World Journal of Educational Studies
Print ISSN: 2959-9989
Online ISSN: 2959-9997
DOI: https://doi.org/10.61784/wjes3121

APPLICATION OF AIGC-ASSISTED VISUAL INSPIRATION IN EARLY CHILDHOOD ART EDUCATION

WangShu Gong[1*], JunRu Ma[2], ZhaoYang Wang[1]
[1]Stony Brook Institute at Anhui University, Hefei 230000, Anhui, China.
[2]Anhui University Kindergarten, Hefei 230000, Anhui, China.
[*]Corresponding Author: WangShu Gong, Email: R32314008@stu.ahu.edu.cn

Abstract: Generative artificial intelligence content (AIGC) is gradually permeating early childhood art education and reshaping teaching models, but its specific impact on key processes of early childhood art creation still lacks scientific empirical support. This study used a controlled variable comparative experimental design, combined with expert scoring, to explore the impact of AIGC software intervention at different stages of the painting course on the two core processes of children's idea generation and design expression in children's paintings. The results showed that passively receiving AIGC inhibited the originality of children's idea generation in painting, but had no significant impact on the fluency of this process. However, actively referring to AIGC during the painting process could improve children's design expression ability, and improve their composition and color expression abilities in this stage. Based on this, the study suggests that AIGC can be used as an auxiliary reference tool in the painting process to help children's artistic abilities develop.
Keywords: Generative AI content; Early childhood art education; Idea generation; Design expression

1 INTRODUCTION

With the rapid development of artificial intelligence technology, its application value in the field of early childhood art education is becoming increasingly prominent. Existing research shows that artificial intelligence can provide multifaceted support for kindergarten art activities, including assisting teachers in lesson preparation, optimizing children's artwork, and expanding creative avenues [1]. In early childhood song performance education, artificial intelligence can also help to personalize teaching, enrich educational resources, and enhance classroom interactivity and fun [2]. It can be seen that the integration of artificial intelligence and early childhood art education has become an emerging teaching model, which can serve as a beneficial supplement to traditional teaching methods, assisting teachers in carrying out personalized teaching and improving the quality of teaching outcomes. However, current research focuses primarily on theoretical discussions and practical descriptions, with limited experimental investigations into the specific implementation effects and mechanisms of artificial intelligence (AI) in kindergarten art education, particularly regarding its impact on children's creative development. Therefore, it is necessary to systematically explore the application effects of AI-generated content (AIGC) in early childhood art education from a scientific and empirical perspective. This study, through field experiments combined with the teaching practice of AIGC software, focuses on analyzing the actual impact of AIGC on various dimensions of creativity in assisting children's painting, aiming to provide empirical evidence for the deep integration of artificial intelligence and early childhood art education.

While there are various opinions in academia regarding the stages of artistic creation, the general consensus is that these stages involve different stages of thinking and behavior. Teng Jing, summarizing previous research on the creative process, points out that creative thinking is divided into a continuous repetition of the "generation-exploration" stage [3]. Botella et al., through interviews with visual arts students, summarized a multi-stage model including Idea Generation, Incubation, Elaboration/Production and Finalization, emphasizing the dynamic development of creation from conception and processing to completion [4]. Yokochi and Okada, in their study comparing experienced artists and novice creators, pointed out that artistic creation often follows a spiral evolutionary path of "theme transformation—technique exploration—process adjustment—form construction," reflecting a continuous transformation from creative conception to finished product presentation [5]. In summary, most scholars agree that artistic creation can be summarized into two core processes: idea generation and design expression. The former refers to the process by which an individual, after identifying a problem, forms a creative image in their mind through divergent and convergent thinking; the latter refers to the process of externalizing the inner concept into a concrete work through visual language, composition, color, and other means.

The abilities relied upon at different creative stages have different emphases. In the "idea generation" stage, creative thinking is mainly reflected in the diversity and novelty of ideas. Runco and Acar et al., as well as Tiansheng Xia et al., pointed out that fluency (the number of ideas generated) and originality (the uniqueness of ideas) can be used as evaluation indicators for this stage [6]. In the "design expression" stage, the focus is on the ability to transform imagery into concrete forms, involving composition, color application, and media manipulation. Goldschmidt believes that this process is accompanied by a "visual-response-adjustment" cycle, with creators optimizing their work based on feedback [7]. Therefore, composition ability and color expression ability can serve as important dimensions for evaluating the design expression stage.

