing research has already implied or discussed the importance of integration in the creative process (Andreasen, 2005; Boden, 2009, 2010; Dubitzky et al., 2012; Heilman et al., 2003; Henriksen et al., 2015; Koestler, 1964; McNally, 1982; Rothenberg, 1976; Simonton, 2011). Our findings further supported their idea. For instance, we found that integrative thinking may differ from divergent and convergent thinking. Also, it may receive the different impacts of social factors and may have predictive power toward real-life creativity. Specifically, the higher predictive power of real-life creativity of FST than AUT and CRAT implies a possibility to map different kinds of creative thinking on the Four-C model of creativity (Kaufman & Beghetto, 2009).

We can also employ the concept of integration to understand domain-specific and domain-general creativity (Baer, 1998a; Feist, 2005). H ere is the logic, our study showed that integrative thinking exhibited predictive power towards engineering creativity. Meanwhile, research showed that integration was required in various disciplines, including arts, sciences, technology, and management (Chen & Vernadat, 2004; Leavy, 2011; Martin & Austen, 1999; Merilinna & Matinlassi, 2006; Sill, 1996). Therefore, we can assume that creativity has both domain-specific and domain-general aspects. Among the two aspects, divergent thinking, convergent thinking, problem-solving, and problem finding are the domain-general aspect and efficiently integrating the thinking skills and knowledge is the domain-specific aspect.

Link creative thinking with social cognition. Another direction of theoretical development would be linking creative thinking with social cognition. For example, studies 2, 3, 4, and 7 showed that comparative social feedback might benefit participants' efforts and performance in creative thinking and product design tasks. Meanwhile, Study 5 – 6 showed that stereotype avoidance could benefit divergent thinking. These results imply that processing social information may share psychological mechanisms with creative thinking. Therefore, looking at social information processing and creative thinking at the cognitive and neuroscience level would be valuable. In other words, we suggest future studies constructing a social cognitive neuroscience approach for creativity.

There are other benefits when linking creative thinking and social cognition. First, linking creative thinking and social cognition may help us understand the inconsistent findings in behavioural studies for the social psychology of creativity at the cognitive level.

More importantly, social cognition may help us differentiate intelligence and creativity. Both definitions of intelligence and creativity include effectiveness. Effectiveness for intelligence could be interpreted as the successful attainment of the domain-specific goals that have been set out a-prior. On the other hand, effectiveness for creativity is for the entity that a-posteriori potentially shows appropriate value in terms of domain-specific functionality or performance (Corazza & Lubart, 2021). In other words, intelligence has a pre-determined objective reference point, but creativity does not have the pre-determined objective reference point and requires unexpected originality. Due to the need for a reference point in producing things, the creative process was more likely to induce mentalising of others' thoughts and compare with others' output (i.e., social cognition) compared to intelligence.

The above hypothesis may also explain the anti-social tendency of creative genius. We assume that an appropriate level of social cognition would benefit creative thinking, but to reach professional and big creativity, the deliberate exclusion of social information may be crucial because people need to save their cognitive resources and avoid "contaminated" social values. Future studies could investigate the possibility and create a threshold model for this.

Methodological implications.

FST. FST could be an effective tool when people want to measure the creative thinking of others (e.g., teacher – students, administrators – employees). FST has several advantages compared to other creative thinking tasks. For example, FST is suitable for online creativity tests - the functions of FST can be flexibly changed, so it is hard for people to find correct answers from online resources (AUT answers are widely spread). Second, FST does not require language fluency (which CRAT requires). Therefore, FST can measure creative thinking without language barriers. Third, FST does not require domain-specific knowledge (which some problem-solving questions require) and is in an everyday context. Therefore, FST can be employed in various contexts, such as entrance examinations and general creative tests. When a university or school wants to measure prospective students' creative thinking without domain-specific knowledge, FST could be a useful indicator of their potential.

Algorithm construction. Also, the FST measures require timely marking, and most aspects of the marking process are subjective. However, Amabile has proposed that the creativity level of a product could be identified by the agreed judgement of a group of experts in the relevant domain (Amabile, 1983a). Nevertheless, it is still hard to deny the limitations. For example, it is challenging to1) recruit judges with the proper expertise, 2) maintain the high quality of judges' marking, and 3) verify the objectivity of judges' opinions. These issues are common for most creative tasks, and creativity researchers are implementing different solutions to solve them. For example, some researchers are creating algorithms and platforms that evaluate AUT ideas based on semantic distance (Beketayev & Runco, 2016). Also, some creativity research centres (Torrance Centre for Creativity and Development) offer training for judges. These attempts aim to make the creativity evaluation as objective and reliable as possible. Considering the novelty and effectiveness of FST, we suggest that future studies put some effort into creating an algorithm and offering training for this new tool.

Creative thinking measures. Our findings also have implications regarding creative thinking measures. For example, Study 1 and Study 7 showed that appropriate fluency and flexibility in AUT and FST shared a high level of variance, which implies that they were replaceable by each other but no other measures. In addition, future studies may add speed and all fluency to FST because these measures may be a useful indicator of the efforts that one puts into the task.

Reflection of novel tool generation. Our research generated several novel tools. Here, we did self-reflection and summarised four skills that may benefit future tool creation - multidisciplinary collaboration, analogical thinking, assumption challenging, and hot topic employment. For instance, we constructed a series of discussions with engineering students and drafted FST based on the discussions. Second, we employed analogical thinking when creating RRCTSF. In detail, we learned from an experimental paradigm in experimental economics - the repeated rounds of online auctions. The paradigm has been used in experimental economics to investigate monetary and social rewards' effects on overbidding (Corrigan et al., 2012). Third, we created the un-stereotype workshop by challenging the assumption of the traditional approach of stereotype avoidance. In detail, we found that traditional stereotype avoidance targets the stereotypes themselves and tries to suppress them (Macrae, Bodenhausen, et al., 1994). Here, we challenged this assumption, targeted the thoughts besides stereotypes, and encouraged people to think as divergent as possible. In addition, we employed an unsolved hot topic – COVID-19 prevention and generated a new creative product design task. This unsolved hot topic allows creative opportunity and knowledge from various disciplines. Finally, we suggest that creativity researchers produce more creative tools in the field, and we hope the reflections help.

Educational implications.

Deliver the concept of integration in engineering education. We suggest that engineering educators address integration when teaching or training engineering creativity. Our findings suggest integrative thinking is crucial in engineering creative product design. We propose that, compared with delivering the concept of creativity, delivering the concept of integration may be more digestible for engineering students. In detail, engineering students were educated to solve problems with constraints (Onarheim, 2012). If we ask the students to be creative with pure divergent thinking, creativity may be disconnected from practical problem-solving.

Furthermore, if we explicitly bring constraints out, the constraints may hinder the divergent thinking of engineering students because the constraints may drive the students to stop divergent thinking soon (David H Cropley, 2015). However, if we talk about engineering creativity as integrative thinking, we incorporate constraints as an element needed to be integrated and encourage novelty as a mandatory outcome. Therefore, discussing creativity as integrative thinking may be an efficient way to balance the weights of imagination (i.e., novelty) and constraints (i.e., usefulness) in engineering education.

Find the optimal length of tasks. For engineering, creativity, effectiveness, and novelty are crucial for problem-solving and product design (David H. Cropley, 2015d). Our research showed that the task length might affect the associations for both. In detail, Study 1 required participants to finish six rounds of FST and found negative associations between effectiveness and novelty. However, Study 7 required participants to finish two rounds of FST and found positive associations between effectiveness and novelty. Although we do not have empirical evidence showing that the task length alters the associations, it would be useful for educators to be aware of the issue.

Promote interdisciplinary education and research. Integrative thinking supports the importance of interdisciplinary education and research. For instance, educational research proposed that creating an integrated literature review that integrates the literature from various disciplines would help researchers creatively find the common ground of different theories and create novel solutions (Clark & Wallace, 2015; Szostak, 2002).

Training programmes and tools based on integrative thinking. We also suggest creating training programmes and tools based on integrative thinking. For example, we can train people to acknowledge the differences and find common grounds in various cultures and ideologies, diverse disciplines, self and others, organisational mechanisms and human beings, technology and human being, nature and human beings. This kind of program could benefit integrative thinking in various contexts (e.g., leadership) and, in turn, help the audience achieve short-term and long-term organisational goals and guide leaders to fulfil employee gifts and efficacy (Reiter-Palmon & Illies, 2004). In addition, we can develop integrative training tools based on integrative thinking. For example, one of the most widely employed tools is brainstorming, which is based on divergent thinking. It would be valuable to create one for integrative thinking.

The delivery of social comparative feedback. Existing research showed that the effect of comparative social feedback on creative thinking is complex and requires further investigation. Our research showed that comparative feedback might increase the effort and time spent but not the creative performance. Therefore, we suggest the education system avoid using comparative social feedback before we develop a clearer idea about it. It is important to note that we are not proposing that comparative social feedback is not good. Instead, we are proposing that the effect is unclear, and we may need to avoid it to maintain students' mental health.

When comparative social feedback is inevitable, we suggest educators employ low-stress social comparative feedback. According to Study 7, ranking feedback on a creativity exam or creative work may benefit engineering students' creativity. When employing a social comparison stimulus, it is essential to be aware of two things. First, the social comparison stimulus should construct a low-stress context rather than a high-stress context. For instance, we may compare engineering students with remote reference points such as award winners rather than their classmates. Second, the social comparison feedback should allow a possibility to improve or encourage self-improvement, which is especially important for people making an upward comparison. For instance, educators may inform students that creativity could be increased for anyone. More importantly, educators may avoid using creativity scores to define students' creativity levels. For example, to encourage self-improvement for students at the bottom of the ranking, educators may deliberately emphasise that creativity scores only represent the performance in one task instead of creativity in general.

Organisational management implications.

Efficient socially responsible communication. The benefit of perceived creativity investigation has already been exhibited in Study 7. For instance, we found that high stereotypical advertising imagery received the highest perceived creativity and purchase intention. However, we also found that it was likely that unexpectedness rather than stereotype-consistent information drove the effect. Therefore, this suggests marketers a way to conduct efficient, socially responsible communication – delivering stereotype inconsistent or irrelevant advertising imagery in an unexpected way.

Also, promoting stereotype avoidance with novel expression should help a brand achieve several goals, including social responsibility fulfilment and a positive reputation establishment (Baker, 2014; Blowfield, 2005; Cho et al., 2017; Herbst et al., 2013; Pollay, 1986), and a sales enhancement (Cravens & Piercy, 2006; Kotler et al., 1990). Accordingly, we suggest future stereotype interventions to emphasise divergent associative thinking because that would help people reach original ideas that are considered more unexpected.

Policy making implications.

Avoid rebound effect in stereotype avoidance. Our research found that divergent thinking and stereotype avoidance were positively correlated with each other. Existing research found that emphasising divergent thinking can avoid the rebound effect of stereotype suppression. In detail, a suppression mechanism controls the stereotype activation towards a specific person or a social group by suppressing the dominant response inclination caused by automatic cognitive processing (Macrae, Bodenhausen, et al., 1994). However, the suppression may promote the thinking of the suppressed stereotypes because it does not reverse a tendency of stereotype-consistent information processing (Monteith et al., 1998; Wang et al., 2020). Priming the creativity mindset may avoid stereotypes and the rebound effect because it leads people to consider alternatives beyond the dominant responses (Sassenberg & Moskowitz, 2005; Sassenberg et al., 2017; Sassenberg et al., 2021). Therefore, policymakers may emphasise divergent thinking/alternative thinking instead of stereotype suppression to construct a society with sustainable "diversity and inclusion".

Say "diversity and inclusion" in an alternative way. This piece of implication may be irrelevant to the research findings, but it came out from the research process. Therefore, we include it in this section. We suggest the policy maker change the name "diversity and inclusion" to "uniqueness and together". Here is the logic. Diversity and inclusion are leader-orientated words. When a leader looks at the people in a society, they may say, "our society has diverse groups and is in an inclusive status". However, these two words may make less sense to the other members of the society because they are the people being classified as part of diverse groups and being included (i.e., passive verbs). If we use "uniqueness and together", society members may be able to perceive the concept actively. Because each of them is a unique being, and they are together in society.

Develop integrative tools for diversity and inclusion. We suggest policymakers encourage integrative thinking tools to realise a diverse and inclusive society. Instead of putting a diverse group of people together and leaving them to interact, the government should think about developing integrative tools that could create a truly acceptant and equal society. For instance, we can develop more tools that integrate the uniqueness of human-being and, in turn, make individuals better persons and society a harmonious space.

Conclusion

Overall, this research project employs integrative and social cognitive perspectives to understand creative thinking and its role in real-life situations. We conducted four pilot studies and seven main studies using quantitative research methods to answer the research questions. We also generated four original tools: FST, RRCTSF, an un-stereotype workshop, and a COVID-19 prevention product design task. Our empirical evidence showed that 1) FST might measure different aspects of AUT and CRAT. 2) Comparative social feedback may change the speed and performance in creative thinking and product design tasks. 3) Advertising stereotype avoidance positively correlates with marketers' divergent thinking and negatively correlated with audience perceived creativity. 4) Divergent and integrative thinking may predict the creative product design of engineering students. Due to the theoretical and methodological limitations, the wide acceptance of our findings and tools may require further replications and validations. However, our original evidence and tools have informed multiple directions for the psychological research and practical development of creativity.

Appendix A: Creativity Psychometric Tools Searching Process

To review existing creativity tasks, we replicated a method from a previous study (Henriksen et al., 2015).

First, we used Ovid search engine and searched in American Psychological Association's (APA) PsycTests ® Database and used the keywords "Creativity", "Creative", "Creativ*", and "Creativity OR Creative". After that, we combined the four results, and we got 629 results from 1900 – April 2021.

Then we looked up the columns in the excel table:

They were all APA PsycTests. Regarding the test type, 359 were original measures, 85 were the revision of original measures, and 51 were the translation of established measures. There were 13 types shown in Figure. Two hundred eighty-seven were available, and 153 did not provide validity (neither test validity, convergent validity, discriminate validity, test reliability, internal consistency, measurement model, structural equation modelling, nor qualitative assessment method).

After that, we summarised the category and numbers:

Battery: 4; Checklist: 6; Coding Scheme: 6; Index/Indicator: 12; Interview/schedule/guide: 7; Inventory/questionnaire: 163; Rating scale: 197; Survey: 21; Task: 23; Test: 76; Vignette/scenario: 3; Blanks: 11.

Then, we excluded:

We excluded tasks without any validity and items available, 220 left. We also excluded measures that were neither tasks nor tests. In addition, we excluded translational measures. More than that, we excluded the tasks that explicitly taught us it was assessing one's "satisfaction", "expression"," perception", "self-reported", "motivation", "traits", "personality", "inquiry", "prototype", "team creativity", "expression", "scale", "attention", "intelligence orientated", explicitly not for creativity (e.g., Stroop test). Twenty-two tasks left.

Finally, we added:

Because synthesis was sometimes not included in creativity, to avoid the neglect of important creative subprocess tasks, we added tasks that have been mentioned in a good review(Lubart, 2001). We chose this reviewed paper for two reasons: 1) it was not old. 2) we shared the definition creative process. We also added TTCT and CRAT even though they were not included in the search results.

Appendix B: Alternative Uses Task – Data Collection, Marking, and Cleaning

Data collection

1. AUT instruction

“Now, you are entering Usages Task. Please read the instructions very carefully. You will only be able to proceed to the official game if you correctly answer the true-or-false questions below. You will have to complete six sessions of the Usages Task.

Each session is made of 2 stages:

Stage 1: Try to devise as many different and unusual uses for an object (e.g., bottle) as possible. Please type uses in the boxes provided. You may also explain the usages to perfect your answers. For instance: if you change the object's shape, you can explain the process; If you feel the use is unusual, you can give an example.

Stage 2: You will then be asked to drop your answers (i.e., uses) into three separate categories: 1) Basic Uses category: The uses you drop in this category are the basic uses for the given object. 2) Alternative Uses category: The uses you drop in this category are the common uses for the given object, but they are NOT the main purpose for which the given object is designed. 3) Unusual Uses category: The uses you drop into this category are unusual when considering the given object's uses. The uses themselves DO NOT have to be unusual.

Using "a bottle" as an example:

A basic use - storing water

An alternative use - storing seeds

An unusual use - flower pot

There is NO right or wrong categorisation. You DO NOT need to put an object into each of the three categories.

Important notes:

- There will be 20 boxes available for each object - you do NOT have to fill them all in but try to fill as many as you can. Only input one use into each box.
- You will have 5 minutes per session, but you do NOT have to wait until the time is complete.
- You should NOT take a break during one session, but you can take a break between sessions.”

2. AUT instruction test

“Please indicate whether the following statements are true or false.

Statement	Correct answer
I should generate as many objects as I can.	True
I will have to complete six sessions of Usages Tasks.	True
I should generate as many uses as possible.	True
I should better explain the usage, if the usage of an item requires the change of the item’s shape or if I feel the use is unusual.	True
I must fill all of 20 boxes provided	False
The uses I drag into Alternative Uses Category must be unusual	False

”

3. AUT

“**Page 1:** Please name as many possible uses as possible in the boxes below. Try to be as creative as possible. (Do not forget to explain the usages to perfect your answers.)

1. Brick	2. Hanger	3. Paperclip
[Insert an image of a brick]	[Insert an image of a hanger]	[Insert an image of a paperclip]
4. Tire	5. Newspaper	6. Mug
[Insert an image of a tire]	[Insert an image of newspaper]	[Insert an image of a mug]

Page 2: Please drop your answers into the appropriate box.

Participants answers that they have wrote in page 1

Basic Uses

Alternative Uses

Unusual Uses

”

4. Notes for experimenters

1) We showed the instruction and instruction test on the same page since the test aimed to ensure participants knew the key rules rather than challenge them. 2) We randomised the instruction test questions. 3) We showed participants the name and images of the objects. 4) We provided participants with 20 blank long textboxes for answers. 3) Participants saw six blocks of AUT questions in random order. In each block, there was one object.

Data marking.

1. Marking criteria training: Judges read the marking sheet template and examples (see next page for template and example) and asked questions if there were any.

2. Answer pool: The experimenter generated an answer pool which included all answers.

3. Frequency marking: The experimenter observed the frequency of occurrences of suggested uses in the answer pool (frequency = the number of suggested uses / the number of answers X 100%). If the frequency exceeded 4%, the experimenter scored the response as 1. If the frequency was between 2% - 4%, the experimenter scored the response as 2. If the frequency was less than 2%, the experimenter scored the response as 3.

4. Categorisation: Experimenter categorised the uses based on personal experience and evaluation. Judges had a chance to adjust it in the marking sheet.

5. Marking sheet: For each FST question, we asked judges to mark on a marking sheet in which uses were in the first row. Therefore, they received six AUT marking sheets in Study 1.

Marking template.

