



Comprehensive Scientific Creativity Assessment (C-SCA): A New Approach for Measuring Scientific Creativity in Secondary School Students

Shiyu Xu^{1,2} · Michael J. Reiss¹ · Wilton Lodge¹

Received: 24 June 2023 / Accepted: 6 May 2024
© The Author(s) 2024

Abstract

This study introduces a Comprehensive Scientific Creativity Assessment (C-SCA) instrument and empirically tests its reliability and validity. While existing instruments to measure scientific creativity generally focus on a single dimension, such as divergent thinking, the C-SCA incorporates scientific knowledge, motivation in scientific creativity and thinking styles within its assessment framework. In this study, the three dimensions of C-SCA were measured using modifications of existing instruments. We provide two versions of the test, to allow tests of the effectiveness of educational interventions on scientific creativity. We incorporated ChatGPT into the scientific creativity scoring process to enhance the objectivity of the scores. Looking ahead, technological advancements hold promise for further improvements in scientific creativity assessment.

Keywords Comprehensive Scientific Creativity Assessment (C-SCA) · Creative Trait Motivation (CTM) · Scientific knowledge · Scientific Creativity Test for Upper Secondary School Students (SCT-USSS) · Thinking styles

In the twenty-first century, fostering students' creativity is increasingly regarded as an essential feature of education (Kupers et al., 2019). Sternberg argued that “creativity is the ability to produce work that is both novel (i.e., original, unexpected) and

✉ Shiyu Xu
s.xu.22@ucl.ac.uk

Michael J. Reiss
m.reiss@ucl.ac.uk

Wilton Lodge
wilton.lodge@ucl.ac.uk

¹ UCL Institute of Education, UCL, London, UK

² IOE, UCL's Faculty of Education and Society, University College London, 20 Bedford Way, London WC1H 0AL, UK

appropriate (i.e., useful, adaptive, concerning task constraints)” (Sternberg, 1999, p. 3). Novelty and appropriateness are two main features of creativity (Nijstad et al., 2010), and these two characteristics should also be involved when we discuss the notion of creativity in the context of science (Hadzigeorgiou et al., 2012).

A feature that perhaps distinguishes scientific creativity from other types of creativity is the strict requirement for the expression of creativity in science to align with the established features of the material world. As Simonton (2004, p. 6) perhaps somewhat optimistically asserted, “once a scientist masters the logic of science and the substance of a particular discipline, creativity is assured”. In addition, scientific discovery can be seen as a form of problem-solving which requires both rationality and strict empirical testing (Kind & Kind, 2007). When novel ideas are verified (at least, not refuted) by experimentation, they can be deemed factually sound and logically acceptable, thus showing their value to the scientific world and society more generally. Therefore, closer inspection is important so as to look beyond surface-level observations and explore the intricate details that contribute to the holistic understanding of scientific creativity. Such inspection reveals that scientific creativity is not only the manifestation of inspiration and imagination, but also the process of transforming creative ideas into scientific knowledge by logical reasoning within the existing intellectual framework for the scientific discipline in question. Lubart et al. (2022) emphasised that employing appropriate scientific creativity assessment tools not only aids in identifying young people’s creative potential in the science field but also, when adequately conceptualised and measured, enables educators to develop effective educational strategies for enhancing students’ creativity and to critique activities that hinder students’ creativity. Accordingly, this article reviews existing assessment tools related to creativity and scientific creativity, and introduces the ‘Comprehensive Scientific Creativity Assessment (C-SCA)’ as a new framework for future research in this field.

Literature Review

Creativity Assessment

The first published statement that creativity can be measured comes from Guilford (1950). He stated that creativity can be evaluated in psychometric tests, marking the establishment of the modern science of creativity. Rhodes (1961) proposed a ‘four Ps’ creativity structure that guided creativity researchers to assess creativity in four different dimensions, namely process, person, product and press. The term ‘processes’ focuses on learning and thinking, and researchers believed that divergent thinking was a major component of creativity, so they developed a number of divergent thinking tests to measure creativity, such as ‘Guilford’s Divergent Production Tests’ (Guilford, 1968), the ‘Torrance Tests of Creative Thinking’ (TTCT) (Torrance, 1966) and the ‘Remote Associates Test’ (RAT) (Mednick, 1962). The term ‘person’ includes information about personality and other traits, and researchers created measurement instruments to assess some creative personalities, such as the ‘Adjective Checklist Creative Personality Scale’ (Gough & Heilbrun, 1965) and

the ‘Creative Personality Scale’ (Kaufman & Baer, 2004). The term ‘product’ is the outcome of creativity, so researchers developed some instruments for evaluating creative products, such as the ‘Consensual Assessment Technique’ (Amabile, 1982), the ‘Lifetime Creativity Scale’ (Richards et al., 1988) and the ‘Creative Achievement Scale’ (Ludwig, 1992). Last of all, the term ‘press’ refers to the relationship between humans and their environment. Some instruments, such as ‘Assessing the Climate for Creativity’ (Tseng & Liu, 2011), the ‘Creative Climate Questionnaire’ (Ekvall, 1996) and the ‘Situational Outlook Questionnaire’ (Isaksen et al., 2001), can be used to measure the creative environment. However, although these instruments can be used in a variety of fields, such as science and art, they lack a specific focus. Furthermore, these assessment instruments only focus on one aspect and do not provide a comprehensive description of creativity. Specifically, creativity may not only require divergent thinking, but also strong intrinsic motivation, defined as “the drive to do something for the sheer enjoyment, interest, and personal challenge of the task itself” (p. 581), thereby triggering the creative process (Hennessey & Amabile, 2010). These are limitations that the scientific creativity instrument introduced below aims to overcome.

In response to these issues, creativity assessment has seen some changes in two aspects. One aspect is creativity’s general versus domain-specific nature. According to Baer (1994), creative performance on a specific task does not serve as a reliable predictor of creative performance on other tasks, and in his empirical studies, the creativity scores of individuals on different tasks confirmed this idea to some extent. As individual creativity may require specific knowledge and skills in different domains, many researchers started to measure creativity in specific domains, such as artistic creativity (Lunke & Meier, 2016), mathematical creativity (Mann, 2009) and scientific creativity (Hu & Adey, 2002; Sak & Ayas, 2013). Another aspect is that researchers since Sternberg (1988) have recognised that no single ability or trait is the key to creativity and that creativity emerges in an integrated, comprehensive process. Sternberg (1988) provided a ‘three-facet model of creativity’, including intelligence, thinking styles and personality, and the ‘Sternberg Triarchic Abilities Test’ pays still more attention to individuals’ intelligence and thinking styles (Koke & Vernon, 2003), though it understates the measurement of personality. Amabile (1996) believed that creative products result from the interaction between the individual and the organisation, but the ‘Consensual Assessment Technique’ developed by Amabile’s research teams relies on subjective judgements (Hennessey et al., 2011). Therefore, existing creativity measurement instruments still have certain limitations, but this holistic perspective leads creativity assessment to a multidimensional stage.

Scientific Creativity Assessment

Scientific creativity is a domain-specific creativity whose assessment should consider scientific knowledge, skills and the existing scientific framework for the science domain. Lubart et al. (2022) reviewed a number of existing instruments for scientific creativity and concluded that they could be divided into three main types:

accomplishment-based measures, science-based competitions and psychometric tests. The first type is based on individual accomplishments, such as a published scientific paper or an invention recognised by peers in the scientific field, but this approach is more suitable for adults engaged in scientific careers than for school students. The second type involves science talent competitions, such as the Science Olympiad in the United States, that seek to foster student interest in science and recognise achievements in science education (Stroup & Thacker, 2007). However, the rules and award structures of such competitions may not facilitate the unveiling of participants' scientific creativity. The final and most commonly used assessment method is a psychometric test, which we now discuss.