In summary, this study divides the quality assessment of early childhood painting into two dimensions: idea generation (including fluency and originality indicators) and design expression (including composition ability and color expression ability indicators). The composition of creativity is shown in Figure 1.

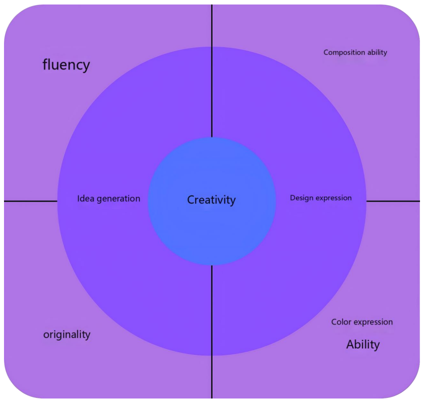


Figure 1 Creativity Composition Diagram

This study aims to explore the influence mechanism of AI-generated content (AIGC) on children's creative process and results in kindergarten art teaching through empirical methods. The study selected 24 children in the middle class of a model kindergarten in a provincial capital city in East China as experimental subjects, and introduced AIGC software for painting-assisted teaching in a natural teaching context. Through field experiments and classroom observations, the behavioral performance and work output of children in the two stages of "idea generation" and "design expression" were systematically collected, and quantitative evaluation was carried out based on four indicators: fluency, originality, composition ability, and color expression ability. This study focuses on analyzing the changes in children's creative ideas and visual expression after AIGC intervention, revealing its specific pathways and effects on various dimensions of children's creativity. This aims to supplement existing research on empirical aspects and provide practical reference for the scientific application of artificial intelligence in early childhood art education.

2 METHODS AND RESULTS

2.1 Research Subjects

This study used 24 children from a model kindergarten in a provincial capital city as research subjects, including 12 boys and 12 girls, aged 5±1 years. All guardians and kindergarten staff provided informed consent in writing and understood the experimental procedure and purpose. Furthermore, based on the students' regular homework assignments, there was no significant difference in their creative abilities, making them comparable.

2.2 Experimental Procedure

The study selected "Draw a car you imagine" as the course theme, designed with three consecutive stages: brainstorming (3 minutes), slide presentation (5 minutes), and independent creation (approximately 25 minutes). In the brainstorming stage, the teacher guided all children to conceive the content of the picture through verbal expression and group discussion; in the presentation stage, the teacher used a PowerPoint presentation to show various types of vehicles to expand children's visual experience and stimulate detailed ideas; in the independent creation stage, children used various drawing tools (such as watercolors, crayons, rulers, etc.) to complete the theme creation.

Based on different AIGC intervention methods, children were randomly divided into three groups:

Group 1 (control group): After participating in brainstorming and slide presentation, they directly entered the independent creation stage without AIGC intervention.

Group 2 (Interactive Generation Group): Building upon the completion of the first two stages as Group 1, children can proactively request creative ideas from the experimental staff during the independent creation stage (e.g., "Draw a car shaped like a cake"). The staff will then generate images using AIGC software for reference. To control the intensity of intervention, each child can refer to AIGC content a maximum of two times.

Group 3 (Passive Reference Group): In addition to completing the first two stages, children will view several pre-generated vehicle drawings from AIGC as inspirational stimuli, but cannot request additional AIGC assistance during the independent creation stage.

2.3 Scoring Procedure

This study measures creativity based on two dimensions: idea generation and design expression. The idea generation dimension is scored on "fluency" and "originality," while the design expression dimension is scored on "composition ability" and "color expression ability." The following is a detailed explanation of these four dimensions and specific scoring rules. The children and teachers were not aware of the scoring criteria.

2.3.1 Fluency evaluation

In his experiment, Tiansheng Xia defined fluency as an indicator of creative output, determined by the total number of creative ideas generated [8]. Fluency scores require counting the total number of complete ideas (excluding incomplete ideas) listed by participants. This experiment uses the same evaluation index, defining fluency as the total number of complete ideas (excluding incomplete ideas) listed by participants, without considering aesthetics, color matching, or other factors unrelated to the number of elements, i.e., the total number of elements in the drawing, including vehicles, paintwork, patterns, and environmental arrangements. Drawings with fewer elements are assigned 1-3 points, and drawings with more elements are assigned 3-5 points. Table 1 (left) shows examples of drawings for each fluency score range:

2.3.2 Originality evaluation

Originality is scored from 1 point (basically similar to the provided materials or everyday items) to 5 points (most original). Expert originality scoring is used because the sample size is too small to determine originality based on the frequency of individual ideas. Expert scoring is based on consensus between the two authors, and any disagreements are resolved through discussion until a consensus is reached. Table 1 (right) shows examples of illustrations for each segment of the originality score:

Table 1 Examples of Fluency and Originality Scores

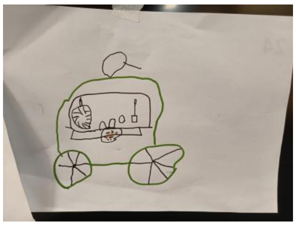
Fluency Score	Example	Reason	Originality Score	Example	Reason
1 point		Too few elements, only vehicle outline and a few internal details.	1 point		Basically the same as the theme and color scheme of the provided "Pumpkin Sports Car" image (Shown in Figure 2).
3 points		Moderate number of elements, full overall vehicle outline, with appropriate external detail decoration.	3 points		Similar to the image of Butterfly Mary in the animation "Happy Cool Baby", but the color scheme and detail outline are different (Shown in Figure 3).
5 points		Abundant number of elements, full vehicle outline, rich external details, and in addition to the vehicle itself, there are appropriate environmental elements as embellishments.	5 points		The vehicle's exterior design is highly original, incorporating various vehicle shapes and matching them well.

Figure 2 Image of a Pumpkin Sports Car Generated by AI Provided in the Experiment

Figure 3 Animation Character "Butterfly Mary"

2.3.3 Evaluation of composition ability

This study evaluates children's composition ability based on the relevant research by Lin Lin and Luo Jinjing on children's composition ability at different developmental stages [9]. This study points out that children's drawing composition exhibits clear stages of evolution with age, specifically progressing from a chaotic style to a parallel style, then to a scattered style, and finally to an occluded style. The core connotations of each stage of composition are defined as follows:

Chaotic composition: Children do not consciously plan the space of the images in the picture, but simply distribute various objects randomly within the picture area, without obvious spatial organization logic;

Parallel composition: Children use single, basic spatial relationships to integrate picture elements. Various objects are mostly presented in parallel or vertical arrangements, with no complex spatial relationships between elements;

Scattered composition: Picture elements begin to form a primary and secondary relationship. The layout shows the characteristic of radiating outwards from the main element as the core, and the spatial organization has a preliminary logic;

Occluded composition: Children can construct spatial layers of depth and distance in the picture through the occlusion relationship between elements, reflecting a preliminary understanding of three-dimensional spatial relationships.

Based on the compositional forms presented in children's classroom paintings, the maturity of their compositional abilities can be judged. This evaluation system divides compositional ability into 5 scoring levels, with the specific correspondence as follows:1 point corresponds to a chaotic composition; 2 points correspond to a parallel composition; 3-4 points correspond to a scattered composition (where the score difference mainly depends on the prominence of the main elements in the picture); 5 points correspond to an obscuring composition. Examples of paintings for each scoring level are shown in Table 2 (left):

2.3.4 Evaluation of color expression ability

This study's assessment of children's color expression ability in the experiment is based on Pian Cen's relevant research on the development of children's painting expression ability [10]. Piancen systematically analyzed the stage-specific characteristics of children's color expression ability, specifically dividing it into different levels. The core performance characteristics of each level are defined as follows:

Weak color expression ability: Children possess basic color control awareness, but their control ability is poor; color selection lacks purpose and logic, exhibiting strong randomness; the overall color harmony of the painting is not considered during the painting process, and color application lacks clear planning;

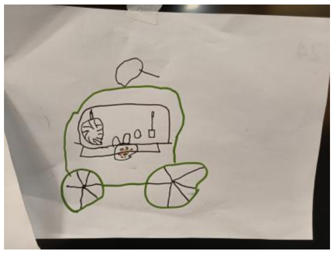
Average color expression ability: Children's fine motor skills have significantly improved, enabling them to complete coloring operations relatively well, and they can consciously limit the color to within the drawn outline, avoiding obvious color overflow problems, but their control over color matching and color richness is weak;

Good color expression ability: Building upon "average color expression ability," children further develop the awareness and ability to apply color evenly, and the number of colors used in the painting is significantly increased compared to the "average level," with a significant improvement in color richness;

Excellent color expression ability: In addition to possessing "good color expression ability," In addition to all the characteristics, young children can also use color as a carrier of emotion, effectively conveying their emotional state when drawing through color selection and application.

