Journal of Educational Sciences
Journal homepage: https://jes.ejournal.unri.ac.id/index.php/JES

P-ISSN 2581-1657
E-ISSN 2581-2203

Development of a Figural Creativity Assessment Instrument Based on Torrance Tests of Creative Thinking (TTCT) for Elementary School Students

Attar Nazaruddin Mirza*, Agung Hastomo, Firmansyah
Department of Elementary Education, Faculty of Education, Universitas Negeri Yogyakarta, Yogyakarta, 55281, Indonesia

ARTICLE INFO

Article history:

Received: 11 March 2026
Revised: 16 March 2026
Accepted: 25 March 2026
Published online: 05 April 2026

Keywords:

Creativity Instrument,
Elementary School,
Instrument Reliability,
Instrument Validity,
Ttct Figural

* *Corresponding author:*

E-mail:
attarnazaruddin.2024@student.uny.ac.id

Article Doi:

https://doi.org/10.31258/jes.10.4.p.424-442

A B S T R A C T

This study aimed to develop a figural creativity assessment instrument based on the Torrance Tests of Creative Thinking (TTCT) for elementary school students and to examine its validity and reliability. The research employed a Research and Development approach using the 4D model, consisting of define, design, develop, and limited disseminate stages. The participants were elementary school students in Yogyakarta. A limited trial involved 16 students, while the field test involved 106 students. Data were collected using the developed figural creativity test, supported by expert validation sheets and student response questionnaires. Content validity was evaluated through expert judgment, internal validity was analyzed using Spearman rank correlation, and inter-rater reliability was examined using the Intraclass Correlation Coefficient (ICC). The results indicated that the instrument demonstrated high feasibility in psychological, linguistic, and instructional aspects. Internal validity analysis showed significant positive correlations among stimuli and creativity dimensions. Inter-rater reliability showed good agreement for fluency (ICC = 0.646), flexibility (ICC = 0.754), and elaboration (ICC = 0.779), while originality showed lower agreement (ICC = 0.194). Overall, the developed instrument is valid and reliable for assessing figural creativity in elementary school students, although further refinement of originality scoring procedures is recommended.

1. Introduction

Creativity has long been recognized as one of the key competencies required in the twenty-first century. Rapid technological development, global competition, and complex social challenges demand individuals who are able to generate innovative ideas, adapt to change, and solve problems creatively. International organizations emphasize that creativity is an essential capability for future generations. UNESCO

highlights creativity as a fundamental component for preparing young people to face uncertainty and global transformation (UNESCO, 2016). Similarly, the Future of Jobs Report published by the World Economic Forum identifies creativity, critical thinking, and complex problem solving as among the most important skills required in the future workforce (WEF, 2023). In education, the twenty-first-century skills framework promoted by the Partnership for 21st Century Learning also positions creativity as a core competency that should be intentionally developed through learning processes rather than treated merely as an outcome of education (GPE, 2020; Trilling & Fadel, 2013).

In Indonesia, the importance of creativity has also been emphasized in national educational policies. The national education system states that education aims to develop learners who are knowledgeable, independent, responsible, and creative citizens. Recent educational reforms further reinforce this orientation through the Pancasila Student Profile, which identifies creativity and critical reasoning as key competencies for Indonesian learners in the Merdeka Curriculum (Kemendikbudristek, 2022). These policy directions demonstrate that creativity is not only a desirable trait but also a fundamental educational objective aimed at preparing students to adapt to rapid social and technological changes (Istiqomah & Na'imah, 2025; Munastiwi, 2023; Sari & Zulaikha, 2020).

The development of creativity is particularly important during elementary education because this stage represents a critical period for children's cognitive and affective growth. During these years, children demonstrate increasing curiosity, imagination, and flexibility in thinking, which provide a strong foundation for creative expression and idea generation (Runco, 2023; Saputri et al., 2024). Students in middle elementary grades gradually develop the ability to express ideas visually, explore alternative possibilities, and elaborate imaginative concepts. These characteristics are closely related to divergent thinking, a cognitive process associated with creativity that involves generating multiple solutions, exploring different perspectives, and producing original ideas (Sternberg & Lubart, 1995; Torrance, 1966). Divergent thinking is typically reflected in several indicators, including fluency, flexibility, originality, and elaboration (Runco & Acar, 2012; Torrance, 1966).

Despite the recognized importance of creativity, classroom practices often remain dominated by convergent approaches that prioritize memorization and single correct answers. Studies indicate that instructional activities in many elementary classrooms still emphasize lower-order cognitive processes rather than encouraging students to explore ideas or engage in creative problem solving (Fauziah et al., 2020; Hidayah et al., 2024; Leasa et al., 2021). Convergent thinking focuses on identifying one correct solution based on established knowledge, whereas divergent thinking encourages exploration of multiple possibilities (Carroll & Guilford, 1968; Runco, 2023). When learning environments primarily emphasize convergent thinking, opportunities for students to develop creativity may become limited.

In educational practice, teachers play a central role in fostering creativity through learning design, classroom interaction, and assessment strategies. Effective

creativity development requires learning environments that encourage idea exploration, imagination, and flexible problem solving (Craft, 2005; Runco, 2023; Tepaso et al., 2025). Assessment is particularly important because it provides information about students' creative abilities and supports teachers in designing appropriate learning interventions (Hickman, 2023; Van Hooijdonk et al., 2024). However, the effectiveness of creativity development depends largely on the availability of valid and reliable assessment instruments capable of capturing different dimensions of creative thinking (Brandt, 2021).

In practice, creativity assessment in elementary schools often relies on approaches that are not specifically designed to measure divergent thinking. Observational assessments, self-assessment questionnaires, and multiple-choice tests are commonly used to evaluate students' creativity. Although these approaches may provide certain insights, they have significant limitations in capturing the complexity of creative thinking processes. Multiple-choice tests primarily measure convergent thinking, while observational and self-report methods are susceptible to subjectivity and response bias (Kanlı, 2021; Runco & Acar, 2024). Consequently, these methods may not accurately represent students' creative potential.