The 'Scientific Creativity Test' (SCT), developed by Hu and Adey (2002), is a widely used instrument. Hu and Adey (2002) started by proposing a new three-dimensional 'scientific structure creativity model', which includes scientific processes, scientific products and personality traits. The SCT consists of seven open-ended questions, designed for secondary school students. Three questions focus on 'unusual uses', 'the sensitivity of a science problem' and 'scientific imagination', and the other four questions assess students' ability, including 'technical products improvement', 'creative science products designing', 'creative science problem-solving ability' and 'creative experimental ability'. Doubts have been raised as to whether the seven-question test can fully assess the complex construct of scientific creativity (Aschauer et al., 2022).

Sak and Ayas's (2013) 'Creative Scientific Ability Test' (C-SAT) provides a more comprehensive and systematic attempt (Aschauer et al., 2022). It consists of five questions developed for students in grade 6 through grade 8, and these questions cover biology, physics, chemistry, ecology, and interdisciplinary science, and are related to hypothesis generation, experiment design and evidence evaluation (Sak & Ayas, 2013). Research on the psychometric properties of the C-SAT shows its acceptable validity and reliability (Ayas & Sak, 2014). However, in a comparative empirical study, both the scores of students' scientific performance and their domain-general creativity reflected their scientific creativity better in the SCT than in the C-SAT, with the authors speculating that the more science-related construct of C-SAT may be related to these findings (Huang & Wang, 2019). As a result, C-SAT's focus on scientific knowledge and skills deserves attention in subsequent scientific creativity assessments.

Another scientific creativity test, developed by Xu (2013), is specifically designed for Chinese upper secondary school students. An advantage of this test is that it assesses both divergent and convergent thinking as students solve science problems. Xu's (2013) study demonstrates its validity in three ways: structural validity, convergent validity and content validity. However, three problems with this test need to be raised. First, it was created in 2013, and some of the questions have contexts, such as acid rain, that are no longer so appropriate in China. Second, to assess convergent thinking, sometimes students are asked to evaluate all the solutions, and sometimes only one solution. As a result, in the first case, if students list too many solutions, they do not have enough time to evaluate all of them. Third, the test has an ambiguous scoring method for convergent thinking. For example, the critique scoring criteria show that when respondents list some reasons, they can be given two points, but

the reasons for this are not clarified (Xu, 2013). Appropriate topic contexts, question design and clear evaluation criteria should be considered for any scientific creativity assessment.

Another instrument for measuring scientific creativity is the ‘divergent problem-solving ability in science test’ (DPAS), which includes two subtests, namely divergent ideation in science tasks (six tasks) and divergent ideation in experimental tasks (seven tasks), targeted at students in grade 5 through grade 13 (Aschauer et al., 2022). However, DPAS only considers physics and chemistry as areas of science, and it does not score novelty, a central criterion of scientific creativity (Hadzigeorgiou et al., 2012). Three other measurement instruments, the ‘Creative Achievement Questionnaire’ (Carson et al., 2005), the ‘Kaufman Domains of Creativity Scale’ (Kaufman, 2012) and the ‘Creative Activity and Accomplishment Checklist’ (Paek & Runco, 2017), are all self-report questionnaires, which ask students to report their creative activities or achievement, e.g. “I often think about ways that scientific problems could be solved” (Carson et al., 2005, p. 49). Self-report assessments raise certain issues, including social desirability bias (students may answer questions in a manner that will be viewed favourably by others) (Krumpal, 2013), so whether these instruments can provide valid measures of students’ scientific creativity is open for debate.

Theoretical Framework for the Comprehensive Scientific Creativity Assessment

The ‘three-facet model of creativity’ was initially proposed by Sternberg (1988), and the model shows that creativity involves intelligence, cognitive style and personality/motivation. The three-facet model was subsequently subsumed within ‘the investment theory of creativity’, which specifies elements (or components) that jointly contribute toward creativity, with these elements being intelligence, knowledge, thinking styles, personality, motivation and environment (Sternberg & Lubart, 1992). The theory mainly emphasises the elements that contribute to creativity, without describing the specific creative processes.

Amabile (1996) not only described the components of creativity, but also how the components contribute to the creative process, thus forming the componential model of creativity. The creative process is seen as involving five stages, namely problem or task identification, preparation, response generation, response validation and communication, and outcome. In the creative process, task motivation plays a crucial role in initiating and sustaining the process, and it determines whether the search for a solution will persist; the domain-relevant skills serve as the materials utilised during the creative operation, and they determine the available pathways and criteria used to evaluate potential responses; the creativity-relevant processes encompass a cognitive style conducive to generating new ideas, an implicit or explicit knowledge of heuristics for generating novel ideas, and a work style that fosters creativity, and these processes collectively influence how the search for responses will proceed (Amabile, 1996).

Based on previous theories, we propose a theoretical framework for Comprehensive Scientific Creativity Assessment (C-SCA). C-SCA covers the subjects of biology, chemistry and physics, and tasks are presented that aim to measure students’ performance in scientific knowledge, motivation, and thinking styles (see Table 1).

Table 1 The components of Comprehensive Scientific Creativity Assessment (C-SCA)

Area of science	Scientific task	Scientific creativity
Biology	(1) Ask scientific questions	Scientific knowledge
Chemistry	(2) Propose solutions	Conceptual knowledge
Physics	(3) Evaluate solutions	Procedural knowledge
	(4) Conduct experiments and verify results	Motivation in scientific creativity
		Intrinsic motivation
		Extrinsic motivation
		Thinking styles
		Divergent thinking
		Convergent thinking

Scientific knowledge is one of the bases of scientific creativity (Hu & Adey, 2002). Specifically, this knowledge not only enables students to discover and solve problems in different ways but also plays an important role in the scientific creativity process, encompassing hypothesis generation, experimental design, and evidence hypotheses (Huang et al., 2017). Scientific knowledge consists of conceptual knowledge and procedural knowledge, with the former referring to ‘knowing that’, which not only involves facts, laws, concepts and theories but also their relationships, and the latter to ‘knowing how’, which refers to the process of problem-solving or understand the methods of scientific inquiry (McCormick, 1997; Millar et al., 1994). These two dimensions of knowledge are also reflected in the Compulsory Education Science Curriculum Standards, and the official document emphasises that the science curriculum aims to cultivate students’ scientific perspectives, involving a comprehensive understanding of objective phenomena through grasping scientific concepts, principles, and laws, alongside fostering their inquiry practices by imparting a general comprehension of the processes and methods inherent in scientific inquiry (Ministry of Education of the People’s Republic of China [MOE], 2022). Therefore, for secondary school students’ scientific creativity, the knowledge dimension means that they can propose new ideas based on conceptual knowledge of science and undertake scientific inquiry based on procedural knowledge to investigate the feasibility of ideas.

