A Psychometric Validation of a Scientific Creativity Measurement Instrument among Year 5 Pupils in Malaysian Primary Schools

Wan Nur Hafizah Wan Hussain, Hidayah Mohd Fadzil and Edy Hafizan Mohd Shahali

Abstract – **Scientific creativity plays a central role in primary science education by nurturing pupils' ability to imagine, innovate, and solve problems using scientific concepts. However, valid and reliable instruments to assess scientific creativity in the Malaysian primary context remain limited. This study aimed to validate a scientific creativity measurement instrument designed for Year 5 pupils. The instrument consists of four open-ended performance tasks that assess creativity through idea generation, design, scientific reasoning, and creative scientific communication. Content Validity Index (CVI) and Face Validity Index (FVI) were used to evaluate expert and respondent agreement. A pilot study was conducted, and pupils' responses were scored using a five-domain analytic rubric. Two independent raters scored all responses, and reliability was examined using the Intraclass Correlation Coefficient (ICC). Results showed high CVI and FVI values. ICC analysis indicated strong agreement between raters, exceeding accepted thresholds for inter-rater reliability. The findings suggest that the instrument has sound psychometric quality and is suitable for assessing scientific creativity among Year 5 pupils in Malaysian primary schools.**

Keywords – **Scientific creativity; instrument validation; primary science education; Year 5 pupils; content validity; face validity; Intraclass Correlation Coefficient (ICC); rubric scoring; psychometric properties; Malaysia**

I. INTRODUCTION

Scientific creativity is increasingly recognised as a fundamental competency in twenty-first century science education, particularly in preparing young learners to think critically, solve real-world problems, and innovate using scientific knowledge (Guo & Woulfin, 2016; Kennedy & Sundberg, 2025; Liu et al., 2026). In the Malaysian context, the Primary School Standard Curriculum (KSSR) emphasises higher-order thinking skills, creative expression, and inquiry-based learning as key outcomes of science instruction (Ismail et al., 2017; Shahali et al., 2017). Despite these curricular aspirations, the measurement of scientific creativity at the primary level remains limited. Most school-based assessments continue to focus on factual recall and procedural tasks rather than creative idea generation or scientific imagination (Fernandez et al., 2025; van Hooijdonk et al., 2020).

Wan Nur Hafizah Wan Hussain, Universiti Malaya, Kuala Lumpur (Email address: 22098238@siswa.um.edu.my).
Hidayah Mohd Fadzil, Universiti Malaya, Kuala Lumpur (Email address:hidayahfadzil@um.edu.my).
Edy Hafizan Mohd Shahali, Universiti Malaya, Kuala Lumpur (Email address:edyhafizan@um.edu.my).

Scientific creativity is generally defined as the ability to generate novel and useful ideas based on scientific concepts and reasoning(Dow, 2022; Shi et al., 2017). Scholars argue that creativity in science is demonstrated through multiple forms, including problem-solving, innovative design, imaginative scientific storytelling, and the integration of scientific principles in real-life scenarios (Chen et al., 2023; Pinar et al., 2025). However, empirical tools to measure these forms of creativity especially among younger pupils remain scarce in Southeast Asian educational systems (Nguyen & Tran, 2025; S. Xu et al., 2025). Most existing instruments were developed in Western contexts and may not fully reflect the linguistic, cultural, and curricular characteristics of Malaysian primary schools (Firdaus et al., 2025; Prahani et al., 2021; Soh, 2015; S. Xu et al., 2025).

Therefore, the development of a valid and reliable scientific creativity instrument is essential to support research, classroom-based assessment, and STEM innovation programmes(Firdaus et al., 2025; O. Hong et al., 2022; Latulipe & Terry, 2008). Valid measurement tools enable teachers to identify pupils' creative strengths, design targeted activities, and monitor learning outcomes aligned with curriculum standards(Ayas & Sak, 2014; Heller, 2007; Oksu Hong et al., 2022; Hu & Adey, 2002; Liu & Lin, 2014; Roth et al., 2022). In psychometric studies, the validation of assessment instruments typically involves evaluating content validity, face validity, construct clarity, and scoring reliability (Almanasreh et al., 2019; Clark & Watson, 2019; Cook & Beckman, 2006; Saw et al., 2025). For open-ended creativity assessments, reliability is often examined through inter-rater scoring consistency using indices such as the Intraclass Correlation Coefficient (ICC), which is considered a robust indicator of rating agreement (Dumas et al., 2023; Lewis et al., 2011).

Given the need for valid scientific creativity assessment at the primary level, this study aims to validate a scientific creativity measurement instrument designed for Year 5 pupils in Malaysian primary schools. The instrument consists of four open-ended tasks that reflect real-world scientific scenarios, creative design, and scientific communication. Content Validity Index (CVI), Face Validity Index (FVI), and inter-rater reliability using ICC were employed to evaluate the psychometric quality of the instrument. The findings provide evidence of validity and reliability, supporting the suitability of the instrument for educational and research applications in the Malaysian context.

