Atfaluna
Journal of Islamic Early Childhood Education
https://journal.iainlangsa.ac.id/index.php/atfaluna

Vol. 8 No. 2 (December, 2025)
pp.1-19

Assessing Divergent and Convergent Creativity in Early Childhood Drawings: A Multi-Task Deep Learning Approach

Anik Indarwati[1]✉, Frangky Tupamahu[2], Suharti[3], Firsta Hannni Enggaring Galih[4], Satya Raj Joshi[5]

[1] Department of Psychology, Universitas Muhammadiyah Gorontalo, Indonesia
[2] Department of Information Systems, Universitas Muhammadiyah Gorontalo, Indonesia
[3] Department of Early Childhood Teacher Education, Universitas Negeri Surabaya, Indonesia
[4] Department of Medical Doctor, Universitas Negeri Gorontalo, Indonesia
[5] Faculty of Humanity and Social Science, Bishwa Shanti Chiran Millan Campus, Tribhuvan University, Nepal

Corresponding e-mail: anikindarwaty@umgo.ac.id
http://dx.doi.org/10.32505/atfaluna.v8i2.1234

Submission : August 14, 2025
Revision : October 20, 2025
Accepted : December 21, 2025

Abstract

This study aims to develop and evaluate a deep learning model for the comprehensive assessment of divergent and convergent creativity dimensions. The dataset comprised 102 digital drawings obtained from Indonesian children aged 4 to 6 years using the Test for Creative Thinking - Drawing Production (TCT-DP). This study employed a quantitative model development approach, where ground-truth labels were derived from the 14 TCT-DP scoring criteria aggregated into divergent and convergent scores through label engineering. Using a Multi-Task Convolutional Neural Network (MT-CNN) based on MobileNetV2 architecture, the study analyzed extracted visual features to predict expert-rated scores.The results revealed a strong positive correlation (r = +0.51) between divergent and convergent thinking scores, challenging the traditional view of these processes as antagonistic and supporting an integrated model of creative cognition. From a technical perspective, the model demonstrated satisfactory predictive capability as a proof-of-concept, achieving a lower error rate for convergent scores (RMSE = 1.52) compared to divergent scores (RMSE = 1.97). This indicates that while structured convergent features are more machine-learnable, the abstract nature of divergent thinking remains a complex challenge. In conclusion, this study validates the feasibility of automated creativity assessment while offering empirical evidence for the synergy between generative and evaluative thinking in early childhood.

Keywords: *Creativity Assessment; Computational Creativity; Deep Learning; Early Childhood; Multi-Task Learning*

p-ISSN: 2655-8572 e-ISSN: 2655-8009

A. INTRODUCTION

Creativity is widely recognized as a vital 21st-century competency, crucial not only for personal development but also for societal progress in an era increasingly defined by complexity and rapid change (Kaufman, 2016; Robinson & Aronica, 2016). In the context of early childhood education, creativity serves as a foundational domain that intersects cognitive, emotional, and social development. Numerous scholars have emphasized that fostering creativity from an early age is critical, as it contributes to children's future capabilities in problem-solving, innovation, and adaptability (Dilnoza, 2023; Mamun, 2024; Sari et al., 2021). Drawing, in particular, is a natural medium through which children externalize their inner worlds, experiment with symbolic representation, and express original ideas (Omatseye & Emeriewen, 2013; Sari et al., 2021).

Despite its pedagogical significance, assessing creativity, especially through children's drawings, remains a persistent challenge. Traditional assessment methods are often subjective, time-consuming, and susceptible to inter-rater inconsistencies (Tarigan & Fadillah, 2022; Xu et al., 2025). Manual scoring typically relies on an assessor's interpretation of visual elements, which introduces variability and reduces reliability. Given these limitations, there is a growing need for more objective, scalable, and data-driven approaches to creativity assessment, particularly in early childhood settings.

Recent advancements in artificial intelligence (AI), and specifically deep learning (DL), offer promising avenues to transform how creativity is assessed. DL's ability to model complex, non-linear relationships in high-dimensional data has shown considerable success in computer vision tasks, including object recognition, scene understanding, and artistic style transfer (Chai et al., 2021; Guo et al., 2021; Tsirtsakis et al., 2025). In educational contexts, Convolutional Neural Networks (CNNs) have been increasingly adopted to analyze children's visual art to infer emotional states (Çiftçi et al., 2025; He et al., 2022; Khare et al., 2024) or psychological traits (Nath & Stevenson, 2024). Furthermore, a growing body of work has demonstrated that CNNs can approximate human judgments in rating visual creativity, reinforcing the feasibility of computational creativity assessment (Gao, 2025; Zhao, 2024).

However, most existing AI-based creativity assessment tools evaluate creativity as a unidimensional construct or focus solely on divergent thinking, typically defined as the capacity to generate multiple, novel, and diverse ideas (Grilli et al., 2024; Hadas & Hershkovitz, 2025). It overlooks the complementary and equally critical role of convergent thinking, which refers to the ability to evaluate, refine, and integrate ideas into coherent, meaningful outcomes (Mhlongo et al., 2023). Contemporary theories of creativity increasingly view it as a dynamic interplay between divergence and convergence, both of which are essential for the production of truly creative outcomes (Rawlings et al., 2025; Raya et al., 2023). Neglecting either aspect may result in an incomplete assessment and an inaccurate representation of a child's creative potential.

Against this backdrop, the scientific novelty of this study lies in the development of an integrated deep learning model capable of simultaneously assessing both divergent and convergent aspects of creativity based on children's digital drawings. Unlike previous

p-ISSN: 2655-8572 e-ISSN: 2655-8009

approaches that assess creativity from a single-dimensional perspective, this study employs a Multi-Task Convolutional Neural Network (MT-CNN), which is specifically designed to handle multiple related outputs within a unified learning framework (Karamizadeh et al., 2025; Marullo et al., 2023; Tao et al., 2025). By doing so, the model not only enhances predictive accuracy but also captures the holistic nature of creativity more effectively.