One of the most widely recognized instruments for assessing creativity internationally is the Torrance Tests of Creative Thinking (TTCT), developed by Ellis Paul Torrance. The TTCT is specifically designed to measure divergent thinking abilities, including fluency, flexibility, originality, and elaboration (Torrance, 1966). The instrument consists of two forms: verbal and figural. The verbal form measures creativity through language-based responses, while the figural form evaluates visual imagination through drawing activities. For elementary school students, the figural form is often considered more suitable because it is less dependent on linguistic ability and allows children to express creative ideas visually (Kim, 2017; Torrance, 1998). Numerous studies have also demonstrated the reliability and validity of TTCT across different contexts (Kim, 2006; Runco & Acar, 2012). Nevertheless, the application of TTCT in Indonesian elementary education remains limited and often partial. Existing creativity assessments are still dominated by subjective or convergent approaches, and standardized instruments designed to measure divergent thinking are rarely used. This situation indicates a gap between the recognized importance of creativity in education and the availability of appropriate assessment tools capable of measuring it comprehensively.

Therefore, this study aims to develop a figural creativity assessment instrument based on the Torrance Tests of Creative Thinking framework that is suitable for elementary school students in Indonesia. The study focuses on designing and evaluating the validity and reliability of the developed instrument in order to provide a contextualized and psychometrically sound tool for assessing students' creative thinking.

2. Methodology

The development process followed the 4D model consisting of define, design, develop, and disseminate stages (Thiagarajan et al., 1974) as illustrated in Figure 1.

Figure 1. Thiagarajan 4D Model

In the define stage, the study identified problems and needs related to the assessment of creativity among elementary school students in Indonesia through literature review and systematic analysis of previous studies on creativity measurement, focusing on the dimensions of fluency, flexibility, originality, and elaboration and the conceptual framework of the figural version of the Torrance Tests of Creative Thinking (TTCT). The design stage involved developing the initial blueprint of the figural creativity instrument by adapting TTCT-based activities while ensuring that the stimuli, language, and context were appropriate for the cognitive and linguistic characteristics of elementary school students.

In the develop stage, the draft instrument underwent expert validation by a panel consisting of a child psychology expert, a language expert, and an elementary school teacher to obtain evidence of content validity. The instrument was then revised based on expert feedback, followed by a limited trial with a small group of students to evaluate readability, clarity of instructions, time allocation, and students' responses to the activities. Subsequently, a field trial was conducted with a larger sample to obtain empirical data for examining internal structure validity and estimating reliability. Finally, the disseminate stage was carried out in a limited form by compiling the final instrument package, including administration guidelines, scoring procedures, and interpretation examples, while broader dissemination was restricted to research reporting and intellectual property registration of the developed instrument (AERA et al., 2014; Torrance, 1974).

Instrument

The primary instrument was a figural creativity test adapted from the TTCT figural format introduced by Ellis Paul Torrance (Torrance, 1974). The instrument consisted of three activities: Circle Activity, Incomplete Figures, and Parallel Lines. These activities were contextually modified to suit the developmental characteristics and learning experiences of Indonesian elementary school students while maintaining the essential structure of figural creativity assessment. The instrument was designed to measure four dimensions of creativity: fluency, flexibility, originality, and elaboration. The complete instrument package consisted of student activity sheets, administration guidelines for teachers, scoring rubrics, and examples of scoring interpretations. The scoring guide was structured

systematically to reduce subjectivity and support inter-rater reliability in evaluating students' figural responses.

Supporting instruments were also developed to collect complementary data. These included expert validation sheets, observation guidelines, and student response questionnaires. The expert validation sheets were designed for three groups of validators: child psychology experts, language experts, and elementary school teachers as practitioners. These instruments assessed aspects such as developmental appropriateness, linguistic clarity, feasibility of implementation in classroom settings, and alignment with creativity constructs (AERA et al., 2014; Piaget, 1952; Usnia et al., 2025; Woolfolk, 2016). Observation sheets were used during the limited trial to record students' responses, comprehension of instructions, engagement during test activities, and potential technical difficulties. In addition, a student response questionnaire using a three-point Likert scale was administered to capture students' perceptions regarding clarity of instructions, attractiveness of activities, level of difficulty, and overall comfort during the test.

Data Collection

Data were collected using both test and non-test techniques. The main data were obtained through the figural creativity test administered in classroom settings under the supervision of the researcher and classroom teacher. Test administration followed standardized procedures to ensure consistency in instructions, time allocation, and testing conditions. Prior to student testing, the instrument underwent expert validation to obtain content validity evidence. Experts evaluated the suitability of indicators with the creativity constructs, clarity of instructions, linguistic appropriateness, developmental compatibility with elementary students, and feasibility of classroom implementation. Their evaluations were recorded using a five-point Likert scale ranging from "very poor" to "very good." During the limited trial stage, classroom observations were conducted to identify students' understanding of instructions, engagement with the tasks, and technical issues encountered during implementation. Additionally, students completed a response questionnaire to provide feedback on the clarity, attractiveness, and perceived difficulty of the activities.

Data Analysis

Data analysis was conducted according to the type of data collected, including expert validation data, creativity test scores, and supporting qualitative data from observations and questionnaires. The analysis aimed to provide evidence of instrument validity and reliability following educational and psychological measurement standards (AERA, APA, & NCME, 2014). Content validity was examined through expert judgment. Scores from expert evaluations were converted into feasibility percentages using the formula proposed by Widoyoko (2018). The percentage results were then interpreted using feasibility criteria ranging from "not feasible" to "highly feasible," as presented in Table 1. Qualitative feedback from experts was also analyzed to refine the instrument.

Table 1. Feasibility Criteria

Feasibility Percentage	Interpretation
85% – 100%	Highly feasible
70% – 84%	Feasible
50% – 69%	Needs revision
< 50%	Not feasible

Evidence of validity based on internal structure was examined through correlation analysis within the framework of classical test theory (Cronbach, 1951; DeVellis, 2017). Item–total correlations were conducted for stimuli within the Incomplete Figures activity to determine their contribution to the measured creativity dimensions. Additionally, correlations among activities measuring similar creativity dimensions were analyzed to assess consistency across tasks. Normality of data distribution was tested using the Kolmogorov–Smirnov test. Pearson Product Moment correlation was used for normally distributed data, while Spearman Rank correlation was applied when the normality assumption was not met. Correlation strength was interpreted using the classification proposed by Cohen (2013) as presented in Table 2.