II. PROBLEM STATEMENT

Scientific creativity is increasingly recognised as an essential outcome of science education, as it enables learners to generate innovative ideas, solve authentic problems, and apply scientific concepts in meaningful ways (Garrett, 1987; Kind & Kind, 2007; Taylor et al., 2008; Xu & Fan, 2025). In Malaysia, the Primary School Standard Curriculum (KSSR) highlights creativity, inquiry, and higher-order thinking as central goals of science learning (Foo, 2025; Mat et al., 2025). However, classroom assessment practices in primary schools continue to focus predominantly on factual recall and procedural knowledge, rather than on creative scientific thinking (Araceli Ruiz-Primo & Furtak, 2006; Klassen, 2006; Osborne, 2013; Rojas-Drummond et al., 1998; Schneider & Plasman, 2011). As a result, teachers often lack valid tools to measure pupils' ability to innovate, reason scientifically, and create original scientific ideas (Adams & Wieman, 2011; DeHaan, 2009; Henze et al., 2009; Liu & Lin, 2014). Although several scientific creativity instruments exist internationally, many were developed in Western contexts and are not fully aligned with the cultural, linguistic, or curriculum requirements of Malaysian primary pupils (Thomas & Watters, 2015; SY Xu et al., 2025). Furthermore, most available assessment tools are designed for secondary school or gifted learners, leaving a gap in measurement at the primary level (Suradin et al., 2025). Without appropriate instruments, efforts to cultivate scientific creativity through STEM programmes, inquiry-based strategies, or creative problem-solving tasks cannot be effectively evaluated or researched. This gap highlights the need for a validated and reliable instrument that measures scientific creativity among Year 5 pupils in Malaysian primary schools. Developing such an instrument requires systematic validation, including content and face validity as well as inter-rater reliability to ensure accurate scoring of open-ended responses (Hossan et al., 2025). Addressing this need will provide a credible assessment tool for teachers and researchers, support curriculum implementation, and contribute to empirical knowledge on scientific creativity in Malaysia.

III. LITERATURE REVIEW

Scientific Creativity in Science Education

Scientific creativity refers to the ability to generate original and useful ideas grounded in scientific knowledge and reasoning (Hu & Adey, 2002). Unlike general creativity, scientific creativity requires imagination combined with scientific concepts, logical explanation, and practical feasibility (Wang et al., 2014). Scholars generally agree that scientific creativity can be expressed through designing new tools, solving scientific problems, imagining scientific scenarios, or producing creative scientific stories (DeHaan, 2009; Smyrnaiou et al., 2020). At the primary school level, scientific creativity is often observed through drawings, inventions, experiments, and creative explanations of scientific phenomena (Oh, 2022; Poddiakov, 2023). Such activities allow pupils to apply inquiry skills and scientific reasoning in innovative ways, making scientific creativity an important element of twenty-first century learning.

Scientific Creativity in Malaysian Primary Schools

In Malaysia, the Primary School Standard Curriculum (KSSR) emphasises creativity and higher-order thinking as desired learning outcomes, particularly in science (Mat Noor, 2022). However, existing classroom assessments primarily measure factual recall and procedural knowledge rather than creative scientific thinking (Osborne, 2013; Stiggins et al., 1989). Teachers often report difficulty in assessing creativity due to the lack of structured instruments and scoring guidelines (Hernández-Ramos & Araya, 2025; Massy et al., 2026). Consequently, pupil's creative scientific potential may be under-represented in current assessment practices(Juriševič & Žerak, 2024; Zbainos & Tziona, 2019). To support curriculum goals, a measurement tool that captures idea generation, problem-solving, and application of scientific concepts is necessary, especially at the upper primary level where cognitive development supports creative expression(Rosen et al., 2020; Yildiz & Guler Yildiz, 2021).

Existing Scientific Creativity Instruments

Several scientific creativity tests have been developed internationally, such as the Scientific Creative Thinking Test (Hu & Adey, 2002) and the Scientific Attitude and Creativity Scale (Koç & Büyük, 2021). While these instruments are well-established, many were designed for secondary or gifted students and are rooted in Western educational contexts, which may not fully align with Malaysian linguistic, cultural, and curricular needs (Cevher et al., 2014; Karademir, 2016; Kim et al., 2016). In addition, many standardized tests rely on closed-ended or forced-choice items, which are less suitable for capturing creative and original ideas(Davis, 1989; Dollinger, 2011; Noddings, 2013). For primary pupils, open-ended performance tasks such as drawing inventions or imagining scientific scenarios provide a more authentic representation of creativity (DeHaan, 2009; Oliveira et al., 2021; Skjelstad Fredagsvik, 2022). However, open-ended assessments require scoring rubrics and strong inter-rater reliability to ensure fairness and objectivity(Bahar & June Maker, 2025; Lewis et al., 2011).

Validity and Reliability in Creativity Measurement

To ensure that an instrument accurately measures scientific creativity, psychometric evidence is essential. Content validity ensures that items represent the intended construct and are appropriate for the target population (Almanasreh et al., 2022; Mastaglia et al., 2003; Sireci, 1998). Face validity confirms that pupils understand the instructions and can respond meaningfully(Yusoff, 2019). For instruments involving open-ended responses, inter-rater reliability is a critical measure of scoring consistency(Jirschitzka et al., 2017; Lim et al., 2012). The

Intraclass Correlation Coefficient (ICC) is widely recommended for assessing the reliability of rubric-scored data due to its sensitivity to absolute agreement between independent raters (Akdeniz & Bangir, 2025; Brookhart & Chen, 2015). Strong ICC values indicate that the scoring rubric is clear and consistently applied, making it suitable for research and classroom assessment(Hansson et al., 2014; Jonsson & Svingby, 2007).