The primary research problem addressed in this study is how to design and validate an AI-based creativity assessment model that is both objective and developmentally sensitive. We hypothesize that an MT-CNN architecture utilizing a transfer learning approach based on MobileNetV2 can successfully predict both divergent and convergent creativity scores from a single image input with reasonable accuracy. Furthermore, this study explores whether these two dimensions of creativity are correlated in early childhood, thus contributing to the ongoing discourse about the developmental structure of creativity.

This study aims to fill a critical gap in both the educational and AI literature by offering a proof-of-concept for a technically robust, theoretically grounded, and practically applicable model for creativity assessment. Rather than serving as a definitive validation, this research functions as a feasibility study exploring whether deep learning architectures can effectively distinguish between divergent and convergent creativity features in early childhood drawings, despite the constraints of limited data availability.

B. METHOD

The methodology for this study was systematically structured to ensure that the development and evaluation process of the deep learning model could be conducted validly and reliably. This section encompasses the research design, the subjects and objects of the study, the instruments employed, the data collection procedures, and the in-depth data analysis techniques.

1. Research Design

This study employed a quantitative research design with a model development approach. The primary focus of this design was to develop and test a computational model, specifically a hybrid deep learning model, capable of comprehensively assessing both the divergent and convergent aspects of creativity. This design is commonly used in AI-based educational research to train and evaluate machine learning models against benchmarked datasets (Chandrasekera et al., 2025; Vandenhende et al., 2021). The research did not involve direct intervention with human subjects; instead, it utilized their creative outputs as the input data to train and evaluate the performance of the artificial intelligence model. The ultimate objective was to produce a predictive model whose accuracy could be statistically measured against assessments provided by human experts.

2. Ethical Considerations

This study was conducted in strict adherence to ethical guidelines for research involving human subjects and artificial intelligence in education. Formal ethical approval was obtained from the Research Ethics Committee of Universitas Muhammadiyah Gorontalo (Approval Number: 018/PK-Pen/LPPM-UMGO/V/2025).

As the research utilized secondary data, the dataset was originally collected during the National Children's Day Program aimed at enhancing the creativity of children aged 4–6 in Pohuwato Regency, conducted at 10 kindergartens supervised by the Department of Education and Culture of Pohuwato Regency. Prior to the primary data collection, written informed consent was obtained from the parents or legal guardians of all participating children. The consent forms explicitly authorized the use of the children's artwork for educational research and subsequent academic analysis. Permission to repurpose this dataset specifically for machine learning training was formally granted by the Department of Education and Culture of Pohuwato Regency and the National Unity and Politics Agency *(Kesbangpol)* of Gorontalo Province.

To ensure child privacy and confidentiality, a rigorous anonymization protocol was implemented before the data were accessed for this study. All personal identifiers (names, dates of birth, class information) were removed and replaced with unique alphanumeric codes (e.g., IMG_001). The research team had access only to the de-identified digital images and their corresponding TCT-DP scores, ensuring that no specific drawing could be traced back to an individual child during the computational analysis.

Data Availability Statement: The de-identified dataset used to support the findings of this study, including the extracted feature vectors and code for the Multi-Task CNN model, is available from the corresponding author upon reasonable request. The raw digital drawings are not publicly archived to protect the privacy of the minor participants, in accordance with the institution's data protection policy.

3. Research Subjects and Objects

The objects of this study were 102 drawings produced by children aged 4 to 6 years. These images were in a black-and-white digital format (JPEG) and resulted from the administration of a standardized creativity test. These drawings were collected through the administration of the standardized Test for Creative Thinking – Drawing Production (TCT-DP), a widely recognized instrument for assessing visual creativity in young children (Desmet et al., 2021; Urban, 2005). The children who created the drawings served as the original data source, while the analysis focused entirely on the image data and corresponding creativity scores. This non-invasive design is consistent with ethical guidelines for AI in education, as no personal identifiers or direct interventions were employed (Ali et al., 2024; Vieriu & Petrea, 2025). The subjects who served as the data source were the early childhood children who participated in the initial data collection process. However, the analysis in this research focused entirely on the digital artifacts, the drawings themselves, and their accompanying score data.

4. Research Instruments

Three primary instruments were utilized in this study. First, the Test for Creative Thinking - Drawing Production (TCT-DP) by Urban (2005) was employed as a standardized instrument to stimulate and elicit creative drawings from the early childhood children. Second, the TCT-DP Scoring Rubric was used by experts (trained psychologists or educators) to quantitatively score each drawing against 14 creativity criteria. This rubric served to generate ground truth label data for research. Third, the primary instrument developed was a Convolutional Neural Network (CNN)-based Deep Learning Model, which adopted a transfer learning approach (e.g., using a base architecture such as MobileNetV2 or ResNet50) and was modified for the specific tasks of this study.

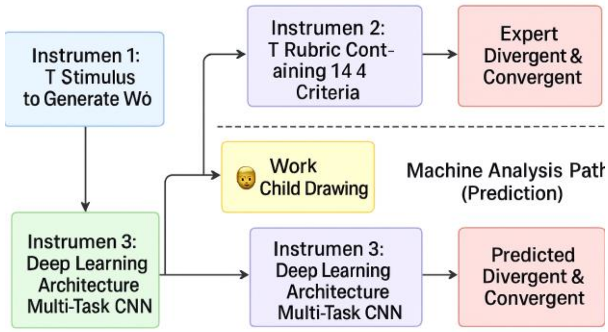


Figure 1. Research Instruments

5. Data Collection

The data utilized in this study were secondary data collected through a systematic procedure. Early childhood subjects initially received the TCT-DP. The artworks were subsequently digitized as black-and-white JPEG files. Subsequently, the digitization process was followed by expert assessment. To ensure high-quality ground truth labels, two qualified experts in educational psychology (each with a minimum of 5 years of experience in early childhood assessment) independently scored all 102 drawings using the standard 14-criterion TCT-DP rubric.

The inter-rater reliability was evaluated using Cronbach's Alpha to measure consistency between the raters. The analysis yielded a coefficient of 0.87, indicating a high level of agreement. Consequently, the final scores used for training were derived by averaging the scores from both experts. The quantitative evaluation outcomes were consolidated into a .csv file that associates each drawing with its fourteen creativity scores, thereby facilitating subsequent computational analysis of the dataset.