Motivation refers to the need or reason for engaging in scientific creativity. Intrinsic motivation refers to the internal drive to engage in an activity purely for the enjoyment, interest or personal challenge it provides, while extrinsic motivation arises when individuals engage in an activity primarily driven by external factors, such as rewards, recognition or the expectations of others (Hennessey & Amabile, 2010). There is a consensus that intrinsic motivation benefits creativity, but the impact of extrinsic motivation on creativity is still unclear (Amabile, 1996). Chinese secondary school students are adolescents, and their learning motivation experiences can change over time – they might become more concerned with peer competition and social relationships because of China’s exam-oriented education system (Wu et al., 2022). Also, patriotism is a central value for Chinese scientists (The State Council of the People’s Republic of China, 2019), which means that they engage in scientific research not only for the sake of science itself, but also to make a contribution to the country. As a result, both intrinsic motivation and extrinsic motivation of Chinese secondary school students might contribute to their scientific creativity. Furthermore, Velayutham et al. (2012) reviewed

previous studies on gender differences in student motivation in science learning and found that girls have lower self-perceptions of their academic ability in science even they perform better than boys, and that during adolescence, girls are more inclined to conform to gender stereotypic roles, implying that boys instead of girls may appear to have higher ability and interest in science. Thus, this study also investigates whether there are gender differences in motivations for scientific creativity.

Thinking styles include divergent thinking and convergent thinking, both of which are crucial components of scientific creativity (Shin & Park, 2021). Divergent thinking leads students to think in different directions from those that are standard, in ways that reflect the novelty of their ideas; it usually relies on three indices: fluency, flexibility and novelty (Acar & Runco, 2019). Fluency reflects the productivity aspect of divergent thinking; flexibility means that individuals can provide a variety of responses; novelty represents those ideas that are uncommon, remote from the everyday, yet clear (Reiter-Palmon et al., 2019). When exploring scientific problems, people with strong divergent thinking could provide more, diverse and novel ideas. By way of contrast, many previous studies argued that convergent thinking conflicts with divergent thinking and hampers creativity (for an early review, see Cropley, 1967). However, as more studies re-examined the effect of convergent thinking on creativity, it became increasingly accepted that creativity necessitates both divergent and convergent thinking (Cropley, 2006; Webb et al., 2017; Zhu et al., 2019).

Convergent thinking is characterised by its focus on identifying the single best or correct answer to a well-defined question (Cropley, 2006). Zhu et al. (2019) discovered that divergent thinking only became important for scientific creativity when individuals possessed a certain level of convergent thinking ability. However, many studies that investigate the role of divergent and convergent thinking on creativity use two separate and unrelated tests (e.g. Webb et al., 2017; Zhu et al., 2019). In addition, convergent thinking tests typically only assess the accuracy of the results (e.g. Mednick, 1962). Such a view of convergent thinking's role in scientific creativity may be inappropriate. According to Cropley (2006), divergent and convergent thinking should be integrated into the creative process. Our study adopts this perspective when measuring students' thinking styles in scientific creativity: divergent thinking is regarded as generating more, various and novel ideas, and convergent thinking is understood as evaluating those ideas and eventually arriving at the most appropriate answer.

Method

Participants

Our empirical test of the Comprehensive Scientific Creativity Assessment (C-SCA) was conducted on 24 April 2023, when sample students undertook the assessment in their classrooms. This study used purposive sampling, and 189 students in grade 10 from four classes in a secondary school were recruited. The sample school was located in Taiyuan, City, Shanxi Province, China. The school was a public institution that adhered to the national curriculum, mirroring the educational practices of most secondary schools in China. Four representative classes were chosen to complete the

Table 2 Participant distribution

Variable		Number	Percent
Class	1	47	24.9
	2	45	23.8
	3	46	24.3
	4	51	27.0
Age	15	57	30.2
	16	124	65.6
	17	8	4.2
Gender	Male	115	60.8
	Female	74	39.2
Total		189	100.0

assessment to ensure that students of different levels could understand the content of the assessment. The C-SCA was tested in two versions: A and B. The difference is that the thinking style test questions are different, as discussed further below. Class 1 and Class 2 completed the A version of the assessment, with students in Class 1 having higher academic performance and students in Class 2 having lower academic performance. Class 3 and Class 4 completed the B version of the assessment, with students in Class 3 having higher academic performance and students in Class 4 having lower academic performance. Details of the sample can be found in Table 2.

Assessment Procedure

First, the principal and teachers reviewed the information sheet and signed the informed consent forms. Next, students read the information sheets to know they were invited to participate in a scientific creativity assessment. Meanwhile, the first author instructed the four teachers responsible for the assessment to familiarise themselves with the whole process. During the assessment, students from a class spent 50 min in the designated classroom, completing the assessment independently, with two teachers supervising and providing support. Participation in the assessment was completely voluntary and anonymous. Before the study commenced, ethical approval was obtained from our university.

Instrument

The C-SCA has three dimensions: scientific knowledge, students' motivation in scientific creativity, and thinking styles.

Scientific Knowledge

Students' scientific knowledge was determined by their performance on tests in biology, chemistry and physics. These tests were organised by the school and

undertaken in March 2023 and were not designed specifically for this study. According to one teacher's description, the three tests covered science content that students had learned in the past, and all students completed the same papers. The total score of the three subjects was each student's final scientific knowledge score.

Students' Motivation in Scientific Creativity

Students' motivation in scientific creativity was measured by the Chinese version of the Creative Trait Motivation (CTM) scale. The original English version of the CTM had satisfactory reliability and validity (Taylor & Kaufman, 2021). The Chinese version was first translated from English by the first author and then given to two other researchers for minor corrections. The revised version was then given to two secondary school science teachers for review to ensure that secondary school students could understand the questions. Cronbach's alphas for the CTM-Science scale were as follows: overall scale ($\alpha=0.827$), intrinsic motivation ($\alpha=0.903$), extrinsic motivation ($\alpha=0.844$), and amotivation¹ ($\alpha=0.888$). The model fit for CTM was deemed acceptable ($X^2(167)=464$, $p<0.001$, CFI=0.852, TLI=0.831, SRMR=0.069, RMSEA=0.097). While the CFI and TLI values fell just below the recommended threshold of 0.90, the SRMR, with a value below 0.08, suggests an adequate fit based on established standardised cutoff criteria (Hu & Bentler, 1999) and in comparison to the results reported for the original version (Taylor & Kaufman, 2021). It is important to note that the relatively small sample size ($N=189<250$) may also contribute to the RMSEA exceeding the ideal cutoff. Despite these nuances, the overall fit indices, coupled with the contextual understanding of the small sample size, support the acceptability of the model fit.

Thinking Styles

Students' thinking styles, namely divergent thinking and convergent thinking, were measured by our 'Scientific Creativity Test for Upper Secondary School Students' (SCT-USSS). SCT-USSS builds on previous creativity and scientific creativity tests, with improvements in question design and scoring methods. In terms of question design, SCT-USSS consists of three tasks, each with four questions. For each task, the first and second questions are designed to assess students' divergent thinking, with the assessment indicators being fluency, flexibility and novelty; the third and fourth questions are designed to assess students' convergent thinking, with the assessment indicators being criticality, elaboration and logicity. The three tasks correspond to scientific knowledge in biology, chemistry and physics, and the four questions cover the process of asking scientific questions, proposing solutions, evaluating solutions and conducting experiments, as well as covering the 'unusual use' and 'imagination' question types frequently used in previous creativity tests. The

¹ Amotivation in this scale means that students have no motivation in scientific creativity, with specific items to determine this, including: "Although I do not see the benefit in it" and "Even though I do not have a good reason for doing so".