Gap in the Literature

Although scientific creativity has been extensively studied in Western contexts, empirical research on assessment tools for Malaysian primary pupils remains limited (Ambo et al., 2019; Omar et al., 2017; Ramly et al., 2022; Siew et al., 2014; Tan et al., 2021). Most available instruments are not validated for Malaysian cultural and curriculum contexts, and few studies have focused on Year 5 pupils. As a result, teachers lack a valid and reliable tool to assess scientific creativity and evaluate learning outcomes aligned with KSSR aspirations. This creates a clear need for an instrument that is age-appropriate, culturally relevant, and psychometrically sound. Developing and validating such an instrument will help strengthen classroom assessment practices, support creative STEM learning, and contribute to the national education agenda.

IV. RESEARCH QUESTIONS

1. To what extent does the scientific creativity measurement instrument demonstrate acceptable content validity based on expert evaluation?
2. Does the scientific creativity measurement instrument show adequate face validity based on feedback from Year 5 pupils in Malaysian primary schools?
3. Does the scientific creativity instrument demonstrate acceptable inter-rater reliability when scored using an analytic rubric and analysed through the Intraclass Correlation Coefficient (ICC)?

V. METHODOLOGY

This study employed a quantitative validation design to examine the psychometric properties of a scientific creativity instrument developed for Year 5 pupils in Malaysian primary schools. The instrument consisted of four open-ended tasks written in Bahasa Malaysia and designed to measure three dimensions of scientific creativity, namely traits, process, and product. The instrument was adapted from established scientific creativity measures developed by Hu and Adey (2002) and Siew et al. (2015). Elements from both instruments were integrated, modified, and contextualised to Malaysian classroom settings while preserving the original dimensions of scientific creativity. Validation procedures included analysis of the Content Validity Index (CVI), Face Validity Index (FVI), and inter-rater reliability using the Intraclass Correlation Coefficient (ICC)(Almanasreh et al., 2019).

Three groups of participants were involved in the validation process. First, eight experts in science education and educational measurement evaluated content relevance, clarity, age appropriateness, and scientific suitability for each task. Second, four language experts reviewed the instrument for vocabulary accuracy, sentence structure, and readability to ensure linguistic suitability for primary pupils. Third, a pilot study was conducted with 150 Year 5 pupils from national primary schools, a sample size that exceeded recommended minimums for pilot validation studies.

The instrument comprised four open-ended scientific creativity tasks. In Question 1, pupils were asked to design an innovative pencil that solves a specific problem related to pencil use, describe the scientific concept applied, draw and label the design, and explain how the new pencil is better than a regular pencil. In Question 2, pupils were required to design a creative fruit-picking machine capable of reaching fruits on tall trees, label the machine's components and functions, and explain how the solution helps farmers and what scientific concepts are utilised. Question 3 required pupils to write a short scientific story titled *"Matahari Hilang Cahaya!"*, explaining the effects of losing sunlight on plants, animals, humans, and weather while integrating scientific facts and imagination. In Question 4, pupils imagined improvements to a normal bicycle, listed two upgrades, drew and labelled the improved bicycle, and explained its usefulness and the scientific principles involved. Responses for all four tasks were scored using an analytic rubric with five criteria, which are originality, scientific accuracy, practicality or usefulness, elaboration and clarity, and creativity of communication, with each criterion scored on a 0–4 scale.

Although the conceptual development of the instrument began in English, the complete version for administration was written in Bahasa Malaysia to ensure comprehension. A forward–backward translation procedure was undertaken. Two bilingual science teachers translated the original instrument into Bahasa Malaysia, after which two independent translators back-translated it into English. Both versions were compared to ensure conceptual and semantic equivalence, and four language experts subsequently refined the final wording to improve clarity and age suitability.

Content validity was conducted using expert ratings on a four-point relevance scale(Mastaglia et al., 2003). The Item-level Content Validity Index (I-CVI) was calculated using the formula:

$$\text{I-CVI} = \frac{\text{Number of experts who rated the item as 3 or 4}}{\text{Total number of experts}}$$

The Scale-level Content Validity Index by averaging (S-CVI/Ave) was calculated as:

$$\text{S-CVI(Ave)} = \frac{\sum \text{I-CVI for all items}}{\text{Total number of items}}$$

Items with I-CVI values below 0.80 were revised based on expert feedback. Face validity was then conducted using ten Year 5 pupils who rated each item as either "clear" (1) or "unclear" (0). The Face Validity Index (FVI) was calculated using the formula:

$$FVI = \frac{\text{Number of pupils who rated the item as clear}}{\text{Total number of pupils}}$$

Items with FVI values of 0.80 or above were retained without modification. The pilot study was administered in regular classroom settings. All 150 pupils completed the four open-ended tasks, and responses were scored independently by two trained raters. Prior to scoring, raters participated in a calibration session using sample responses to ensure consistent interpretation of the rubric descriptors. Inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC), based on a two-way random effects model and absolute agreement(Gisev et al., 2013). ICC was selected because it is suitable for continuous rubric-based scoring and provides a robust indicator of scoring consistency(Yousef & Ayyoub, 2024). Interpretation followed established criteria in which values below 0.50 indicate poor reliability, values between 0.50 and 0.75 indicate moderate reliability, values between 0.75 and 0.90 indicate good reliability, and values above 0.90 indicate excellent reliability.