6. Data Analysis

Data analysis constitutes the central phase of this research, encompassing a series of computational processes ranging from initial data preparation to final model evaluation. The computational framework was implemented using the Python 3.9 programming language (Van Rossum & Drake, 2009). Deep learning architectures were constructed using TensorFlow 2.10 (Abadi et al., 2016), while data processing and performance evaluation metrics were managed using Scikit-learn 1.2 (Pedregosa et al., 2011). The model development and analysis process was systematically organized into three primary stages.

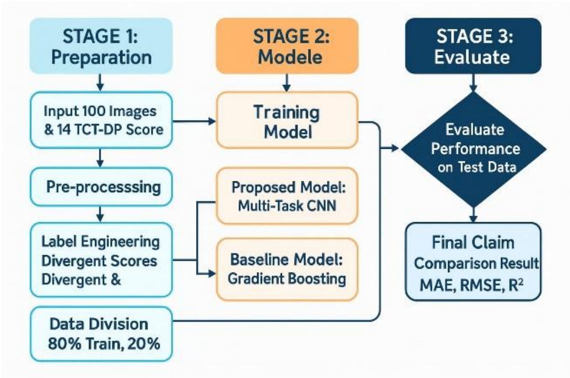


Figure 2. Overall Research Flow Diagram

a. Data Preprocessing and Feature Development Stage

The initial step was data preparation to ensure its suitability and optimality for model training. First, a process of *label engineering* was conducted, wherein the 14 raw score criteria from the TCT-DP were aggregated into two dependent target variables: a Divergent Creativity Score and a Convergent Creativity Score.

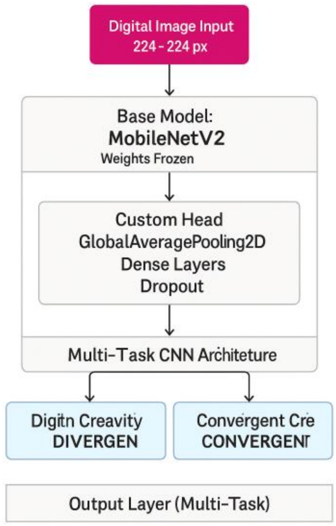


Figure 3. Deep Learning Model Structure

This aggregation was accomplished by summing the scores from the criteria that were theoretically relevant to each respective construct. Subsequently, the image data were uniformly processed by resizing their resolution to 224x224 pixels to match the pre-trained model's input requirements and by normalizing the pixel values to a range of [0, 1]. Given the limited size of the dataset (N=102), data augmentation techniques were extensively applied exclusively to the training set. This augmentation included random rotations (up to 10 degrees), horizontal and vertical shifts (up to 10% of the total dimension), and zooming (up to 10%) to increase data variance and mitigate the risk of overfitting artificially. Finally, the entire dataset was proportionally split into a training set (80%) and a test set (20%), using a stratified sampling method where possible to preserve the original score distribution.

b. Performance Evaluation Process and Comparative Analysis

Model Architecture Specification: The final classification layers of the MobileNetV2 base model were removed, and the weights of the remaining feature extractor were frozen. A custom head designed for Multi-Task Learning was appended directly after the base model's Global Average Pooling 2D layer. This custom head consists of a shared Dense layer with 256 neurons utilizing the ReLU (Rectified Linear Unit) activation function, followed by a Dropout layer with a rate of 0.5 to mitigate overfitting.

Architecture Branching: Immediately following the Dropout layer, the architecture branches into two distinct, independent output layers. Each branch consists of a single Dense neuron with a linear activation function, designed to predict the continuous values for the Divergent Creativity Score and Convergent Creativity Score, respectively. This shared-bottom, multi-tower design allows the model to learn common visual representations before specializing in each specific creativity task.

Hyperparameter Selection and Training Configuration. The training process was configured with specific hyperparameters selected through systematic empirical tuning on a validation subset. We conducted a grid search evaluating learning rates between 1e-3 and 1e-5 and batch sizes of 8, 16, and 32. The optimal configuration identified was the Adam optimizer with an initial learning rate of 1e-4, which offered the best balance between convergence speed and stability.

Training was conducted with a batch size of 16 to accommodate GPU memory constraints while ensuring sufficient gradient updates. The loss function employed for both outputs was Mean Squared Error (MSE). To prevent overfitting, given the small dataset, training was set for a maximum of 100 epochs, incorporating an Early Stopping mechanism configured to halt training if validation loss did not improve for 15 consecutive epochs (patience), automatically restoring the best-performing model weights.

c. Performance Evaluation & Comparative Analysis

The final stage consisted of a quantitative evaluation of the model's performance. First, to provide context and to benchmark the efficacy of the deep learning approach, a baseline model was established using a non-neural network algorithm, such as a Gradient Boosting Regressor. In contrast to the primary model, this baseline was not trained on raw pixel data, but rather on a set of 10-15 hand-crafted features extracted from each image, such as the number of contours, Hu moments, and aspect ratio.

The performance of the proposed deep learning model was subsequently evaluated on the 20% test set, comprising data the model had not previously seen, and its results were directly benchmarked against the performance of the baseline model. The evaluation metrics employed were Mean Absolute Error (MAE) to measure the average prediction error, Root Mean Squared Error (RMSE) to measure the magnitude of the error, and the coefficient of determination (R-squared, or R^2) to measure the proportion of variance in the expert scores that was predictable from the model. This comparison of metrics between the proposed model and the baseline model formed the basis for drawing conclusions regarding the effectiveness and validity of the developed model.

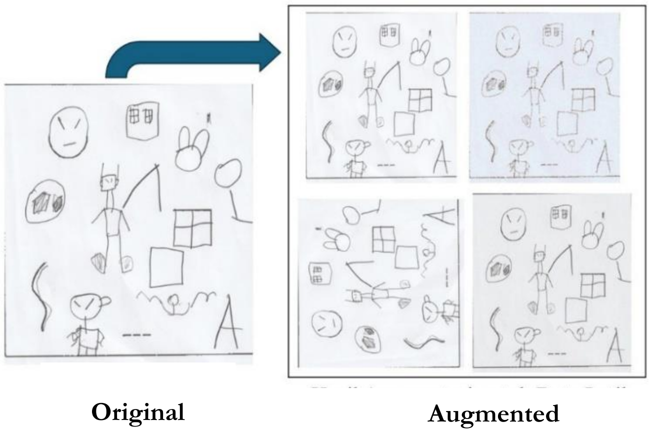


Figure 4. Results of data augmentation.