Based on the above research findings, this study divides children's color expression ability into 5 scoring levels, with the specific correspondence as follows:1 point corresponds to "the picture is basically uncolored"; 2 points correspond to the characteristic of "weak color expression ability" mentioned above; 3 points correspond to the characteristic of "average color expression ability" mentioned above; 4 points correspond to the characteristic of "good color expression ability" mentioned above; 5 points correspond to the characteristic of "excellent color expression ability" mentioned above. Examples of pictures for each scoring are shown in Table 2 (right):

Table 2 Examples of scoring for composition ability and color expression ability

Composition ability score	Example	Reason	Color expression ability score	Example	Reason
1 point		The elements in the picture are randomly arranged and do not reflect good spatial organization logic.	1 point		The picture is basically uncolored.
2 points		The picture presents a simple parallel layout. Although the elements differ in size and shape, there is no complex spatial relationship.	2 points		The hand control during the coloring process is weak, resulting in irregular or overflowing coloring in some areas.
3 points		The picture presents a scattered layout, with a relatively clear central subject and a preliminary sense of spatial organization.	3 points		The coloring is regular and basically within the framed area, but the color selection is rather random and the aesthetics are poor.
4 points		The picture presents a scattered layout, with a very clear central subject and a preliminary sense of spatial organization.	4 points		The coloring is regular and basically within the framed area. The color selection is roughly consistent with the color of a "watermelon" in real life, and the aesthetics are good.

5 points	None	None	5 points		The coloring is regular and basically within the framed area. The color richness is good, and the emotional tension is high.

2.4 Scoring Data Analysis

This study collected the artwork of 24 students after participating in this course as the experimental sample. To evaluate the artwork performance, this study invited two experts to score the artwork. Both experts have master's degrees in education and have worked in early childhood education for many years, possessing rich experience in early childhood art education and the ability to score children's artwork rigorously and meticulously.

2.4.1 Validation of Judges' scoring data

This paper conducts an inter-rater consistency test (ICC test) on the scoring results of two judges [11]. The intra-group correlation coefficient (ICC) is an indicator used to measure the degree of consistency in scoring when multiple raters rate the same group of subjects. It not only measures correlation but also assesses whether the scores are interchangeable. ICC is based on the analysis of variance (ANOVA) framework, which decomposes the total variance into variance among subjects, variance among raters, and variance of random error, thereby quantifying the reliability index of inter-rater consistency. Its core idea is that if the inter-rater consistency is high, the variance among subjects should be much greater than the variance caused by rater differences or random error. The formula is:

$$\mathrm{ICC} = \frac{\sigma^2_{\mathrm{subject}}}{\sigma^2_{\mathrm{subject}} + \sigma^2_{\mathrm{rater}} + \sigma^2_{\mathrm{error}}} \tag{1}$$

where, $\sigma^2_{\mathrm{subject}}$ is the variance component among the evaluated subjects (i.e., the variance component of the children's scores), $\sigma^2_{\mathrm{rater}}$ is the variance component among the scorers, $\sigma^2_{\mathrm{error}}$ where is the residual variance component. The ICC value ranges from 0 to 1, and the closer to 1, the higher the consistency. Generally, ICC < 0.50 indicates poor consistency, 0.50–0.75 indicates moderate consistency, 0.75–0.90 indicates good consistency, and > 0.90 indicates excellent consistency.

Further, to test whether the scoring standards used by the two judges on each scoring dimension have similar variance, that is, whether the difference in their scoring volatility reaches a significant level, this paper conducts a Levene test on the scoring differences on the four evaluation dimensions[12]. The Levene test is used to test whether the variances of two or more groups are equal, i.e., whether the homogeneity of variance assumption is satisfied, to verify the robustness of the data. The test statistic W is constructed as the ratio of the between-group mean square to the within-group mean square, which follows an F-distribution under the null hypothesis (equal variances).

$$\mathrm{T} = \frac{\mathrm{SSB}}{\mathrm{SSW}/(2\mathrm{n} - 2)} = \frac{(2\mathrm{n} - 2) \cdot \mathrm{SSB}}{\mathrm{SSW}} \tag{2}$$

The larger the value, the more significant the difference in the dispersion (variance) of the scores of the two judges is relative to the random fluctuations within their respective scores.

Levene test p The value represents the probability of observing a difference as in the current sample if the variances among the judges are truly equal. The formula is as follows: $p = P(T \geq t_{obs} | H_0)$. Where, T is the test statistic, t_{obs} is the specific observed value of the test statistic calculated from the current sample data, $P(T \geq t_{obs} | H_0)$ represents the probability calculated under the condition that the null hypothesis H_0 is true, that is, the probability of obtaining a test statistic observation value that is at least as extreme as or more extreme than the current sample statistic.