Data from the pilot study were analysed using descriptive statistics, CVI and FVI computations, and ICC reliability indices, with results presented in tables to summarise item-level validity and scoring consistency.

VI. FINDINGS

Findings from the content validity analysis demonstrated that the scientific creativity instrument possessed strong expert agreement regarding relevance, clarity, and scientific suitability. All four open-ended items recorded Item-level Content Validity Index (I-CVI) values of 0.90, while the Scale-level CVI (S-CVI/Ave) was also 0.90. These values exceeded the recommended minimum cut-off of 0.80, indicating that the items were suitable for assessing scientific creativity among Year 5 pupils. Only minor adjustments in vocabulary simplicity were suggested by experts. Table 1 presents the CVI and FVI results.

TABLE 1 CONTENT VALIDITY INDEX (I-CVI AND S-CVI/AVE) AND FACE VALIDITY INDEX (FVI) FOR SCIENTIFIC CREATIVITY ITEMS

Item (Scientific Creativity Task)	Experts Rating 3–4 (n/8)	I-CVI	Pupils Rating "Clear" (n/10)	FVI
Item 1: Innovative Pencil	7 / 8	0.90	8 / 10	0.85
Item 2: Fruit-Picking Machine	7 / 8	0.90	8 / 10	0.85
Item 3: Creative Science Story	7 / 8	0.90	8 / 10	0.85
Item 4: Improved Bicycle Design	7 / 8	0.90	8 / 10	0.85
Scale-Level Average	—	0.90	—	0.85

Face validity results indicated that pupils generally understood the instructions and were able to complete the tasks. However, the FVI for all items was 0.85, which, although above the acceptable threshold of 0.80, suggests slightly lower clarity than content validity. Pupils' feedback indicated that the term *"konsep sains"* in Items 1 and 4 was not immediately familiar to some respondents, and a few pupils needed teacher prompts to understand what scientific ideas could be applied. Item 3 was perceived as the most cognitively demanding because it required integrating scientific facts with imagination across multiple components (plants, animals, humans, and weather). Despite these minor challenges, all items met the required clarity level and were retained without structural modification. Inter-rater reliability testing further confirmed the stability of the scoring rubric. Using a two-way random effects model with absolute agreement, the overall Intraclass Correlation Coefficient (ICC) was 0.887, which is interpreted as *good* reliability. Task-level ICC values ranged from 0.86 to 0.90, demonstrating consistent scoring for both written and diagram-based responses. The ICC results are presented in Table 2.

TABLE 2 INTER-RATER RELIABILITY FOR SCIENTIFIC CREATIVITY SCORING

Task	ICC Value	Interpretation
Innovative Pencil	0.86	Good
Fruit-Picking Machine	0.89	Good
Creative Science Story	0.90	Excellent
Improved Bicycle Design	0.87	Good
Overall ICC	0.887	Good Reliability

Overall, the findings confirm that the scientific creativity instrument demonstrates strong content validity, acceptable face validity, and good inter-rater scoring reliability. These results support the suitability of the instrument for assessing scientific creativity among Year 5 pupils in Malaysian primary schools.

VII. DISCUSSION

This study was conducted to validate a scientific creativity instrument that was adapted from the work of Hu and Adey (2002) and Siew et al. (2015) and then contextualised for Year Five pupils in Malaysian primary schools. Elements from both instruments were integrated and modified to suit the Malaysian science curriculum, language requirements, and classroom culture, while the original dimensions of scientific creativity were preserved. The findings demonstrate that the adapted instrument is valid and reliable for assessing scientific creativity among primary pupils.

The content validity results confirmed that the tasks were relevant, scientifically accurate, and developmentally appropriate. All items recorded an Item Content Validity Index value of 0.90 and the Scale Content Validity Index by averaging was also 0.90. These values are higher than the recommended minimum of 0.80 suggested by Polit and Beck (2006) which indicates strong expert agreement regarding the suitability of the items. The use of real situations such as designing a pencil, creating a fruit picking machine, improving a bicycle and writing a creative science story helped to connect the assessment tasks with pupils' daily experiences. This supports previous claims that meaningful and realistic contexts can encourage pupils to think creatively in science.

The Face Validity Index was 0.85, which shows that the pupils generally understood the instructions and were able to complete the tasks. The slightly lower FVI, compared with the CVI, suggests that some wording may have been challenging for a few pupils. In particular, several pupils found it difficult to interpret the phrase "konsep sains". The writing task in Item Three was also more demanding because it required pupils to combine imagination with scientific reasoning. Although these challenges existed, all items still exceeded the minimum value of 0.80, which means the tasks were clear enough to be used without further modification.