Marking template	Criteria explanation	Answer 1	Answer 2	…
Appropriateness for tire	If the use is appropriate, we press 1—otherwise, press 0.			
Originality	Suppose the appropriate dimension was 0, press 0. Otherwise: if the answer is novel and unexpected, we press 3. Otherwise: if the answer is novel and not unexpected, press 2. Otherwise, if the answer is easy to think of, press 1.			
Adjusted category	The experimenter has classified uses in different categories, which is the default value in this row. Please change the category's name if you feel the default category is inappropriate.	Category name 1	Category name 1	…

Marking examples

Marking example	Swing set	Weight lift
Appropriateness for **tire**	The judge thought the tire could be used as a swing set. S/he pressed 1.	The judge thought we could use a tire to do weight lift. S/he pressed 1.
Originality	The judge thinks "swing set" is novel and expected, she marked 2.	The judge thought weight lift was novel and expected. S/he pressed 2.
Adjusted category	The judge agreed with the default category of swing set as “entertainment”, and s/he did not change it.	The judge disagreed with the default category of exercise as “entertainment”, and s/he changed it to “exercise”.
(There is no right or wrong answer in marking, please mark based on the template explanation rather than the experimenter’s opinion in the example)		

Data cleaning.

Variables	Excel functions
Appropriate fluency	Countif (Appropriateness = 1)
Appropriate flexibility	If Appropriateness = 1 & if Adjusted category is a new name, 1. Otherwise, 0.
Appropriate originality	If Appropriateness = 1, novelty.
Appropriate frequency	If Appropriateness = 1, frequency.

Appendix C: Compound Remote Associate Test – Data Collection and Marking

Data collection.

1. CRAT instruction

"Now, you are entering Relatedness Task.

Please read the instructions very carefully. You will only be able to proceed to the official game if you correctly answer the true-or-false questions below.

You will have to complete six sessions of the Relatedness Task. Each session is made of five word-puzzle:

In each puzzle, you will see three stimulus words. Then, you should come up with the fourth word that makes up a common compound word or phrase with each of the three stimulus words.

Here is an example, if you see "cottage/swiss/cake", you should come up with the fourth word - "cheese" ("cottage cheese", "Swiss cheese", "cheesecake").

Important notes:

- Each word puzzle has a single correct answer. Please input the answer into the answer box.
- There will be five puzzles per session - you do NOT have to answer them all but try to fill as many as possible.
- You will have three minutes per session, but you do NOT have to wait until the time is complete.
- You should NOT take a break during one session, but you can take a break between sessions.

2. CRAT instruction test

“Please indicate whether the following statements are true or false.

Statement	Correct answer
My answer should be able to make up a common word or phrase with each stimulus word.	True
I will see two stimulus words.	False
I cannot take a break between different sessions, but I can stop during one session.	False
I have 8 minutes to complete each session.	False
I will encounter six sessions; each session is made up with ten puzzles	False

“

3. CRAT

“Please come up with the fourth word that makes up a common compound word or phrase with each of the three stimulus words. “

(See Table in the next page)

4. Notes for experimenters

1) We showed the instruction and instruction test on the same page since the test aimed to ensure participants knew the key rules rather than challenge them. 2) We randomised the instruction test questions. 3) Participants saw six blocks of CRAT questions in random order. In each block, there were three sessions. Each participant was randomly allocated only one session of CRAT questions.

	Session 1					Session 2					Session 3				
Block 1	forward/flush/razor	man/glue/star	pea/shell/chest	trip/house/goal	loser/throat/spot	hold/print/stool	pine/crab/sauce	fly/clip/wall	carpet/alert/ink	age/mile/sand	end/line/lock	note/chain/master	time/blown/nelson	wheel/hand/shoppin	high/district/house
Block 2	Self/attorney/spending	Keg/puff/room	Mouse/bear/sand	Break/bean/cake	Cane/daddy/plum	Nose/stone/bear	Grass/king/meat	Home/arm/room	Right/cat/carbon	Sense/courtesy/place	Line/fruit/drunk	Nuclear/feud/album	Mail/board/lung	Pile/market/room	Fox/man/peep
Block 3	Hungry/order/belt	Wise/work/tower	Cry/front/ship	Mill/tooth/dust	Fountain/baking/pop	Master/toss/finger	Date/alley/fold	Illness/bus/computer	Down/question/chec	Dress/dial/flower	Bump/egg/step	Baby/spring/cap	Iron/shovel/engine	Sleeping/bean/trash	Piece/mind/dating
Block 4	Mate/shoes/total	Tooth/potato/heart	Roll/bean/fish	Guy/rain/down	Dream/break/light	Pet/bottom/garden	Spoon/cloth/card	Oil/bar/tuna	Home/sea/bed	Fish/mine/rush	Off/military/first	French/car/shoe	Fork/dark/man	Eight/skate/stick	Water/mine/shaker
Block 5	Land/hand/house	Fence/card/master	House/thumb/pepper	Rain/test/stomach	Flake/mobile/cone	Control/place/rate	Cut/cream/war	Test/runner/map	Basket/eight/snow	Political/surprise/line	Tomato/bomb/picker	Chamber/mask/natur	Lift/card/mask	Force/line/mail	Tank/hill/secret
Block 6	Board/blade/back	Type/ghost/screen	Hammer/gear/hunter	Horse/human/drag	Safety/cushion/point	Fight/control/machin	Shock/shave/taste	Wet/law/business	Dust/cereal/fish	Fur/rack/tail	Rope/truck/line	Dive/light/rocket	Teeth/arrest/start	Health/taker/less	Stick/maker/point

Data marking & cleaning.

1. If the answer was correct as in the CRAT standard version, the experimenter marked it as 1. Otherwise, 0.

2. Variables

a. Accurate fluency = the number of correct answers

b. Rated Correct / Speed = the number of correct answers / time length

Appendix D: Function Synthesis Task – Data Collection, Marking, and Cleaning

Data collection.

1. FST instruction and instruction test.

"Now, you are entering Function Synthesis Task. Please read the instructions very carefully. You will only be able to proceed to the official game if you correctly answer the true-or-false questions below. You will have to complete six sessions of the Function Synthesis Task.

Each session is made of 2 stages. **In stage 1,** you will see three functions. You should come up with objects that fulfil the three functions. Name as many objects as possible. Please type objects in the boxes provided. You may also explain the objects to perfect your answers. For instance: if you change the shape of objects to satisfy the functions, you can explain the process; If you feel the object is unusual, you can briefly introduce it. In stage 2, you will be asked to drop your answers (i.e., objects) into three separate categories: 1) Basic objects category: The objects you drop into this category are designed to fulfil the three functions given. 2) Alternative objects category: The objects you drop into this category are frequently considered when considering the three functions. 3) Unusual objects category: The objects you drop into this category are unusual to think of when considering fulfilling the three functions given. The objects themselves DO NOT have to be unusual.

Using functions "sound, entertain, perform" as an example:

A basic object - piano

An alternative object - speaker

An unusual object - firework

There is NO right or wrong categorisation. You DO NOT need to put an object in the three categories.

Important notes:

- There will be 20 boxes available for each object - you do NOT have to fill them all in but try to fill as many as you can.
- Only input one object into each box.
- You will have 7 minutes per session, but you do NOT have to wait until the time is complete before moving on.
- You should NOT take a break during one session, but feel free to take breaks between sessions.

2. FST instruction test

“According to the instruction, please indicate whether the following statements are true or false.

Statement	Correct answer
My answer must fulfil all the three functions given.	True
I should generate as many objects as I can.	True
I have 15 minutes to complete each session.	False
I can give unusual objects and change the shape of original objects, but I should better explain.	True
The objects I drop into the Unusual Objects Category must be unusual in our daily lives.	False
I should generate as many uses as possible.	False

”

3. FST

“**Page 1:** Please name as many objects that can fulfil all three functions as possible. Try to be as creative as possible. (Do not forget to explain your rationale to perfect your answers.)

- Interact/immerse/recreate
- Profit/advertise/decorate
- Customise/comfort/sanitise
- Illuminate/alarm/contain
- Protect/entertain/comfort
- Store/package/disseminate

Page 2: Please drop your answers into the appropriate box

<table>
<tr><td rowspan="3">Participants answers that they have wrote in page 1</td><td>Basic Objects</td></tr>
<tr><td>Alternative Objects</td></tr>
<tr><td>Unusual Objects</td></tr>
</table>

”

4. Notes for experimenters

1) We showed the instruction and instruction test on the same page since the test aimed to ensure participants knew the key rules rather than challenge them. 2) We randomised the instruction test questions. 3) We provided participants with 20 blank long textboxes for answers. 4) Participants saw six blocks of FST questions in random order. In each block, there was one object.

Data marking.

1. Functions training: For each FST question (e.g., protect/entertain/comfort), judges read definitions and examples of the given function. The definitions and examples were extracted from Oxford English Dictionary. For example, the function *entertain* has three definitions: 1) *provide (someone) with amusement or enjoyment (e.g., a tremendous game that thoroughly entertained the crowd)*, 2) *receive (someone) as a guest and provide them with food and drink (e.g., a private dining room where members could entertain groups of friends)*, and 3) *given attention or considerations to an idea or feeling (e.g., Washington entertained little hope of an early improvement in relations)*. After learning the functions, judges proposed questions to the experimenter if they had any.

2. Marking criteria training: Judges read the marking sheet template and examples (see next page for template and example) and asked questions.

3. Answer pool: The experimenter generated an answer pool which included all answers.

4. Frequency marking: The experimenter observed the frequency of occurrences of suggested objects in the answer pool (frequency = the number of a suggested object / the number of answers X 100%). If the frequency exceeded 4%, the experimenter scores the response as 1. If the frequency was between 2% - 4%, the experimenter scored the response as 2. If the frequency was less than 2%, the experimenter scored the response as 3.

5. Categorisation: Experimenter categorised the objects based on personal experience and evaluation. Judges had a chance to adjust it on the marking sheet.

6. Marking sheet: For each FST question, we asked judges to mark on a marking sheet in which all the answers (in the answer pool) were in the first row. Therefore, they received six FST marking sheets in Study 1.

Marking template.

Marking template	Criteria explanation	Answer 1	Answer 2	…
MNC for function 1	Suppose the answer is not an object, press 0. Otherwise: If the object is designed for “function 1” (main purpose; M), press 3. Otherwise: If the object naturally (N) provides the function ‘function 1”, press 2. Otherwise: If the object is designed to help “function 1” or can provide the function “interact” with a little bit of assistance (C), press 1. Otherwise, do not press.			
MNC for function 2	Same instruction as above but replace function 1 with function 2.			
MNC for function 3	Same instruction as above but replace function 2 with function 3.			
Appropriateness	Suppose we pressed 0 for any of the above MNC dimensions, press 0. Otherwise: press 1.			
Originality	Suppose the appropriate dimension was 0, press 0. Otherwise: if the answer is novel and unexpected, we press 3. Otherwise: if the answer is novel and not unexpected, we press 2. Otherwise, if the answer is easy to think of, press 1.			

Accuracy	Suppose the appropriate dimension was 0, press 0. Otherwise, If the answer individually provides three functions to the same objects, press 1. Otherwise, press 0.			
Adjusted category	The experimenter has classified objects in different categories, which is the default value in this row. Please change the category's name if you feel the default name is inappropriate.	Category name 1	Category name 1	…

Marking examples.

Marking example	Billboard	Flower
MNC for **interact**	The judge thought “billboard” could provide the function “interact” with little assistance. S/he pressed 1.	0
MNC for **immerse**	The judge thought “billboard” could provide the function “interact” with little assistance. S/he pressed 1.	1
MNC for **recreate**	The judge thought people designed “billboards” to “recreate” – entertainment. S/he pressed 2.	1
Appropriateness	Since none of the above dimensions was 0, the appropriateness was automatically set up as 1.	Since none of the above dimensions were 0, the appropriateness automatically set up as 0.
Originality	The judge thought the billboard was novel and expected. S/he pressed 2.	Since appropriateness was 0, originality automatically set up as 0.
Accuracy	The judge thought "billboard" was not a correct answer for the two functions of "decoration" and "profit" since billboards could decorate the city but makes a profit for a company. Therefore, it is not providing the two functions to the same object. Therefore, she pressed 0.	Since appropriateness was 0, accuracy automatically set up as 0.
Adjusted category	The judge	Since appropriateness was 0, category

<table>
<tr><td></td><td>agreed with the default category of the billboard as “advertising”. Therefore, s/he did not change it.</td><td>automatically changed to 0.</td></tr>
<tr><td colspan="3">(There is no right or wrong answer in marking, please mark based on the template explanation rather than the experimenter’s opinion in this example)</td></tr>
</table>

Data cleaning.

Variable	Excel functions
Appropriate effectiveness	If Appropriateness = 1, Average (MNC for Function 1, MNC for Function 2, MNC for Function 3). Otherwise, 0.
Appropriate fluency	Countif (Appropriateness = 1)
Appropriate flexibility	If Appropriateness = 1 & if Adjusted category is a new name, 1. Otherwise, 0.
Appropriate novelty	If Appropriateness = 1, novelty.
Appropriate frequency	If Appropriateness = 1, frequency.
Accurate effectiveness	If Accuracy = 1, Average (MNC for Function 1, MNC for Function 2, MNC for Function 3). Otherwise, 0.
Accurate fluency	Countif (Accuracy = 1)
Accurate flexibility	If Accuracy = 1 & if Adjusted category is a new name, 1. Otherwise, 0.
Accurate novelty	If Accuracy = 1, novelty.
Accurate frequency	If Accuracy = 1, frequency.

Appendix E: Creative Self-efficacy Inventory

“The statements list 28 different activities. Please rate how confident you are that you can do them as of now. Rate your degree of confidence by dragging the sliders to a point that within 0 and 100. 0 means Not at all confident; 100 means Highly certain that you can do the task.

- Get a large number of different ideas or responses
- Come up with many possible solutions to a situation.
- Arrive at a variety of conclusions given a difficult situation.
- Think of many answers to a difficult problem or situation.
- Come up with responses that from different categories, not just different responses.
- Answer problems in different ways, each of which are unique and special.
- Think of many types of ideas while considering a problem.
- Answer problems in different forms or styles.
- Think of ways to defend a crazy thought, by thinking back on what you already know.
- Talk to your friends about wild ideas and make them sound reasonable.
- Tell stories based on dreams you had, even if you need to fill in answers.
- Connect daydreams or new ideas to things you have already learned.
- Be the first in a group to come up with an original idea.
- Arrive at a novel solution before other people.
- Beat other people in imaging a brand-new idea first.
- Think of ideas no one else has.
- Make sense of something you want to learn to do.
- Start to learn to do something, even if there are obstacles to doing so.
- Teach yourself how to do something new.
- Choose to do something that is more important within your culture.
- Create novelty that people will choose, over other novelties available.
- Find an audience that is well-connected to others in society.
- Network with people to convince them that what you made is best.
- Convinced other that you have made a valuable contribution.
- Be motivated to come up with new ideas.
- Have fun coming up with new ideas, after having learned from others.

- Wake up feeling like you can come up with new ideas if you want.
- Sustain wonder about something, even after working with it for years or decades.”

Appendix F: Study 1 Script

1. Welcome

1.1. Work ID: “Please enter your HIT work ID here”.

1.2. Language: “To save time, you must be an English Native Speaker to join this experiment. If you are not an English Native Speaker, the experiment will not make sense to you”. Participants chose “I am an English Native Speaker” or “I am not an English Native Speaker”. Only people who chose the former choice proceeded to the next stage.

1.3. Participant’s information sheet and consent sheet. Only the participants who gave consent proceeded to the next sheet.

1.4. Brief introduction to creativity: “What is creativity? How to measure one’s potential to be creative? Research suggested that creativity correlates with one’s performance in the Relatedness Task, the Usage Task, and the Function Synthesis Task. You will be doing these tasks now!”

2. Main experiment

2.1. Participants saw AUT (Appendix B), CRAT (Appendix C), and FST (Appendix D) in random order.

2.2. Manipulation check: “For the three tasks you just finished, where did you get the answers from (multiple choices)?” For FST, AUT, and CRAT, participants chose “I came up with the answers by my original ideation”, “I extracted the answers from my memory, since I have done the task before”, or “I searched the answer online.”

2.3. Creative activity inventory questionnaire (Appendix E).

2.4. Demographic information

2.4.1. Age, country of origin, country of residence, gender, and working field (“Architecture, Arts, Biological and life science, Business and Management, Computer science, Economics, politics, and social sciences, Education, Engineering, History, Language and Culture, Law, Mathematics and Statistics, Medicine, Physical science, Psychology, Chemistry”).

Appendix G: Motives of Social Comparison

Items
Self-improvement
So, I can get better.
To give me a goal.
To improve my own situation.
To learn what to do or what not to do.
Because they serve as role models
Common bound
For empathy and support.
So, I won't feel along or isolated.
To share experiences.
Because we have things in common.
Altruism
To help them.
Because I feel sorry for them.
To make them feel better.
Self-enhancement
To make myself feel better.
To feel good about my own situation.
To convince myself that I am not like them.
So, I can attribute the illness to something else besides me.
To reassure me about my own situation.
Self-destruction
To confirm my fear of getting worse.
To prove how helpless the situation really is.
To confirm my belief that I am in trouble.
So, I can give up.
Self-evaluation
To see how I am doing.
To provide insight into my own situation.
To see if I am recovering fast enough.

Appendix H: Individualism and Collectivism

Items
Horizontal individualism
I would rather depend on myself than others.
I rely myself most of the time; I rarely reply on others.
I often do “my own thing”.
My personal identity, independent of others, is very important to me.
Vertical individualism
It is important that I do my job better than others.
Winning is everything.
Competition is the law of nature.
When another person does better than I do, I get tense and aroused.
Horizontal collectivism
If a co-worker gets a prize, I will feel proud.
The well-being of my co-workers is important to me.
To me, pleasure is spending time with others.
I feel good when I cooperate with others.
Vertical collectivism
Parents and children must stay together as much as possible.
It is my duty to take care of my family, even when I must sacrifice what I want.
Family members should stick together, no matter what sacrifices are required.
It is important to me that I respect the decisions made by my groups.

Appendix I: Study 2 Script

1. Welcome
    1.1. "Please make sure you are an English NATIVE-SPEAKER. A built-in system will tell the experimenter whether you are an English native speaker. We cannot redeem your effort if you are NOT an English native speaker since the game would not make sense to you. Thanks for your interest in our game."
    1.2. Participants information sheet and consent
2. Instruction and understanding test

"Please read very carefully. We will proceed to the official game only if you answer all true-or-false questions correctly. You have three chances.

In this game, you will see three stimulus words. Then, you should come up with the fourth word that makes up a common compound word or phrase with each of the three stimulus words. So, please type the fourth word in a given box.

For instance, if you see 'cottage/Swiss/cake', you should come up with the fourth word 'cheese' ("cottage cheese", "Swiss cheese", "cheesecake") and type "cheese" in a box.

The game consists of 3 blocks. In different blocks, you will have chances to compete with another player from our participant pool, play on your own, and be evaluated by our star-rating system. You can take a break between different blocks, but you cannot stop during the blocks. There are six rounds in each block. You will have 180 seconds (3 minutes) per round, consisting of three puzzles. You do not have to run out of time. Try to be as accurate and fast as you can.

According to the Instruction, please indicate whether the statements are true or false.

- I should write the fourth word in a blank box.
- I should write three compound words or phrases in a blank box.
- I will see two stimulus words.
- I will have a chance to compete with another player from a participant pool.
- I can take a break between different blocks, but I cannot stop during the blocks.
- I do not have a chance to play on my own.
- There are three blocks of which there are six rounds in each."