Table 3 Scientific creativity test for upper secondary school students

Subject	Task	Question type	Thinking style
Biology	Biodiversity and environment A: Chinese Sturgeon B: South China tiger	Q1: Ask science questions	Divergent thinking
		Q2: Propose solutions	
		Q3: Evaluate solutions	Convergent thinking
		Q4: Conduct experiment	
Chemistry	Chemistry and sustainable development A: Plastic bottles B: Plastic bags	Q5: Unusual uses	Divergent thinking
		Q6: Propose solutions	
		Q7: Evaluate solutions	Convergent thinking
		Q8: Conduct experiment	
Physics	Sound and Light A: Roadway noise B: Light pollution	Q9: Imagination	Divergent thinking
		Q10: Propose solutions	
		Q11: Evaluate solutions	Convergent thinking
		Q12: Conduct experiment	

Note. Q1-Q12 indicates the order in which each question appears in the test

three tasks were also based on the learning content students were studying, with issues they are likely to encounter in everyday life. In addition, two versions of the test, A and B, were designed to evaluate students' scientific creativity, with minor modifications to the content of the questions. Table 3 presents the specific design of questions. The 12 questions for version A of C-SCA and an example of one student's answer for task 1 can be found in the Electronic Supplementary Material.

Scoring Process The SCT-USSS consisted of three tasks, each with four questions, which needed to be graded according to the scoring criteria (see Table 4).

Scoring for Divergent Thinking The scoring for divergent thinking was based on three aspects: fluency, flexibility and novelty, assessed through the first and second questions of each task. The specific process was as follows:

First, fluency was determined by the number of scientific questions or solutions provided, with one mark awarded for each appropriate response. In this context, appropriateness referred to the direct relevance of students' responses to the question. For example, the second question asked, "If Chinese sturgeons haven't become extinct yet, what solutions do you have to protect them? Please list as many solutions as you can"; a response such as "*People should protect the Chinese sturgeon*" was not considered appropriate and did not receive any marks. At the same time, some strange responses also did not receive any marks, such as "*Consider the Chinese sturgeon as God*", along with any clearly spoofed responses. The scoring reflects that scientific creativity needs to be appropriate – products or ideas should be useful, adaptive and concerned with task constraints (Sternberg, 1999).

Second, scoring flexibility required all responses to be categorised. Categorisation allowed evaluation of a student's ability to think about a question from various perspectives. If a student's responses corresponded to several categories, they could score higher on flexibility, reflecting their ability to think flexibly about the

Table 4 The scoring criteria for the SCT-USSS

	Dimension	Scoring criteria
Divergent thinking	Fluency	Score of 1 point for each response
	Flexibility	Score of 1 point for each category of response
	Novelty	Score is 2 points if probability of each sub-category is less than 5%, 1 point if between 5 and 10%, and 0 points if greater than 10%; the final score is the sum of all sub-categories
Convergent thinking	Critique	1 point: only gives advantages of this solution, without disadvantages 2 points: gives both advantages and disadvantages of this solution, but without explanations 3 points: gives both advantages and disadvantages of this solution, and also explains the reasons for not choosing other solutions
	Elaboration	1 point: very brief or no elaboration 2 points: detailed elaboration 3 points: provides both text and illustrations to further explain the process and details of the experiment and solution
	Logicity	1 point: the whole solution and experiment has many scientific and logical errors 2 points: only a few parts of the solution and experiment are not scientific and logical, and the majority of them are reasonable 3 points: the solution and the experiment are very scientific and logical, and they exist in reality and have a realistic basis

question. This categorisation process involved three rounds and was carried out by the first author, another researcher and ChatGPT. To elaborate on the role of ChatGPT, it is an artificial intelligence model with a notable skill in ‘Function calling’, allowing developers to get structured data back from input text (Open AI, 2023). As demonstrated by Lo (2023) across various domains, ChatGPT has shown notable effectiveness in content analysis, such as economics and programming, but yields unsatisfactory results in mathematics. Therefore, during the categorisation process, it was imperative to continuously refine and modify the results generated by ChatGPT based on researchers’ suggestions until achieving a satisfactory outcome. In the context of this study, ChatGPT was chosen due to its proven ability to handle thematic categorisation tasks effectively. The model was employed to complement the manual thematic classification carried out by the first author. For example, for the first question in version A, a total of 350 appropriate responses were given by the 92 students. There were three steps to categorise these responses: 1) The first author took a thematic approach to classify the responses and identified six categories. 2) The 350 responses were fed into ChatGPT, asking it to undertake a thematic categorisation. Then, the first author adapted the categorisation to the answers given by ChatGPT to arrive at a revised categorisation. 3) This revised categorisation was passed to another researcher for review to ensure the validity of the categories, resulting in the final categorisation. An example of the categorisation for Q1 version A can be found in the Electronic Supplementary Material.

Third, scoring novelty required calculating the frequency of students’ responses. In this study, each question received an average of 344 responses; it was unrealistic to calculate the frequency for each response, and it was also difficult to determine the differences among many responses. One example arose in the case of the scientific use of plastic bottles. Some students said, “*Plastic bottles can be used for drinking water*”, and others, “*Plastic bottles can be used for domestic water*”. There are subtle differences between the two, but in both responses, plastic bottles were essentially used to hold water. Therefore, to make it easier to calculate the frequency of responses, the first author carried out a second categorisation of the students’ responses. The sub-categories represented further subdivisions of categories for flexibility, and the re-categorisation process also involved three rounds and was completed by the first author, another researcher and ChatGPT. An example of the sub-categories for Q1 version A can be found in the Electronic Supplementary Material. A lower frequency for a sub-category indicated that fewer students answered the question from that perspective. Therefore, this study not only assessed students’ ability to think flexibly from different perspectives but also determined whether these perspectives were substantively novel.

Scoring for Convergent Thinking The scoring for convergent thinking was based on critique, elaboration and logicity. The scores for the third question of each task were based on the critique scoring criteria, and the final score for the fourth question was the sum of the elaboration and logicity scores. The specific scoring criteria can be found in Table 4. Two scorers completed this part. After reading the scoring criteria, the two scorers first attempted to mark the results of ten samples. The two scorers’ marking results were then reviewed together, and any areas that seemed

unclear were discussed. After ensuring that both scorers had understood the scoring criteria in the same way, they began to work independently through the entire grading process. The final scores for convergent thinking for all samples were determined by the average of the two scorers.

Results

Reliability of the SCT-USSS

Internal consistency and inter-scorer scorer reliability were used to test the reliability of the SCT (Hu & Adey, 2002) and the C-SAT (Sak & Ayas, 2013). This study also used these two reliability tests for the SCT-USSS.

Internal Consistency

The two versions each contained 12 questions, with similar question types but different content, and they showed acceptable internal consistency (A: $\alpha=0.730$; B: $\alpha=0.653$), indicating that they examined a common trait: scientific creativity.

Internal consistency analysis was also conducted for two types of thinking. Table 5 shows the results. Both versions of convergent thinking showed an acceptable internal consistency (A: Cronbach's $\alpha=0.679$; B: Cronbach's $\alpha=0.681$). Version A's internal consistency for divergent thinking was also acceptable (Cronbach's $\alpha=0.600$), but version B's was only 0.521, which was questionable.

Q1, Q2, Q5, Q6, Q9 and Q10 were designed for assessing divergent thinking. To further see which question in divergent thinking disturbed version B's internal consistency, Cronbach's α coefficients were calculated for the five remaining questions when a particular question was removed (see Table 6). This process was also carried out on version A to see if the issue arose on the same question. The results indicated that after removing Q5 in B, Cronbach's α coefficient for the remaining five questions (0.554) was higher than for the original six questions (0.521). Similarly, after removing Q5 in version A, Cronbach's α coefficient for the remaining five questions (0.611) was slightly higher than for the original six questions (0.600). As a result, it was concluded that Q5 might need to be revised.