Table 2. Interpretation of Correlation Coefficient

Correlation Coefficient	Interpretation
0.81 – 1.00	Very strong correlation
0.61 – 0.80	Strong correlation
0.41 – 0.60	Moderate correlation
0.21 – 0.40	Weak correlation
0.00 – 0.20	Very weak correlation

Reliability was examined through inter-rater reliability because scoring involved the interpretation of students' figural products. Each student's response was independently evaluated by three trained raters. Agreement among raters was analyzed using the Intraclass Correlation Coefficient (ICC) with a two-way random effects model and absolute agreement. ICC values were interpreted following the criteria proposed by Koo and Li (2016) as presented in Table 3.

Table 3. Coefficient Interpretation of Inter-Rater Reliability (ICC)

ICC Value	Interpretation
≥ 0.75	Excellent reliability
0.60 – 0.74	Good reliability
0.40 – 0.59	Moderate reliability
< 0.40	Poor reliability

Supporting data from observations were analyzed descriptively through qualitative categorization to identify patterns related to instruction clarity, student engagement, and implementation feasibility. Student questionnaire responses were analyzed using descriptive percentages to capture students' perceptions of the developed instrument. These qualitative and quantitative findings were triangulated to provide a comprehensive evaluation of the instrument's feasibility and psychometric quality.

3. Results and Discussion

This study developed a figural creativity assessment instrument based on the principles of the Torrance Tests of Creative Thinking (TTCT) for elementary school students in phases B and C. The instrument was developed using the 4D model consisting of define, design, develop, and limited disseminate stages. The final product is designed to measure students' creative thinking through figural tasks that encourage open-ended responses. The instrument consists of three main drawing activities administered sequentially, each with an estimated completion time of approximately 15 minutes (as shown in Figures 2, 3, and 4). Each activity uses different visual stimuli intended to elicit creative responses from students. The tasks are designed to measure four dimensions of creativity, namely fluency, flexibility, originality, and elaboration.

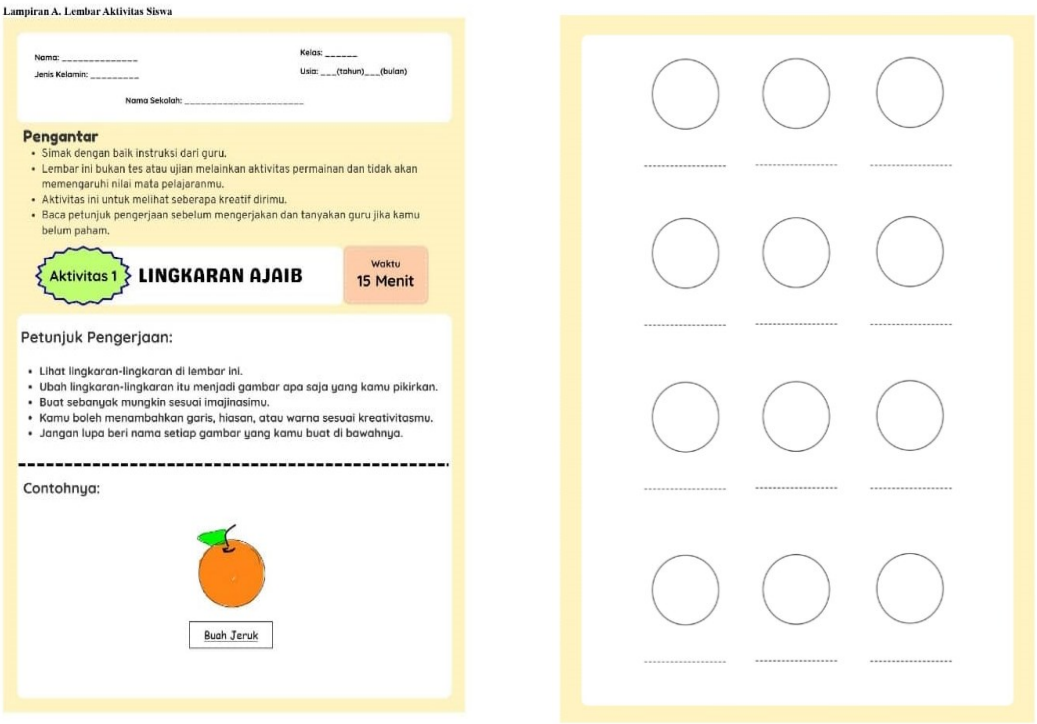

Figure 2. Activity 1: Circle Activity

The final product includes four main components: (1) a test administration guide, (2) student activity guidelines containing the three figural tasks, (3) a scoring guide explaining the assessment criteria for each creativity dimension, and (4) instrument appendices consisting of student worksheets and scoring sheets. The test administration guide provides instructions regarding test objectives, required materials, time allocation, environmental conditions, and the role of teachers during implementation. The student activity section contains the sequence of drawing tasks and stimulus images, while the scoring guide provides operational definitions and scoring procedures for each creativity dimension.