2.1. If participants did not pass the instruction understanding test: “Sorry, you did not pass the test of instruction, so we cannot proceed you to the game. You can re-enter the survey and try it again. Thank you very much!”

2.2. If participants pass the instruction understanding test: “Congratulations! You pass the test. Let’s start!”

3. Trial round

3.1. Question screen:

“Here is a trial round. Please come up with the fourth word for the following word puzzles:

“political/surprise/line”

“rocking/wheel/high”

“measure/worm/video””

3.2. Feedback screen:

“Number of correct answers:

(The number of actual numbers of correct answers)”

4. Main experiment

*The blocks for competition, star-rating, and control occur randomly.

4.1. Competition block

4.1.1. Break screen:

“If you feel tired now, please take a break before clicking on the button below, which will proceed you to the next block. If you stop during the next six rounds, your responses will be invalid, and we cannot redeem your work in this situation.”

4.1.2. Introduction screen:

“You will compete with another player in this block (including six rounds). We will match you with another participant in our participant pool. The more correct answers you get, the higher likelihood you win the round; the fewer correct answers you get, the higher likelihood you lose the round. The competition outcome of each round is independent of other rounds.”

4.1.3. Competitor allocation screen:

“We are allocating you a competitor from our participant pool ......”

4.1.4. Question screen:

For example: Peach/arm/tar, Fox/man/peep, Iron/shove/engine

4.1.5. Feedback screen:

You win in this round!

Your number of correct answers: 1

Competitor's number of correct answers: 0

OR

You lose in this round.

Your number of correct answers: 1

Competitor's number of correct answers: 2

OR

Draw in this round.

Your number of correct answers: 1

Competitor's number of correct answers: 1

**Competitor's number of correct answer was randomly selected from 0-3.*

4.2. Star-rating block

4.2.1. Break screen: see 4.1.1.

4.2.2. Introduction screen:

"In this block (including six rounds), your performance will be evaluated by our 'star grading system'. When giving an evaluation, we consider your results, other players' results, and the difficulties of questions. Three-star indicates a great performance relative to all the other players, and a one-star indicates poor performance. The more correct answers you get, the higher likelihood you are evaluated as a three-star player; the fewer correct answers you get, the higher likelihood you are evaluated as a one-star player. The evaluation of each round is independent of other rounds."

4.2.3. Question screen: see 4.1.4.

4.2.4. Feedback screen:

★

You are evaluated as a one-star player in this round.

Number of correct answers: 1

OR

★★

You are evaluated as a two-star player in this round.

Number of correct answers: 1

OR

★★★

You are evaluated as a three-star player in this round.

Number of correct answers: 1

4.3. Control block

4.3.1. Break screen: see 4.1.1.

4.3.2. Introduction screen:

"You are playing independently in this block (including six rounds). Therefore, you will see the number of correct answers at the end of each round."

4.3.3. Question screen: see 4.1.4.

4.3.4. Feedback screen:

Number of correct answers: 1

5. Questionnaires

5.1. Motives for social comparison (see Appendix G)

5.2. Horizontal and vertical individualism and collectivism (see Appendix H)

6. Demographic information and debrief

6.1. Age, gender, working field.

6.2. Debrief.

Appendix J: Scoring and Star Rating Algorithm for CRAT (Study 2)

There are eight possibilities of outcomes. We gave participants star ratings based on the outcome and difficulties of questions. For instance, if a participant answered the easiest and the most difficult items correctly (True) and answered the medium-difficult item incorrectly (False), then we rate the participant a one-star player in that round.

For each round,	If			Then
	The easiest item	The medium difficult item	The most difficult item	Star rating
1	False	False	False	1-star
2	True	False	False	1-star
3	False	True	False	1-star
4	False	False	True	2-star
5	True	True	False	2-star
6	Ture	False	True	2-star
7	False	True	True	3-star
8	True	True	True	2-star

Appendix K: Study 3 Script

Study 3 replicated the procedure in Study 2. Therefore, we point out only the changes that were made based on Study 2 Script.

1. Changes in Instructions and Understanding Tests

"Please read very carefully. You will only be able to proceed to the official game if you correctly answer the true-or-false questions below. You have three chances. One session of the "Many Uses Task" is made of 2 stages:

Stage 1: Try to devise as many different and unusual uses for an object (e.g., bottle) as possible. Please type the uses in the boxes provided. Only input one use into each box. There will be fifteen boxes available- you don't have to fill them all in but try to fill as many as you can.

Stage 2: You will then be asked to place your ideas (i.e., uses) into three separate categories:

- Drop the basic uses in the "Basic Uses" category,
- Drop what you consider an alternative use in the "Alternative Uses" category,
- Drop the most unusual uses in the "Unusual Uses" category.

Using a bottle as an example:

A Basic use - "storing water"

An Alternative use -"storing seeds"

An Unusual use - "flowerpot"

After you have done a session, you will receive a score that represents the creative level of your ideas. We consider both the number of uses and how original they are. You cannot score highly by just focusing on one of these criteria and neglecting the other. The scoring is only based on your performance in stage 1 (not stage 2). Answers that are either from online sources or are inappropriate (i.e., random words) will be marked as invalid and will be rejected. As a result, we will not be able to redeem your work.

The game consists of 3 blocks. Each block will consist of 2 sessions (6 sessions in total). In different blocks, you will have chances to compete with another player, play on your own, and be evaluated relative to our whole participant pool. You can take a break between different blocks, but you cannot stop during the blocks. You will have 5 minutes per session. You do not have to wait until the time is complete.

According to the Instruction, please indicate whether the statements are true or false.

- The scoring will be based on the number of uses only.
- I should write multiple answers in a single box.
- There are six sessions in the whole game.
- I will have a chance to compete with another player that is from a participant pool.
- I can take a break between different blocks but i cannot stop during the blocks.
- I do not have a chance to play on my own.
- The scoring will be based on your performance in Stage 1 and 2."

2. Changes in Trial Rounds

   2.1. Question screen:

   "Here is a trial round. The object is a "Knife". Please name as many uses as possible that you can think of in the boxes below. Try to be as creative as possible. (Here, we provide participants with 15 long textboxes)"

   2.2. Self-evaluation screen:

   "Please drop your answers in the appropriate boxes. (Here, participants see the answers they have come up with in the Question screen and three boxes (i.e., basic uses, alternative uses, and unusual uses) that they can drop the answers in)."

   2.3. Feedback screen:

   Your score is 0.40/1

   (This means that you are scored 0.40 out of 1.00)

3. Changes in Competitor Allocation Screen in Competition Block

   3.1. Allocating screen: "We are allocating you a competitor from our participant pool."

   3.2. Allocated screen: "You will be competing against player 10. Please wait a few moments until you are forwarded to the start of the round."

4. Changes in Question Screens

   We employed AUT questions (rather than CRAT questions) in this study. Please see Changes in Trial rounds for an example.

5. Changes in Feedback Screens
    5.1. For all the feedback screens in AUT, the word “score” replaced the “number of correct answers” in CRAT, considering that AUT did not have single correct answers. For an example, please see the trial round above.
    5.2. In evaluation screen, the one-star player was presented as ★☆☆ rather than ★ and the two-star player was presented as ★★☆, to remind people this star-rating system had a maximum of three stars.

Appendix L: Scoring and Star Rating Algorithm for AUT (Study 3)

The algorithms for AUT feedback: since AUT did not have pre-determined correct answer(s), we generated algorithms that determine the scores of participants and competitors and star ratings and increase the trustworthiness of the social feedback. Our algorithms did not aim to provide an accurate evaluation to participants but to provide more credible feedback manipulation compared to random scores and star ratings.

The underlying algorithm participant' scores:

If fluency =	Then score was in the range of
0	0.00
1 – 3	0.20 – 0.35
4 – 6	0.45 – 0.69
7 +	0.70 – 0.90

The algorithm underlying the competition condition:

Fluency	Participant	Competitor	Competition outcome
0	0.00	0.20 – 0.80	Lose
1 – 3	0.20 – 0.35	0.20 – 0.80	Win *or* Lose *or* Draw
4 – 6	0.45 – 0.69	0.20 – 0.80	Win *or* Lose *or* Draw
7 +	0.70 – 0.90	0.20 – 0.80	Win *or* Lose *or* Draw

The algorithm underlying the star-rating system:

If fluency =	And if the score was in the range of	Then star rating
0	0.00	★☆☆
1 – 3	0.20 – 0.33	★☆☆
	0.34 – 0.36	★★☆
4 – 6	0.45 – 0.50	★☆☆
	0.51 – 0.60	★★☆
	0.61 – 0.69	★★★
7 +	0.70 – 0.79	★★☆
	0.80 – 0.90	★★★

Appendix M: Study 4 Script

Study 4 replicated the procedure in Study 2. Therefore, we point out only the changes that have been made based on Study 2 Script.

1. Changes in Instructions and Understanding Tests

"Please read very carefully. You will only be able to proceed to the official game if you correctly answer the true-or-false questions below. You have three chances. One session of the "Function Synthesis Task" is made of 2 stages:

Stage 1: You will see three functions. Please come up with objects that fulfil the three functions: name as many objects as possible. Please type objects in the boxes provided. You may also explain the objects to perfect your answers. For instance: You can explain the process if you change the object's shape to satisfy the functions; If you feel the object is unusual, you can briefly introduce it.

Stage 2: You will then be asked to drop your answers (i.e., objects) into three separate categories:

- Drop the basic objects in the "Basic Objects" category: the objects you drop into this category are designed to fulfil the three functions.
- Drop what you consider an alternative object in the "Alternative Objects" category: The objects you drop into this category are frequently thought of when considering the three functions given.
- Drop the most unusual objects in the "Unusual Objects" category: the objects you drop into this category are unusual to think of when considering fulfilling the three functions given. The objects themselves DO NOT have to be unusual.

Using the functions "sound, entertain, perform" as an example:

A Basic Object - "piano"

An Alternative object -"speaker"

An Unusual Object - "firework"

There is NO right or wrong categorisation. It is UNNECESSARY to put an object in a category. There will be 20 boxes available for each object – you do NOT have to fill them all in but try to fill as many as you can. Only input one object into each box. You will have seven minutes per question, but you do NOT have to wait until the time is complete before moving on.

Structure: the game consists of three blocks; each block will consist of two sessions (six sessions). In different blocks, you will have chances to compete with another player, play on your own, and be evaluated relative to our whole participant pool. You can take a break between different blocks, but you should NOT stop during the blocks.

Scoring: After you have done a session, you will receive a score that represents the creative level of your ideas. We consider both the number of uses and how original they are. Therefore, you cannot score highly by just focusing on one of the criteria and neglecting the others. *The scoring is only based on your performance in Stage 1 (not Stage 2). *Answers from online resources or inappropriate (i.e., random words) will be marked invalid and rejected. As a result, we may not be able to redeem your work.

According to the Instruction, please indicate whether the statements are true or false.

- My answers must fulfil all the three functions given.
- I should generate as many objects as I can.
- I have 15 minutes to complete each session.
- I can give unusual objects and change the shape of original objects, but I should better explain.
- The objects I drop into the Unusual Objects category must be unusual in our daily lives.
- I should write multiple answers in a single box.
- The scoring will be based on the number of objects only.
- I can take a break between different blocks, but I should not stop during the blocks.
- I do not have a chance to play on my own."

2. Changes in Trial Rounds

   We did not offer trial round in Study 5.

3. Changes in Competitor Allocation Screen in Competition Block

   3.1. Allocating screen: "We are allocating you a competitor from our participant pool."

   3.2. Allocated screen: "You will be competing against player 10. Please wait a few moments until you are forwarded to the start of the round."

4. Changes in Question Screens

We employed FST questions (rather than CRAT questions) in this study. For example：

"Functions:

- Interact
- Immerse
- Recreate

Please name as many objects as possible that can fulfil all three functions. Try to be as creative as possible (do not forget to explain your rationale to perfect your answers)."

5. Changes in Questionnaires

Instead of questionnaires on Horizontal and Vertical Individualism and Collectivism, and Motives for Social Comparison, we asked participants to finish questionnaires on the Creative self-efficacy inventory (see Appendix E).

Appendix N: Scoring and Star Rating Algorithm for FST (Study 4)

Participants' scores: to determine participants' scores, we generated a pool for appropriate answers (FST appropriate answer pool) based on the dataset in Study 1. If a participant's answer is in the FST-appropriate answer pool, the participant's score pluses 1. In the end, the participant's FST scores equal the number of FST answers that match the answer in the FST appropriate answer pool.

Competitors' scores: a random number from 3 to 7.

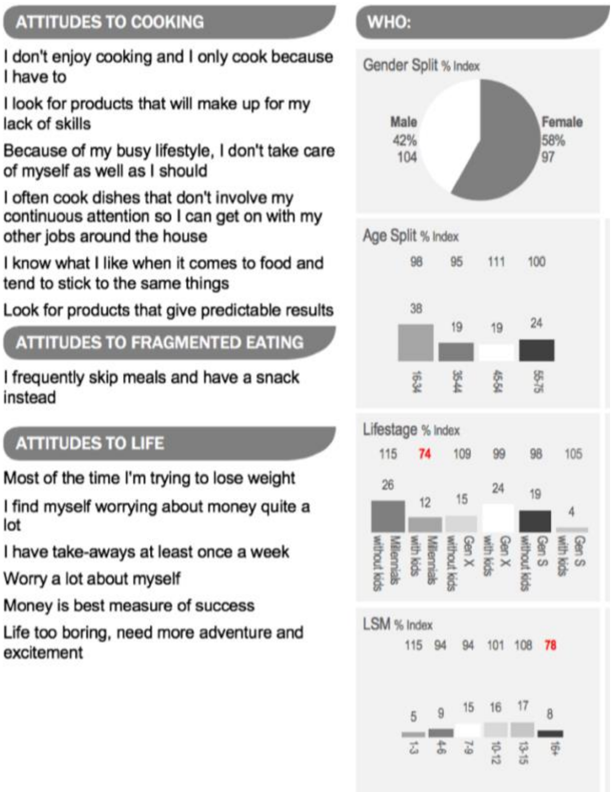
Star rating system: participants' score determines the star ratings that they receive. In detail:

Number of appropriate answers	Probability to get each star-rating (%)			
	0 star	1 star	2 stars	3 stars
0	100	0	0	0
1	0	90	10	0
2	0	70	30	0
3	0	30	60	10
4	0	10	60	30
5	0	0	30	70
6	0	0	10	90
>=7	0	0	0	100

Appendix O: Consumer Labelling Task

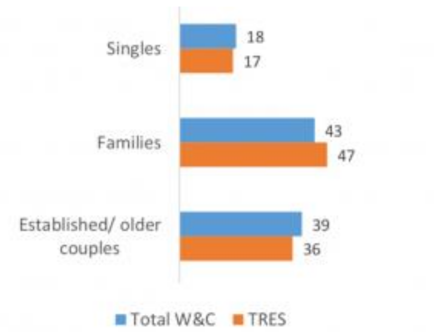
"Please read the instructions carefully. Then, we would like you to investigate some consumer information, build a description of the consumer from an existing list of attributes and behaviours, and write about this audience in your own words. "

(The consumer information for Consumer Labelling Task 1)



(The consumer information for Consumer Labelling Task 2)

Statements – top 2 box agreement	Gen Pop	Our Buyers
I Challenge Myself To Be A Better Me	71%	78%
I Make The Extra Effort To Look Good	60%	71%
I Do What Is Required To Stay On Top Of Things	63%	71%
My Looks Are Important To Me	73%	81%
I Like To Try The Newest Beauty Products	40%	55%
I Keep Up With Trends And Fashion	35%	46%



Cable TV Networks	TRES Buyers index vs. G.P.
TeenNick	152
Toon Disney	145
Nickelodeon	141
Disney Channel	138
Nick At Nite	136
Cartoon Network	127
MTV (Music Television)	125
VH1 (Video Hits One)	124

Basketball, football, softball, volleyball, baseball, tennis

Theme park, zoo, aquarium, camping trips

Go-karting, archery, skating, horse-riding, bowling

“Q1: Based on the data you have read, think about attributes, traits, interests etc., that might help bring this audience to life. Here are some descriptors to get you started. You can drag as many or as few (or none) as you wish.

Creative Eco-Friendly Risk Seeker Active Loyal Unhealthy Price-Driven Anxious Adventurous Lazy Security Preferred Convenience Preferred Self-Consciousness Socialising Status Driven Fickle	Drag the items to here

Q1b: Your ideas are very welcome! Please write more attributes/descriptors of the target customers not listed above.

Q2. In the text box below, start to bring this consumer to life in a way that would inform the creative development process. Please write an INSIGHT below in the first box.

Q3. Please write a TARGET CONSUMER description in the second box as you would in a Brand Love Key under ‘the people we serve or in an agency brief to the creative team (please write 150-200 words).

Q4. Do you think you received enough information to complete this task? Yes / No. If you answered no, please let us know in the box below what other information you would like to have. “

Appendix P: Study 5 Script

Pre-intervention assessment.

1. Welcome

"Please generate your unique anonymized code using the first two letters of your mother's first name and the day you were born. If, for your example, your mother's name is JOY and you were born on the 14th, your code is JO14."

2. DNA test: Predicting Your Results.

2.1. "Please list the key regions from which you think you descend (choose as many as you think are relevant).

Europe: Ashkenazi, East European, Sardinian, North Mediterranean, Finnish, Western European, Basque.

America: Amazonian, Central South American, North American.

Africa: Nilotic, Central African, West African, Bantu, Khoisan.

Oceano: Australia, New Zealand, Solomon Island, PNG, Fiji.

Middle East, North Africa: Levantine, Caucasus, Northwest African, Arabian.

Asia: Austronesian, Northeast Asian, South Asian, Central Asian, Siberia, Southeast Asian.

2.2. Please indicate what percentage of your DNA you think will be captured by these regions. For instance, if you expect 50% of your DNA to be East European, you should enter 50 in the box beside East European."

3. Consumer Labelling Task 1 (see Appendix O).

4. Alternative Uses Task (see Appendix B).

5. Self-Efficacy questionnaire and demographic information (age, gender, job title).

Manipulation.

The experimental group joined the un-stereotype workshop, and the control group did not join the workshop.

Post-intervention assessment.

1. Welcome

"Please generate your unique anonymized code using the first two letters of your mother's first name and the day you were born. If, for your example, your mother's name is JOY and you were born on the 14th, your code is JO14."

2. Consumer Labelling Task 1 & 2 (see Appendix O).

3. Alternative Uses Task (see Appendix B).

4. Self-Efficacy questionnaire and demographic information (age, gender, job title).

5. DNA test: reaction to results.

5.1. “In the box below, please tell us your DNA results, where you are from, and what you have discovered. We also appreciate it if you do not want to share this and feel uncomfortable doing so, but please be reassured that all data remains anonymous. - How surprised were you by your DNA results? Very unsurprised / somewhat unsurprised / do not know / somewhat surprised / very surprised. “

Appendix Q: List of Advertising Brands and Products

Here, we list the video advertisements we employed in studies 6a and 6b. Please be aware that, 1) for study 6a, we employed 66 advertisements in the study, but the video of two advertisements denied access at the time of writing. Therefore, we reported the information on 64 video advertisements. 2) For study 6b, the advertisements in high stereotypicality condition were labelled with a red background. The advertisements in medium stereotypicality condition were labelled in a yellow background colour. The advertisements in low stereotypicality condition were labelled in a green background colour. 3) The industry/product category column was based on authors' common sense and internet searches and did not have scientific support. The column was used to help the readers quickly grasp a broad picture of products.