C. RESULT AND DISCUSSION

This chapter presents a comprehensive analysis of the research findings, which encompasses the characteristics of the dataset used, a quantitative evaluation of the developed deep learning model's performance, and an in-depth discussion regarding the implications of these findings.

1. Result

a. Analyze the characteristics and relationships of dataset variables

Prior to evaluating the model's performance, an exploratory data analysis was conducted on the curated dataset (N=102) to understand the fundamental characteristics of the creativity data. To investigate the internal relationships between the TCT-DP criteria, a correlation analysis was performed. As illustrated in the simplified correlation matrix in Figure 5 (with the full detailed matrix provided in Supplementary Material), a distinct pattern of positive intercorrelations was observed. Notably, criteria associated with structural elaboration, such as Completion and Thematic Connection (r = +0.52), showed strong clustering, empirically supporting the distinctness of the convergent dimension. Furthermore, the overall positive correlation between most criteria reinforces the internal consistency and construct validity of the TCT-DP instrument utilized in this dataset.

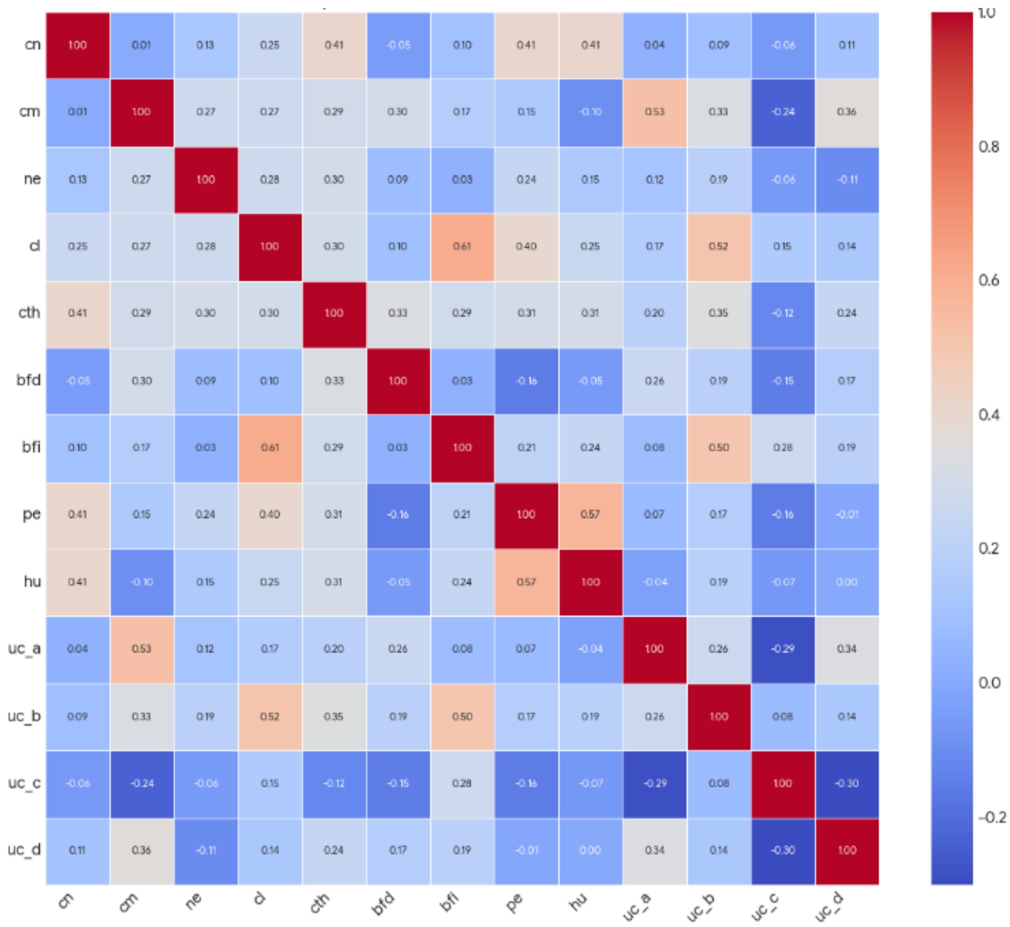


Figure 5. TCT-DP Feature Correlation Heatmap.

Further analysis of the relationship between the aggregated divergent and convergent scores revealed the most significant finding of this stage. As visualized in Figure 6, a strong positive correlation (r = +0.51) was found between these two aspects of creativity. The most significant theoretical contribution of this research, therefore, stems from this discovery of a positive relationship between divergent and convergent thinking. This finding provides empirical evidence that challenges classical views, which often contrast these two processes (Xia et al., 2021), and instead supports a more contemporary and holistic perspective on creativity. Several recent studies have begun to move in a similar direction, suggesting that to achieve high-level creative products, the divergent ability to generate ideas and the convergent ability to evaluate and implement them must work in synergy. Our findings align with this perspective and specifically demonstrate it within the context of early childhood. It implies that mature creativity at this developmental stage is not a choice between being "wildly creative" or "structurally organized," but rather an integrated capacity where the ability to generate original ideas goes hand-in-hand with the ability to organize them into a meaningful whole. Thus, the AI model we have developed does not merely assess two separate aspects, but effectively begins to measure the degree of creative integration in a child.