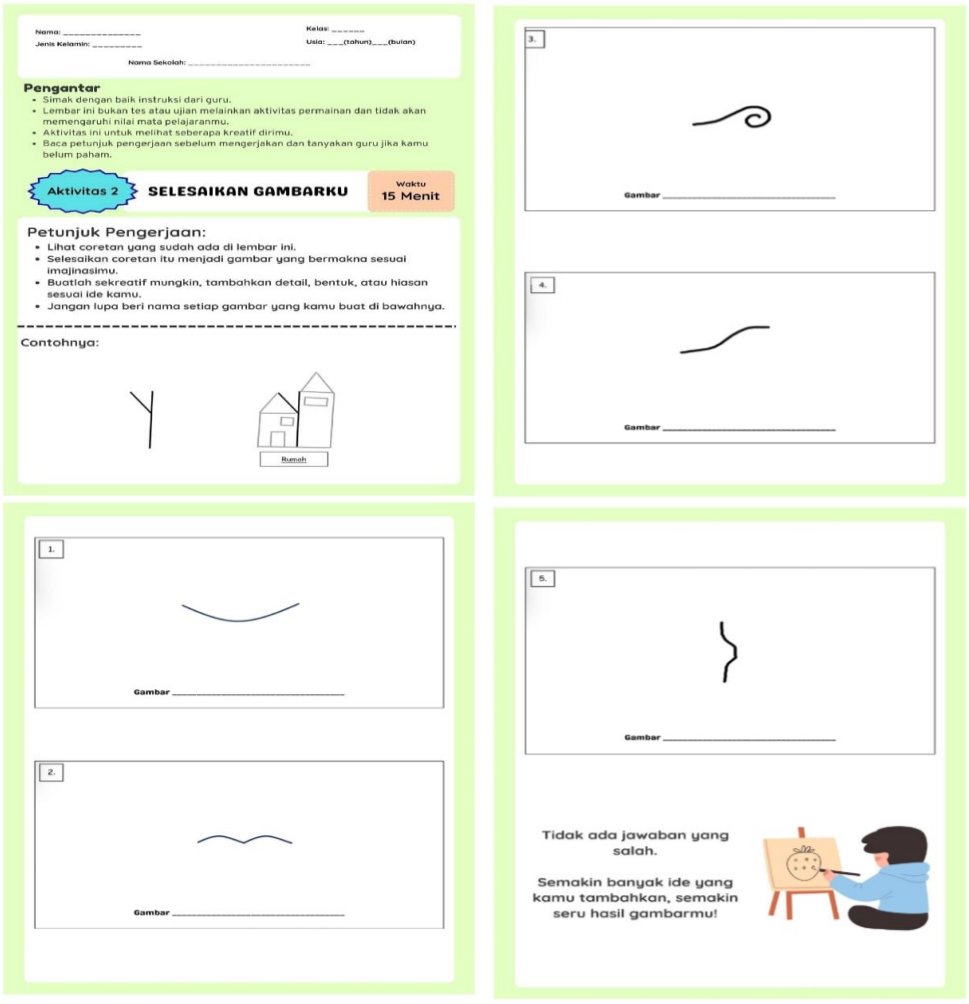

Figure 3. Activity 2: Incomplete Figures

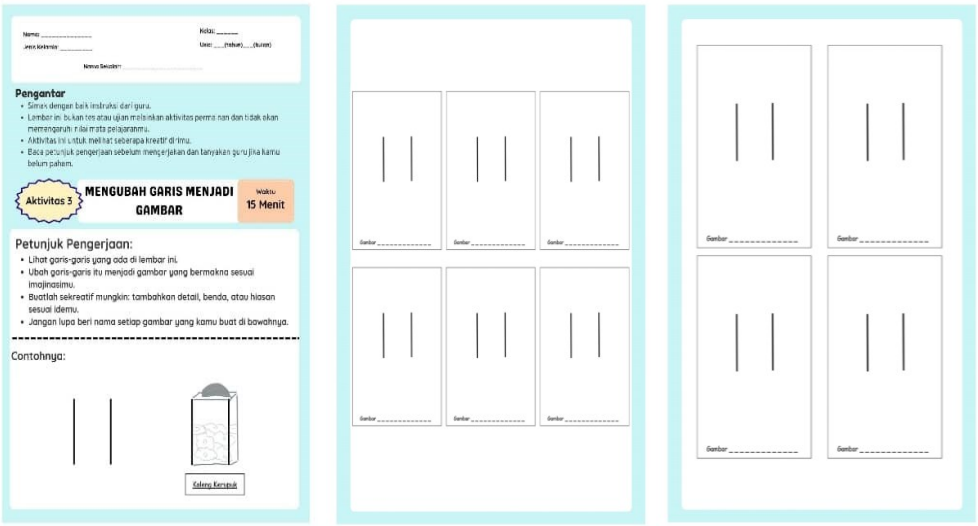

Figure 4. Activity 3: Parallel Lines

The structure of the instrument was designed to ensure clarity of instructions, appropriateness of tasks for elementary school students, and systematic scoring procedures to support consistent assessment of creative responses. The development results presented in this section correspond to the stages of the 4D model applied in this study. The define and design stages produced the conceptual framework of the instrument, including the selection of creativity dimensions, the design of figural tasks, and the preparation of scoring rubrics. The develop stage involved expert validation and readability testing to evaluate the feasibility, clarity, and practicality of the instrument prior to classroom implementation. Subsequently, empirical testing was conducted to examine the internal validity and reliability of the instrument. These analyses functioned as part of the empirical evaluation before the limited dissemination stage, ensuring that the instrument met basic psychometric requirements prior to broader application in educational contexts.

Content Validity

Content validity was evaluated through expert judgment involving child psychology experts, language experts, and elementary school teachers. The validation process assessed several aspects of the instrument, including psychological suitability, language clarity, instructional feasibility, and scoring guidelines, as presented in Table 4.

Table 4. Content Validity Results

Validation Aspect	Assessment Indicator	Percentage	Category
Child Psychology Expert Validation	• Appropriateness for children's cognitive development	93.33%	Highly Feasible
	• Alignment with children's creativity development stage	100%	Highly Feasible
	• Clarity of instructions for children	86.67%	Highly Feasible
	• Reasonable cognitive load	80.00%	Feasible
	• Emotional and motivational aspects	100%	Highly Feasible
	• Appropriateness of creativity indicators	100%	Highly Feasible
	• Fairness for students with different abilities	80.00%	Feasible
Language Validation (Student Worksheet)	• Readability	100%	Highly Feasible
	• Clarity of instructions	93.33%	Highly Feasible
	• Vocabulary appropriateness	93.33%	Highly Feasible
	• Language consistency	100%	Highly Feasible
	• Language attractiveness	100%	Highly Feasible
Language Validation (Test Administration Guide)	• Language clarity	100%	Highly Feasible
	• Presentation structure	93.33%	Highly Feasible
	• Suitability for classroom context	80.00%	Feasible
	• Language tone	100%	Highly Feasible
Language Validation (Scoring Guide)	• Clarity of score descriptions	100%	Highly Feasible
	• Language practicality	100%	Highly Feasible
	• Terminology consistency	100%	Highly Feasible

Validation Aspect	Assessment Indicator	Percentage	Category
Classroom Teacher Validation	• Explanatory examples	100%	Highly Feasible
	• Clarity of instructions	80.00%	Feasible
	• Feasibility of activities	73.33%	Feasible
	• Clarity of teacher guidelines	73.33%	Feasible
	• Suitability with student characteristics	80.00%	Feasible
	• Conciseness and practicality	73.33%	Feasible
	• Ease of scoring	73.33%	Feasible

The results indicate that the instrument met the feasibility criteria across all assessed aspects. The indicators, stimulus design, and scoring system were considered appropriate for the developmental characteristics of elementary school students and aligned with the objectives of measuring figural creativity. Expert evaluation from the psychological perspective confirmed that the tasks were suitable for students' cognitive and creative development stages.