The inter-rater reliability analysis provided additional support for the stability of the scoring system. The overall Intraclass Correlation Coefficient was 0.887, which is considered a good level of agreement. This shows that the scoring rubric allowed different raters to interpret and score pupils' work in a consistent manner. High agreement across written and drawing based responses indicates that the analytic rubric succeeded in defining originality, scientific accuracy, practicality, elaboration and communication clearly. Even though the ICC did not reach the level that is classified as excellent, it still shows that the instrument is reliable for classroom use and for research purposes.

Taken together, the results demonstrate that the scientific creativity instrument has strong content validity, acceptable face validity and good inter-rater reliability. The instrument is suitable for primary pupils because it encourages the use of creative ideas, scientific knowledge, problem solving, and communication skills. Therefore, the instrument can be used by teachers to assess scientific creativity in classroom activities, and by researchers who wish to study creativity in science learning in Malaysian schools.

Future research may involve a larger and more diverse sample of pupils, as well as the use of Rasch analysis or confirmatory factor analysis to further examine the measurement structure. The instrument may also be adapted for other age groups or used to explore relationships between creativity and other variables such as curiosity, inquiry skills, motivation or classroom climate.

VIII. CONCLUSION

This study has validated a scientific creativity instrument that was adapted from the work of Hu and Adey (2002) and Siew et al. (2014) for use among Year Five pupils in Malaysian primary schools. The results demonstrated that the instrument has strong content validity, acceptable face validity and good inter-rater reliability. Expert review confirmed that the items were relevant, accurate and suitable for pupils. Pupils also reported that the instructions were clear, while the scoring rubric allowed consistent evaluation by independent raters. These findings indicate that the instrument is appropriate for assessing scientific creativity in Malaysian classrooms.

The instrument has several strengths. It encourages pupils to generate creative ideas, apply scientific knowledge, solve practical problems and communicate their thinking through writing and drawing. The use of real and familiar contexts also makes the tasks meaningful, engaging and suitable for the developmental level of primary pupils. This supports the goal of fostering creativity in science education, which is consistent with current curriculum aspirations that encourage inquiry, problem solving and innovative thinking.

Although the findings are positive, there are limitations that should be acknowledged. The pilot study involved pupils from a limited number of schools. The results therefore cannot be generalised to all primary pupils in Malaysia. In addition, the instrument uses open ended tasks, which require considerable time for administration and scoring. Some pupils also needed guidance to interpret certain scientific terms. Future refinement of wording and examples may help younger pupils understand the expectations more easily.

Suggestions for further research include validating the instrument with a larger and more diverse sample, including rural and urban schools. Researchers may also conduct Rasch analysis or confirmatory factor analysis to examine the measurement structure in greater detail. In addition, the instrument can be applied to different age groups to compare creativity across year levels. Future studies may also explore how scientific creativity relates to other variables such as curiosity, inquiry skills, classroom climate or teaching approaches.

In conclusion, this study provides evidence that the scientific creativity instrument is a valid and reliable tool for use with Year Five pupils. It can support teachers, curriculum planners and researchers who aim to measure and encourage creativity in science learning. The instrument contributes to the development of assessment tools that reflect the goals of innovative and transformative science education in Malaysia.

ACKNOWLEDGEMENT

The authors would like to express sincere appreciation to the experts in science education, educational measurement and language studies who provided valuable comments during the validation process. Special thanks are extended to the Year Five pupils and teachers from the participating primary schools for their cooperation and involvement in the pilot study. The authors also acknowledge the support of the Faculty of Education, University of Malaya, for the guidance and resources provided throughout this research.

REFERENCES

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289–1312.

Akdeniz, H., & Bangir, G. (2025). The Impact of Emotional Intelligence on Gifted and Talented Students' Creative Problem-Solving Performance. *Gifted Education International*, 02614294251379619.

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in social and administrative pharmacy*, *15*(2), 214–221.

Almanasreh, E., Moles, R. J., & Chen, T. F. (2022). A practical approach to the assessment and quantification of content validity. In *Contemporary research methods in pharmacy and health services* (pp. 583–599). Elsevier.

Ambo, N., Siew, N., & Han, C. (2019). Development and evaluation of a project-based STEM learning module on the scientific creativity of gred five Students in Malaysia. In *Empowering Science and Mathematics for Global Competitiveness* (pp. 283–294). CRC Press.

Araceli Ruiz-Primo, M., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, *11*(3-4), 237–263.

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.

Bahar, K., & June Maker, C. (2025). A new perspective on scoring children's originality: a standards-based criterion-referenced assessment approach. *Frontiers in Psychology*, *16*, 1545396.

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, *67*(3), 343–368.

Cevher, A. H., Ertekin, P., & Koksal, M. S. (2014). Investigation of scientific creativity of eighth grade gifted students. *International Journal of Innovation, Creativity and Change*, *1*(4), 1–8.

Chen, Y. T., Liu, M. J., & Cheng, Y. Y. (2023). Discovering Scientific Creativity with Digital Storytelling [Article]. *Journal of Creativity*, *33*(1), Article 100041. https://doi.org/10.1016/j.yjoc.2022.100041

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological assessment*, *31*(12), 1412.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine*, *119*(2), 166. e167–166. e116.

Davis, G. A. (1989). Testing for creative potential. *Contemporary Educational Psychology*, *14*(3), 257–274.