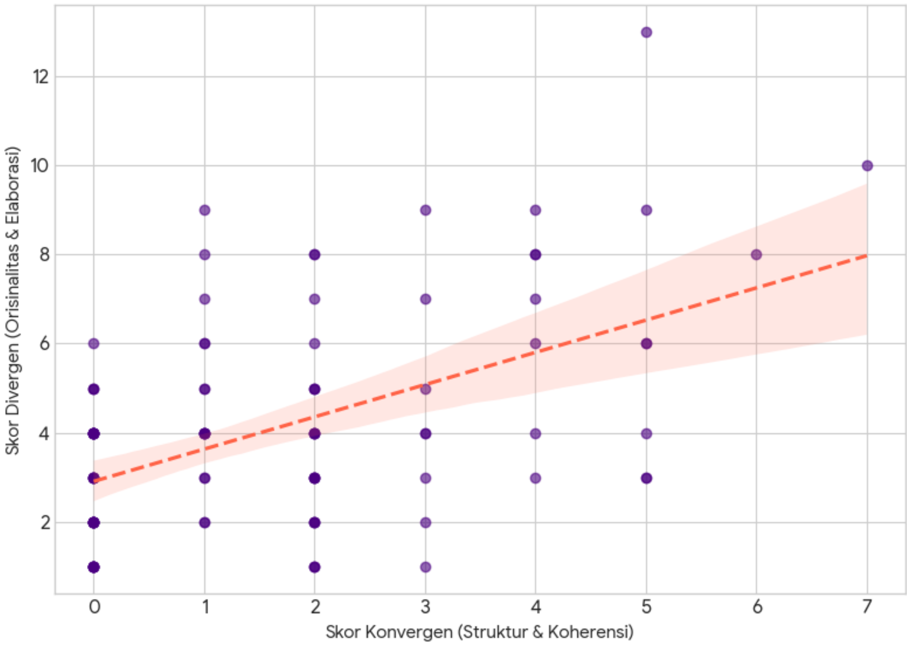


Figure 6. Correlation Between Divergent and Convergent Creativity Scores.

b. Predictive Model Evaluation

The performance of the proposed Multi-Task CNN model was evaluated on the 20% test set (N=21), comprising data the model had not previously seen. To ensure a comprehensive assessment as outlined in the methodology, the evaluation utilized three key metrics: Mean Absolute Error (MAE) to measure the average prediction error, Root Mean Squared Error (RMSE) to measure the magnitude of the error, and the Coefficient of Determination (R2) to measure the proportion of variance in expert scores explained by the model.

Model Performance Analysis. As presented in Table 1, the proposed model demonstrated satisfactory predictive capability. Specifically, it achieved a significantly lower error rate for Convergent Creativity Scores (RMSE = 1.5211) compared to Divergent Creativity Scores (RMSE = 1.9743). Crucially, the R2 scores provide insight into the model's explanatory power. The results indicate that the model explains approximately 65% of the variance (R2= 0.65) in the expert-rated convergent scores, whereas it captures about 48% of the variance (R2= 0.48) for divergent scores. It confirms the hypothesis that the model more consistently learns the structural elements of convergent thinking than the highly variable divergent features.

p-ISSN: 2655-8572 e-ISSN: 2655-8009

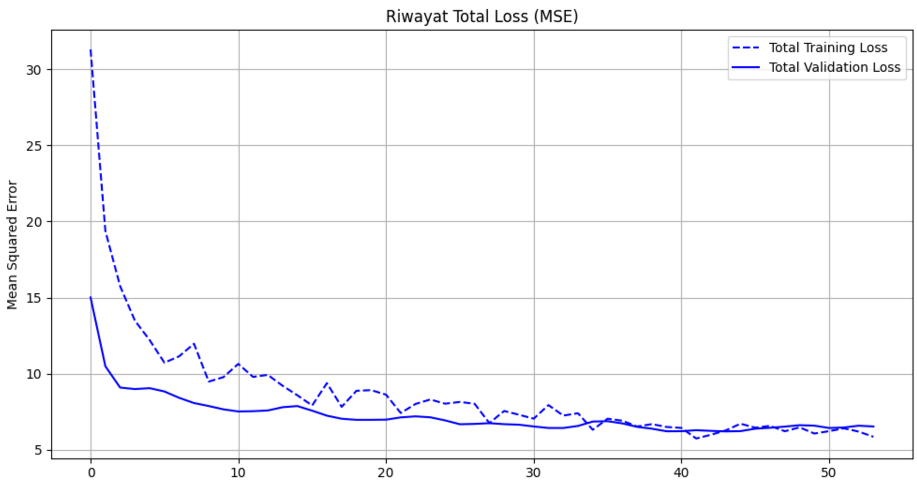
Table 1. Model Evaluation Results on Test Data

Metric	Divergent Score	Convergent Score	Interpretation
MAE	1.652	1.185	Average absolute prediction error (lower is better)
RMSE	1.9743	1.5211	Magnitude of error (lower is better)
R^2 Score	0.48	0.65	Variance explained by the model (higher is better)
Loss (MSE)	3.3081	2.1370	Direct loss function value from training

Benchmarking Against Baseline Model To validate the efficacy of the deep learning approach, these results were directly benchmarked against the baseline model (Gradient Boosting Regressor) described in the Method section. As shown in Table 2, the proposed MT-CNN model consistently outperformed the baseline across all metrics.

Table 2. Performance Comparison: Baseline Model vs. Proposed Multi-Task CNN

Model	Task Target	MAE ↓	RMSE ↓	R2 Score ↑
Baseline	Divergent Score	2.105	2.4510	0.21
(Gradient Boosting)	Convergent Score	1.620	1.9840	0.42
Proposed	**Divergent Score**	**1.652**	**1.9743**	**0.48**
(Multi-Task CNN)	**Convergent Score**	**1.185**	**1.5211**	**0.65**