	Advertising brand/product	**Industry / product category**		**Advertising brand/product**	**Industry / product category**
1	888 poker	Gambling	33	Weetabix	Cereals product
2	Army jobs	Recruitment	34	N.7	Personal care product
3	Ask Your Derm	Health care	35	Novetel	Hotel
4	Bakers	Dog food product	36	Nytol	Medication
5	Batiste	Hair care product	37	Royal mail	Postal service
6	Betway	Gambling	38	Royal mail	Postal service
7	Bisto	Food product	39	Royal British Legion	Charity
8	Bud light	Alchol drink	40	Richmond	Food product
9	Co-op	Food retail	41	Rowse	Honey manufacturer
10	Deezer	Online music streaming service	42	Sixt	Car rental
11	Doritos	Food product	43	Sky mobile	Phone
12	Glenfiddich	Alchol drink	44	Sofology	Furniture
13	Go compare	Financial service comparison website	45	Xperia Sony	Phone
14	Godiva	Food product	46	Specsaver	Glasses

15	Goodness knows	Food product	47	Stella Artois	Beer
16	Greenflag	Roadside assistance and vehicle recovery provider	48	Subway	Fast food
17	Guphoria	Food product	49	Sunny	Loan
18	Remington air plates	Personal care product	50	Swinton	Insurance
19	Hula Hoops	Food product	51	Sainsbury	Food retail
20	HSBC	Bank	52	Taylors	Coffee bags
21	IRN Bru	Soft drink	53	Tena	Pads
22	Crunchy Nut	Food product	54	F&F	Clothing
23	Kelly's	Food product	55	Freesat	Television
24	KP	Food product	56	Thorpe Park	Amusement Park
25	Topman	Clothing	57	Twinings	Tea bags
26	Martini	Alcohol drink	58	Tyrrells	Chips
27	Mazda	Car	59	uSwitch	Price comparison service and switching website
28	McCafe	Fast food company	60	Vanquis	Bank
29	McDonalds	Fast food company	61	Westland	Natrual lawn feed
30	McVitie's	Snack food brand	62	Typhoon	Tea bags
31	Moonpig	Card product	63	Wren	Kitchen retail
32	A toast to the world	Alcohol drink	64	Yorkshire tea	Tea bags

Appendix R: Study 6 Script

1. Welcome

1.1. Participant Information Sheet and Consent.

1.2. Please type in today's date

1.3. Before you start, please switch off phone/email/music so you can focus on the study.

1.4. For how long (in years) have you been a resident of UK?

2. Instruction

You will now be directed to videos showing different advertisements. In the next page, you will see blue hyperlinked text that has to be clicked to watch the video. The link will take you to a new window where you can play the video. Please watch the video carefully. After watching it, you can close the window and return to the survey window. You can then click submit to continue to the questions regarding the video. There will be multiple blocks of such videos and questions.

3. Main task

Please watch the video and then submit to proceed.

(Playing video)

Q1. Have you seen the advertisement before? Yes/no/maybe.

Q2. How familiar are you with the advertisement? 1 – 7

Q3. Does this advertisement depict a LGBTQIA+ (People who identify as lesbian, gay, bisexual, transgender, queer, intersex, asexual and other such sexual or gender identities) person/people? Yes/no.

If yes,

Q3.1. How explicit do you think the membership of this character/characters is to their social category (LGBTQIA+)?

Q3.2. How significant is the role played by the member/s of this social category (LGBTQIA+)? Lead: main character; Incidental: A background character, not noticeable.

Q3.3. How stereotypical is the role being performed by the member/s of this social category (LGBTQIA+) appear?

Q3.4. How would you describe the interaction, as shown in the ad, between - the member/s of this social category (LGBTQIA+) and other characters (who do not belong in this category)?

If no, proceed to Q4.

Q4. Does this advertisement depict a BAME (Black, Asian, and other Minority Ethnic groups not identifying as White British) person/people? Yes/no.

If yes, repeat the question Q3.1-Q3.4 but replace LGBTQIA+ with BAME.

If no, proceed to Q5.

Q5. Does this advertisement depict a Person/People with Disability? Yes/no.

If yes, repeat the question Q3.1-Q3.4 but replace LGBTQIA+ with Disability.

If no, proceed to Q6.

Q6. Does this advertisement depict a woman/woman? Yes/no.

If yes, repeat the question Q3.1-Q3.4 but replace LGBTQIA+ with woman/women.

If no, proceed to Q7.

Q7. Based on the advert, to what extent does the brand appear to be inclusive of people based on their ethnicity, gender, sexual orientation, or disability? 1 – 7.

Q8. How positive does the advertisement make you feel? 1 – 7.

Q9. How negative does the advertisement make you feel? 1 – 7.

Q10. How much do you agree/disagree with the following statement? 1 – 7.

The advertisement was different.

The advertisement was uncommon.

The advertisement was relevant to you.

The advertisement was meaningful to you.

Q11. Overall, how creative do you think the advertisement was? 1 – 7.

*Proceed the above process in Main Task section for 22 advertisements

4. Demographic information (age, ethnic, and gender)

Appendix S: Compound Remote Associate Test in Different Languages

*The CRAT questions were randomly selected by the experimenter (i.e., the author).

	Session 1					Session 2				
Chinese	条/原/压	升/问/费	源/将/信	歌/友/养	留/图/随	印/号/帐	官/案/针	播/导/得	全/代/移	航/产/运
Dutch	Vlokken/ketting/pet	val/meloen/lelie	vis/mijn/geel	worm/kast/legger	trommel/beleg/mes	goot/kool/bak	schommel/klap/rol	licht/dromen/maan	bed/zee/school	kop/boon/pause
English	pet/bottom/garden	spoon/cloth/card	oil/bar/tuna	home/sea/bed	fish/mine/rush	board/blade/back	type/ghost/screen	hammer/gear/hunter	horse/human/drag	safety/cushion/point
Japanese	員/芸/折	抗/刷/象	火/輪/数	禁/筋/相	刷/徴/価	力/睡/考	病/走/車	答/潔/良	想/的/本	来/縁/食
Romania	Principala`/Ma˘tur˘ator/Lumin˘	Companie/Vˆana˘toare/Nev˘az˘	Munte/Lance/Ascuţit	Raza˘/Floare/Rupt	Pretext/Acoperire/Egida	Vizual/Tragere/Semantic	Mandat/Oficiu/Timbru	Scris/Cusut/Ulei	Perete/Ultimul/Electronic	Perete/Miez/Cocos
Russian	свежая/английская/новости	неожиданно/человек/улица	умная/косы/свежая	прошлый/время/трудный	дедушка/очки/добрая	слон/дом/великан	вечерняя/бумага/стенная	цвет/заяц/сахар	много/чепуха/прямо	Вода/течь/высота
Spanish	caídas/choques/rayos	manchas/miedos/esmalte	manteles/pantallas/vidas	cabezas/hielos/olas	volea/cesto/pie	chimeneas/cristales/parabrisa	secretario/rector/presidente	carro/sierra/bomba	arriba/manga/bajo	tensión/calórico/ventilado

Appendix T: Engineering Creative Design – Data Collection, Marking, and Cleaning

Data Collection

1. Welcome

“You are now entering a creative design task for COVID-19 prevention. Please take a 2-minute break if you are tired of thinking. Please read the background information very carefully.

2. Background information: two ways that COVID-19 is transmitted

Direct Transmission: COVID-19 is transmitted by respiratory particles of someone infected with COVID-19 that reach healthy people directly. The particles are emitted through sneezing, coughing, or even talking. These droplets from an infected person are packed with millions of viral particles on whom fall in close range and infect whatever they land.

Fomite Transmission: COVID-19 is transmitted by touching an infective surface (e.g., keys, telephones, power buttons) and touching your face.”

3. Ideation Task (IT)

“Please develop as many novel and useful COVID-19 prevention solutions for a cosy restaurant (we provide 30 long textboxes below). Of course, we will strictly protect the copyright of all your ideas.

Here is an initial brainstorming, so do not criticise any ideas in your mind. Just write down as many ideas as you can. You have 5 minutes, but you do not have to wait until the time is complete before moving on.”

4. Product Design Task (PDT)

“Please design a COVID-19 prevention product for a cosy restaurant. The product must satisfy four requirements. We strictly protect the copyright of your designs. Requirements for the product: 1) The product aims to prevent people from being infected with COVID-19. 2) The product will be used in a cosy restaurant. 3) The product must be effective and novel. 4) We appreciate elegant products which can be used in other contexts. 5) There is no word limit, and you have 25 minutes to finish it. They are listed at the bottom of this page if you want to access the background information and your brainstorming ideas.

Please describe your product in four aspects:

1> What is your product?

2> What are the functions of this product?

3> What are the materials needed?

4> How can we generalize this product to other contexts?”

5. Product self-evaluation

“Please evaluate your product based on the indicators below. Then, drag the sliders to a point within 0 and 100. 0 means not at all; 100 means highly certain that my product satisfies.

1> The products accurately reflect conventional knowledge or techniques.

2> The product could do what it is supposed to do.

3> The product offers a fundamentally new perspective on possible solutions.

4> The solution is environmentally friendly.

5> The solution offers ideas for solving apparently unrelated problems.”

6. Notes for experimenters

1) We showed participants the background information throughout the task to ensure they had access to it whenever needed. 2) we randomised the self-evaluation questions.

Data marking and cleaning

Data marking and cleaning for IT replicate the process of FST. Please refer to Appendix D. We employed the Creative Solution Diagnosis Scale (CSDS) (Cropley & Kaufman, 2012) to measure participants’ product design creativity.

Appendix U: Short Scale of Creative Self

Below you will find several sentences used by people to describe themselves. Please decide to what extent each of these statements describes you (definitely no, somewhat no, neither yes nor no, somewhat yes, definitely yes).

1. I think I am a creative person.
2. My creativity is important for who I am.
3. I know I can efficiently solve even complicated problems.
4. I trust my creative abilities.
5. My imagination and ingenuity distinguished me from my friends.
6. Many times, I have proved that I can cope with difficult situations.
7. Being a creative person is important to me.
8. I am sure I can deal with problems requiring creative thinking.
9. I am good at proposing original solutions to problems.
10. Creativity is an important part of myself.
11. Ingenuity is a characteristic that is important to me.

Appendix V: Grit

Below you will find several sentences used by people to describe themselves. Please decide to what extent each of these statements describes you.

1. I often set a goal but later choose to pursue a different one.
2. New ideas and projects sometimes distract me from previous ones.
3. I become interested in new pursuits every few months.
4. My interests change from year to year.
5. I have been obsessed with a certain idea or project for a short time but later lost interests.
6. I have difficulty maintaining my focus on projects that take more than a few months to complete.
7. I have achieved a goal that took years of work.
8. I have overcome setbacks to conquer an important challenge.
9. I finish whatever I begin.
10. Setbacks do not discourage me.
11. I am a hard worker.
12. I am diligent.

Appendix W: Study 7 (Pilot) Script

Welcome

"A letter to participants: Dear UCL engineering students, thanks very much for your interest in our study. There are two goals of the study. The first one is to validate brand-new measurements of engineering creativity, so your considerate responses play an important role in the investigation surrounding human creativity. The second one is to help you understand your level of creative thinking and inspire you to improve it. Therefore, we would like to request you to concentrate when doing our tasks fully. Our study consists of three thinking tasks and one product design task, which will take you 40-60 minutes to complete. After completing these tasks, you can request creativity performance feedback (e.g., creativity scores, rankings compared to other participants, marking sheets from experts). You will also be entered into our random draw to receive a £30 Amazon Voucher (we award the voucher to 10% of all participants). Please feel free to quit at any time point. We still appreciate your interests and willingness to join us. We hope you will enjoy the tasks by employing your creative thinking, have fun!"

Participant information sheet and consent.

Main experiment

Participants saw AUT, CRAT, and FST randomly.

Status check. 1) please refer to the manipulation check in. 2) how much effort did you put in while doing three tasks (single choice) "I tried my best." "I put some efforts." "I was not focusing."

Participants did IT and PDT.

Status check. 1) & 2) please refer to the manipulation check in stage 2.2. 3) "While doing the design task, did you expect to receive a performance evaluation later?"

Participants saw creative-self questionnaire and grit questionnaire randomly.

Demographic information

Age, country of origin, gender, engineering educational stage, year of study, engineering domain.

Participants who wanted feedback from our study leave their emails.

Appendix X: Compound Remote Associate Test Questions in Study 7

pet / bottom / garden
spoon / cloth / card
oil / bar / tuna
home / sea / bed
fish / mine / rush
board / blade / back
type / ghost / screen
hammer / gear / hunter
horse / human / drag
safety / cushion / point

Appendix Y: Self-improvement Motivation Scale

Please rate to what degree the following statements describe your thoughts (1 = not at all, 7 = very much).

1. In my opinion, I should have down more for these tasks.
2. If I put in effort, I can perform better.
3. I would like to receive help and perform better.
4. Whatever rank I receive in creative potential is enough for me.
5. There is nothing I can do to improve my performance.
6. I should have done something differently in order to perform better.
7. I will do the same things in another creativity task.
8. In my opinion, I did enough for these tasks.
9. I would like to receive a good rank in creativity.

Appendix Z: Sensitivity to Social Comparison Scale

Please indicate whether the following 13 statements are true or false (1 = always false, 5 = always true).

1. It is my feeling that if everyone else in a group is behaving in a certain manner, this must be the proper way to behave.
2. I actively avoid wearing clothes that are not in style.
3. At parties I usually try to behave in a manner that makes me fit in.
4. When I am uncertain how to act in a social situation, I look to the behaviour of others for cues.
5. I try to pay attention to the reactions of others to my behaviour in order to avoid being out of place.
6. I find that I tend to pick up slang expressions from others and use them as part of my own vocabulary.
7. I tend to pay attention to what others are wearing.
8. The slightest look of disapproval in the eyes of a person with whom I am interacting is enough to make me change my approach.
9. It's important to me to fit in to the group I'm with.
10. My behaviour often depends on how I feel others wish me to behave.
11. If I am the least bit uncertain as to how to act in a social situation, I look to the behaviour of others for cues.
12. I usually keep up with clothing style changes by watching what others wear.

Appendix AA: Perceived Threat Scale

Please indicate how the following statements relate to you (1 = not at all, 7 = very much).

1. I feel nervous or distressed thinking about my rank compared to others.
2. I view other participants as competitors.
3. I feel like other participants may easily outperform me.

Appendix AB: Study 7 Script

1. Welcome

1.1. “A letter to participants: Dear all, Thanks very much for your interest in our study. The study’s main goal is to validate brand-new measurements of engineering creativity, so your considerate responses play an important role in the investigation surrounding human creativity. Therefore, we request you to concentrate when doing our tasks fully. Our study consists of four thinking tasks, which take 45 minutes to complete. After completing these tasks, you will receive £6.50 for your participation. Please feel free to quit at any time point. We still appreciate your interest and willingness to join our study. We hope you will enjoy the tasks by employing your creative thinking, have fun! Please enter your unique Prolific ID to begin:”

1.2. Participants information sheet and consent.

2. Pre-manipulation

2.1. Participants saw AUT, FST, and CRAT in random order. We replicated the process in the Pilot study. The only difference is that we did not ask people to drop their uses or objects into basic, alternative, and unusual categories.

2.2. Status check: please refer to the status check in Study 7 (Pilot).

2.3. Expectation check: “We administered this task to 1500 undergraduates and created a ranking pool for their performance. Please select which range you think you achieved compared to them: 1) Best performance range, 2) High-performance range, 3) Medium performance range, 4) Low-performance range, 5) Worst performance range.” (See Figure 19a for screenshot)

Figure 19a

Participants' Ranking	Performance Range
1	Best Performance
50	
100	
150	High Performance
200	
250	
300	
350	
400	
450	Medium Performance
500	
550	
600	
650	
700	
750	
800	
850	Low Performance
900	
950	
1000	
1050	
1100	
1150	
1200	
1250	
1300	Worst Performance
1350	
1400	
1450	
1500	

3. Manipulation

For downward and upward comparison groups:

"3.1. Thank you for sharing your experience. Once you click on the arrow below, our creativity scoring system will proceed with your task Responses and evaluate your performance by comparing it with the other 1500 undergraduates.

3.2. Please wait… A creativity scoring system is processing and evaluating your performance. In detail, the system evaluates all your responses' novelty, originality, usefulness, and accuracy. The outcome has been approved as an accurate predictor of your creative potential. The process takes up to 1 minute (we show them a waiting page to make it more believable).

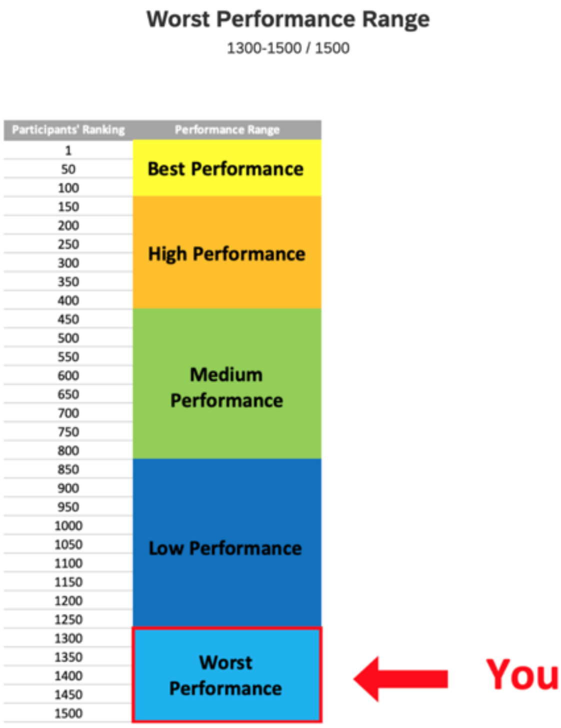
3.3. Compared to the performance of other 1500 university undergraduates, your performance in three creative thinking tasks is ranking in:

For the downward comparison group (see Figure 19b for screenshot):

Figure 19b

For the upward comparison group (See Figure 19c for screenshot):

Figure 19c



"

For the control group:

Thank you for sharing your experience. We will now proceed to the next stage of creative tasks.

4. post-manipulation

4.1. Participants did IT and PDT. We replicated the process in the Pilot study.

4.2. Status check: please refer to the status check in the Pilot study.

4.3. Participants did questionnaires for Short Scale of Creative Self, motivations for self-improvement, sensitivity to social comparison, and perceived threat randomly.

4.4. Demographic information (age, country of origin, gender, language, and specific engineering background) and debrief.

Reference

Aaker, J. L., Brumbaugh, A. M., & Grier, S. A. (2000). Nontarget markets and viewer distinctiveness: The impact of target marketing on advertising attitudes. *Journal of consumer psychology, 9*(3), 127-140.