DeHaan, R. L. (2009). Teaching creativity and inventive problem solving in science. *CBE—Life Sciences Education*, *8*(3), 172–181.

Dollinger, S. J. (2011). "Standardized minds" or individuality? Admissions tests and creativity revisited. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(4), 329.

Dow, G. T. (2022). Defining Creativity. In *Creativity and Innovation: Theory, Research, and Practice, Second Edition* (pp. 5–21). Taylor and Francis. https://doi.org/10.4324/9781003233923-2

Dumas, D., Acar, S., Berthiaume, K., Organisciak, P., Eby, D., Grajzel, K., Vlaamster, T., Newman, M., & Carrera, M. (2023). What makes children's responses to creativity assessments difficult to judge reliably? *The Journal of Creative Behaviour*, *57*(3), 419–438.

Fernandez, S. M., Madelo, P. K., Lu Suico, R. A., Cane, J. F., Magsayo, J., Capuyan, M., Siew, N. M., & Acut, D. (2025). Thinking What No One Else Has Thought: Investigating the Scientific Creativity of Primary School Students in a Science Class [Article]. *Center for Educational Policy Studies Journal*, *15*(3), 95–124. https://doi.org/10.26529/cepsj.1514

Firdaus, F., Wiyanto, W., Putra, N. M. D., & Isnaen, W. (2025). Design of instruments for scientific creative thinking skills and creative thinking digital skills: Rasch models and confirmatory factor analysis [Article]. *Eurasia Journal of Mathematics, Science and Technology Education*, *21*(5), Article em2632. https://doi.org/10.29333/ejmste/16310

Foo, S. Y. (2025). Investigating gifted students' higher-order thinking skills in a differentiated learning environment: A case study. *Gifted Education International*, *41*(2), 236–259.

Garrett, R. (1987). Issues in science education: problem-solving, creativity and originality. *International Journal of Science Education*, *9*(2), 125–137.

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in social and administrative pharmacy*, *9*(3), 330–338.

Guo, J., & Woulfin, S. (2016). Twenty-first century creativity: An investigation of how the partnership for 21st century instructional framework reflects the principles of creativity. *Roeper Review*, *38*(3), 153–161.

Hansson, E. E., Svensson, P. J., Strandberg, E. L., Troein, M., & Beckman, A. (2014). Inter-rater reliability and agreement of rubrics for assessment of scientific writing. *Education*, *4*(1), 12–17.

Heller, K. A. (2007). Scientific ability and creativity. *High Ability Studies*, *18*(2), 209–234.

Henze, I., Van Driel, J. H., & Verloop, N. (2009). Experienced science teachers' learning in the context of educational innovation. *Journal of teacher education*, *60*(2), 184–199.

Hernández-Ramos, J., & Araya, R. (2025). Do School Activities Foster Creative Thinking? An Analysis of PISA Results. *Education Sciences*, *15*(2), 133.

Hong, O., Park, M.-H., & Song, J. (2022). The assessment of science classroom creativity: scale development. *International Journal of Science Education*, *44*(8), 1356–1377.

Hong, O., Park, M. H., & Song, J. (2022). The assessment of science classroom creativity: scale development [Article]. *International Journal of Science Education*, *44*(8), 1356–1377. https://doi.org/10.1080/09500693.2022.2077466

Hossan, D., Wolfs, B., & Petkovic, M. (2025). Questionnaire Validity and Reliability: A Review with Practical Guidelines. *Journal of Entrepreneurship, Business and Economics*, *13*(1), 135–186.

Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education*, *24*(4), 389–403.

Ismail, M. H., Abdullah, N., Salleh, M. F. M., & Ismail, M. (2017). Higher order thinking skills (HOTS): Teacher training and skills in assessing science learning [Article]. *Advanced Science Letters*, *23*(4), 3259–3262. https://doi.org/10.1166/asl.2017.7732

Jirschitzka, J., Oeberst, A., Göllner, R., & Cress, U. (2017). Inter-rater reliability and validity of peer reviews in an interdisciplinary field. *Scientometrics*, *113*(2), 1059–1092.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, *2*(2), 130–144.

Juriševič, M., & Žerak, U. (2024). Creative potential profiles of primary school students. *Learning and Individual Differences*, *113*, 102496. https://doi.org/https://doi.org/10.1016/j.lindif.2024.102496

Karademir, E. (2016). Investigation the scientific creativity of gifted students through project-based activities. *International Journal of Research in Education and Science*, *2*(2), 416–427.

Kennedy, T. J., & Sundberg, C. W. (2025). Twenty-First Century Skills. In *Science Education in Theory and Practice: An Introductory Guide to Learning Theory* (pp. 461–486). Springer.

Kim, M. K., Roh, I. S., & Cho, M. K. (2016). Creativity of gifted students in an integrated math-science instruction. *Thinking Skills and Creativity*, *19*, 38–48.

Kind, P. M., & Kind, V. (2007). Creativity in science education: Perspectives and challenges for developing school science.

Klassen, S. (2006). Contextual assessment in science education: Background, issues, and policy. *Science Education*, *90*(5), 820–851.