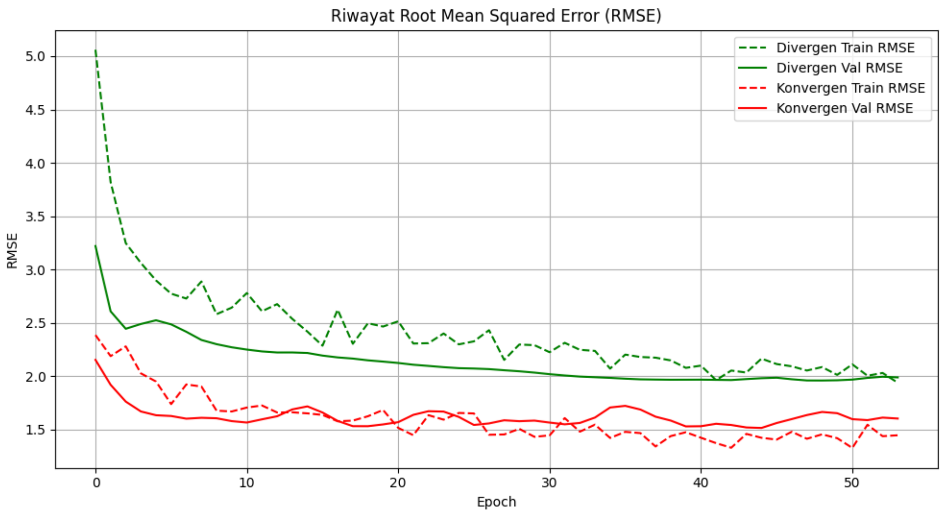


Figure 7. Model Training History

Figure 7. Training and validation performance curves across 100 epochs. (A) The Model Loss graph, where the blue solid line represents the training loss and the orange dashed line represents the validation loss, indicating stable convergence without significant overfitting. (B) The Accuracy graph for divergent (green line) and convergent (red line) tasks, showing how the model learns structural features faster than abstract features.

These results justify the use of the deep learning architecture for this task. The baseline model, which relied on hand-crafted features (e.g., contour counts and Hu moments), yielded significantly lower R^2 scores, particularly for divergent thinking (R^2 = 0.21). In contrast, the MT-CNN's ability to learn hierarchical feature representations directly from raw pixels allowed it to approximate better the complex expert judgments required for creativity assessment.

c. Qualitative Analysis of Model Predictions

To provide a more tangible interpretation of the model's performance, Figure 8 visualizes a selection of six representative test samples alongside their ground truth (expert) scores and the model's predictions. These samples were selected to illustrate both the model's strengths and its limitations.

As shown in the top row of Figure 8, the model demonstrates high accuracy in assessing drawings with clear structural boundaries and explicit thematic elements, resulting in low prediction errors (e.g., Sample A: Actual Convergent=8.0, Predicted= 17.8). It aligns with the quantitative finding that convergent features are more easily learnable. Conversely, the bottom row illustrates cases where the model struggled, particularly in detecting subtle divergent features such as "Humour" or abstract "Perspective." For instance, in Sample D, the model underestimated the Divergent score (Actual = 25.0, Predicted = 19.4), likely because the unconventionality of the drawing was interpreted as noise rather than creativity. This qualitative inspection confirms that while the model captures the general distribution of creativity scores, it occasionally misses nuanced semantic cues that are obvious to human experts.

To further analyze the training stability of the best-performing model (MT-CNN), its training history was examined as presented in Figure 7. The Total Loss curve demonstrates a healthy learning process without significant overfitting. The first curve in Figure 7 represents the Total Loss (MSE), showing the combined error from both predictive tasks. It is observable that both the Training Loss curve (blue dashed line) and the Validation Loss curve (blue solid line) consistently decrease and begin to plateau in the final epochs. This pattern is a strong indicator of a healthy training process, wherein the model successfully learned to minimize its errors without succumbing to significant overfitting.

The second curve, the RMSE-per-task, offers a more granular insight by translating the abstract loss value into the more readily interpretable RMSE metric. The most critical observation lies in the comparison between the solid green line (Validation RMSE for the Divergent Score) and the solid red line (Validation RMSE for the Convergent Score). The solid red line consistently settles at a lower level than the solid green line, providing strong visual evidence for the numerical results presented in Table 1. It definitely confirms that the model is inherently more accurate in predicting convergent creativity scores.

2. Discussion

a. Theoretical and Pedagogical Implications

The findings of this study carry several significant implications across theoretical, technical, and practical domains. At a technical level, the successful performance of the Multi-Task Convolutional Neural Network (CNN) in simultaneously predicting both divergent and convergent creativity scores lends strong support to the hypothesis that deep learning models can serve as valid and reliable tools for assessing complex psychological constructs such as creativity. The lower RMSE value for convergent scores, compared to divergent scores, suggests that structured, thematic, and logically coherent elements in children's drawings are more consistently detectable by the model. These elements may manifest in more regular and interpretable visual patterns, such as symmetry, object placement, or thematic consistency, which align well with CNNs' capacity to identify spatial hierarchies and visual coherence (Xia et al., 2021; Gao, 2025).

By contrast, divergent creativity typically characterized by originality, spontaneity, and idea fluency, often displays greater visual variability. This complexity renders divergent thinking more difficult to model, as it is less likely to adhere to consistent visual features. Prior studies have shown that divergent thinking tends to be more context-dependent and subjective, often reflecting a child's unique background, mood, and developmental stage (Gerwig et al., 2021; Rawlings et al., 2025). As such, while the model was able to learn generalizable patterns associated with convergent traits, its comparatively higher error in divergent prediction underscores the challenge of capturing fluid and abstract dimensions of creativity through image-based machine learning alone.

One of the most conceptually significant contributions of this study is the empirical discovery of a strong positive correlation between the aggregated divergent and convergent creativity scores ($r = +0.51$). This finding provides compelling evidence against traditional dichotomous models that often portray these two cognitive processes as oppositional or mutually exclusive (Xia et al., 2021). Instead, it supports emerging perspectives in creativity research that emphasize their interdependence. Scholars such as Duval et al. (2023) and Wang et al. (2023) argue that creative cognition should be understood as an integrated process in which idea generation (divergence) and idea refinement (convergence) operate in a complementary manner. Our study contributes empirical validation to this paradigm shift, specifically within the developmental context of early childhood.

b. Practical Implementation and Workflow

To bridge the gap between algorithmic development and classroom application, this study proposes a comprehensive deployment framework designed specifically for resource-constrained educational settings. While deep learning models often require substantial computational power, our decision to utilize the MobileNetV2 architecture ensures that the system is lightweight and optimized for mobile or embedded vision applications. Consequently, high-end hardware such as dedicated graphics processing units is not required for the inference phase. The trained model can be deployed effectively on standard teacher laptops, mid-range tablets, or even via a cloud-based web interface accessible through basic smartphones. This hardware flexibility is crucial for ensuring equitable access to advanced assessment tools in diverse school environments, including those in developing regions.