Abbott, D. H. (2010). Constructing a creative self-efficacy inventory: A mixed methods inquiry. *Open Access Theses and Dissertations from the College of Education and Human Sciences*, 68.

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science, 53*(10), 1659-1674.

Abraham, A. (2018). *The neuroscience of creativity*. Cambridge University Press.

Abraham, A., & Windmann, S. (2007). Creative cognition: The diverse operations and the prospect of applying a cognitive neuroscience perspective. *Methods, 42*(1), 38-48.

Agnoli, S., Mastria, S., Zanon, M., & Corazza, G. E. (2022). Dopamine supports idea originality: The role of spontaneous eye blink rate on divergent thinking. *Psychological Research*, 1-11.

Agnoli, S., Zanon, M., Mastria, S., Avenanti, A., & Corazza, G. E. (2020). Predicting response originality through brain activity: An analysis of changes in EEG alpha power during the generation of alternative ideas [Electrophysiology 2530]. *Neuroimage, 207*. https://doi.org/http://dx.doi.org/10.1016/j.neuroimage.2019.116385

Akbari Chermahini, S., & Hommel, B. (2012). More creative through positive mood? Not everyone! *Frontiers in Human Neuroscience, 6*, 319.

Åkestam, N. (2017). *Understanding advertising stereotypes* Diss., Stockholm School of Economics, Ekonomiska institutionen. https://www …].

Åkestam, N., Rosengren, S., & Dahlen, M. (2017a). Advertising “like a girl”: Toward a better understanding of “femvertising” and its effects. *Psychology & Marketing, 34*(8), 795-806.

Åkestam, N., Rosengren, S., & Dahlen, M. (2017b). Think about it–can portrayals of homosexuality in advertising prime consumer-perceived social connectedness and empathy? *European Journal of Marketing*.

Akinola, M., Kapadia, C., Lu, J. G., & Mason, M. F. (2019). Incorporating physiology into creativity research and practice: The effects of bodily stress responses on creativity in organizations. *Academy of Management Perspectives, 33*(2), 163-184.

Al’tshuller, G. S. (1999). *The innovation algorithm: TRIZ, systematic innovation and technical creativity*. Technical innovation center, Inc.

Alexandria, V. *General Information Concerning Patent*. United States Patent And Trademark Office. Retrieved 15 Feb from https://www.uspto.gov/patents/basics/general-information-patents

Amabile, T. (2011). *Componential theory of creativity*. Harvard Business School Boston, MA.

Amabile, T. M. (1982a). Children's artistic creativity: Detrimental effects of competition in a field setting. *Personality and Social Psychology Bulletin, 8*(3), 573-578.

Amabile, T. M. (1982b). Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology, 43*(5), 997.

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology, 45*(2), 357.

Amabile, T. M. (1993). Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human resource management review, 3*(3), 185-201.

Amabile, T. M. (1996). Creativity and innovation in organizations.

Amabile, T. M. (2018). *Creativity in context: Update to the social psychology of creativity*. Routledge.

Amabile, T. M., Goldfarb, P., & Brackfleld, S. C. (1990). Social influences on creativity: Evaluation, coaction, and surveillance. *Creativity Research Journal, 3*(1), 6-21.

Amabile, T. M., & Mueller, J. S. (2008). Studying creativity, its processes, and its antecedents: An exploration of the componential theory of creativity. *Handbook of organizational creativity, 3162*.

Amoussou, G.-A., Porter, M., & Steinberg, S. J. (2011). Assessing creativity practices in design. 2011 Frontiers in Education Conference (FIE),

An, D., Song, Y., & Carr, M. (2016). A comparison of two models of creativity: Divergent thinking and creative expert performance. *Personality and Individual Differences, 90*, 78-84.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological review, 98*(3), 409.

Anderson, N., Potočnik, K., & Zhou, J. (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of management, 40*(5), 1297-1333.

Andreasen, N. C. (2005). *The creating brain: The neuroscience of genius*. Dana Press.

Andrews, E. G. (1930). The Development of Imagination in the Pre-School Child: University of Iowa Studies in Character. *Iowa City, Iowa: University of Iowa*.

Andrews, F. M. (1967). Creative ability, the laboratory environment, and scientific performance. *IEEE Transactions on Engineering Management*(2), 76-83.

Andriopoulos, C. (2001). Determinants of organisational creativity: a literature review. *Management Decision*.

Ang, S. H., Lee, Y. H., & Leong, S. M. (2007). The ad creativity cube: Conceptualization and initial validation. *Journal of the Academy of Marketing Science, 35*(2), 220-232.

Ang, S. H., Leong, S. M., Lee, Y. H., & Lou, S. L. (2014). Necessary but not sufficient: Beyond novelty in advertising creativity. *Journal of Marketing Communications, 20*(3), 214-230.

Ang, S. H., & Low, S. Y. (2000). Exploring the dimensions of ad creativity. *Psychology & Marketing, 17*(10), 835-854.

Astor, P. J., Adam, M. T., Jähnig, C., & Seifert, S. (2013). The joy of winning and the frustration of losing: A psychophysiological analysis of emotions in first-price sealed-bid auctions. *Journal of Neuroscience, Psychology, and Economics, 6*(1), 14.

Aziz-Zadeh, L., Liew, S.-L., & Dandekar, F. (2013). Exploring the neural correlates of visual creativity. *Social cognitive and affective neuroscience, 8*(4), 475-480.

Baack, D. W., Wilson, R. T., & Till, B. D. (2008). Creativity and memory effects: Recall, recognition, and an exploration of nontraditional media. *Journal of Advertising, 37*(4), 85-94.

Baer, J. (1997). Gender differences in the effects of anticipated evaluation on creativity. *Creativity Research Journal, 10*(1), 25-31.

Baer, J. (1998a). The case for domain specificity of creativity. *Creativity Research Journal, 11*(2), 173-177.

Baer, J. (1998b). Gender differences in the effects of extrinsic motivation on creativity [Cognitive & Perceptual Development 2820]. *The Journal of Creative Behavior, 32*(1), 18-37. https://doi.org/http://dx.doi.org/10.1002/j.2162-6057.1998.tb00804.x

Baer, J. (2014). *Creativity and divergent thinking: A task-specific approach*. Psychology Press.

Baer, J., Kaufman, J. C., & Gentile, C. A. (2004). Extension of the consensual assessment technique to nonparallel creative products. *Creativity Research Journal, 16*(1), 113-117.

Baggen, Y., Kampen, J. K., Naia, A., Biemans, H. J., Lans, T., & Mulder, M. (2018). Development and application of the opportunity identification competence assessment test (OICAT) in higher education. *Innovations in Education and Teaching International, 55*(6), 735-745.

Bai, X., Ramos, M. R., & Fiske, S. T. (2020). As diversity increases, people paradoxically perceive social groups as more similar. *Proceedings of the National Academy of Sciences, 117*(23), 12741-12749.

Baker, M. J. (2014). *Marketing strategy and management*. Macmillan International Higher Education.

Balietti, S., Goldstone, R. L., & Helbing, D. (2016). Peer review and competition in the Art Exhibition Game [Cognitive Processes 2340]. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 113*(30), 8414-8419. https://doi.org/http://dx.doi.org/10.1073/pnas.1603723113

Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of personality and social psychology, 60*(6), 941.

Barbot, B., & Lubart, T. (2012). Creative thinking in music: Its nature and assessment through musical exploratory behaviors. *Psychology of Aesthetics, Creativity, and the Arts, 6*(3), 231.

Barron, F. (1955). The disposition toward originality. *The Journal of Abnormal and Social Psychology, 51*(3), 478.

Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual review of psychology, 32*(1), 439-476.

Batey, M., Chamorro-Premuzic, T., & Furnham, A. (2009). Intelligence and personality as predictors of divergent thinking: The role of general, fluid and crystallised intelligence. *Thinking Skills and Creativity, 4*(1), 60-69.

Batey, M., & Furnham, A. (2006). Creativity, intelligence, and personality: A critical review of the scattered literature. *Genetic, social, and general psychology monographs, 132*(4), 355-429.

Beaty, R. E., Benedek, M., Wilkins, R. W., Jauk, E., Fink, A., Silvia, P. J., Hodges, D. A., Koschutnig, K., & Neubauer, A. C. (2014). Creativity and the default network: A functional connectivity analysis of the creative brain at rest [Neuropsychology & Neurology 2520]. *Neuropsychologia, 64*, 92-98. https://doi.org/https://dx.doi.org/10.1016/j.neuropsychologia.2014.09.019

Beaty, R. E., Chen, Q., Christensen, A. P., Kenett, Y. N., Silvia, P. J., Benedek, M., & Schacter, D. L. (2020). Default network contributions to episodic and semantic processing during divergent creative thinking: A representational similarity analysis. *Neuroimage, 209*, 116499.

Beaty, R. E., Nusbaum, E. C., & Silvia, P. J. (2014). Does insight problem solving predict real-world creativity? *Psychology of Aesthetics, Creativity, and the Arts, 8*(3), 287.

Beaty, R. E., Seli, P., & Schacter, D. L. (2019). Network neuroscience of creative cognition: mapping cognitive mechanisms and individual differences in the creative brain. *Current opinion in behavioral sciences, 27*, 22-30.

Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts, 6*(4), 309.

Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., & Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Memory & cognition, 42*(7), 1186-1197.

Behm-Morawitz, E., & Mastro, D. (2009). The effects of the sexualization of female video game characters on gender stereotyping and female self-concept. *Sex roles, 61*(11-12), 808-823.

Beketayev, K., & Runco, M. A. (2016). Scoring divergent thinking tests by computer with a semantics-based algorithm. *Europe's journal of psychology, 12*(2), 210.

Belski, I. (2017). Engineering creativity-how to measure it? 28th Annual Conference of the Australasian Association for Engineering Education (AAEE 2017),

Benedek, M., Bergner, S., Könen, T., Fink, A., & Neubauer, A. C. (2011). EEG alpha synchronization is related to top-down processing in convergent and divergent thinking. *Neuropsychologia, 49*(12), 3505-3511.

Benedek, M., & Fink, A. (2019). Toward a neurocognitive framework of creative cognition: The role of memory, attention, and cognitive control. *Current Opinion in Behavioral Sciences, 27*, 116-122.

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence, 46*, 73-83.

Benedek, M., Kenett, Y. N., Umdasch, K., Anaki, D., Faust, M., & Neubauer, A. C. (2017, 2017/04/03). How semantic memory structure and intelligence contribute to creative thought: a network science approach. *Thinking & Reasoning, 23*(2), 158-183. https://doi.org/10.1080/13546783.2016.1278034

Benedek, M., Könen, T., & Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts, 6*(3), 273.

Benedek, M., & Neubauer, A. C. (2013). Revisiting Mednick's model on creativity-related differences in associative hierarchies. Evidence for a common path to uncommon thought. *The Journal of Creative Behavior, 47*(4), 273-289.

Bernstein, D. (1974). *Creative advertising: For this you went to Oxford?: A personal textbook of advertising*. Longman Publishing Group.

Besemer, S. P., & Treffinger, D. J. (1981). Analysis of creative products: review and synthesis. *The Journal of Creative Behavior*.

Bilder, R. M., & Knudsen, K. S. (2014). Creative cognition and systems biology on the edge of chaos [Psychological Disorders 3210]. *Frontiers in Psychology, 5*. https://doi.org/https://dx.doi.org/10.3389/fpsyg.2014.01104

Bissell, K., & Rask, A. (2010). Real women on real beauty: Self-discrepancy, internalisation of the thin ideal, and perceptions of attractiveness and thinness in Dove's Campaign for Real Beauty. *International Journal of Advertising, 29*(4), 643-668.

Bittner, J. V., & Heidemeier, H. (2013). Competitive mindsets, creativity, and the role of regulatory focus. *Thinking Skills and Creativity, 9*, 59-68.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*(6), 1142.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*(5), 828.

Blowfield, M. (2005). Corporate social responsibility: reinventing the meaning of development? *International Affairs, 81*(3), 515-524.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Routledge.

Boden, M. A. (2009). Conceptual spaces. In *Milieus of creativity* (pp. 235-243). Springer.

Boden, M. A. (2010). *Creativity and art: Three roads to surprise*. Oxford University Press.

Boeree, C. G. (2006). Carl Jung. *Archived from the original on, 6*.

Boot, N., Baas, M., van Gaal, S., Cools, R., & De Dreu, C. K. (2017). Creative cognition and dopaminergic modulation of fronto-striatal networks: Integrative review and research agenda. *Neuroscience & Biobehavioral Reviews, 78*, 13-23.

Botella, M., Glaveanu, V., Zenasni, F., Storme, M., Myszkowski, N., Wolff, M., & Lubart, T. (2013). How artists create: Creative process and multivariate factors. *Learning and Individual Differences, 26*, 161-170.

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior research methods, instruments, & computers, 35*(4), 634-639.

Bowen, L., & Schmid, J. (1997). Minority presence and portrayal in mainstream magazine advertising: An update. *Journalism & Mass Communication Quarterly, 74*(1), 134-146.

Brehm, J. W. (1966). A theory of psychological reactance.

Brophy, D. R. (1998). Understanding, measuring, and enhancing individual creative problem-solving efforts. *Creativity Research Journal, 11*(2), 123-150.

Brougher, S. J., & Rantanen, E. M. (2009). Creativity and design: Creativity's new definition and its relationship to design. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,

Brumbaugh, A. M., & Grier, S. A. (2006). Insights from a" failed" experiment: Directions for pluralistic, multiethnic advertising research. *Journal of Advertising, 35*(3), 35-46.

Bruner, J. S. (1962). The conditions of creativity. Contemporary Approaches to Creative Thinking, 1958, University of Colorado, CO, US; This paper was presented at the aforementioned symposium.,

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes, 102*(1), 3-21.

Byron, K., Khazanchi, S., & Nazarian, D. (2010). The relationship between stressors and creativity: a meta-analysis examining competing theoretical models. *Journal of Applied Psychology, 95*(1), 201.

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological review, 67*(6), 380.

Carpenter, W. A. (2016). *Engineering creativity: Toward an understanding of the relationship between perceptions and performance in engineering design* University of Akron].

Carson, D. K. (1999). The importance of creativity in family therapy: A preliminary consideration. *The family journal, 7*(4), 326-334.

Cattell, R. B., & Butcher, H. J. (1968). *The prediction of achievement and creativity*. Ardent Media.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81*(1), 1-8.

Chassell, L. M. (1916). Tests for originality. *Journal of educational psychology, 7*(6), 317.

Chen, D., & Vernadat, F. (2004). Standards on enterprise integration and engineering—state of the art. *International Journal of Computer Integrated Manufacturing, 17*(3), 235-253.

Cheng, C.-Y., & Hong, Y.-Y. (2017). Kiasu and creativity in Singapore: An empirical test of the situated dynamics framework [Culture & Ethnology 2930]. *Management and Organization Review, 13*(4), 871-894. https://doi.org/http://dx.doi.org/10.1017/mor.2017.41

Cheng, C.-Y., Sanchez-Burks, J., & Lee, F. (2008). Connecting the dots within: Creative performance and identity integration. *Psychological science, 19*(11), 1178-1184.

Chermahini, S. A., Hickendorff, M., & Hommel, B. (2012). Development and validity of a Dutch version of the Remote Associates Task: An item-response theory approach. *Thinking Skills and Creativity, 7*(3), 177-186.

Chermahini, S. A., & Hommel, B. (2010). The (b) link between creativity and dopamine: spontaneous eye blink rates predict and dissociate divergent and convergent thinking. *Cognition, 115*(3), 458-465.

Cho, Y.-J., Fu, P.-W., & Wu, C.-C. (2017). Popular research topics in marketing journals, 1995–2014. *Journal of Interactive Marketing, 40*, 52-72.

Clapham, M. M. (1997). Ideational skills training: A key element in creativity training programs. *Creativity Research Journal, 10*(1), 33-44.

Clapham, M. M., & Schuster, D. H. (1992). Can engineering students be trained to think more creatively? *The Journal of Creative Behavior*.

Clark, R. A., & Goldsmith, R. E. (2006). Interpersonal influence and consumer innovativeness. *International Journal of Consumer Studies, 30*(1), 34-43.

Clark, S. G., & Wallace, R. L. (2015). Integration and interdisciplinarity: concepts, frameworks, and education. *Policy Sciences, 48*(2), 233-255.

Clement, D. E., & Schiereck, J. J. (1973). Sex composition and group performance in a visual signal detection task. *Memory & cognition, 1*(3), 251-255.

Collard, P., & Looney, J. (2014). Nurturing creativity in education. *European Journal of Education, 49*(3), 348-364.

Colvin, S. S., & Meyer, I. (1906). Imaginative elements in the written work of school children. *The Pedagogical Seminary, 13*(1), 84-93.

Conti, R., Collins, M. A., & Picariello, M. L. (2001). The impact of competition on intrinsic motivation and creativity: considering gender, gender segregation and gender role orientation. *Personality and Individual Differences, 31*(8), 1273-1289.

Cools, R., & D'Esposito, M. (2011). Inverted-U–shaped dopamine actions on human working memory and cognitive control. *Biological psychiatry, 69*(12), e113-e125.

Corazza, G. E., & Lubart, T. (2021). Intelligence and creativity: Mapping constructs on the space-time continuum. *Journal of Intelligence, 9*(1), 1.

Corbett, A. C. (2007). Learning asymmetries and the discovery of entrepreneurial opportunities. *Journal of business venturing, 22*(1), 97-118.

Corrigan, J. R., Drichoutis, A. C., Lusk, J. L., Nayga Jr, R. M., & Rousu, M. C. (2012). Repeated rounds with price feedback in experimental auction valuation: An adversarial collaboration. *American Journal of Agricultural Economics, 94*(1), 97-115.

Cortes, R. A., Weinberger, A. B., Daker, R. J., & Green, A. E. (2019). Re-examining prominent measures of divergent and convergent creativity. *Current Opinion in Behavioral Sciences, 27*, 90-93.

Craft, A. (2003). The limits to creativity in education: Dilemmas for the educator. *British journal of educational studies, 51*(2), 113-127.

Cravens, D. W., & Piercy, N. (2006). *Strategic marketing* (Vol. 6). McGraw-Hill Columbus.

Crockenberg, S. B. (1972). Creativity tests: A boon or boondoggle for education? *Review of Educational Research, 42*(1), 27-45.

Crocker, J., Thompson, L. L., McGraw, K. M., & Ingerman, C. (1987). Downward comparison, prejudice, and evaluations of others: Effects of self-esteem and threat. *Journal of personality and social psychology, 52*(5), 907.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281.

Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal, 18*(3), 391-404.

Cropley, A. (2012). Creativity and education: An Australian perspective. *IJCPS-International Journal of Creativity and Problem Solving, 22*(1), 9.

Cropley, A. J. (2001). *Creativity in education & learning: A guide for teachers and educators*. Psychology Press.

Cropley, D., & Cropley, A. (2005). Engineering creativity: A systems concept of functional creativity. In *Creativity across domains* (pp. 187-204). Psychology Press.

Cropley, D., & Cropley, A. (2010). Recognizing and fostering creativity in technological design education. *International Journal of Technology and Design Education, 20*(3), 345-358.