High ratings on creativity, emotional engagement, and motivational indicators suggest that the figural activities provide opportunities for students to express ideas without excessive pressure. These findings are consistent with research emphasizing the effectiveness of open-ended figural tasks in stimulating creative responses (Potters et al., 2023). Although the cognitive load and fairness indicators were categorized as feasible, the complexity of visual stimuli must still be carefully controlled to avoid overburdening some students. According to Cognitive Load Theory (Sweller, 2011), learning tasks should be designed in accordance with learners' working memory capacity. Differentiation of tasks can further enhance creativity while maintaining equitable access for students (Witarsa & Sofiarni, 2024). Therefore, proportional cognitive load management becomes an important consideration in the design of inclusive assessment instruments.

From a linguistic perspective, the instrument achieved a high level of readability and clarity, allowing students to understand instructions independently. Clear language contributes to the accuracy and validity of assessments because students are able to interpret tasks consistently (Kouamé, 2010; Rabbani & Reinita, 2025; Rahmawati & Sulistyono, 2021). In addition, the scoring guide received very high feasibility ratings due to its clear score descriptions, consistent terminology, and illustrative examples. Clear rubrics are essential in creativity assessment because subjective interpretation can influence scoring decisions. Explicit and operational scoring criteria help improve consistency among raters (Ling, 2024; Windyariani & Setiono, 2024; W. Xu & Tognolini, 2022). Teacher validation results also indicated that the instrument was practical for classroom use, although teachers suggested providing concise guidance to facilitate implementation. This finding highlights the importance of teacher literacy and perception in determining the effectiveness of assessment practices in classroom contexts (Maggi et al., 2025; Nissa et al., 2024; Quainoo et al., 2025).

Readability and Feasibility Testing

The readability test was conducted to evaluate students' understanding of the instructions and the feasibility of implementing the instrument in classroom settings. Observations during the testing process indicated that students were able to complete the tasks with minimal additional explanation from teachers. The results demonstrate a very high level of comprehension and task implementation. Students showed active participation and were willing to attempt all activities provided. The absence of task rejection suggests that the instrument created a sense of comfort and psychological safety during the testing process. Instructional clarity plays an important role in improving response quality and student engagement (Oschwald et al., 2025; Roy et al., 2018).

The allocated time was generally sufficient for students to complete the activities, indicating that the complexity of the tasks was appropriately balanced. However, variations in individual working pace were observed, which is a natural characteristic of creative processes. Differences in response tempo reflect the diverse ways students generate and elaborate ideas (Beck et al., 2016; Syafrial & Kadir, 2025) Classroom observations also revealed occasional discussions among students and attempts to imitate peers' responses. Such behavior highlights the importance of effective classroom management to maintain the authenticity of student responses during creativity assessments (Zhao et al., 2022). Despite these dynamics, the diversity of ideas and the emergence of original drawings demonstrate that the instrument is capable of capturing a wide range of creative expressions. Student response questionnaires further support these findings, as shown in Table 5, showing high percentages for clarity of instructions, comfort during task completion, and perceived usefulness of the activities. When students perceive assessments as supportive of learning, their motivation and confidence tend to increase, encouraging greater creative expression (Shao et al., 2024).

Table 5. Results of the Student Response Questionnaire

Indicator	Percentage
Clarity of instructions	100%
Attractiveness of the activities	96%
Absence of perceived difficulty	93%
Comfort during the activities	99%
Perceived usefulness of the activities	100%

Beyond its psychometric properties, the developed instrument also provides practical value for classroom assessment. By analyzing students' responses across the dimensions of fluency, flexibility, originality, and elaboration, teachers can obtain a clearer profile of students' creative thinking. This multidimensional information helps teachers identify strengths and limitations in students' creativity and design learning activities that better support the development of creative thinking in classroom settings (Alfarisa et al., 2025).

Internal Structure Validity of the Instrument

Internal validity evidence was obtained by examining the empirical relationships between stimulus scores, activity scores, and the overall creativity dimensions measured by the instrument, as presented in Table 6.

Table 6. Spearman Rank Correlation Results for Stimulus Validity in Activity 2

Stimulus	Creativity Dimension	Spearman's rho	p-value	Interpretation
Stimulus 1	Total Originality Score	0.573	< .001	Valid
Stimulus 2	Total Originality Score	0.618	< .001	Valid
Stimulus 3	Total Originality Score	0.510	< .001	Valid
Stimulus 4	Total Originality Score	0.442	< .001	Valid
Stimulus 5	Total Originality Score	0.409	< .001	Valid
Stimulus 1	Total Elaboration Score	0.608	< .001	Valid
Stimulus 2	Total Elaboration Score	0.574	< .001	Valid
Stimulus 3	Total Elaboration Score	0.742	< .001	Valid
Stimulus 4	Total Elaboration Score	0.739	< .001	Valid
Stimulus 5	Total Elaboration Score	0.744	< .001	Valid

The validity test of stimuli in Activity 2 showed that all stimuli were significantly and positively correlated with the total scores of originality and elaboration dimensions. This indicates that the stimuli contribute meaningfully to the measurement of the intended creativity constructs. The results suggest that the level of difficulty and openness of the stimuli was sufficient to generate varied creative responses from students. The use of open-ended stimuli is consistent with the theoretical design principles of creativity assessments, which aim to balance task structure with freedom of response (Silvia et al., 2008). This design approach aligns with the conceptual framework of the TTCT, which encourages divergent thinking through flexible visual prompts.

The task-level validity test also revealed significant correlations between activity scores and the total scores of each creativity dimension, as presented in Table 7. These results indicate measurement stability across different tasks within the instrument. Consistency across activities supports the stability of the measured construct and strengthens the internal structure of the assessment (Xu et al., 2025). Nevertheless, variations in correlation strength across certain dimensions were observed. Such variations reflect the multidimensional and contextual nature of creativity, which may manifest differently depending on the cognitive demands of each task (Vuichard et al., 2023). Rather than indicating measurement inconsistency, these variations suggest that the instrument is sensitive to differences in task characteristics and student creative processes.