In addition, Pearson's correlation coefficients were calculated between several variables: Q1, Q2, Q5, Q6, Q9 and Q10, with the corrected divergent thinking scores. The correlation coefficients were calculated between Q3, Q4, Q7, Q8, Q11 and Q12 with the corrected convergent thinking scores (see Table 7). The corrected scores, as used here, were the total scores for the remaining questions when a particular question was removed. For example, in the version A, Pearson's r between Q1 and the corrected divergent thinking scores (Q2 + Q5 + Q6 + Q9 + Q10) was 0.417. Corrected scores were used to investigate the relationship between each question and the thinking scores while removing the potential bias introduced by that question. The results showed that all questions, except Q5, correlated significantly with the corrected corresponding thinking style scores. Therefore, it was again verified that Q5 needed to be modified.

Table 5 Internal consistency measures for SCT-USSS

Version	Thinking style	Indicator	Cronbach's α		
A	Divergent thinking	Fluency	0.680	0.600	0.730
		Flexibility	0.472		
		Novelty	0.371		
	Convergent thinking	Critique	0.662	0.679	
		Elaboration	0.659		
		Logicity	0.593		
B	Divergent thinking	Fluency	0.640	0.521	0.653
		Flexibility	0.397		
		Novelty	0.294		
	Convergent thinking	Critique	0.666	0.681	
		Elaboration	0.561		
		Logicity	0.453		

Table 6 Internal consistency of divergent thinking if a specific question is removed

Version	Thinking style	Question	Cronbach's α if specific question removed
A	Divergent thinking	Q1	0.521
		Q2	0.591
		Q5	0.611
		Q6	0.551
		Q9	0.517
		Q10	0.529
B	Divergent thinking	Q1	0.502
		Q2	0.434
		Q5	0.554
		Q6	0.442
		Q9	0.484
		Q10	0.422

Inter-Scorer Reliability

To ensure the reliability of the scoring system for convergent thinking (Q3, Q4, Q7, Q8, Q11, Q12), an assessment of the interpretability was conducted by independent scorers. The scorers included one individual who had no involvement in the research project and the first author. Scoring for 189 students was independently performed by both scorers, and the Pearson product-moment correlation coefficients between the two sets of scores are presented in Table 8. The correlations ranged from 0.814 to 0.919, with total convergent thinking question correlations of 0.908 and 0.903 for versions A and B, respectively. These results indicated that the scoring procedure demonstrated satisfactory agreement.

Table 7 Pearson's correlation coefficients between each question and corresponding thinking styles

Version	Question	Pearson's r with corrected divergent thinking scores	Pearson's r with corrected convergent thinking scores
A	Q1	0.417***	
	Q2	0.242*	
	Q5	0.187	
	Q6	0.351**	
	Q9	0.436***	
	Q10	0.408***	
	Q3		0.280**
	Q4		0.364***
	Q7		0.280**
	Q8		0.562***
	Q11		0.436***
	Q12		0.468***
B	Q1	0.212*	
	Q2	0.340**	
	Q5	0.115	
	Q6	0.326**	
	Q9	0.259*	
	Q10	0.355***	
	Q3		0.415***
	Q4		0.344**
	Q7		0.377***
	Q8		0.464***
	Q11		0.379***
	Q12		0.307**

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Validity of the SCT-USSS

Factor analysis is commonly utilised by researchers to evaluate the construct validity of their scientific creativity instruments (Hu & Adey, 2002; Sak & Ayas, 2013). Additionally, Hu and Adey (2002) employed face validity in their study, asking 35 science education researchers and teachers if the SCT items looked like they were assessing scientific creativity.

The design of the SCT-USSS drew upon four previously established creativity instruments, namely the TTCT (Torrance, 1966), the SCT (Hu & Adey, 2002), the C-SAT (Sak & Ayas, 2013) and another scientific creativity instrument for high school students (Xu, 2013). Empirical studies have demonstrated the ability of these instruments to measure creativity or scientific creativity effectively. In this study, the SCT-USSS integrated question types from these instruments, while also incorporating adaptations specific to the science learning content of Chinese upper secondary

Table 8 Inter-scorer correlations for the convergent thinking questions

Version	Q3	Q4	Q7	Q8	Q11	Q12	All
A	0.838***	0.899***	0.828***	0.855***	0.858***	0.919***	0.908***
B	0.843***	0.814***	0.847***	0.818***	0.902***	0.895***	0.903***

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

school students (as shown in Table 3). Consequently, based on the theoretical alignment, it can be suggested that the SCT-USSS has the potential to validly assess scientific creativity.

Before testing students, the designed instrument was given to two science education researchers, three PhD students in education and two Chinese science teachers for review; all of them indicated that the 12 questions were consistent with the science learning content taught to upper secondary school students. Importantly, they also concurred that the questions effectively assessed students' scientific creativity, allowing students to apply their creative thinking skills in providing responses. As a result, the instrument demonstrated good face validity, corroborating its appropriateness and suitability for evaluating scientific creativity among upper secondary school students.

After collecting the data, factor analysis with principal components was undertaken. When factor analysis was conducted on 12 questions simultaneously, the Kaiser–Meyer–Olkin (KMO) value was found to be 0.616 for version A and 0.590 for version B. Bartlett's Test of Sphericity reached statistical significance for both versions, supporting the factorability of the correlation matrix. However, the results showed that both versions generated five components instead of only one component, which might have been expected based on previous studies (Hu & Adey, 2002; Sak & Ayas, 2013). This discrepancy may be attributed to the complexity of the question design (as shown in Table 3), which encompassed various question types and involved divergent and convergent thinking, as well as three different subjects. In contrast, the SCT's seven questions did not involve variations in subjects or convergent thinking (Hu & Adey, 2002), while C-SAT only assessed students' divergent thinking (Sak & Ayas, 2013). In order to test the validity of the SCT-USSS, validity tests for divergent thinking and convergent thinking were conducted separately to see if the corresponding questions assessed two different thinking styles.

For divergent thinking, exploratory factor analysis was conducted for version A. The analysis revealed a KMO measure of sampling adequacy at 0.596, indicating its suitability for factor analysis. Initially, a two-factor solution was obtained, accounting for a cumulative variance of 52.3%. However, due to the inappropriate grouping of Q2 and Q5, which deviated from the study's design, a subsequent factor analysis was performed, limiting the factors to one. This approach assumed that all questions were assessing divergent thinking. The one-factor solution accounted for 34.3% of the variance. Notably, the factor loadings ranged from 0.327 to 0.710, indicating that all questions effectively measured divergent thinking. Table 9 provides more detailed findings.

Table 9 Factor analysis results for the divergent thinking questions

Question	Factor loadings of each question	
	Version A	Version B
Q1	0.681	0.393
Q2	0.420	0.667
Q5	0.327	0.201
Q6	0.595	0.691
Q9	0.710	0.509
Q10	0.673	0.691

For version B, exploratory factor analysis was also conducted. The analysis revealed a KMO measure of sampling adequacy at 0.611, confirming its suitability for factor analysis. Initially, a two-factor solution was obtained, explaining a cumulative variance of 49.0%. However, the rotated component matrix did not yield satisfactory factor outcomes. Subsequently, the number of factors was restricted to one, resulting in a one-factor solution that accounted for 30.9% of the variance. The factor loadings ranged from 0.201 to 0.691, as presented in Table 9. These findings also suggest the need to modify Q5 to ensure validity.

For convergent thinking, Q3, Q7 and Q11 were designed to evaluate critique, while Q4, Q8 and Q12 focused on elaboration and logicity. Thus, the individual scores for each task needed to be combined to derive a total convergent thinking score. Q3 and Q4 contributed to the biology task's convergent thinking score, Q7 and Q8 to the chemistry task, and Q11 and Q12 to the physics task. Subsequently, these scores for the three tasks were subjected to factor analysis to assess the effectiveness of the questions in measuring convergent thinking.