Cropley, D., & Cropley, A. (2012). A psychological taxonomy of organizational innovation: Resolving the paradoxes. *Creativity Research Journal, 24*(1), 29-40.

Cropley, D. H. (2015a). Chapter 3 - Phases: Creativity and the Design Process. In D. H. Cropley (Ed.), *Creativity in Engineering* (pp. 35-61). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-800225-4.00003-3

Cropley, D. H. (2015b). Chapter 4 - Product: The Creativity of Things. In D. H. Cropley (Ed.), *Creativity in Engineering* (pp. 63-85). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-800225-4.00004-5

Cropley, D. H. (2015c). Chapter 5 - Process: Generating Creative Ideas. In D. H. Cropley (Ed.), *Creativity in Engineering* (pp. 87-129). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-800225-4.00005-7

Cropley, D. H. (2015d). *Creativity in engineering : novel solutions to complex problems / David H. Cropley*. London, England : Academic Press.

Cropley, D. H. (2015). Promoting creativity and innovation in engineering education. *Psychology of Aesthetics, Creativity, and the Arts, 9*(2), 161.

Cropley, D. H. (2016). Creativity in engineering. In *Multidisciplinary contributions to the science of creative thinking* (pp. 155-173). Springer.

Cropley, D. H., & Kaufman, J. C. (2012). Measuring functional creativity: Non-expert raters and the Creative Solution Diagnosis Scale. *The Journal of Creative Behavior, 46*(2), 119-137.

Csikszentmihalyi, M. (2014). Society, culture, and person: A systems view of creativity. In *The systems model of creativity* (pp. 47-61). Springer.

Csikszentmihalyi, M. (2015). *The systems model of creativity: The collected works of Mihaly Csikszentmihalyi*. Springer.

Dahlén, M., Sjödin, H., Thorbjørnsen, H., Hansen, H., Linander, J., & Thunell, C. (2013). "What will 'they'think?": Marketing leakage to undesired audiences and the third-person effect. *European Journal of Marketing*.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8-15.

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*(12), 1615-1628.

Davis, G. A., & Belcher, T. L. (1971). How Shall Creativity Be Measured? Torrance Tests, RAT, Alpha Biographical, and IQ.

Davis, S. F. (2008). *Handbook of research methods in experimental psychology*. John Wiley & Sons.

De Dreu, C. K., Nijstad, B. A., Baas, M., Wolsink, I., & Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin, 38*(5), 656-669.

De Vet, A. J., & De Dreu, C. K. (2007). The influence of articulation, self-monitoring ability, and sensitivity to others on creativity. *European Journal of Social Psychology, 37*(4), 747-760.

Dearborn, G. V. (1898). A study of imaginations. *The American Journal of Psychology, 9*(2), 183-190.

Der-Karabetian, A., & Balian, N. (1992). Ingroup, outgroup, and global-human identities of Turkish-Armenians. *The Journal of social psychology, 132*(4), 497-504.

DeYoung, C. G. (2020). Intelligence and personality.

Diamond, J. (2014). *The third chimpanzee: on the evolution and future of the human animal*. Simon and Schuster.

Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychonomic bulletin & review, 11*(6), 1011-1026.

Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin, 136*(5), 822.

Direito, I., Chance, S., & Malik, M. (2021). The study of grit in engineering education research: a systematic literature review. *European Journal of Engineering Education, 46*(2), 161-185.

Drevdahl, J. E. (1956). Factors of importance for creativity. *Journal of Clinical Psychology*.

Dubitzky, W., Kötter, T., Schmidt, O., & Berthold, M. R. (2012). Towards creative information exploration based on Koestler's concept of bisociation. In *Bisociative Knowledge Discovery* (pp. 11-32). Springer.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of personality assessment, 91*(2), 166-174.

Duke, L. (2001). Like an idea, only better: How do advertising educators and practitioners define and use the creative concept? *Journal of Advertising Education, 5*(1), 10-22.

[Record #340 is using a reference type undefined in this output style.]

Eisenberg, J., & Thompson, W. F. (2011). The effects of competition on improvisers' motivation, stress, and creative performance. *Creativity Research Journal, 23*(2), 129-136.

Eisenberger, R., Armeli, S., & Pretz, J. (1998). Can the promise of reward increase creativity? *Journal of personality and social psychology, 74*(3), 704.

Eisenberger, R., & Rhoades, L. (2001). Incremental effects of reward on creativity. *Journal of personality and social psychology, 81*(4), 728.

Eisend, M. (2015). Have we progressed marketing knowledge? A meta-meta-analysis of effect sizes in marketing research. *Journal of Marketing, 79*(3), 23-40.

Eisend, M., & Hermann, E. (2019). Consumer responses to homosexual imagery in advertising: A meta-analysis. *Journal of Advertising, 48*(4), 380-400.

Eisend, M., & Langner, T. (2010). Immediate and delayed advertising effects of celebrity endorsers' attractiveness and expertise. *International Journal of Advertising, 29*(4), 527-546.

Eisend, M., Plagemann, J., & Sollwedel, J. (2014). Gender roles and humor in advertising: The occurrence of stereotyping in humorous and nonhumorous advertising and its consequences for advertising effectiveness. *Journal of Advertising, 43*(3), 256-273.

Ellemers, N., Kortekaas, P., & Ouwerkerk, J. W. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European journal of social psychology, 29*(2-3), 371-389.

Elliott, R., & Elliott, C. (2005). Idealized images of the male body in advertising: a reader-response exploration. *Journal of Marketing Communications, 11*(1), 3-19.

Epstein, R. A., Becker, G. S., Coase, R. H., Miller, M. H., & Posner, R. A. (1997). The roundtable discussion. *The University of Chicago Law Review, 64*(4), 1132-1165.

Erat, S., & Gneezy, U. (2016). Incentives for creativity [Motivation & Emotion 2360]. *Experimental Economics, 19*(2), 269-280. https://doi.org/http://dx.doi.org/10.1007/s10683-015-9440-5

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. *The Cambridge handbook of expertise and expert performance, 38*(685-705), 2-2.3.

Eveleens, C. (2010). Innovation management; a literature review of innovation process models and their implications. *Science, 800*(2010), 900-916.

Farid, F., El-Sharkawy, A. R., & Austin, L. K. (1993). Managing for creativity and innovation in A/E/C organizations. *Journal of Management in Engineering, 9*(4), 399-409.

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and social psychology review, 2*(4), 290-309.

Feist, G. J. (2005). Domain-specific creativity in the physical sciences. In *Creativity Across Domains* (pp. 141-156). Psychology Press.

Feist, G. J. (2010). The function of personality in creativity: The nature and nurture of the creative personality.

Felder, R. M. (1987). On creating creative engineers. *Engineering education, 77*(4), 222-227.

Festinger, L. (1954). A theory of social comparison processes. *Human relations, 7*(2), 117-140.

Fink, A., Grabner, R. H., Benedek, M., Reishofer, G., Hauswirth, V., Fally, M., Neuper, C., Ebner, F., & Neubauer, A. C. (2009). The creative brain: Investigation of brain activity during creative problem solving by means of EEG and fMRI [Physiological Psychology & Neuroscience 2500]. *Human Brain Mapping, 30*(3), 734-748. https://doi.org/https://dx.doi.org/10.1002/hbm.20538

Fink, A., Schwab, D., & Papousek, I. (2011). Sensitivity of EEG upper alpha activity to cognitive and affective creativity interventions. *International Journal of Psychophysiology, 82*(3), 233-239.

Finke, R. A., & Slayton, K. (1988). Explorations of creative visual synthesis in mental imagery. *Memory & cognition, 16*(3), 252-257.

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). Creative cognition: Theory, research, and applications.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. *The handbook of social psychology, 2*(4), 357-411.

Fiske, S. T., Harris, L. T., & Cuddy, A. J. (2004). Why ordinary people torture enemy prisoners.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: ten years later.

Fiske, S. T., & Neuberg, S. L. (1989). Category-based and individuating processes as a function of information and motivation: Evidence from our laboratory. In *Stereotyping and prejudice* (pp. 83-103). Springer.

Forgeard, M. J., & Benson, L. (2019). Extracurricular involvement and psychological adjustment in the transition from adolescence to emerging adulthood: The role of mastery and creative self-efficacy. *Applied Developmental Science, 23*(1), 41-58.

Fredrickson, B. L., & Roberts, T. A. (1997). Objectification theory: Toward understanding women's lived experiences and mental health risks. *Psychology of women quarterly, 21*(2), 173-206.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological science, 17*(2), 172-179.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534.

Furnham, A., Crump, J., Batey, M., & Chamorro-Premuzic, T. (2009). Personality and ability predictors of the "Consequences" Test of divergent thinking in a large non-student sample. *Personality and Individual Differences, 46*(4), 536-540.

Furnham, A., & Paltzer, S. (2010). The portrayal of men and women in television advertisements: An updated review of 30 studies published since 2000. *Scandinavian journal of Psychology, 51*(3), 216-236.

Gabora, L. (2018). The Neural Basis and Evolution of Divergent and Convergent Thought. In O. Vartanian & R. E. Jung (Eds.), *The Cambridge handbook of the neuroscience of creativity* (pp. 58-70). Cambridge University Press. https://doi.org/DOI: 10.1017/9781316556238.005

Gaither, S. E., Remedios, J. D., Sanchez, D. T., & Sommers, S. R. (2015). Thinking outside the box: Multiple identity mind-sets affect creative problem solving. *Social Psychological and Personality Science, 6*(5), 596-603.

George, T., & Wiley, J. (2020). Need something different? Here's what's been done: Effects of examples and task instructions on creative idea generation. *Memory & Cognition, 48*(2), 226-243.

Getzels, J. W., & Jackson, P. W. (1961). Family environment and cognitive style: A study of the sources of highly intelligent and of highly creative adolescents. *American Sociological Review*, 351-359.

Getzels, J. W., & Jackson, P. W. (1962). Creativity and intelligence: Explorations with gifted students.

Gilhooly, K. J., Fioratou, E., Anthony, S. H., & Wynn, V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology, 98*(4), 611-625.

Gilhooly, K. J., Georgiou, G. J., Garrison, J., Reston, J. D., & Sirota, M. (2012). Don't wait to incubate: Immediate versus delayed incubation in divergent thinking. *Memory & Cognition, 40*(6), 966-975.

Glăveanu, V. P., Gillespie, A., & Karwowski, M. (2019). Are people working together inclined towards practicality? A process analysis of creative ideation in individuals and dyads. *Psychology of Aesthetics, Creativity, and the Arts, 13*(4), 388.

Gocłowska, M. A., Baas, M., Crisp, R. J., & De Dreu, C. K. (2014). Whether social schema violations help or hurt creativity depends on need for structure. *Personality and Social Psychology Bulletin, 40*(8), 959-971.

Gocłowska, M. A., & Crisp, R. J. (2013). On counter-stereotypes and creative cognition: When interventions for reducing prejudice can boost divergent thinking. *Thinking skills and creativity, 8*, 72-79.

Gocłowska, M. A., & Crisp, R. J. (2014). How dual-identity processes foster creativity. *Review of general psychology, 18*(3), 216-236.

Gocłowska, M. A., Crisp, R. J., & Labuschagne, K. (2013). Can counter-stereotypes boost flexible thinking? *Group Processes & Intergroup Relations, 16*(2), 217-231.

Gough, H. G. (1961, Oct. 13-17). Techniques for identifying the creative research scientist. In Proceedings, Conference on "The creative person", Lake Tahoe, California.

Grau, S. L., & Zotos, Y. C. (2016). Gender stereotypes in advertising: a review of current research. *International Journal of Advertising, 35*(5), 761-770.

Grier, S. A., & Brumbaugh, A. M. (1999). Noticing cultural differences: Ad meanings created by target and non-target markets. *Journal of Advertising, 28*(1), 79-93.

Grier, S. A., Brumbaugh, A. M., & Thornton, C. G. (2006). Crossover dreams: Consumer responses to ethnic-oriented products. *Journal of Marketing, 70*(2), 35-51.

Groborz, M., & Necka, E. (2003). Creativity and cognitive control: Explorations of generation and evaluation skills. *Creativity Research Journal, 15*(2-3), 183-197.

Gruber, H. E. (1988). The evolving systems approach to creative work. *Creativity Research Journal, 1*(1), 27-51.

Guilford, J. (1981). Higher-order structure-of-intellect abilities. *Multivariate Behavioral Research, 16*(4), 411-435.

Guilford, J. P. (1950). Creativity. *American psychologist, 5*(9), 444-454. https://doi.org/10.1037/h0063487

Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior, 1*(1), 3-14.

Guilford, J. P. (1967). The nature of human intelligence.

Gulas, C. S., & McKeage, K. (2000). Extending social comparison: An examination of the unintended consequences of idealized advertising imagery. *Journal of Advertising, 29*(2), 17-28.

Haberland, G. S., & Dacin, P. A. (1992). The development of a measure to assess viewers' judgments of the creativity of an advertisement: A preliminary study. *ACR North American Advances*.

Hadis, B. F. (2005). Why are they better students when they come back? Determinants of academic focusing gains in the study abroad experience. *Frontiers: The Interdisciplinary Journal of Study Abroad, 11*, 57-70.

Haertel, T., Terkowsky, C., & Jahnke, I. (2012). Where have all the inventors gone?: Is there a lack of spirit of research in engineering education curricula? 2012 15th International Conference on Interactive Collaborative Learning (ICL),

Hakmiller, K. L. (1966). Threat as a determinant of downward comparison. *Journal of Experimental Social Psychology, 1*, 32-39.

Halliwell, E., & Dittmar, H. (2004). Does size matter? The impact of model's body size on women's body-focused anxiety and advertising effectiveness. *Journal of Social and Clinical Psychology, 23*(1), 104-122.

Hammock, T., & Brehm, J. W. (1966). The attractiveness of choice alternatives when freedom to choose is eliminated by a social agent. *Journal of personality*.

Harris, D. (1960a). The development and validation of a test of creativity in engineering. *Journal of Applied Psychology, 44*(4), 254.

Harris, D. (1960b). Test of Creativity in Engineering [Organizational, Occupational, and Career Development 7000]. https://doi.org/http://dx.doi.org/10.1037/t04279-000

Harris, L. T. (2021). Leveraging cultural narratives to promote trait inferences rather than stereotype activation during person perception. *Social and Personality Psychology Compass, 15*(6), e12598.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological science, 17*(10), 847-853.

Harris, L. T., & Fiske, S. T. (2007). Social groups that elicit disgust are differentially processed in mPFC. *Social cognitive and affective neuroscience, 2*(1), 45-51.

Harris, L. T., & Fiske, S. T. (2009). Social neuroscience evidence for dehumanised perception. *European review of social psychology, 20*(1), 192-231.

Harris, L. T., & Harris, L. T. (2017). *Invisible mind: Flexible social cognition and dehumanization*. MIT Press.

Hatchuel, A., & Weil, B. (2003). A new approach of innovative Design: an introduction to CK theory. DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm,

Hawkins, D. I., & Mothersbaugh, D. L. (2010). *Consumer behavior: Building marketing strategy*. Boston: McGraw-Hill Irwin.

Heilman, K. M., Nadeau, S. E., & Beversdorf, D. O. (2003). Creative innovation: possible brain mechanisms. *Neurocase, 9*(5), 369-379.

Helgeson, V. S., & Mickelson, K. D. (1995). Motives for social comparison. *Personality and Social Psychology Bulletin, 21*(11), 1200-1209.

Henderson-King, D., Henderson-King, E., & Hoffmann, L. (2001). Media images and women's self-evaluations: Social context and importance of attractiveness as moderators. *Personality and Social Psychology Bulletin, 27*(11), 1407-1416.

Hennessey, B. A. (1989). The effect of extrinsic constraints on children's creativity while using a computer. *Creativity Research Journal, 2*(3), 151-168.

Hennessey, B. A. (1994). The consensual assessment technique: An examination of the relationship between ratings of product and process creativity. *Creativity Research Journal, 7*(2), 193-208.

Hennessey, B. A. (2003). The social psychology of creativity. *Scandinavian journal of educational research, 47*(3), 253-271.

Henriksen, D., Mishra, P., & Mehta, R. (2015). Novel, effective, whole: Toward a NEW framework for evaluations of creative products. *Journal of Technology and Teacher Education, 23*(3), 455-478.

Herbst, K. C., Hannah, S. T., & Allan, D. (2013). Advertisement disclaimer speed and corporate social responsibility:"Costs" to consumer comprehension and effects on brand trust and purchase intention. *Journal of business ethics, 117*(2), 297-311.

Hernández-Torrano, D., & Ibrayeva, L. (2020). Creativity and education: A bibliometric mapping of the research literature (1975–2019). *Thinking Skills and Creativity, 35*, 100625.

Hickey, M. (2001). An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions. *Journal of Research in Music Education, 49*(3), 234-244.

Hirshfield, L. J., & Koretsky, M. D. (2020). Cultivating creative thinking in engineering student teams: Can a computer-mediated virtual laboratory help? *Journal of Computer Assisted Learning*.

Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of personality assessment, 45*(5), 450-464.

Hong, Y.-y., Morris, M. W., Chiu, C.-y., & Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist, 55*(7), 709.

Howard, T., Culley, S. J., & Dekoninck, E. (2007). Creativity in the engineering design process. DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 28.-31.07. 2007,

Howard, T. J., Culley, S. J., & Dekoninck, E. (2008). Describing the creative design process by the integration of engineering design and cognitive psychology literature. *Design studies, 29*(2), 160-180.

Isaksen, S. G., & Parnes, S. J. (1985). Curriculum planning for creative thinking and problem solving. *The Journal of Creative Behavior*.

Jackson, P. W., & Messick, S. (1965). The person, the product, and the response: conceptual problems in the assessment of creativity. *Journal of personality*.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017, 2017/12/06). When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology, 17*(1), 162. https://doi.org/10.1186/s12874-017-0442-1

Jauk, E., Benedek, M., & Neubauer, A. C. (2012). Tackling creativity at its roots: Evidence for different patterns of EEG alpha activity related to convergent and divergent modes of task processing. *International Journal of Psychophysiology, 84*(2), 219-225.

Jaussi, K. S., Randel, A. E., & Dionne, S. D. (2007). I am, I think I can, and I do: The role of personal identity, self-efficacy, and cross-application of experiences in creativity at work. *Creativity Research Journal, 19*(2-3), 247-258.

Johnson, C. S., & Stapel, D. A. (2007). No pain, no gain: The conditions under which upward comparisons lead to better performance. *Journal of personality and social psychology, 92*(6), 1051.

Johnson, G. D., & Grier, S. A. (2012). What about the intended consequences? *Journal of Advertising, 41*(3), 91-106.

Jonides, J., & Smith, E. E. (1997). The architecture of working. *Cognitive neuroscience, 243*.

Jöreskog, K. G., & Lawley, D. N. (1968). New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology, 21*(1), 85-96.

Jung, C. G. (2014). *THE COLLECTED WORKS OF CG JUNG: Mysterium Coniunctionis (Volume 14): An Inquiry into the Separation and Synthesis of Psychic Opposites in Alchemy*. Routledge.

Karwowski, M. (2011). It doesn't hurt to ask… But sometimes it hurts to believe: Polish students' creative self-efficacy and its predictors. *Psychology of Aesthetics, Creativity, and the Arts, 5*(2), 154.