2.4.2 Analysis of scoring results

To explore whether there are systematic differences in the scores of different groups on each assessment dimension, this study first tests the basic distribution of the scoring data of each group on each dimension, and selects an appropriate statistical comparison method based on the test results.

This study uses the Shapiro-Wilk test to assess the normality of the scores of each group on each dimension [13]. This test is applicable to situations with limited sample size. Its statistic W quantifies the degree of deviation by measuring the correlation between the sample data and the order statistic of the ideal normal distribution. The calculation formula is as follows:

$$\mathrm{W} = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1} (x_i - \bar{x})^2} \tag{3}$$

where, $x_1 \leq x_2 \leq \cdots \leq x_n$ represents the sample observations arranged in ascending order, $\bar{x}$ represents the sample mean, coefficient a_i The weight is derived from the expected value of the standard normal distribution order statistic. The value of W ranges from (0, 1]. The closer the value is to 1, the greater the likelihood that the data follows a normal distribution.

Based on the premise of normality and homogeneity of variance (Levene test results), this study follows the decision-making process to select the inter-group difference test method:

If the data simultaneously satisfies normality and homogeneity of variance, then the parametric test method of one-way ANOVA (for comparisons of three or more groups) is used;

If the data satisfies the normal distribution but the variance is unequal, then Welch's ANOVA, which is robust to heteroscedasticity, is used.

If the data violates the normality assumption, then regardless of the result of homogeneity of variance, a nonparametric test is used.

For comparisons of multiple independent samples, the Kruskal-Wallis H test is used [14]. The formula for calculating the test statistic H is:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \tag{4}$$

where, k represents the number of groups, N represents the total sample size, n_i represents the sample size of theigroup,$R_i$represents the sample size of theirepresents the rank sum of all observations within the group.HThe statistic approximately follows an empirical distribution of the permutation statistic with $k-1$ Chi-square distribution. If the Kruskal-Wallis test result is significant, further post-hoc pairwise comparisons will be performed to determine the specific source of the difference. Meanwhile, considering the small sample size in this experiment, the permutation test was used as a supplementary method [15]. Its principle is to mix all N observations, randomly redistribute them to each group, and keep the sample size unchanged. Calculate the Kruskal-Wallis statistic for the data after each permutation$H^{(b)}$degrees of freedom, repeatedbtimes. After permutation testpThe value formula is:

$$p_{perm} = \frac{\sum_{b=1}^{B} I\left(H^{(b)} \geq H_{obs}\right) + 1}{B+1} \tag{5}$$

To explore which groups have differences in dimensions, this paper uses Dunn's Test post-hoc tests on groups with significant differences in each dimension under non-parametric tests. The principle is to adjust the standard error and use multiple comparison correction to make it applicable to multiple group comparison scenarios. The core idea is to compare the average rank of each group to determine which specific group pairs have significant differences [16]. Itspvalue calculation formula is:

$$p_{ij}^{Dunn} = \min\left(1, m \cdot \left[2 \times \left(1 - \Phi\left(\frac{\bar{R}_i - \bar{R}_j}{\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)C}}\right)\right)\right]\right) \tag{6}$$

If the p-value is less than the significance level a, then it is considered that there is a significant difference between groups.

3 RESULTS AND DISCUSSION

3.1 Experimental Results

3.1.1 Validation of Judges' scoring data

This paper first performs an inter-rater consistency test (ICC test) on the scoring results of the two judges. The test results show that the ICC value is 0.706 and the p-value is 0.002, indicating that there is a high degree of consistency between the two judges, and the scoring results have good reliability. Therefore, it can be considered that the review results of these two judges have strong reference value and are suitable as the basis for subsequent experimental analysis.

Furthermore, this paper conducts Levene tests on the scores given by the two judges. The results show that the Levene test statistic for fluency is 0.052 (p-value 0.820); for originality, it is 3.538 (p-value 0.066); for composition ability, it is 0.447 (p-value 0.507); and for color expression, it is 1.574 (p-value 0.216). These results indicate that the scoring criteria used by the two judges across all scoring dimensions have similar variances, and the differences in score volatility are not statistically significant, further demonstrating the objectivity of the scoring criteria in this study.