Koç, A., & Büyük, U. (2021). Effect of robotics technology in science education on scientific creativity and attitude development. *Journal of Turkish Science Education*, *18*(1), 54–72.

Latulipe, C., & Terry, M. (2008). Evaluation instruments for creativity support tools. Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, BCS HCI 2008,

Lewis, S. E., Shaw, J. L., & Freeman, K. A. (2011). Establishing open-ended assessments: investigating the validity of creative exercises. *Chemistry Education Research and Practice*, *12*(2), 158–166.

Lim, S. M., Palethorpe, N., & Rodger, S. (2012). Understanding the common interrater reliability measures. *International Journal of Therapy and Rehabilitation*, *19*(9), 488–496.

Liu, Q., Cui, T., Xiao, L., & Liu, J. (2026). Exploring the synergy of social-individual enablers of creativity in science: A multidimensional study of fourth graders [Article]. *Thinking Skills and Creativity*, *59*, Article 101988. https://doi.org/10.1016/j.tsc.2025.101988

Liu, S.-C., & Lin, H.-s. (2014). Primary teachers' beliefs about scientific creativity in the classroom context. *International Journal of Science Education*, *36*(10), 1551–1567.

Massy, G., Didier, J., & Besançon, M. (2026). Assessing creativity in early childhood: How future teachers balance conformity and originality. *Thinking Skills and Creativity*, *59*, 102023. https://doi.org/https://doi.org/10.1016/j.tsc.2025.102023

Mastaglia, B., Toye, C., & Kristjanson, L. J. (2003). Ensuring content validity in instrument development: challenges and innovative approaches. *Contemporary Nurse*, *14*(3), 281–291.

Mat, H., Mustakim, S. S., & Razali, F. (2025). Proficiency of Higher-Order Thinking Skills in Malaysian Primary Science Curriculum: Insights From Year 5. *International Journal of Education and Social Science Research*, 26–40.

Mat Noor, M. (2022). An insight into primary science education in Malaysia. *ASE International*, *16*, 34–41.

Nguyen, T. V., & Tran, V. N. (2025). Researching creativity in education from ASEAN countries: bibliometric analysis [Article]. *International Journal of Evaluation and Research in Education*, *14*(4), 2593–2604. https://doi.org/10.11591/ijere.v14i4.33669

Noddings, N. (2013). Standardized curriculum and loss of creativity. *Theory Into Practice*, *52*(3), 210–215.

Oh, J.-Y. (2022). Understanding the scientific creativity based on various perspectives of science. *Axiomathes*, *32*(6), 907–929.

Oliveira, A. W., Brown, A. O., Zhang, W. S., LeBrun, P., Eaton, L., & Yemen, S. (2021). Fostering creativity in science learning: The potential of open-ended student drawing. *Teaching and Teacher Education*, *105*, 103416.

Omar, S. S., Harun, J., Halim, N. D. A., Surif, J., & Muhammad, S. (2017). Investigating the level of scientific creativity of science students. *Advanced Science Letters*, *23*(9), 8247–8250.

Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, *10*, 265–279.

Pinar, F. I. L., Panergayo, A. A. E., Sagcal, R. R., Acut, D. P., Roleda, L. S., & Prudente, M. S. (2025). Fostering scientific creativity in science education through scientific problem-solving approaches and STEM contexts: a meta-analysis [Review]. *Disciplinary and Interdisciplinary Science Education Research*, *7*(1), Article 18. https://doi.org/10.1186/s43031-025-00137-9

Poddiakov, A. (2023). Creativity of creativity researchers: invention of problems and experimental objects to study thinking. *Integrative Psychological and Behavioural Science*, *57*(1), 43–64.

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in nursing & health*, *29*(5), 489–497.

Prahani, B. K., Suprapto, N., Rachmadiarti, F., Sholahuddin, A., Mahtari, S., & Siswanto, J. (2021). Online Scientific Creativity Learning (OSCL) in Science Education to Improve Students' Scientific Creativity in Covid-19 Pandemic [Article]. *Journal of Turkish Science Education*, *18*(Special Issue), 77–90. https://doi.org/10.36681/tused.2021.73

Ramly, S. N. F., Ahmad, N. J., & Yakob, N. (2022). Development, validity, and reliability of chemistry scientific creativity test for pre-university students. *International Journal of Science Education*, *44*(14), 1–16.

Rojas-Drummond, S., Hernández, G., Vélez, M., & Villagrán, G. (1998). Cooperative learning and the appropriation of procedural knowledge by primary school children. *Learning and Instruction*, *8*(1), 37–61.

Rosen, Y., Stoeffler, K., & Simmering, V. (2020). Imagine: Design for Creative Thinking, Learning, and Assessment in Schools. *J Intell*, *8*(2). https://doi.org/10.3390/jintelligence8020016

Roth, T., Conradty, C., & Bogner, F. X. (2022). Testing creativity and personality to explore creative potentials in the science classroom. *Research in science education*, *52*(4), 1293–1312.

Saw, Z. K., Yuen, J. J. X., Ashari, A., Ibrahim Bahemia, F., Low, Y. X., Nik Mustapha, N. M., & Lau, M. N. (2025). Forward-backward translation, content validity, face validity, construct validity, criterion validity, test-retest reliability, and internal consistency of a questionnaire on patient acceptance of orthodontic retainer. *PLoS ONE*, *20*(1), e0314853.