Karwowski, M. (2014). Creative mindsets: Measurement, correlates, consequences. *Psychology of Aesthetics, Creativity, and the Arts, 8*(1), 62.

Karwowski, M., Lebuda, I., & Wisniewska, E. (2018). Measuring Creative Self-efficacy and Creative Personal Identity. *The Journal of Creativity and Problem Solving, 28*, 45-57.

Karwowski, M., Lebuda, I., & Wiśniewska, E. (2012). Measurement of creative self-efficacy and creative role-identity. *High Ability Studies, 22*, 291-231.

Kaufman, J. C., Baer, J., & Cole, J. C. (2009). Expertise, domains, and the consensual assessment technique. *The Journal of Creative Behavior, 43*(4), 223-233.

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology, 13*(1), 1-12.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology, 78*(5), 871.

Kenett, Y. N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in human neuroscience, 8*, 407.

Kenett, Y. N., Levy, O., Kenett, D. Y., Stanley, H. E., Faust, M., & Havlin, S. (2018). Flexibility of thought in high creative individuals represented by percolation analysis. *Proceedings of the National Academy of Sciences, 115*(5), 867-872.

Kharkhurin, A. (2011). Conceptual freedom of the globalized mind: Multicultural experiences enhance human cognition through the expansion of conceptual categories. *Journal of Consciousness Studies, 18*(3-4), 66-83.

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal, 18*(1), 3-14.

King, L., & Gurland, S. T. (2007). Creativity and experience of a creative task: Person and environment effects [Personality Traits & Processes 3120]. *Journal of Research in Personality, 41*(6), 1252-1259. https://doi.org/http://dx.doi.org/10.1016/j.jrp.2007.01.005

King, R., & Ruiz-Gelices, E. (2003). International student migration and the European 'year abroad': effects on European identity and subsequent migration behaviour. *International Journal of Population Geography, 9*(3), 229-252.

Kneller, G. F. (1965). *The art and science of creativity*. Holt, Rinehart and Winston.

Knoll, S., Eisend, M., & Steinhagen, J. (2011). Gender roles in advertising: Measuring and comparing gender stereotyping on public and private TV channels in Germany. *International Journal of Advertising, 30*(5), 867-888.

Koestler, A. (1964). The act of creation.

Kotler, P., Armstrong, G., Harker, M., & Brennan, R. (1990). *Marketing: an introduction* (Vol. 1). Prentice-Hall Englewood Cliffs, NJ.

Kover, A. J., Goldberg, S. M., & James, W. L. (1995). Creativity vs. effectiveness? An integrating classification for advertising. *Journal of Advertising Research, 35*(6), 29-41.

Kover, A. J., James, W. L., & Sonner, B. S. (1997). To whom do advertising creatives write? An inferential answer. *Journal of Advertising Research, 37*(1), 41-54.

Kudrowitz, B., & Dippo, C. (2013). Getting to the novel ideas: exploring the alternative uses test of divergent thinking. ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference,

Kurman, J. (2006). Self-enhancement, self-regulation and self-improvement following failures. *British Journal of Social Psychology, 45*(2), 339-356.

Landers, R. N., Collmus, A. B., & Williams, H. (2019). The greatest battle is within ourselves: An experiment on the effects of competition alone on task performance [Engineering & Environmental Psychology 4000]. *International Journal of Human-Computer Studies, 127*, 51-61. https://doi.org/http://dx.doi.org/10.1016/j.ijhcs.2018.09.011 (International Journal of Man-Machine Studies)

Lawley, D., & Maxwell, A. (1973). Regression ana factor analysis. *Biometrika, 60*(2), 331-338.

Lawshe, C. H., & Harris, D. (1960). *Purdue creativity test*. Purdue University, University book store.

Leavy, B. (2011). Roger Martin explores three big ideas: customer capitalism, integrative thinking and design thinking. *Strategy & Leadership*.

Lee, E.-J., & Schumann, D. W. (2004). Explaining the special case of incongruity in advertising: Combining classic theoretical approaches. *Marketing theory, 4*(1-2), 59-90.

Lee, V., & Harris, L. (2013). How social cognition can inform social decision making. *Frontiers in neuroscience, 7*, 259.

Lee, V. K., & Harris, L. T. (2013). How social cognition can inform social decision making. *Frontiers in neuroscience, 7*, 259.

Lehnert, K., Till, B. D., & Carlson, B. D. (2013). Advertising creativity and repetition: Recall, wearout and wearin effects. *International Journal of Advertising, 32*(2), 211-231.

Lei, W., Deng, W., Zhu, R., Runco, M. A., Dai, D. Y., & Hu, W. (2020). Does cognitive style moderate expected evaluation and adolescents' creative performance: An empirical study [Personality Psychology 3100]. *The Journal of Creative Behavior*, No-Specified. https://doi.org/http://dx.doi.org/10.1002/jocb.439

Lennox, R. D., & Wolfe, R. N. (1984). Revision of the self-monitoring scale.

Leon, S. A., Altmann, L. J., Abrams, L., Gonzalez Rothi, L. J., & Heilman, K. M. (2014). Divergent task performance in older adults: declarative memory or creative potential? *Creativity Research Journal, 26*(1), 21-29.

Leonard, B. L. (2014). *Stereotype transfer in advertising: Process and moderators* University of Colorado at Boulder].

Leong, C.-H., & Ward, C. (2000). Identity conflict in sojourners. *International journal of intercultural relations, 24*(6), 763-776.

Lin, W.-L., & Lien, Y.-W. (2013). The different role of working memory in open-ended versus closed-ended creative problem solving: a dual-process theory account. *Creativity Research Journal, 25*(1), 85-96.

Lingard, R., & Barkataki, S. (2011). Teaching teamwork in engineering and computer science. 2011 Frontiers in education conference (FIE),

Livne, N. L., Livne, O. E., & Wight, C. A. (2008). Enhancing Mathematical Creativity through Multiple Solution to Open-Ended Problems Online. *Diperoleh dari http://www. iste. org/Content/NavigationMenu/Research/NECC_Research_P aper_Archives/NECC2008/Livne. pdf*.

LOKŠA, J., & LOKŠOVÁ, I. (2000). ANALYSIS OF THE EFFECT OF SOCIO-PSYCHOLOGICAL. *Studia Psychologica, 42*(3), 209.

Lowenfeld, V., & Beittel, K. (1959). Interdisciplinary criteria of creativity in the arts and sciences: A progress report. *Research yearbook: National art education association, 19*, 35-44.

Lubart, T. I. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal, 13*(3-4), 295-308.

Lysonski, S., & Pollay, R. W. (1990). Advertising sexism is forgiven, but not forgotten: Historical, cross-cultural and individual differences in criticism and purchase boycott intentions. *International Journal of Advertising, 9*(4), 317-329.

Mace, M.-A., & Ward, T. (2002). Modeling the creative process: A grounded theory analysis of creativity in the domain of art making. *Creativity Research Journal, 14*(2), 179-192.

MacKinnon, D. W. (1965). Personality and the realization of creative potential. *American psychologist, 20*(4), 273.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of personality and social psychology, 67*(5), 808.

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of personality and social psychology, 66*(1), 37.

Mannix, E., & Neale, M. A. (2005, 2005/10/01). What Differences Make a Difference?: The Promise and Reality of Diverse Teams in Organizations. *Psychological Science in the Public Interest, 6*(2), 31-55. https://doi.org/10.1111/j.1529-1006.2005.00022.x

Martin, M. C., & Kennedy, P. F. (1993). Advertising and social comparison: Consequences for female preadolescents and adolescents. *Psychology & Marketing, 10*(6), 513-530.

Martin, R., & Austen, H. (1999). The art of integrative thinking. *Rotman Management, 12*(17), 4-9.

Martindale, C. (1999). Biological bases of creativity. *Handbook of creativity, 2*, 137-152.

Martins, E.-C., & Terblanche, F. (2003). Building organisational culture that stimulates creativity and innovation. *European journal of innovation management*.

Mastro, D., Lapinski, M. K., Kopacz, M. A., & Behm-Morawitz, E. (2009). The influence of exposure to depictions of race and crime in TV news on viewer's social judgments. *Journal of Broadcasting & Electronic Media, 53*(4), 615-635.

Mastro, D. E., & Stern, S. R. (2003). Representations of race in television commercials: A content analysis of prime-time advertising. *Journal of Broadcasting & Electronic Media, 47*(4), 638-647.

Mayseless, N., Eran, A., & Shamay-Tsoory, S. G. (2015). Generating original ideas: The neural underpinning of originality. *Neuroimage, 116*, 232-239.

Mayseless, N., Uzefovsky, F., Shalev, I., Ebstein, R. P., & Shamay-Tsoory, S. G. (2013). The association between creativity and 7R polymorphism in the dopamine receptor D4 gene (DRD4). *Frontiers in human neuroscience, 7*, 502.

McCloy, W., & Meier, N. C. (1939). Re-creative imagination. *Psychological Monographs, 51*(5), 108.

McKenzie, M., Bugden, M., Webster, A., & Barr, M. (2018). Advertising (in) equrality: The impacts of sexist advertising on women's health and wellbeing. *Women's Health Issues Paper*(14), i.

McLeod, D. M., Detenber, B. H., & Eveland Jr, W. P. (2001). Behind the third-person effect: Differentiating perceptual processes for self and other. *Journal of Communication, 51*(4), 678-695.

McNally, C. (1982). *The experience of being sensitive* The Union for Experimenting Colleges and Universities].

Mead, G. H., & Schubert, C. (1934). *Mind, self and society* (Vol. 111). University of Chicago press Chicago.

Mednick, S. (1962). The associative basis of the creative process. *Psychological review, 69*(3), 220.

Meeker, M., Meeker, R., & Roid, G. (1985). Structure of intellect learning abilities test (SOI-LA). *Los Angeles, CA: Western Psychological Services*.

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. *Personality and Social Psychology Bulletin, 36*(4), 512-523.

Merilinna, J., & Matinlassi, M. (2006). State of the art and practice of opensource component integration. 32nd EUROMICRO Conference on Software Engineering and Advanced Applications (EUROMICRO'06),

Merrifield, P., Gardner, S. F., & Cox, A. B. (1964). *Aptitudes and personality measures related to creativity in seventh-grade children*. Psychological Laboratory, University of Southern California.

Merten, T., & Fischer, I. (1999). Creativity, personality and word association responses: Associative behaviour in forty supposedly creative persons. *Personality and Individual Differences, 27*(5), 933-942.

Michinov, N., Jamet, E., Métayer, N., & Le Hénaff, B. (2015). The eyes of creativity: Impact of social comparison and individual creativity on performance and attention to others' ideas during electronic brainstorming. *Computers in Human Behavior, 42*, 57-67.

Michinov, N., & Primois, C. (2005). Improving productivity and creativity in online groups through social comparison process: New evidence for asynchronous electronic brainstorming [Human Factors Engineering 4010]. *Computers in Human Behavior, 21*(1), 11-28. https://doi.org/http://dx.doi.org/10.1016/j.chb.2004.02.004

Milgram, R. M., Milgram, N. A., Rosenbloom, G., & Rabkin, L. (1978). Quantity and quality of creative thinking in children and adolescents. *Child Development*, 385-388.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49-100.

Modig, E., & Rosengren, S. (2013). More than price? Exploring the effects of creativity and price in advertising. *ACR North American Advances*.

Monteith, M. J., Spicer, C. V., & Tooman, G. D. (1998). Consequences of stereotype suppression: Stereotypes on AND not on the rebound. *Journal of Experimental Social Psychology, 34*(4), 355-377.

Moreau, C. P., & Dahl, D. W. (2005). Designing the solution: The impact of constraints on consumers' creativity. *Journal of Consumer Research, 32*(1), 13-22.

Mumford, M. D. (2000). Managing creative people: Strategies and tactics for innovation. *Human resource management review, 10*(3), 313-351.

Mumford, M. D., & Gustafson, S. B. (1988). Creativity syndrome: Integration, application, and innovation. *Psychological Bulletin, 103*(1), 27.

Muscatell, K. A., Morelli, S. A., Falk, E. B., Way, B. M., Pfeifer, J. H., Galinsky, A. D., Lieberman, M. D., Dapretto, M., & Eisenberger, N. I. (2012). Social status modulates neural activity in the mentalizing network. *NeuroImage, 60*(3), 1771-1777.

Nabi, R. L., & Keblusek, L. (2014). Inspired by hope, motivated by envy: Comparing the effects of discrete emotions in the process of social comparison to media figures. *Media Psychology, 17*(2), 208-234.

Nathans, L. L., Oswald, F. L., & Nimon, K. (2012). Interpreting multiple linear regression: A guidebook of variable importance. *Practical Assessment, Research, and Evaluation, 17*(1), 9.

Nelson, L. J., & Klutas, K. (2000). The distinctiveness effect in social interaction: creation of a self-fulfilling prophecy. *Personality and Social Psychology Bulletin, 26*(1), 126-135.

Neuhaus, J. (2011). *Housework and housewives in American advertising: Married to the mop*. Springer.

Newell, A., Shaw, J. C., & Simon, H. A. (1962). The processes of creative thinking. Contemporary Approaches to Creative Thinking, 1958, University of Colorado, CO, US; This paper was presented at the aforementioned symposium.,

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2*(2), 175-220.

Niiniluoto, I. (2016). Science vs. Technology: Difference or Identity? In *Philosophy of technology after the empirical turn* (pp. 93-106). Springer.

Nijstad, B. A., De Dreu, C. K., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European review of social psychology, 21*(1), 34-77.

Nimon, K. F., & Oswald, F. L. (2013). Understanding the results of multiple linear regression: Beyond standardized regression coefficients. *Organizational Research Methods, 16*(4), 650-674.

Niu, W., & Sternberg, R. (2002). Contemporary studies on the concept of creativity: The East and the West. *The Journal of Creative Behavior, 36*(4), 269-288.

Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2017). Ha ha? Assessing individual differences in humor production ability. *Psychology of Aesthetics, Creativity, and the Arts, 11*(2), 231.

Oakenfull, G. K., McCarthy, M. S., & Greenlee, T. B. (2008). Targeting a minority without alienating the majority: Advertising to gays and lesbians in mainstream media. *Journal of Advertising Research, 48*(2), 191-198.

Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real-world problems. *Journal of Psychoeducational assessment, 9*(1), 45-53.

Olteţeanu, A.-M., Taranu, M., & Ionescu, T. (2019). Normative data for 111 compound Remote Associates Test problems in Romanian. *Frontiers in Psychology*, 1859.

Olteţeanu, A.-M., & Zunjani, F. H. (2020). A visual remote associates test and its validation. *Frontiers in Psychology, 11*, 26.

Onarheim, B. (2012). Creativity from constraints in engineering design: lessons learned at Coloplast. *Journal of Engineering Design, 23*(4), 323-336.

Pahl, G., & Beitz, W. (1984). Engineering design, The design council. *London, Springer, 12*, 221-226.

Park, B., Judd, C. M., & Ryan, C. S. (1991). Social categorization and the representation of variability information. *European review of social psychology, 2*(1), 211-245.

Peláez-Alfonso, J. L., Pelegrina, S., & Lechuga, M. T. (2020). Normative data for 102 Spanish remote associate problems and age-related differences in performance. *Psicológica Journal, 41*(1), 39-65.

Pereira, L. Q. (1999). Divergent thinking and the design process. International Conference on Design and Technology Educational Research and Curriculum Development Conference Book,

Perkins, D. N. (1998). In the country of the blind an appreciation of Donald Campbell's vision of creative thought. *The Journal of Creative Behavior*.

Petty, R. E., & Cacioppo, J. T. (1979). Effects of forwarning of persuasive intent and involvement on cognitive responses and persuasion. *Personality and Social Psychology Bulletin, 5*(2), 173-176.

Phinney, J. S., & Devich-Navarro, M. (1997). Variations in bicultural identification among African American and Mexican American adolescents. *Journal of research on Adolescence, 7*(1), 3-32.

Pick, D., & Eisend, M. (2014). Buyers' perceived switching costs and switching: a meta-analytic assessment of their antecedents. *Journal of the Academy of Marketing Science, 42*(2), 186-204.

Pick, D. F., Sweeney, J., & Clay, J. A. (1991). Creative advertising and the von Restorff effect. *Psychological Reports, 69*(3), 923-926.

Pickover, C. A. (2015). Strange Brains and Genius: The Secret Lives of Eccentric Scientists and Madmena, b. *Culture of Chemistry: The Best Articles on the Human Side of 20th-Century Chemistry from the Archives of the Chemical Intelligencer*, 93.

Pieters, R., Warlop, L., & Wedel, M. (2002). Breaking through the clutter: Benefits of advertisement originality and familiarity for brand attention and memory. *Management Science, 48*(6), 765-781.

Plakoyiannaki, E., & Zotos, Y. (2009). Female role stereotypes in print advertising: Identifying associations with magazine and product categories. *European Journal of Marketing*.

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational psychologist, 39*(2), 83-96.

Pollay, R. W. (1986). The distorted mirror: Reflections on the unintended consequences of advertising. *Journal of Marketing, 50*(2), 18-36.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism, 13*(2), 253.

Radel, R., Davranche, K., Fournier, M., & Dietrich, A. (2015). The role of (dis) inhibition in creativity: Decreased inhibition improves idea generation. *Cognition, 134*, 110-120.

Raina, M. (1968). A study into the effect of competition on creativity. *Gifted Child Quarterly, 12*(4), 217-220.

[Record #1424 is using a reference type undefined in this output style.]

Redifer, J. L., Bae, C. L., & Zhao, Q. (2021). Self-efficacy and performance feedback: Impacts on cognitive load during creative thinking. *Learning and Instruction, 71*, 101395.

Reiter-Palmon, R., & Illies, J. J. (2004). Leadership and creativity: Understanding leadership from a creative problem-solving perspective. *The Leadership Quarterly, 15*(1), 55-77.

Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan, 42*(7), 305-310.

Richards, L. G. (1998). Stimulating creativity: teaching engineers to be innovators. FIE'98. 28th Annual Frontiers in Education Conference. Moving from'Teacher-Centered'to'Learner-Centered'Education. Conference Proceedings (Cat. No. 98CH36214),

Richards, R., Kinney, D. K., Benet, M., & Merzel, A. P. (1988). Assessing everyday creativity: Characteristics of the Lifetime Creativity Scales and validation with three large samples. *Journal of personality and social psychology, 54*(3), 476.

Richins, M. L. (1991). Social comparison and the idealized images of advertising. *Journal of consumer research, 18*(1), 71-83.

Roe, A. (1953). A psychological study of eminent psychologists and anthropologists, and a comparison with biological and physical scientists. *Psychological Monographs: General and Applied, 67*(2), 1.

Rosengren, S., Dahlén, M., & Modig, E. (2013). Think outside the ad: Can advertising creativity benefit more than the advertiser? *Journal of Advertising, 42*(4), 320-330.

Rosengren, S., Eisend, M., Koslow, S., & Dahlen, M. (2020). A Meta-Analysis of When and How Advertising Creativity Works. *Journal of Marketing, 84*(6), 39-56.

Rothenberg, A. (1976). Janusian thinking and creativity. *Psychoanalytic Study of Society*.