3.1.2 Analysis of Judges' scoring data

According to the established scoring rules, for each child's painting, scores were assigned to each dimension, and the average score of the two judges was used as the final score for the corresponding dimension of the child. Subsequently, according to the pre-defined group classifications, the mean and standard deviation of the scores for each dimension between groups were calculated. The results are shown in Table 3:

Table 3 Mean and Standard Deviation Table for Each Dimension of Children's Painting

Dimension	Group	Mean	Standard Deviation

	Control Group	3.19	0.3720
Fluency	Interactive Generation Group	3.36	1.1851
	Passive Reference Group	2.69	0.8839
	Control Group	3.63	1.2748
Originality	Interactive Generation Group	3.13	0.9161
	Passive Reference Group	2.63	0.7440
	Control Group	3.13	0.9161
Composition Ability	Interactive Generation Group	3.56	0.3204
	Passive Reference Group	2.75	0.4629
	Control Group	2.75	1.0690
Color Expression Ability	Interactive Generation Group	3.44	0.7289
	Passive Reference Group	2.44	1.0155

To explore the significance level of differences between groups across dimensions, this study first conducts a normality test on the scores between groups for each dimension. We conduct Shapiro-Wilk tests on the scores for each dimension of each group, and the results are shown in Table4.

Table 4 Shapiro-Wilk Test p-Values For Each Dimension

Control Group	0.0270	0.0846	0.8278	0.7833
Interactive Generation Group	0.0846	0.8278	0.0370	0.2861
Passive Reference Group	0.2914	0.4283	0.5224	0.9250

Due to the small sample size, this paper uses p>0.1 as the criterion for judging whether the data conforms to a normal distribution. The results show that only the fluency and originality scores of the control group and the fluency and composition ability scores of the interactive generation group conform to a normal distribution, while the others do not. Therefore, this paper uses a nonparametric test for multiple independent samples to test the significance. This paper uses the Kruskal-Wallis H test as a nonparametric test for multiple independent samples and uses the permutation test as a supplementary method. The results of the nonparametric tests for multiple independent samples in each dimension are shown in Table 5.

Table 5 Nonparametric Test Results for Multiple Independent Samples for Each Dimension

	Statistic (H)	p-value
Fluency	2.3906	0.3026
Originality	6.3478	0.0418
Composition Ability	11.0808	0.0039
Color Expression Ability	5.6062	0.0606

This paper uses p>0.1 as the criterion for judging whether there is a significant difference. The results revealed significant differences among the three groups in originality, composition ability, color expression ability, and total score. Next, we performed a post-hoc Dunn's Test on the groups with significant differences to determine which groups showed significant differences within each dimension. The results are shown in Figure 4:

	originality	Composition ability	Color expression ability
Group 1 vs. Group 2	0.8968	0.057	0.1994
Group I vs. Group 3	0.0365	1	1
Group 2 vs. Group 3	0.4259	0.0038	0.0807

Figure 4 Heatmap of Dunn's Test post-hoc test p-values

3.2 Discussion of Results

3.2.1 Description of phenomena

Based on the data in Table 5, differences exist in the performance of the control group, interactive generation group, and passive reference group across four dimensions: fluency, originality, composition ability, and color expression ability. Specifically, the interactive generation group scored higher than both the control group and the passive reference group in fluency, composition ability, and color expression ability; while the control group showed an advantage in originality, scoring higher than both the interactive generation group and the passive reference group. The passive reference group did not show a relative advantage in any of the above dimensions.

Combining the heatmap of p-values from the Dunn post-hoc test results (Figure 4), the following key phenomena can be observed:

In the comparison between the control group (Group 1) and the interactive generation group (Group 2), the differences between the two in the dimensions of composition ability and color expression ability are significant, indicating that the composition ability and color expression ability of the interactive generation group are better than those of the control group;

In the comparison between the control group (Group 1) and the passive reference group (Group 3), the difference in originality is significant, indicating that the originality of the control group is better than that of the passive reference group;

In the comparison between the interactive generation group (Group 2) and the passive reference group (Group 3), the differences between the two in the two dimensions of composition ability and color expression ability are significant, indicating that the composition ability and color expression ability of the interactive generation group are better than those of the passive reference group.