Table 7. Activity-Level Validity Test Results

Creativity Dimension	Activity	Correlated Score	Spearman's rho	p-value	Inter-pretation
Fluency	Activity 1	Total Fluency Score	0.869	< .001	Valid
Fluency	Activity 3	Total Fluency Score	0.460	< .001	Valid
Flexibility	Activity 1	Total Flexibility Score	0.731	< .001	Valid
Flexibility	Activity 3	Total Flexibility Score	0.711	< .001	Valid
Originality	Activity 2	Total Originality Score	0.773	< .001	Valid

Originality	Activity 3	Total Originality Score	0.745	< .001	Valid
Elaboration	Activity 2	Total Elaboration Score	0.901	< .001	Valid
Elaboration	Activity 3	Total Elaboration Score	0.888	< .001	Valid

Inter-Rater Reliability of Creativity Scoring

Reliability analysis was conducted using inter-rater reliability to determine the consistency of scoring among evaluators. The analysis employed the Intraclass Correlation Coefficient (ICC) across the four creativity dimensions measured in the instrument.

Table 8. Inter-rater Reliability Results

Creativity Dimension	ICC (2,k) Point Estimate	Lower 95% CI	Upper 95% CI	Interpretation
Fluency	0.646	0.363	0.818	Good
Flexibility	0.754	0.554	0.874	Good
Originality	0.194	-0.090	0.480	Poor
Elaboration	0.779	0.151	0.924	Good

The results indicate that three creativity dimensions: fluency, flexibility, and elaboration, achieved good reliability levels. This suggests that score variations in these dimensions primarily reflect differences in students' creative abilities rather than inconsistencies among raters. These findings align with the principles of Generalizability Theory, which emphasize the importance of multiple raters in identifying and minimizing measurement errors associated with rater effects (Kusumadewi & Rosnawati, 2025; Mancar & Gulleroglu, 2022).

In addition, the use of structured analytic rubrics contributes to improving scoring consistency because clear and operational criteria guide evaluators in interpreting student responses (Limgomolvilas & Sukserm, 2025). The relatively high ICC values therefore indicate that the scoring rubric developed in this study functions effectively in supporting reliable assessment. However, the originality dimension produced a relatively low ICC coefficient, indicating that agreement among raters was not optimal. This result can be explained by the conceptual nature of originality as one of the most abstract and context-dependent dimensions of creativity. Evaluating originality often requires comparative judgment regarding the relative novelty of a response among a set of responses, which increases the potential for subjective interpretation (Bahar & June Maker, 2025).

The relatively wide confidence interval for originality further indicates variability among raters. Previous studies suggest that rater training, calibration sessions, and the provision of representative scoring examples can improve consistency in rubric-based assessments (Ling, 2024). Rater orientation and calibration meetings prior to formal scoring may help align evaluators' interpretations of scoring criteria, although their effectiveness may vary depending on the complexity of the task (Halliday & Patterson, 2025). These findings also indicate the need for further refinement of the originality scoring procedure. Because originality judgments rely on the relative novelty of responses within a group, insufficiently operationalized criteria may lead to differences in raters' interpretations. Improving the scoring rubric by providing clearer descriptors, representative examples of student

responses, and more explicit decision rules may reduce ambiguity during evaluation. In addition, systematic rater training and calibration sessions are important to align evaluators' understanding of the originality criteria and to improve scoring consistency. Such refinement would help ensure that originality scores more accurately reflect students' creative potential rather than variations in rater judgment (Pellegrino et al., 2025). Therefore, the lower reliability observed in the originality dimension does not necessarily indicate a weakness in the construct being measured. Instead, it highlights the need for further refinement of the scoring guidelines and the implementation of systematic rater training procedures to ensure more consistent application of rubric criteria.

4. Conclusion

This study developed a figural creativity assessment instrument based on the Torrance Tests of Creative Thinking approach for elementary school students in phases B and C. The instrument was produced through a systematic research and development process using the 4D model, including the stages of define, design, develop, and limited dissemination. The final product consists of structured figural tasks designed to measure four key dimensions of creativity: fluency, flexibility, originality, and elaboration.

The development process demonstrated that the instrument is feasible for use in elementary school contexts. Expert validation confirmed that the tasks, instructions, and scoring procedures are appropriate for the cognitive and creative characteristics of elementary students. The language used in the instrument was considered clear and accessible, allowing students to understand and complete the activities independently. Classroom testing further indicated that the activities were engaging and manageable within the allocated time, suggesting that the instrument can be practically implemented in real classroom settings. The internal validity analysis showed that the stimuli and tasks function coherently in representing the construct of figural creativity. Relationships among stimuli, activities, and creativity dimensions indicate that the instrument consistently captures variations in students' creative responses. In addition, the scoring rubric supports relatively consistent evaluations among raters for most creativity dimensions, indicating that the scoring procedures are sufficiently operational for classroom assessment.

Overall, the results indicate that the developed instrument successfully fulfills its objective as a tool for assessing figural creativity among elementary school students. The instrument may serve as a diagnostic and evaluative tool to help teachers better understand students' creative thinking patterns and to support more creative and responsive learning practices in the classroom. Future studies may extend this work by implementing the instrument in broader educational contexts, refining scoring procedures, and conducting further psychometric testing to strengthen its measurement properties.

References

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Alfarisa, F., Arisetyawan, A., Maulana, F. R., Lathifah, N. Y., Afriani, S., & Mashudi, M. (2025). Creative Thinking in Elementary Education: Reliable Assessment Instruments for Science and Social Studies. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *29*(1), 13–25. https://doi.org/10.21831/pep.v29i1.79429

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*. https://doi.org/10.3389/fpsyg.2025.1545396

Beck, S. R., Williams, C., Cutting, N., Apperly, I. A., & Chappell, J. (2016). Individual Differences in Children's Innovative Problem-Solving Are Not Predicted by Divergent Thinking or Executive Functions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1690). https://doi.org/10.1098/rstb.2015.0190

Brandt, W. C. (2021). *Measuring student success skills: A review of the literature on creativity*. www.nciea.org

Carroll, J. B., & Guilford, J. P. (1968). The Nature of Human Intelligence. *American Educational Research Journal*, *5*(2). https://doi.org/10.2307/1161820

Cohen, Jacob. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates. https://doi.org/https://doi.org/10.4324/9780203771587

Craft, A. (2005). *Creativity in Schools* (A. Craft, Ed.). Routledge. https://doi.org/10.4324/9780203357965

Cronbach, L. J. (1951). Coefficient Alpha and The Internal Structure of Tests. *Psychometrika*, *16*(3), 297–334.

DeVellis, R. F. (2017). *Scale Development : Theory and Applications*. SAGE.