In the case of version A, exploratory factor analysis was conducted, revealing a KMO measure of sampling adequacy of 0.597, indicating its suitability for factor analysis. A one-factor solution was obtained, which could not be rotated. The results, presented in Table 10, demonstrated substantial loadings (ranging from 0.694 to 0.847) of the three tasks onto the single factor, explaining 58.8% of the total variance. Similarly, for version B, exploratory factor analysis was performed, yielding a KMO measure of sampling adequacy of 0.554, confirming its appropriateness for factor analysis. A one-factor solution was obtained, which also could not be rotated. The three tasks exhibited significant loadings (ranging from 0.593 to 0.831) onto the single factor, explaining 53.3% of the total variance. These results indicated that both versions' convergent thinking questions possess good construct validity, measuring a single factor: convergent thinking.

Modifications to the SCT-USSS

Based on the SCT-USSS reliability and validity analysis, it was recommended to modify Q5. Q5, known as the 'Unusual Use' question, was originally introduced by Torrance (1966) in TTCT and later adapted by Hu and Adey (2002) in the

Table 10 Factor analysis results for the convergent thinking

Task	Question	Factor loadings of each question	
		Version A	Version B
Biology	Q3	0.694	0.745
	Q4		
Chemistry	Q7	0.847	0.831
	Q8		
Physics	Q11	0.752	0.593
	Q12		

Table 11 Gender differences in intrinsic motivation, extrinsic motivation and amotivation

Variable	N	Intrinsic motivation		Extrinsic motivation		Amotivation		
		M ± SD	t	M ± SD	t	M ± SD	t	
Gender	Males	115	39.781 ± 7.177	-0.030	15.544 ± 5.577	-1.012	10.375 ± 4.467	1.243
	Females	74	39.813 ± 6.849		16.371 ± 5.338		9.585 ± 3.929	

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

SCT. However, the inclusion of the term ‘scientific’ in the SCT seemed to distract our participants in completing the SCT-USSS. More specifically, participants’ responses to Q5 were analysed, with 92 students providing 264 valid responses in Version A and 97 students providing 303 valid responses in Version B. In Version A, 49 responses indicated that plastic bottles could be utilised for various scientific research purposes, such as “*Storing reagents that are non-reactive with plastic*”, “*Creating biological models*”, or “*Conducting physical experiments related to pressure*”. In Version B, 76 responses suggested that plastic bags could serve multiple scientific research functions, including “*Acting as protective sleeves for instruments*”, “*Observation bags*”, or “*Useful tools for studying thermoplastics*”. It may be that the term ‘scientific’ in the question misled some students who may have thought they should use the plastic bottles or bags in some scientific experiments, potentially constraining their creativity. For this reason, Q5 was changed to something closer to the TTCT and SCT, namely “Please write down as many uses for plastic bottles (plastic bags) as possible. Do not be limited by their size. You may use as many of them as you like”.

Descriptive Statistics of the C-SCA

Table 11 shows no significant difference in the scores of intrinsic, extrinsic and amotivation scores between genders. In both versions of the thinking style test, there was no significant difference in the divergent thinking scores between genders. There was

Table 12 Gender differences in overall thinking style, divergent thinking and convergent thinking

Variable	N	Overall thinking style		Divergent thinking		Convergent thinking	
		M±SD	t	M±SD	t	M±SD	t
Version A	Males	60.809±13.050	-1.006	45.691±11.114	-0.481	15.118±3.781	-2.303*
	Females	63.757±14.818		46.919±13.223		16.838±3.064	
Version B	Males	71.283±15.948	-0.588	55.000±14.375	-0.381	16.283±3.501	-1.056
	Females	73.054±11.463		56.054±11.115		17.000±2.776	

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13 Correlational matrix for measured indicators ($N = 189$)

	1	2	3	4
1 Scientific knowledge	1			
2 Intrinsic motivation	0.183*	1		
3 Extrinsic motivation	0.073	0.256***	1	
4 A motivation	0.062	-0.130	0.075	1

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14 Correlational matrix for measured indicators in version A ($N=92$)

	1	2	3	4	5	6
1 Scientific knowledge	1					
2 Intrinsic motivation	0.243*	1				
3 Extrinsic motivation	0.140	0.315**	1			
4 Amotivation	0.073	-0.070	0.088	1		
5 Divergent thinking	-0.075	-0.008	-0.003	-0.062	1	
6 Convergent thinking	-0.239*	-0.060	-0.027	-0.021	0.399***	1

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15 Correlational matrix for measured indicators in version B ($N=97$)

	1	2	3	4	5	6
1 Scientific knowledge	1					
2 Intrinsic motivation	0.072	1				
3 Extrinsic motivation	-0.028	0.164	1			
4 Amotivation	0.041	-0.227*	0.055	1		
5 Divergent thinking	0.099	-0.015	0.149	0.056	1	
6 Convergent thinking	0.302**	0.052	0.088	0.002	0.259*	1

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a significant difference in the convergent thinking scores between genders in version A ($t = -2.303$, $p < 0.05$), with females (16.838) scoring slightly higher than males (15.118), but this difference was not observed in version B (see Table 12).

Correlation Analysis of the Three Dimensions of the C-SCA

Table 13 presents the Pearson correlations between scientific knowledge and motivation in scientific creativity for all the samples. Scientific knowledge exhibited a significant and positive correlation with intrinsic motivation ($r = 0.183$). Intrinsic motivation displayed a significant and positive correlation with extrinsic motivation ($r = 0.256$).

Table 14 shows the results of the Pearson correlation analysis for Version A. Scientific knowledge significantly and positively correlated with intrinsic motivation ($r = 0.243$). Intrinsic motivation showed a significant positive correlation with extrinsic motivation ($r = 0.315$). Convergent thinking exhibited a significant negative correlation with scientific knowledge ($r = -0.239$) but a significant positive correlation with divergent thinking (0.399).

Table 15 presents the results of the Pearson correlation analysis for Version B. Convergent thinking showed a significant positive correlation with scientific knowledge ($r = 0.302$), contradicting the results of Version A, but showed a significant positive correlation with divergent thinking ($r = 0.259$), consistent with version A. In addition, amotivation demonstrated a significant and negative correlation with intrinsic motivation ($r = -0.227$).

Discussion

Scientific creativity has remained a central focus and challenge for researchers, given its significance as a key competency in the twenty-first century. Many existing instruments mostly focus on one dimension, such as the most widely used – divergent thinking (Reiter-Palmon et al., 2019) – or students' self-reported achievement (Lubart et al., 2022) or motivation in scientific creativity (Taylor & Kaufman, 2021), rather than taking a comprehensive view of this complex ability. This study employed a comprehensive framework for assessing scientific creativity, encompassing three key dimensions: scientific knowledge, motivation in scientific creativity, and thinking styles.

The three dimensions of the C-SCA were evaluated using distinct measures. The Chinese version of the CTM demonstrated satisfactory reliability and validity. The SCT-USSS also proved to be a reliable and valid measure for assessing students' thinking styles. The correlation analysis revealed a positive relationship between divergent thinking and convergent thinking, aligning with findings from other studies (Cromptley, 2006; Zhu et al., 2019). When confronted with scientific tasks presented in the C-SCA, students demonstrated the ability to generate a greater number and variety of novel questions and plans through divergent thinking, while utilising convergent thinking to critically evaluate these ideas and articulate their plans in a clear and logical manner.