3.2.2 Phenomenon analysis

(1) Pre-painting AIGC projection has an inhibitory effect on children's originality in painting

The experimental results show that the originality of painting in the control group is significantly higher than that in the passive reference group. This phenomenon indicates that pre-painting AIGC image projection has an inhibitory effect on children's originality in painting, and this result is consistent with the cognitive laws of creative generation. From a cognitive perspective, AIGC-generated images often possess strong visual impact (such as exaggerated shapes and novel elements). This type of content is more likely to occupy children's cognition through strong sensory stimulation, leading to a deep memory of the reference image. This "high-intensity visual input" causes children's creative ideas to become overly focused on the themes and forms presented by AIGC, weakening their exploration of other creative sources—reducing their observation and thinking about the basic forms of "vehicles" in the real world, and ignoring the possibility of creatively combining them with previously encountered cartoons, animations, and other scenes. Ultimately, children's creativity falls into "path dependence," tending to directly reuse the visual paradigms provided by AIGC, rather than integrating diverse experiences through independent thinking to form unique ideas, resulting in a significant decrease in originality. Further comparison of the originality performance of the interactive generation group and the passive reference group reveals that, using AIGC as a reference, the interactive generation group exhibits superior originality. This difference reveals the crucial role of the "thought process in creative generation" in originality: In the interactive generation group, children submit creative requests to AIGC (such as "draw a car shaped like a cake") as needed during the drawing process. This process requires children to first clarify their own creative intentions and then transform those intentions into reference images through interaction with AIGC. AIGC serves only as an auxiliary tool for realizing creative ideas, not as the "leading source" of the ideas themselves. In contrast, children in the passive reference group passively receive images from AIGC before drawing, resulting in a lack of independent thinking and directly using reference images as creative templates, thus suppressing originality. From an educational perspective, the reduced originality in the passive reference group does more harm than good: while the static pre-input of AIGC provides visual inspiration for children, over-reliance on this type of "direct inspiration stimulation" weakens children's "autonomous exploration" and "deep thinking" abilities during the creative generation process, which is detrimental to the long-term development of their creative thinking.

In summary, this phenomenon indicates that in early childhood art education, the early stimulation of AIGC directly affects its role in creative generation. Pre-emptive, static, passive reference easily inhibits originality, while simultaneous thinking and reference during the creative process allows children to leverage the advantages of AIGC while preserving their autonomy in creative generation and maintaining or even enhancing their level of originality.

(2) AIGC can significantly improve children's design expression ability in painting

This study shows that young children who interact with artificial intelligence (AI) and refer to its generated content during the drawing process have significantly better design expression abilities than children who do not interact with AI throughout the process. This difference is clearly reflected in both the composition and color expression abilities, two core dimensions.

From the perspective of composition ability, experimental data shows that the average composition ability of middle-class children is generally in a transitional stage from "parallel composition" to "scattered composition." The formation of this stage characteristic is directly related to the current level of children's mental development: on the one hand, children's understanding of the spatial distribution patterns of objects in the real world is insufficient, and they have not yet established a systematic physical spatial logic; on the other hand, their understanding and focus on the core content of the painting theme is limited, making it difficult to construct a spatial relationship of "prominent subject and clear hierarchy" in the picture, only achieving a basic listing or preliminary integration of elements. In contrast, AI-generated content (AIGC) has two major advantages: first, it can accurately analyze the spatial distribution patterns of objects in the real world (although occasional cognitive biases or generation errors may occur due to algorithmic limitations); second, it can accurately identify children's creative intentions and generate reference images that conform to physical spatial logic and children's expressive needs. Experimental results show that children in the interactive generation group who referenced AIGC maintained a relatively stable "scattered composition" level in their drawing ability—this phenomenon indicates that the immediate reference of AIGC not only helped children establish a clear understanding of the physical form (such as shape features and size proportions) of the creative theme "vehicles," but also strengthened their clarity regarding their creative goals. It is noteworthy that the passive reference group, which only statically encountered AIGC during the PPT presentation, did not show a significant difference in drawing ability compared to the control group that did not encounter AIGC ($p > 0.1$). This result reveals that the static image display stage before drawing cannot significantly improve children's drawing ability. From the perspective of early childhood cognitive development characteristics, this is mainly due to the fact that the cognitive comprehension ability of middle-class children is still in the developmental stage: even if they are exposed to reference images generated by AIGC in the early stage, it is difficult for them to effectively transform the spatial logic and composition experience contained therein into practical operational ability in the painting process.