Fauziah, M., Marmoah, S., & Murwaningsih, T. (2020). The Effectiveness of Thematic Learning with Thinking Actively in a Social Context and Creative Problem Solving Learning Models on Divergent Thinking Skills. *International Journal for Educational and Vocational Studies*, *2*(3). https://doi.org/10.29103/ijevs.v2i3.2197

Global Partnership for Education. (2020). *21st Century Skills: What potential role for the Global Partnership for Education?*

Halliday, D. A., & Patterson, B. J. (2025). Rater Training and Assessment of Student Clinical Judgment: An Experimental Inquiry. *Nursing Education Perspectives*, *46*(4), 209–215. https://doi.org/10.1097/01.NEP.0000000000001399

Hickman, R. (2023). Assessment, creativity and learning: A personal perspective. *Future in Educational Research*, *1*(2), 104–114. https://doi.org/10.1002/fer3.19

Hidayah, N., Gunarhadi, & Karsono. (2024). Elementary School Students' Learning Styles: Perspective of Creative Thinking Skills. *Jurnal Ilmiah Sekolah Dasar*, *8*(2), 202–211. https://doi.org/10.23887/jisd.v8i2.68385

Istiqomah, N., & Na'imah. (2025). Education Transformation: Paradigm Shifts in Learning in the Age of Technology. *Journal of Educational Sciences*, *9*(3), 1546–1557. https://doi.org/10.31258/jes.9.3.p.1546-1557

Kanlı, E. (2021). Assessment of Creativity: Theories and Methods. In *Creativity - A Force to Innovation*. IntechOpen. https://doi.org/10.5772/intechopen.93971

Kemendikbudristek. (2022). *Dimensi, Elemen, dan Subelemen Profil Pelajar Pancasila pada Kurikulum Merdeka*.

Kim, K. H. (2006). Can We Trust Creativity Tests? A Review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, *18*(1), 3–14. https://doi.org/10.1207/s15326934crj1801_2

Kim, K. H. (2017). The Torrance Tests of Creative Thinking - Figural or Verbal: Which One Should We Use? *Creativity. Theories – Research - Applications*, *4*(2), 302–321. https://doi.org/10.1515/ctra-2017-0015

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kouamé, J. B. (2010). Using Readability Tests to Improve the Accuracy of Evaluation Documents Intended for Low-Literate Participants. *Ideas to Consider Journal of MultiDisciplinary Evaluation*, *6*(14). http://www.jmde.com/

Kusumadewi, P. D. A., & Rosnawati, R. (2025). Assessing the Reliability of Performance-Based Evaluation in Fashion Education: A Generalizability Theory Approach. *Journal of Education Research and Evaluation*, *9*(3), 631–641. https://doi.org/10.23887/jere.v9i3.100248

Leasa, M., Batlolona, J. R., & Talakua, M. (2021). Elementary students' creative thinking skills in science in the Maluku Islands, Indonesia. *Creativity Studies*, *14*(1), 74–89. https://doi.org/10.3846/cs.2021.11244

Limgomolvilas, S., & Sukserm, P. (2025). Examining Rater Reliability When Using an Analytical Rubric for Oral Presentation Assessments. *LEARN Journal: Language Education and Acquisition Research Network*, *18*(1), 110–134. https://doi.org/10.70730/JQGY9980

Ling, J. H. (2024). Pedagogy: Indonesian Journal of Teaching and Learning Research A Review of Rubrics in Education: Potential and Challenges. *OPEN ACCESS JOURNAL Pedagogy: Indonesian Journal of Teaching and Learning Research*, *2*(1), 1–14.

Maggi, D. P. T., Yun, P. T., Aqil, M., Zulnuraini, & Fasli, M. (2025). Analysis of Teachers' Creativity in SBDP Learning Class V SDN 24 Palu. *Journal of Educational Sciences*, *9*(3), 1334–1341. https://doi.org/10.31258/jes.9.3.p.1334-1341

Mancar, S. A., & Gulleroglu, H. D. (2022). Comparison of Inter-Rater Reliability Techniques in Performance-Based Assessment. *International Journal of Assessment Tools in Education*, *9*(2), 515–533. https://doi.org/10.21449/ijate.993805

Munastiwi, E. (2023). Merdeka Belajar: Solution for Developing Creativity and Independence in Early Childhood in the Era of the Industrial Revolution 4.0. *JPI (Jurnal Pendidikan Indonesia)*, *12*(4), 615–626. https://doi.org/10.23887/jpiundiksha.v12i4.61500

Nissa, K., Nurbadriyah, F., Jayanti, S. N., Firdausa, R. A., Sa'diyah, H., Darmawan, P., & Kusumawardani, A. (2024). Persepsi Guru terhadap Implementasi Asesmen Diagnostik dalam Pembelajaran yang Berpusat pada Siswa Sekolah Dasar. *Journal of Innovation and Teacher Professionalism*, *2*(3), 309–319. https://doi.org/10.17977/um084v2i32024p309-319

Oschwald, A., Moeller, J., Kracke, B., Viljaranta, J., & Dietrich, J. (2025). Student Motivation and Instructional Clarity: Linking Experience Sampling Method Data to Objective Behavioural Observations. *British Journal of Educational Psychology*, *95*(S1), S281–S299. https://doi.org/10.1111/bjep.12775

Pellegrino, G., Saretzki, J., & Benedek, M. (2025). Controlling Rater Effects in Divergent Thinking Assessment: An Item Response Theory Approach to Individual Response and Snapshot Scoring. *Journal of Intelligence*, *13*(6). https://doi.org/10.3390/jintelligence13060069

Piaget, J. (1952). *The origins of intelligence in children*. W W Norton & Co. https://doi.org/10.1037/11494-000

Potters, O. T. A., van Schijndel, T. J. P., Jak, S., & Voogt, J. (2023). Two Decades of Research on Children's Creativity Development During Primary Education in Relation to Task Characteristics. In *Educational Research Review* (Vol. 39). Elsevier Ltd. https://doi.org/10.1016/j.edurev.2023.100532

Quainoo, E. A., Asamoah, D., Baidoo-Anu, D., Ye, J., & Akuamoah Boateng, C. (2025). Examining Profiles of Classroom Assessment Literacy among Basic School Teachers. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2558163

Rabbani, S., & Reinita. (2025). Analysis of the Validity of Sparkol Videoscribe Media in Pancasila Education Learning in Grade IV of Elementary School. *Journal of Educational Sciences*, *9*(6), 5635–5646. https://doi.org/10.31258/jes.9.6.p.5635-5646