The findings aligned with previous research suggesting that students' creative abilities are not necessarily linked to their academic performance (Batey & Furnham, 2006). On the other hand, it is possible that because the assessment of scientific knowledge relied on general school examinations rather than targeting specific knowledge in the scientific creativity tasks, this explains why any associations are weak. Moreover, the descriptive results indicated that the average scores for amotivation ranged from 'corresponds a little' to 'corresponds moderately', suggesting that Chinese upper secondary students may not perceive a necessity to demonstrate creativity in the field of science, as China's exam-centric environment places emphasis on providing correct answers rather than novel but unexpected ones in examinations (Wang & Greenwood, 2013). Consequently, this attitude may disrupt the interplay among scientific knowledge, motivation in scientific creativity and thinking styles. Additionally, it is essential to note that each version of the assessment was employed with a limited sample size, and that substantial variations in academic performance between students may impact the correlations between scientific knowledge and other variables.

The assessment of creativity, particularly in relation to divergent thinking, has always posed challenges for researchers due to the substantial time required, which acts as a significant barrier to further exploration in this area (Reiter-Palmon et al., 2019). For the tasks involving divergent thinking in this study, each question received an average of 344 responses, and the researchers were required to tally each student's valid responses, categorise all the responses, and match each student's responses to the corresponding categories one by one, and then calculate the frequency of each response in the total sample. Throughout

this process, the categorisation of student responses was primarily determined by the researchers, introducing the possibility of subjective factors influencing the results, as also highlighted by Reiter-Palmon et al. (2019). Furthermore, the level of categorisation, whether detailed or broad, may also impact the final outcomes. For the first issue, this study utilised ChatGPT to help ensure a rigorous categorisation process. However, the second issue was only identified in this study during the practical implementation of scientific creativity scoring; no substantial evidence was obtained to confirm whether the extent of categorisation affects the results, thereby necessitating further research to address it. Additionally, we posit that secondary students' responses exhibit a relatively constrained range, which is predominantly circumscribed by the scope of their educational curriculum and aligns with established scientific frameworks, as research supports that domain-specific knowledge is foundational to scientific creativity (Huang et al., 2017; Sun et al., 2020). By compiling an answer repository through extensive sampling and subjecting it to a series of processing steps, we can assign a corresponding score to each response. This approach suggests that future users of the C-SCA can efficiently score responses based on predetermined criteria, eliminating the need for repetitive assessments of flexibility and novelty.

Conclusions

The aim of this study was to design a comprehensive approach to assessing secondary students' scientific creativity, specifically upper secondary students in China. Building upon established theories of creativity, we developed a theoretical framework for a Comprehensive Scientific Creativity Assessment (C-SCA). Its primary objective was to evaluate students' scientific knowledge, motivation in scientific creativity, and thinking styles. The assessment of these three dimensions was conducted using separate approaches, including school science tests, the Chinese version of the CTM, and the newly designed SCT-USSS. The empirical study demonstrated the satisfactory reliability and validity of the C-SCA, and the A and B versions of the SCT-USSS can assist educators and researchers in evaluating the effectiveness of interventions on scientific creativity. Measuring creativity has always posed a complex and time-consuming challenge, which we also encountered in this study. Despite incorporating multiple researchers and utilising new technology, ChatGPT, to help ensure objectivity in the scoring process, there is still a need to explore more rigorous and efficient methods of assessing scientific creativity. In the future, we believe that the integration of artificial intelligence and natural language recognition technology could play a role in enhancing the scoring process, leading to advancements in scientific creativity assessment.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10763-024-10469-z>.

Acknowledgements The first author received financial support for the research from University College London.

Data Availability The data that support the findings of this study are available on request from the corresponding author.

Declarations

Disclosure Statement Ethical approval was obtained from University College London.

Conflict of Interest The authors declare that there is no conflict of interest.

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

References

- Acar, S., & Runco, M. A. (2019). Divergent thinking: New methods, recent research, and extended theory. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 153–158. <https://doi.org/10.1037/aca0000231>
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*(5), 997–1013. <https://doi.org/10.1037/0022-3514.43.5.997>
- Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Westview Press.
- Aschauer, W., Haim, K., & Weber, C. (2022). A contribution to scientific creativity: A validation study measuring divergent problem solving ability. *Creativity Research Journal*, *34*(2), 195–212. <https://doi.org/10.1080/10400419.2021.1968656>
- Ayas, M. B., & Sak, U. (2014). Objective measure of scientific creativity: Psychometric validity of the Creative Scientific Ability Test. *Thinking Skills and Creativity*, *13*, 195–205. <https://doi.org/10.1016/j.tsc.2014.06.001>
- Baer, J. (1994). Divergent thinking is not a general trait: A multidomain training experiment. *Creativity Research Journal*, *7*(1), 35–46. <https://doi.org/10.1080/10400419409534507>
- 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. <https://doi.org/10.3200/MONO.132.4.355-430>
- Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*, *17*(1), 37–50. https://doi.org/10.1207/s15326934crj1701_4
- Cropley, A. J. (1967). *Creativity*. Longman.
- Cropley, A. J. (2006). In Praise of Convergent Thinking. *Creativity Research Journal*, *18*(3), 391–404. https://doi.org/10.1207/s15326934crj1803_13
- Ekvall, G. (1996). Organizational climate for creativity and innovation. *European Journal of Work and Organizational Psychology*, *5*(1), 105–123. <https://doi.org/10.1080/13594329608414845>
- Gough, H. G., & Heilbrun, A. B. (1965). *The adjective check list manual*. Consulting Psychologists Press. <https://doi.org/10.1037/t02310-000>
- Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*(9), 444–454. <https://doi.org/10.1037/h0063487>
- Guilford, J. P. (1968). *Creativity, intelligence, and their educational implications*. Robert R. Knapp.
- Hadzigeorgiou, Y., Fokialis, P., & Kabouropoulou, M. (2012). Thinking about creativity in science education. *Creative Education*, *3*(5), 603–611. <https://doi.org/10.4236/ce.2012.35089>

- Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, 61(1), 569–598. <https://doi.org/10.1146/annurev.psych.093008.100416>
- Hennessey, B. A., Amabile, T. M., & Mueller, J. S. (2011). Consensual assessment. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of Creativity (Second Edition)* (pp. 253–260). Academic Press. <https://doi.org/10.1016/B978-0-12-375038-9.00046-7>
- Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education*, 24(4), 389–403. <https://doi.org/10.1080/09500690110098912>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, C. F., & Wang, K. C. (2019). Comparative analysis of different creativity tests for the prediction of students' scientific creativity. *Creativity Research Journal*, 31(4), 443–447. <https://doi.org/10.1080/10400419.2019.1684116>
- Huang, P. S., Peng, S. L., Chen, H. C., Tseng, L. C., & Hsu, L. C. (2017). The relative influences of domain knowledge and domain-general divergent thinking on scientific creativity and mathematical creativity. *Thinking Skills and Creativity*, 25, 1–9. <https://doi.org/10.1016/j.tsc.2017.06.001>
- Isaksen, S. G., Lauer, K. J., Ekvall, G., & Britz, A. (2001). Perceptions of the best and worst climates for creativity: Preliminary validation evidence for the situational outlook questionnaire. *Creativity Research Journal*, 13(2), 171–184. https://doi.org/10.1207/S15326934CRJ1302_5
- Kaufman, J. C. (2012). Counting the muses: Development of the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, 6(4), 298–308. <https://doi.org/10.1037/a0029751>
- Kaufman, J. C., & Baer, J. (2004). Sure, I'm creative—but not in mathematics!: Self-reported creativity in diverse domains. *Empirical Studies of the Arts*, 22(2), 143–155. <https://doi.org/10.2190/26HQ-VHE8-GTLN-BJIM>
- Kind, P. M., & Kind, V. (2007). Creativity in science education: Perspectives and challenges for developing school science. *Studies in Science Education*, 43(1), 1–37. <https://doi.org/10.1080/03057260708560225>
- Koke, L. C., & Vernon, P. A. (2003). The Sternberg Triarchic Abilities Test (STAT) as a measure of academic achievement and general intelligence. *Personality and Individual Differences*, 35(8), 1803–1807. [https://doi.org/10.1016/S0191-8869\(03\)00005-9](https://doi.org/10.1016/S0191-8869(03)00005-9)
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4), 2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>
- Kupers, E., Lehmann-Wermser, A., McPherson, G., & van Geert, P. (2019). Children's creativity: A theoretical framework and systematic review. *Review of Educational Research*, 89(1), 93–124. <https://doi.org/10.3102/0034654318815707>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410–425. <https://doi.org/10.3390/educsci13040410>
- Lubart, T., Kharkhurin, A. V., Corazza, G. E., Besançon, M., Yagolkovskiy, S. R., & Sak, U. (2022). Creative potential in science: Conceptual and measurement issues. *Frontiers in Psychology*, 13, 750224. <https://doi.org/10.3389/fpsyg.2022.750224>
- Ludwig, A. M. (1992). The creative achievement scale. *Creativity Research Journal*, 5(2), 109–119. <https://doi.org/10.1080/10400419209534427>
- Lunke, K., & Meier, B. (2016). Disentangling the impact of artistic creativity on creative thinking, working memory, attention, and intelligence: Evidence for domain-specific relationships with a new self-report questionnaire. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01089>
- Mann, E. L. (2009). The search for mathematical creativity: Identifying creative potential in middle school students. *Creativity Research Journal*, 21(4), 338–348. <https://doi.org/10.1080/10400410903297402>
- McCormick, R. (1997). Conceptual and procedural knowledge. *International Journal of Technology and Design Education*, 7(1), 141–159. <https://doi.org/10.1023/A:1008819912213>
- Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69(3), 220–232. <https://doi.org/10.1037/h0048850>
- Millar, R., Lubben, F., Got, R., & Duggan, S. (1994). Investigating in the school science laboratory: Conceptual and procedural knowledge and their influence on performance. *Research Papers in Education*, 9(2), 207–248. <https://doi.org/10.1080/0267152940090205>
- Ministry of Education of the People's Republic of China (MOE). (2022). *Compulsory Education Science Curriculum Standards*. Beijing Normal University Press [in Chinese].

- Nijstad, B. A., De Dreu, C. K. W., 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. <https://doi.org/10.1080/10463281003765323>
- Open AI. (2023). *Function calling and other API updates*. Retrieved December 27, 2023, from <https://openai.com/blog/function-calling-and-other-api-updates>
- Paek, S. H., & Runco, M. A. (2017). Dealing with the criterion problem by measuring the quality and quantity of creative activity and accomplishment. *Creativity Research Journal*, 29(2), 167–173. <https://doi.org/10.1080/10400419.2017.1304078>
- Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, 13(2), 144–152. <https://doi.org/10.1037/aca0000227>
- Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, 42(7), 305–310. <https://www.jstor.org/stable/20342603>
- 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), 476–485. <https://doi.org/10.1037/0022-3514.54.3.476>
- Sak, U., & Ayas, M. (2013). Creative Scientific Ability Test (C-SAT): A new measure of scientific creativity. *Psychological Test and Assessment Modeling*, 55(3), 316–329. Retrieved June 20, 2023, from <https://www.proquest.com/scholarly-journals/creative-scientific-ability-test-c-sat-new/docview/1468914888/se-2>.
- Shin, W., & Park, J. (2021). Developing a list of behavioral characteristics of creative physicists during their growth period. *International Journal of Science and Mathematics Education*, 19, 701–725. <https://doi.org/10.1007/s10763-020-10082-w>
- Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139165358>
- Sternberg, R. J. (1988). A three-facet model of creativity. In R. J. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives* (pp. 125–147). Cambridge University Press.
- Sternberg, R. J. (Ed.). (1999). *Handbook of creativity*. Cambridge University Press.
- Sternberg, R. J., & Lubart, T. I. (1992). Buy low and sell high: An investment approach to creativity. *Current Directions in Psychological Science*, 1(1), 1–5. <https://doi.org/10.1111/j.1467-8721.1992.tb00002.x>
- Stroup, D. F., & Thacker, S. B. (2007). Epidemiology and education: Using public health for teaching mathematics and science. *Public Health Reports*, 122(3), 283–291. <https://doi.org/10.1177/003335490712200301>
- Sun, M., Wang, M., & Wegerif, R. (2020). Effects of divergent thinking training on students' scientific creativity: The impact of individual creative potential and domain knowledge. *Thinking Skills and Creativity*, 37, 100682. <https://doi.org/10.1016/j.tsc.2020.100682>
- Taylor, C. L., & Kaufman, J. C. (2021). The creative trait motivation scales. *Thinking Skills and Creativity*, 39, 100763. <https://doi.org/10.1016/j.tsc.2020.100763>
- The State Council of the People's Republic of China. (2019). *Opinions on further promoting the spirit of scientists and strengthening ethical and academic conduct in the context of Chinese culture*. Author. Retrieved Jun. 20, 2023, from http://www.gov.cn/zhengce/2019-06/11/content_5399239.htm. [in Chinese].
- Torrance, E. P. (1966). *Torrance tests of creative thinking—norms technical manual research edition—verbal tests, forms A and B—figural tests, forms A and B*. Personnel Press.
- Tseng, H. M., & Liu, F. C. (2011). Assessing the Climate for Creativity (KEYS): Confirmatory factor analysis and psychometric examination of a Taiwan version. *International Journal of Selection and Assessment*, 19(4), 438–441. <https://doi.org/10.1111/j.1468-2389.2011.00572.x>
- Velayutham, S., Aldridge, J. M., & Fraser, B. (2012). Gender differences in student motivation and self-regulation in science learning: A multi-group structural equation modeling analysis. *International Journal of Science and Mathematics Education*, 10, 1347–1368. <https://doi.org/10.1007/s10763-012-9339-y>
- Wang, B., & Greenwood, K. M. (2013). Chinese students' perceptions of their creativity and their perceptions of Western students' creativity. *Educational Psychology*, 33(5), 628–643. <https://doi.org/10.1080/01443410.2013.826345>
- Webb, M. E., Little, D. R., Cropper, S. J., & Roze, K. (2017). The contributions of convergent thinking, divergent thinking, and schizotypy to solving insight and non-insight problems. *Thinking & Reasoning*, 23(3), 235–258. <https://doi.org/10.1080/13546783.2017.1295105>

- Wu, F., Jiang, Y., Liu, D., Konorova, E., & Yang, X. (2022). The role of perceived teacher and peer relationships in adolescent students' academic motivation and educational outcomes. *Educational Psychology, 42*(4), 439–458. <https://doi.org/10.1080/01443410.2022.2042488>
- Xu, X. (2013). *The construct and type analysis on scientific creativity of High-school students* [Unpublished master dissertation]. Beijing Normal University. (in Chinese).
- Zhu, W., Shang, S., Jiang, W., Pei, M., & Su, Y. (2019). Convergent thinking moderates the relationship between divergent thinking and scientific creativity. *Creativity Research Journal, 31*(3), 320–328. <https://doi.org/10.1080/10400419.2019.1641685>