In terms of operational efficiency, the proposed system offers a transformative advantage over traditional manual assessment. Standard manual scoring of the TCT-DP is notoriously labor-intensive, typically requiring a trained psychologist or educator approximately 15 to 20 minutes to evaluate a single drawing across 14 distinct criteria meticulously. For a class of 30 students, this amounts to nearly 10 hours of administrative work. In sharp contrast, our automated pipeline processes a digitized drawing and generates prediction scores in approximately 0.4 seconds on a standard consumer-grade CPU. This drastic reduction in processing time shifts the educator's role from a scorer to a facilitator, allowing them to allocate time toward interpreting results rather than generating them.

The recommended workflow for practitioners is structured as follows: (1) Digitization: Teachers capture photos of children's drawings using a smartphone or scanner; (2) Processing: Images are uploaded to the application, which automatically crops and normalizes the input; (3) Assessment: The model generates immediate Divergent and Convergent scores; and (4) Verification: Teachers review the automated scores alongside the original drawing in a "human-in-the-loop" validation process.

Ultimately, this streamlined workflow allows teachers to obtain rapid, nuanced creative profiles of their students. Instead of spending hours on grading, educators can focus on interpreting the results to inform personalized pedagogical interventions. For instance, as suggested by recent studies (de Vink et al., 2022; Su & Yang, 2022), children identified with strong divergent tendencies but weaker convergent structuring can be immediately targeted for scaffolded support in organizing ideas, while the reverse profile may suggest a need to encourage imaginative exploration.

c. Sample Size Implications and Study Framing

A critical discussion is warranted regarding the dataset size (N=102). In the context of deep learning, which typically benefits from thousands of samples, our use of 82 training images and 21 test images presents a risk that the model might partially "memorize" training samples rather than learning purely generalizable features. Although we implemented rigorous data augmentation and Transfer Learning (using MobileNetV2) to mitigate this risk, the model performance metrics reported should be interpreted with caution. Consequently, this study is framed as a proof-of-concept demonstrating the feasibility of automated assessment, rather than a final commercial-ready validation. However, despite these computational constraints, the study provides robust scientific value in two key areas:

1). **Theoretical Robustness:** The discovery of a strong positive correlation (r = +0.51) between divergent and convergent thinking is statistically robust as it is derived from the full dataset (N=102). This finding stands independently of the deep learning model's performance and significantly contributes to the developmental psychology literature by challenging the traditional dichotomy of creativity.

2). **Feasibility Evidence:** The fact that the model could achieve reasonable predictive accuracy (RMSE ≈ 1.5 - 1.9) even with limited data suggests that the visual features of creativity are distinct enough to be learned. It paves the way for future large-scale studies to refine this approach with greater confidence.

d. Limitations and Future Directions

Methodologically, it is important to acknowledge that this study utilized a single stratified train-test split rather than a k-fold cross-validation framework, primarily due to computational resource constraints. While we employed bootstrap resampling to estimate statistical uncertainty (Confidence Intervals), relying on a single data split carries an inherent risk. Future research with expanded datasets should prioritize k-fold cross-validation to provide a more exhaustive assessment of model stability.

Moreover, the exclusive use of black-and-white images may have limited the richness of creative features available for analysis. Color is a vital component of children's artistic expression (Andari, 2025; Boyatzis & Varghese, 1994). Integrating color channels into future versions of the model may enable the learning of more nuanced features. An additional avenue for future research involves the incorporation of Explainable AI (XAI) techniques. Methods such as Gradient-weighted Class

Activation Mapping (Grad-CAM) could visualize which areas of a drawing contribute most to each prediction, bridging the gap between automated assessment and human intuition (Panfilova et al., 2024). Such transparency is particularly important in educational settings, where trust in automated systems hinges on their perceived interpretability.

D. CONCLUSION

This study bridges the gap between developmental psychology and artificial intelligence, offering two primary contributions. First and most significantly, the research provides empirical evidence regarding the nature of creative cognition in early childhood. The robust positive correlation ($r = +0.51$) found between divergent and convergent scores challenges dichotomous models that view these processes as mutually exclusive. Instead, the findings support an integrated perspective, suggesting that mature creativity in children involves a synergy between the ability to generate novel ideas (divergence) and the ability to structure them coherently (convergence). Second, from a technical perspective, this study serves as a successful proof-of-concept for the digitization of creativity assessment. Despite the constraints of a limited dataset (N=102), the Multi-Task Deep Learning model demonstrated the feasibility of learning visual creativity features directly from raw pixels. The model achieved a prediction accuracy of $R^2 = 0.65$ for convergent scores and $R^2 = 0.48$ for divergent scores, outperforming a traditional Gradient Boosting baseline. The performance gap indicates that while AI can effectively capture structural regularity, the nuances of abstract divergent thinking require further architectural refinement.

Future research should address the study's limitations by expanding the dataset size to enable k-fold cross-validation and by integrating color information to capture a richer array of creative features. Ultimately, this research paves the way for AI-assisted tools that can support teachers in providing more holistic and personalized creativity assessments in early childhood education.

E. REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2016). Tensorflow : A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, 265–283.

Ali, O., Murray, P. A., Momin, M., Dwivedi, Y. K., & Malik, T. (2024). technological forecasting & social change the effects of artificial intelligence applications in educational settings : Challenges and strategies. *Technological Forecasting & Social Change*, *199*(December 2023), 123076. https://doi.org/10.1016/j.techfore.2023.123076

Andari, I. A. M. Y. (2025). Coloring as an activity for the development of artistic creativity in early childhood. *Widya Sundaram : Jurnal Pendidikan Seni Dan Budaya*, *03*(02), 122–133.

Boyatzis, C. J., & Varghese, R. (1994). Children's emotional associations with colors. *The Journal of Genetic Psychology*, *155*(1), 77–85. https://doi.org/10.1080/00221325.1994.9914760

Chai, J., Zeng, H., Li, A., & Ngai, E. W. T. (2021). Machine learning with applications deep learning in computer vision : A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, *6*(August). https://doi.org/10.1016/j.mlwa.2021.100134

Çiftçi, S., Karci, H., Karaca, N., Söylemez, B., & Koçakoğlu, H. (2025). Emotion analysis of children ' s drawings. *IEEE Access*, *13*(August). https://doi.org/10.1109/ACCESS.2025.3606359

Desmet, O. A., Weerdenburg, M. Van, Poelman, M., Hoogeveen, L., & Yang, Y. (2021). Validity and utility of the test of creative thinking drawing production for dutch adolescents. *Journal of Advanced Academics*, 1–24. https://doi.org/10.1177/1932202X21990099

Dilnoza, Q. (2023). The essence of developing creative abilities in primary school pupils. *Current Research Journal of Pedagogics*, *04*(06), 95–100.