Rahmawati, L. E., & Sulistyono, Y. (2021). Assessment and Evaluation on Text Readability in Reading Test Instrument Development for BIPA-1 to BIPA-3. *Asian Journal of University Education*, *17*(3), 51–57. https://doi.org/10.24191/ajue.v17i3.14522

Usnia, Lagandasa, Y. R., Lapasere, S., Dewi, A. I., & M, A. R. (2025). The Role of Teachers in Increasing Motivation to Learn Arts and Culture Subjects for Class IV SDN 1 Tanantovea. *Journal of Educational Sciences*, *9*(5), 3256–3266. https://doi.org/10.31258/jes.9.5.p.3256-3266

Roy, S., Beer, C., & Lawson, C. (2018). The Importance of Clarity in Written Assessment Instructions. *Journal of Further and Higher Education*, *44*(2), 143–155. https://doi.org/10.1080/0309877X.2018.1526259

Runco, M. A. . (2023). *Creativity : Rsearch, development and practice*. Elsevier.

Runco, M. A. ., & Acar, Selcuk. (2024). *Handbook of creativity assessment*. Edward Elgar Publishing.

Runco, M. A., & Acar, S. (2012). Divergent Thinking as an Indicator of Creative Potential. *Creativity Research Journal*, *24*(1), 66–75. https://doi.org/https://doi.org/10.1080/10400419.2012.652929

Saputri, W. D., Mahardani, A. J., & Wulansari, B. Y. (2024). Enhancing Early Childhood Creativity through Project-Based Learning: Indonesian

Traditions at KJRI Penang. *Lectura : Jurnal Pendidikan*, *15*(2), 497–508. https://doi.org/10.31849/lectura.v15i2.20631

Sari, N., & Zulaikha, S. (2020). Vocational School Curriculum Development Design In The Era Of The 4.0 Industry Revolution. *EDUCATIO : Journal Of Education*, *5*(2). https://www.

Shao, Y., Liu, Q., Dong, Y., & Liu, J. (2024). Perceived formative assessment practices in homework and creativity competence: The mediating effects of self-confidence in learning and intrinsic motivation. *Studies in Educational Evaluation*, *83*, 101376. https://doi.org/10.1016/j.stueduc.2024.101376

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing Creativity With Divergent Thinking Tasks: Exploring the Reliability and Validity of New Subjective Scoring Methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 68–85. https://doi.org/10.1037/1931-3896.2.2.68

Sternberg, R. J., & Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity*. Free Press.

Sweller, J. (2011). Cognitive Load Theory. In *The psychology of learning and motivation: Cognition in education* (pp. 37–76). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Syafrial, & Kadir, A. (2025). Exploration of Differences In Students Creative Thinking Abilities Viewed from Perspective Gender and Domicile: Rasch Model Analysis. *Journal of Biology Learning*, *85*(2), 2623–1476. https://doi.org/10.32585/jbl.v7i2.7642

Tepaso, A., Yun, R. L., Aqil, M., Hariana, K., & Pahriadi. (2025). Analysis of the Role of Teachers in Arts and Culture Learning for Grade IV Students of SD Inpres 8 Mamboro. *Journal of Educational Sciences*, *9*(5), 3438–3448. https://doi.org/10.31258/jes.9.5.p.3438-3448

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*.

Torrance, E. P. (1966). Torrance Tests of Creative Thinking. *Educational and Psychological Measurement*.

Torrance, E. P. (1998). *The Torrance Tests of Creative Thinking Norms—Technical Manual Figural (Streamlined) Forms A & B*.

Torrance, E. Paul. (1974). *Torrance Tests of Creative Thinking : Directions Manual and Scoring Guide, Figural Test Booklet B*. Personnel Press.

Trilling, Bernie., & Fadel, Charles. (2013). *21st century skills: Learning for life in our times*. Jossey-Bass.

UNESCO. (2016). *Fostering creativity for the 21st century*. http://www.unesco.org/open-access/terms-use-ccbysa-en

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2024). Can elementary school teachers assess students' creative problem solving abilities? *Teaching and Teacher Education*, *146*, 104644. https://doi.org/10.1016/j.tate.2024.104644

Vuichard, A., Botella, M., & Capron Puozzo, I. (2023). Creative Process and Multivariate Factors through a Creative Course "Keep Calm and Be Creative." *Journal of Intelligence*, *11*(5). https://doi.org/10.3390/jintelligence11050083

WEF. (2023). Future of jobs report 2023. In *World Economic Forum* (Number January).

Widoyoko, E. P. (2018). *Teknik Penyusunan Instrumen Penelitian* (7th ed.). Pustaka Pelajar.

Windyariani, S., & Setiono, S. (2024). Performance Assessment Creative Thinking Rubric in Science Technology Engineering and Mathematics (STEM) Learning. *JISE*, *13*(2), 64–73. https://journal.unnes.ac.id/journals/jise/

Witarsa, R., & Sofiarni, E. (2024). Kreativitas Siswa Sekolah Dasar melalui Pembelajaran Berdiferensiasi. In *Journal of Education Research* (Vol. 5, Number 2).

Woolfolk, A. (2016). *Educational Psychology*. Pearson.

Xu, S., Reiss, M. J., & Lodge, W. (2025). Validation and Use of the Comprehensive Scientific Creativity Assessment (C-SCA) Instrument for Secondary School Students. *Creativity Research Journal*. https://doi.org/10.1080/10400419.2024.2448995

Xu, W., & Tognolini, J. (2022). Build an Assessment Rubric of Student Creativity in Higher Education. *International Conference on Higher Education Advances*, *2022-June*, 135–142. https://doi.org/10.4995/HEAd22.2022.14695

Zhao, L., Mao, H., Compton, B. J., Peng, J., Fu, G., Fang, F., Heyman, G. D., & Lee, K. (2022). Academic Dishonesty and Its Relations to Peer Cheating and Culture: A Meta-Analysis of the Perceived Peer Cheating Effect. In *Educational Research Review* (Vol. 36). Elsevier Ltd. https://doi.org/10.1016/j.edurev.2022.100455

How to cite this article:

Mirza, A. N., Hastomo, A., & Firmansyah. (2026). Development of a Figural Creativity Assessment Instrument Based on Torrance Tests of Creative Thinking (TTCT) for Elementary School Students. *Journal of Educational Sciences*, *10(4)*, 424-442.