Rowe, G., Hirsh, J. B., & Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences, 104*(1), 383-388.

Rudowicz, E., & Yue, X. D. (2000). Concepts of creativity: Similarities and differences among mainland, Hong Kong and Taiwanese Chinese. *The Journal of Creative Behavior, 34*(3), 175-192.

Runco, M. A. (2008). Commentary: Divergent thinking is not synonymous with creativity.

Runco, M. A. (2010). Divergent thinking, creativity, and ideation.

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66-75.

Runco, M. A., & Acar, S. (2019). Divergent Thinking. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (2 ed., pp. 224-254). Cambridge University Press. https://doi.org/DOI: 10.1017/9781316979839.013

Runco, M. A., & Acar, S. (2019). *Divergent thinking*. Cambridge University Press.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92-96.

Sak, U., & Ayas, M. B. (2013). Creative Scientific Ability Test (C-SAT): A new measure of scientific creativity. *Psychological Test and Assessment Modeling, 55*(3), 316.

Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain research reviews, 35*(2), 146-160.

Sassenberg, K., & Moskowitz, G. B. (2005). Don't stereotype, think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology, 41*(5), 506-514.

Sassenberg, K., Moskowitz, G. B., Fetterman, A., & Kessler, T. (2017). Priming creativity as a strategy to increase creative performance by facilitating the activation and use of remote associations. *Journal of Experimental Social Psychology, 68*, 128-138.

Sassenberg, K., Winter, K., Becker, D., Ditrich, L., Scholl, A., & Moskowitz, G. B. (2021). Flexibility mindsets: Reducing biases that result from spontaneous processing. *European review of social psychology*, 1-43.

Sawyer, R. K. (2006). Educating for innovation. *Thinking Skills and Creativity, 1*(1), 41-48.

Scharrer, E. (2002). Third-person perception and television violence: The role of out-group stereotyping in perceptions of susceptibility to effects. *Communication Research, 29*(6), 681-704.

Schexnayder, C., & Anderson, S. (2011). Construction engineering education: History and challenge. *Journal of construction engineering and management, 137*(10), 730-739.

Scopelliti, I., Cillo, P., Busacca, B., & Mazursky, D. (2014). How do financial constraints affect creativity? *Journal of Product Innovation Management, 31*(5), 880-893.

Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology, 80*(4), 645.

Seki, O., Nakatsuka, T., Shibata, H., & Kawamura, K. (2010). A compound-specific n-alkane δ13C and δD approach for assessing source and delivery processes of terrestrial organic matter within a forested watershed in northern Japan. *Geochimica et Cosmochimica Acta, 74*(2), 599-613.

Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2003). Metrics for measuring ideation effectiveness. *Design studies, 24*(2), 111-134.

Shalley, C. E. (1991). Effects of productivity goals, creativity goals, and personal discretion on individual creativity. *Journal of Applied Psychology, 76*(2), 179.

Shalley, C. E. (1995). Effects of coaction, expected evaluation, and goal setting on creativity and productivity. *Academy of Management Journal, 38*(2), 483-503.

Shalley, C. E., & Oldham, G. R. (1997). Competition and creative performance: Effects of competitor presence and visibility. *Creativity Research Journal, 10*(4), 337-345.

Shalley, C. E., & Perry-Smith, J. E. (2001). Effects of social-psychological factors on creative performance: The role of informational and controlling expected evaluation and modeling experience. *Organizational Behavior and Human Decision Processes, 84*(1), 1-22.

Shalley, C. E., Zhou, J., & Oldham, G. R. (2004). The effects of personal and contextual characteristics on creativity: Where should we go from here? *Journal of management, 30*(6), 933-958.

Shamay-Tsoory, S. G., Adler, N., Aharon-Peretz, J., Perry, D., & Mayseless, N. (2011). The origins of originality: the neural bases of creative thinking and originality. *Neuropsychologia, 49*(2), 178-185.

Shaw, M. P. (1989). The eureka process: A structure for the creative experience in science and engineering. *Creativity Research Journal, 2*(4), 286-298.

Sheinin, D. A., Varki, S., & Ashley, C. (2011). The differential effect of ad novelty and message usefulness on brand judgments. *Journal of Advertising, 40*(3), 5-18.

Shepard, R. N., & Cooper, L. A. (1986). *Mental images and their transformations*. The MIT Press.

Shepherd, M. M., Briggs, R. O., Reinig, B. A., Yen, J., & Nunamaker Jr, J. F. (1995). Invoking social comparison to improve electronic brainstorming: Beyond anonymity. *Journal of Management Information Systems, 12*(3), 155-170.

Sherman, J. W., & Frost, L. A. (2000). On the encoding of stereotype-relevant information under cognitive load. *Personality and Social Psychology Bulletin, 26*(1), 26-34.

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75*(3), 589.

Sill, D. J. (1996). Integrative thinking, synthesis, and creativity in interdisciplinary studies. *The Journal of General Education, 45*(2), 129-151.

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (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*(2), 68.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science, 22*(11), 1359-1366.

Simonton, D. K. (1999). *Origins of genius: Darwinian perspectives on creativity*. Oxford University Press.

Simonton, D. K. (2008). Creativity and genius.

Simonton, D. K. (2010). Creative thought as blind-variation and selective-retention: Combinatorial models of exceptional creativity. *Physics of life reviews, 7*(2), 156-179.

Simonton, D. K. (2011). Creativity and discovery as blind variation: Campbell's (1960) BVSR model after the half-century mark. *Review of General Psychology, 15*(2), 158-174.

Slijkhuis, J., Rietzschel, E. F., & Van Yperen, N. W. (2013). How evaluation and need for structure affect motivation and creativity. *European Journal of Work and Organizational Psychology, 22*(1), 15-25.

Smith, R. E., Chen, J., & Yang, X. (2008). The impact of advertising creativity on the hierarchy of effects. *Journal of Advertising, 37*(4), 47-62.

Smith, R. E., MacKenzie, S. B., Yang, X., Buchholz, L. M., & Darley, W. K. (2007). Modeling the determinants and effects of creativity in advertising. *Marketing science, 26*(6), 819-833.

Smith, R. E., & Yang, X. (2004). Toward a general theory of creativity in advertising: Examining the role of divergence. *Marketing theory, 4*(1-2), 31-58.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797.

Stein, M. I. (1953). Creativity and culture. *The journal of psychology, 36*(2), 311-322.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of personality and social psychology, 49*(3), 607.

Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized*. Cambridge University Press.

Sternberg, R. J. (2006). The nature of creativity. *Creativity Research Journal, 18*(1), 87.

Sternberg, R. J. (2020). What's wrong with creativity testing? *The Journal of Creative Behavior, 54*(1), 20-36.

Sternberg, R. J., & Davidson, J. E. (2005). *Conceptions of giftedness* (Vol. 2). Cambridge University Press New York, NY.

Sternberg, R. J., & Lubart, T. I. (1991). An investment theory of creativity and its development. *Human development, 34*(1), 1-31.

Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American psychologist, 51*(7), 677.

Stouffer, W., Russell, J. S., & Oliva, M. G. (2004). Making the strange familiar: Creativity and the future of engineering education. American Society for Engineering Education Annual Conference and Exposition,

Strong, S. R., & Gray, B. L. (1972). Social comparison, self-evaluation, and influence in counseling [Personality Psychology 3100]. *Journal of Counseling Psychology, 19*(3), 178-183. https://doi.org/http://dx.doi.org/10.1037/h0032715

Swencionis, J. K., & Fiske, S. T. (2014). How social neuroscience can inform theories of social comparison. *Neuropsychologia, 56*, 140-146.

Szostak, R. (2002). How to do interdisciplinarity: Integrating the debate.

Tadmor, C. T., Tetlock, P. E., & Peng, K. (2009). Acculturation strategies and integrative complexity: The cognitive implications of biculturalism. *Journal of Cross-Cultural Psychology, 40*(1), 105-139.

Talbot, R., Cooper, C., & Barrow, S. (1992). Creativity and stress. *Creativity and Innovation Management, 1*(4), 183-193.

Tan, A.-G. (2015). Convergent Creativity: From Arthur Cropley (1935-) Onwards. *Creativity Research Journal, 27*(3), 271-280.

Taylor, C. R., & Stern, B. B. (1997). Asian-Americans: Television advertising and the "model minority" stereotype. *Journal of Advertising, 26*(2), 47-61.

Taylor, S. E., & Lobel, M. (1989). Social comparison activity under threat: downward evaluation and upward contacts. *Psychological review, 96*(4), 569.

Tellis, G. J. (1997). Effective frequency: One exposure or three factors? *Journal of Advertising Research*, 75-80.

Terman, L. M. (1940). Psychological approaches to the biography of genius. *Science, 92*(2388), 293-301.

Thissen, D. (1983). Timed testing: An approach using item response theory. In *New horizons in testing* (pp. 179-203). Elsevier.

Thompson, G., & Lordan, M. (1999). A review of creativity principles applied to engineering design. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, 213*(1), 17-31.

Till, B. D., & Baack, D. W. (2005). Recall and persuasion: does creative advertising matter? *Journal of Advertising, 34*(3), 47-57.

Toivainen, T., Olteteanu, A.-M., Repeykova, V., Likhanov, M., & Kovas, Y. (2019). Visual and linguistic stimuli in the remote associates test: a cross-cultural investigation. *Frontiers in Psychology*, 926.

Torrance, E. P. (1962). Cultural discontinuities and the development of originality of thinking. *Exceptional Children, 29*(1), 2-13.

Torrance, E. P. (1972). Predictive validity of the torrance tests of creative thinking. *The Journal of Creative Behavior*.

Treffinger, D. J. (1995). Creative problem solving: Overview and educational implications. *Educational Psychology Review, 7*(3), 301-312.

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of personality and social psychology, 74*(1), 118.

Ueda, Y., Tominaga, A., Kajimura, S., & Nomura, M. (2016). Spontaneous eye blinks during creative task correlate with divergent processing. *Psychological Research, 80*(4), 652-659.

Urban, K. K. (2005). Assessing Creativity: The Test for Creative Thinking-Drawing Production (TCT-DP). *International Education Journal, 6*(2), 272-280.

Van de Ven, N. (2017). Envy and admiration: Emotion and motivation following upward social comparison. *Cognition and Emotion, 31*(1), 193-200.

Van de Ven, N., Zeelenberg, M., & Pieters, R. (2011). Why envy outperforms admiration. *Personality and Social Psychology Bulletin, 37*(6), 784-795.

Van Knippenberg, A., Wilke, H., & de Vries, N. K. (1981). Social comparison on two dimensions [Psychosocial & Personality Development 2840]. *European Journal of Social Psychology, 11*(3), 267-283. https://doi.org/http://dx.doi.org/10.1002/ejsp.2420110303

Van Leeuwen, W., & Baas, M. (2017). Creativity under attack: how people's role in competitive conflict shapes their creative idea generation. *Creativity Research Journal, 29*(4), 354-369.

Vartanian, O. (2009). Variable attention facilitates creative problem solving. *Psychology of Aesthetics, Creativity, and the Arts, 3*(1), 57-59. https://doi.org/10.1037/a0014781

Vartanian, O. (2016). Attention, cognitive flexibility, and creativity: Insights from the brain [Cognitive Processes 2340]. *Creativity and reason in cognitive development., 2nd ed.*, 246-258. https://doi.org/https://dx.doi.org/10.1017/CBO9781139941969.013 (Current perspectives in social and behavioral sciences.)

Verene, D. P. (1981). *Vico's science of imagination*. Cornell University Press.

Vygotsky, L. S. (1997). *The collected works of LS Vygotsky: Problems of the theory and history of psychology* (Vol. 3). Springer Science & Business Media.

Walker, B. R., & Jackson, C. J. (2014). How the five factor model and revised reinforcement sensitivity theory predict divergent thinking. *Personality and Individual Differences, 57*, 54-58.

Wallace, D. B., & Gruber, H. E. (1989). *Creative people at work: Twelve cognitive case studies*. Oxford University Press on Demand.

Wallach, M. A., & Kogan, N. (1965a). Modes of thinking in young children.

Wallach, M. A., & Kogan, N. (1965b). A new look at the creativity-intelligence distinction 1. *Journal of personality, 33*(3), 348-369.

Wallas, G. (1926). The art of thought.

Wan, F., Ansons, T. L., Chattopadhyay, A., & Leboe, J. P. (2013). Defensive reactions to slim female images in advertising: The moderating role of mode of exposure. *Organizational Behavior and human decision processes, 120*(1), 37-46.

Wang, D., Liu, T., & Shi, J. (2020). Neural dynamic responses of monetary and social reward processes in adolescents. *Frontiers in Human Neuroscience, 14*, 141.

Wang, J., Wang, L., Liu, R.-D., & Dong, H.-Z. (2017). How expected evaluation influences creativity: Regulatory focus as moderator [Personality Psychology 3100]. *Motivation and Emotion, 41*(2), 147-157. https://doi.org/http://dx.doi.org/10.1007/s11031-016-9598-y

Wang, K., & Holahan, P. (2017). The effect of monetary reward on creativity: The role of motivational orientation. Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS),

Ward, T. B., Smith, S. M., & Vaid, J. (1997). Conceptual structures and processes in creative thought [Cognitive Processes 2340]. *Creative thought: An investigation of conceptual structures and processes.*, 1-27. https://doi.org/https://dx.doi.org/10.1037/10227-001

Wei, D., Yang, J., Li, W., Wang, K., Zhang, Q., & Qiu, J. (2014). Increased resting functional connectivity of the medial prefrontal cortex in creativity by means of cognitive stimulation. *cortex, 51*, 92-102.

Weinberger, A. B., Iyer, H., & Green, A. E. (2016). Conscious augmentation of creative state enhances "real" creativity in open-ended analogical reasoning. *PloS one, 11*(3), e0150773.

Weisberg, R. (1986). *Creativity: Genius and other myths*. WH Freeman/Times Books/Henry Holt & Co.

Wen, F., Zuo, B., Xie, Z., & Gao, J. (2019). Boosting Creativity, but Only for Low Creative Connectivity: The Moderating Effect of Priming Stereotypically Inconsistent Information on Creativity. *Frontiers in psychology, 10*, 273.

West, D., Koslow, S., & Kilgour, M. (2019). Future directions for advertising creativity research. *Journal of Advertising, 48*(1), 102-114.

Wheeler, L. (1966). Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology, 1*, 27-31.

Wigboldus, D. H., Sherman, J. W., Franzese, H. L., & Knippenberg, A. v. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition, 22*(3), 292-309.

Wijaya, A. (2018). How do open-ended problems promote mathematical creativity? A reflection of bare mathematics problem and contextual problem. Journal of Physics: Conference Series,

Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin, 50*(5), 362.

Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). Unusual Uses Test [Personality 7200]. https://doi.org/http://dx.doi.org/10.1037/t50605-000

Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological bulletin, 106*(2), 231.

Wood, J. V. (1996). What is social comparison and how should we study it? *Personality and social psychology bulletin, 22*(5), 520-537.

Worthen, B. R., & Clark, P. M. (1971). Toward an improved measure of remote associational ability. *Journal of Educational Measurement, 8*(2), 113-123.

Wu, C.-L., & Chen, H.-C. (2017). Normative data for Chinese compound remote associate problems. *Behavior Research Methods, 49*(6), 2163-2172.

Wulf, W. A. (1998). The urgency of engineering education reform. Proceedings, Conference on Realizing the New Paradigm for Engineering Education,

Yang, X., & Smith, R. E. (2009). Beyond attention effects: Modeling the persuasive and emotional effects of advertising creativity. *Marketing science, 28*(5), 935-949.

Yeh, Y.-c., Lai, G.-J., Lin, C. F., Lin, C.-W., & Sun, H.-C. (2015). How stress influences creativity in game-based situations: Analysis of stress hormones, negative emotions, and working memory. *Computers & Education, 81*, 143-153.

Youn, S., Faber, R. J., & Shah, D. V. (2000). Restricting gambling advertising and the third-person effect. *Psychology & Marketing, 17*(7), 633-649.

Yuan, F., & Zhou, J. (2008). Differential effects of expected external evaluation on different parts of the creative idea production process and on final product creativity [Cognitive Processes 2340]. *Creativity Research Journal, 20*(4), 391-403. https://doi.org/http://dx.doi.org/10.1080/10400410802391827

Zabelina, D. L. (2018a). Attention and creativity. In *The Cambridge handbook of the neuroscience of creativity.* (pp. 161-179). Cambridge University Press. https://doi.org/10.1017/9781316556238.010

Zabelina, D. L. (2018b). Attention and Creativity. In O. Vartanian & R. E. Jung (Eds.), *The Cambridge handbook of the neuroscience of creativity* (pp. 161-179). Cambridge University Press. https://doi.org/DOI: 10.1017/9781316556238.010

Zabelina, D. L., Colzato, L., Beeman, M., & Hommel, B. (2016). Dopamine and the creative mind: Individual differences in creativity are predicted by interactions between dopamine genes DAT and COMT. *PLoS ONE, 11*(1), e0146768.

Zabelina, D. L., & Ganis, G. (2018, 2018/09/01/). Creativity and cognitive control: Behavioral and ERP evidence that divergent thinking, but not real-life creative achievement, relates to better cognitive control. *Neuropsychologia, 118*, 20-28. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2018.02.014

Zabelina, D. L., O'Leary, D., Pornpattananangkul, N., Nusslock, R., & Beeman, M. (2015, 2015/03/01/). Creativity and sensory gating indexed by the P50: Selective versus leaky sensory gating in divergent thinkers and creative achievers. *Neuropsychologia, 69*, 77-84. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2015.01.034

Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal, 23*(1), 24-37.

Zhai, L., & Scheer, S. D. (2004). Global perspectives and attitudes toward cultural diversity among summer agriculture students at the Ohio State University. *Journal of Agricultural Education*.

Zhou, J., & George, J. M. (2001). When job dissatisfaction leads to creativity: Encouraging the expression of voice. *Academy of Management Journal, 44*(4), 682-696.

Zhou, J., & Oldham, G. R. (2001). Enhancing creative performance: Effects of expected developmental assessment strategies and creative personality. *The Journal of Creative Behavior, 35*(3), 151-167.

Zhu, W., Chen, Q., Xia, L., Beaty, R. E., Yang, W., Tian, F., Sun, J., Cao, G., Zhang, Q., & Chen, X. (2017). Common and distinct brain networks underlying verbal and visual creativity. *Human brain mapping, 38*(4), 2094-2111.

Zimmerman, A., & Dahlberg, J. (2008). The sexual objectification of women in advertising: A contemporary cultural perspective. *Journal of Advertising Research, 48*(1), 71-79.

Zittoun, T., Baucal, A., Cornish, F., & Gillespie, A. (2007). Collaborative research, knowledge and emergence. *Integrative Psychological and Behavioral Science, 41*(2), 208-217.

Zuberi, T., & Bonilla-Silva, E. (2008). *White logic, white methods: Racism and methodology*. Rowman & Littlefield Publishers.

Zuo, B., Wen, F., Wang, M., & Wang, Y. (2019). The Mediating Role of Cognitive Flexibility in the Influence of Counter-Stereotypes on Creativity. *Frontiers in psychology, 10*, 105.