Schneider, R. M., & Plasman, K. (2011). Science teacher learning progressions: A review of science teachers' pedagogical content knowledge development. *Review of Educational Research*, *81*(4), 530–565.

Shahali, E. H. M., Ismail, I., & Halim, L. (2017). STEM education in Malaysia: Policy, trajectories and initiatives. *Asian Research Policy*, *8*(2), 122–133.

Shi, B., Cao, X., Chen, Q., Zhuang, K., & Qiu, J. (2017). Different brain structures associated with artistic and scientific creativity: A voxel-based morphometry study [Article]. *Scientific Reports*, *7*, Article 42911. https://doi.org/10.1038/srep42911

Siew, N. M., Chong, C. L., & Chin, K. O. (2014). Developing a scientific creativity test for fifth graders. *Problems of Education in the 21st Century*, *62*, 109.

Siew, N. M., Chong, C. L., & Lee, B. N. (2015). Fostering Fifth Graders' Scientific Creativity through Problem-Based Learning. *Journal of Baltic Science Education*, *14*(5), 655–669.

Sireci, S. G. (1998). The construct of content validity. *Social indicators research*, *45*(1), 83–117.

Skjelstad Fredagsvik, M. (2022). Student approaches to creative processes when participating in an open-ended project in science. *International Journal of Science Education*, *44*(10), 1583–1600.

Smyrnaiou, Z., Georgakopoulou, E., & Sotiriou, S. (2020). Promoting a mixed-design model of scientific creativity through digital storytelling—the CCQ model for creativity. *International Journal of STEM Education*, *7*(1), 25.

Soh, K. (2015). Creativity fostering teacher behaviour around the world: Annotations of studies using the CFTIndex [Article]. *Cogent Education*, *2*(1), Article 1034494. https://doi.org/10.1080/2331186X.2015.1034494

Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, *26*(3), 233–246.

Suradin, S., Rasul, M., Omar, M., Ab Hamid, N., & Othman, N. (2025). Scientific Creativity in TVET Product Design: Integrating Scientific Knowledge and Imagination. *Journal of Technical Education and Training*, *17*(3), 43–63. https://doi.org/10.30880/jtet.2025.17.03.004

Tan, C.-S., Tan, S.-A., Cheng, S.-M., Hashim, I. H. M., & Ong, A. W.-H. (2021). Development and preliminary validation of the 20-item Kaufman domains of Creativity Scale for use with Malaysian populations. *Current Psychology*, *40*(4), 1946–1957.

Taylor, A. R., Jones, M. G., Broadwell, B., & Oppewal, T. (2008). Creativity, inquiry, or accountability? Scientists' and teachers' perceptions of science education. *Science Education*, *92*(6), 1058–1075.

Thomas, B., & Watters, J. J. (2015). Perspectives on Australian, Indian and Malaysian approaches to STEM education. *International Journal of Educational Development*, *45*, 42–53.

van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & van Tartwijk, J. (2020). Creative Problem Solving in Primary Education: Exploring the

Role of Fact Finding, Problem Finding, and Solution Finding across Tasks [Article]. *Thinking Skills and Creativity*, *37*, Article 100665. https://doi.org/10.1016/j.tsc.2020.100665

Wang, C.-C., Ho, H.-C., Wu, J.-J., & Cheng, Y.-Y. (2014). Development of the scientific imagination model: A concept-mapping perspective. *Thinking Skills and Creativity*, *13*, 106–119.

Xu, C., & Fan, D. (2025). Creative-Becoming Holism: Reflections on and Development of Creative Holism in the Case of Science Education. *Systems Research and Behavioural Science*, *42*(1), 98–110.

Xu, S., Reiss, M., & Lodge, W. (2025). Comprehensive Scientific Creativity Assessment (C-SCA): A New Approach for Measuring Scientific Creativity in Secondary School Students. *International Journal of Science and Mathematics Education*, *23*(2), 293–319. https://doi.org/10.1007/s10763-024-10469-z

Xu, S., Reiss, M. J., & Lodge, W. (2025). Comprehensive Scientific Creativity Assessment (C-SCA): A New Approach for Measuring Scientific Creativity in Secondary School Students [Article]. *International Journal of Science and Mathematics Education*, *23*(2), 293–319, Article 750224. https://doi.org/10.1007/s10763-024-10469-z

Yildiz, C., & Guler Yildiz, T. (2021). Exploring the relationship between creative thinking and scientific process skills of preschool children. *Thinking Skills and Creativity*, *39*, 100795. https://doi.org/https://doi.org/10.1016/j.tsc.2021.100795

Yousef, A., & Ayyoub, A. (2024). Rubric development and validation for assessing educational robotics skills. Frontiers in Education,

Yusoff, M. S. B. (2019). ABC of response process validation and face validity index calculation. *Educ Med J*, *11*(10.21315).

Zbainos, D., & Tziona, A. (2019). Investigating Primary School Children's Creative Potential Through Dynamic Assessment [Original Research]. *Frontiers in Psychology*, *Volume 10 - 2019*. https://doi.org/10.3389/fpsyg.2019.00733