Gao, S. (2025). Creative generation and evaluation system of art design based on artificial intelligence. *Discover Artificial Intelligence Research*, *5*(118).

Grilli, L., Pedota, M., & Department. (2024). Creativity and artificial intelligence : A multilevel perspective. *Creativity and Innovation Management Published*, *33*, 234–247. https://doi.org/10.1111/caim.12580

Guo, Z., Huang, Y., Hu, X., Wei, H., & Zhao, B. (2021). A survey on deep learning based approaches for scene understanding in autonomous driving. *Electronics 2021, 10*. https://doi.org/0.3390/ electronics10040471

Hadas, E., & Hershkovitz, A. (2025). Assessing creativity across multi-step intervention using generative AI models. *Journal of Learning Analytics*, *12*(1), 91–109.

He, T., Li, C., Wang, J., Wang, M., Wang, Z., & Jiao, C. (2022). An emotion analysis in learning environment based on theme-specified drawing by convolutional neural network. *Public Health*. https://doi.org/10.3389/fpubh.2022.95887

Karamizadeh, S., Chaeikar, S. S., & Salarian, H. (2025). Combining MTCNN and enhanced facenet with adaptive feature fusion for robust face recognition. *Technologies*, 1–25.

Kaufman, J. C. (2016). *Creativity 101 (2nd ed.)*. Springer Publishing.

Khare, S. K., Blanes-vidal, V., Nadimi, E. S., & Acharya, U. R. (2024). Emotion recognition and artificial intelligence : A systematic review (2014 – 2023) and research recommendations. *Information Fusion*, *102*(August 2023), 102019. https://doi.org/10.1016/j.inffus.2023.102019

Mamun, F. Al. (2024). Fostering creativity and critical thinking in the classroom : strategies for 21st-century education. *International Journal for Multidisciplinary Research*, *6*(4), 1–12.

Marullo, G., Tanzi, L., Ulrich, L., Porpiglia, F., & Vezzetti, E. (2023). A multi-task convolutional neural network for semantic segmentation and event detection in laparoscopic surgery. *Journal of Personalized Medicine*, *13*, 1–12.

Mhlongo, S., Mbatha, K., Ramatsetse, B., & Dlamini, R. (2023). Challenges , opportunities , and prospects of adopting and using smart digital technologies in learning environments : An iterative. *Heliyon*, *9*(6), e16348. https://doi.org/10.1016/j.heliyon.2023.e16348

Nath, S. S., & Stevenson, C. E. (2024). Pencils to Pixels : A systematic study of creative drawings across children , Adults and AI. *ArXiv*. https://doi.org/10.48550/arXiv.2502.05999

Omatseye, B. O. J., & Emeriewen, K. O. (2013). Every picture tells a story : stages of art development in children. *AFRREV IJAH*, *2*(3), 36–52.

Pedregosa, F., Weiss, R., & Brucher, M. (2011). Scikit-learn : Machine learning in Python. *Journal Of Machine Learning Research*, *12*, 2825–2830.

Rawlings, S., Chetwynd-talbot, D., Husband, E., Nuttall, A., Quinn, E., Taggart, R., Roome, H. E., Rawlings, S., Chetwynd-talbot, D., Husband, E., & Nuttall, A. (2025). Divergent thinking is linked with convergent thinking ; implications for models of creativity. *Thinking & Reasoning*, 1–23. https://doi.org/10.1080/13546783.2025.2485059

Raya, M. A., Ogunyemi, A. O., Carstensen, V. R., Broder, J., Illanes-Manrique, M., & Rankin, K. P. (2023). The reciprocal relationship between openness and creativity : from neurobiology to multicultural environments. *Frontiers in Neurology*, *October*. https://doi.org/10.3389/fneur.2023.1235348

Robinson, K., & Aronica, L. (2016). *Creative schools: The grassroots revolution that's transforming education*. Penguin Books.

Sari, N. P., Setiawan, M. A., & Makaria, E. C. (2021). Can science develop creativity in early children? In *Proceedings of the 2nd Progress in Social Science, Humanities and Education Research Symposium (PSSHERS 2020)*, *563*(Psshers 2020), 202–205. https://doi.org/10.2991/assehr.k.210618.040

Tao, S., Fu, H., Yang, R., & Wang, L. (2025). A multi-task learning framework with enhanced cross-level semantic consistency for multi-level land cover classification. *Remote Sensing*, 1–27. https://doi.org/10.3390/rs17142442

Tarigan, M., & Fadillah, F. (2022). Inter-rater and intra-rater reliability test with goodenough-harris drawing Test. *The Open Psychology Journal*, *15*, 1–5. https://doi.org/10.2174/18743501-v15-e2207130

Tsirtsakis, P., Zacharis, G., Maraslidis, G. S., & Fragulis, G. F. (2025). Deep learning for object recognition : A comprehensive review of models and algorithms. *International Journal of Cognitive Computing in Engineering*, *6*(December 2023), 298–312. https://doi.org/10.1016/j.ijcce.2025.01.004

Urban, K. K. (2005). Assessing creativity : The test for creative thinking - drawing production (TCT-DP). *International Education Journal*, *6*(2), 272–280.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CA: CreateSpace.

Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences*, *15*(3), 1–12. https://doi.org/10.3390/educsci15030343

Xu, Y., Nazri, M., & Rahman, A. (2025). Evaluating children ' s creativity through drawing : Designing and validating an assessment tool for originality. *Journal of Education and Learning*, *15*(1), 101–112. https://doi.org/10.5539/jel.v15n1p101

Zhao, F. (2024). Deep reinforcement learning enhances artistic creativity : The case study of program art students integrating computer deep learning. *Journal of Intelligent Systems*, *33*.