In the dimension of color expression ability, the immediate interactive effect of AIGC is also significant. Data from the control group showed that the fine motor skills of kindergarten children were still relatively weak, and their performance in terms of the rationality of color selection and the accuracy of coloring was average. However, the interactive generation group of children who received real-time AIGC references showed significant improvement in both color selection ability and coloring quality. This difference can be explained from two aspects: First, kindergarten children have limited experience in color matching, and the reference images generated by AIGC provided them with concrete examples of color matching, making up for their lack of experience. Second, the mode of referring to AIGC as needed during the painting process helped children clarify the direction of color selection and painting, enhancing their confidence and patience in painting. Some children who originally had weak painting foundations and were afraid of the coloring process or even the painting process gradually got rid of their fear after interacting with AI and were able to complete the coloring and painting tasks more attentively. Some children said after class, "AI told me what colors to use, and drawing wasn't difficult at all." Furthermore, there was no significant difference in color expression ability between the passive reference group and the control group, further confirming the positive effect of AIGC on color ability. This improvement needs to be based on "dynamic reference synchronized with the creative process," as static, pre-emptive visual input cannot extend this influence to the coloring stage.

The core conclusion from the above analysis is that in early childhood art education, when children can access AIGC content on demand, AI can more fully realize its auxiliary value—not only can it effectively guide children to plan their creative content according to the logic of real physical space and artistic expression (such as clarifying "what to draw next," "how to arrange elements," and "what colors to choose"), thus promoting the improvement of composition and color expression abilities; it can also deepen children's physical cognition of the real world and their logical understanding of the content of their paintings, stimulating their interest in painting and enhancing their creative confidence. This finding further illustrates that compared to static, pre-emptive visual input, the dynamic reference mode of AIGC synchronized with the creative process aligns with children's cognitive patterns and can more accurately and effectively assist children's painting.

4 CONCLUSION

This study compares the differences in fluency, originality, composition, and color expression in children's paintings before and after the introduction of AI in different stages of their painting process. The main conclusions are as follows: First, AI-generated content stimulation significantly enhances children's composition and color expression abilities in the design and expression stage. AI-generated content stimulation not only helps children correctly establish their understanding of real-world physical logic and painting layout logic, but also enhances their confidence in the painting process. Second, direct stimulation with AI-generated content without children's critical thinking does not effectively maintain the originality of children's paintings.

However, due to the small sample size of this experiment, more large-scale standardized experiments are needed to corroborate the above conclusions. Also, since this paper only selected kindergarten students from a provincial capital city in eastern China for the experiment, and did not conduct in-depth research on their family backgrounds, the practical applicability of the results to young children has not yet been verified.

The conclusions of this study can provide a reference for the design of kindergarten painting courses. In the context of the gradual penetration of artificial intelligence technology into educational scenarios, guiding children to establish a correct understanding and reasonable use of AI is an important issue that needs attention in preschool education and is also one of the important teaching goals of preschool teachers. Although AI-generated content can significantly improve children's painting skills and broaden their horizons, it is still necessary to emphasize the children's active role in painting; the process of human creative generation cannot be replaced by AI.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work was supported by the National College Student Innovation and Entrepreneurship Training Program of China under Grant 4023001.This work was supported by Mr Zhengyong Gong and Xianghong Cai.

REFERENCE

[1] Shao Mengting. Application of Artificial Intelligence in Kindergarten Art Activities. Shanghai Fashion, 2024(08): 107-109.
[2] Li Jing, Jiang Luzhen. Artificial Intelligence Assists Children's Art: A New Perspective on Song Performance Education. Yellow River Voice, 2025(06): 185-188.
[3] Teng Jing. A Study on the Cognitive and Neural Basis of Visual Art Creativity from the Perspective of Specialty. East China Normal University, 2021.
[4] Botella M, Zenasni F, Lubart T. What are the stages of the creative process? What visual art students are saying. Frontiers in psychology, 2018, 9: 2266.
[5] Yokochi S, Okada T. The process of art-making and creative expertise: An analysis of artists' process modification. The Journal of Creative Behavior, 2021, 55(2): 532-545.
[6] Runco A M, Acar S. Divergent Thinking as an Indicator of Creative Potential. Creativity Research Journal, 2012, 24(1): 66-75.
[7] Goldschmidt G. The dialectics of sketching.Creativity Research Journal, 2009, 4(2): 123-143.
[8] Tiansheng Xia, Sun Y, An Y, et al. The influence of music environment on conceptual design creativity. Frontiers in Psychology, 2023, 14: 1052257.
[9] Lin Lin, Luo Jinjing. Development and Guidance of Children's Painting Composition Ability. Preschool Education Research, 2005(12): 12-14.
[10] Pian Cen. A Study on the Development of Painting Expression Ability in Children Aged 3-6. Shanghai Normal University, 2014.
[11] Koo K T, Li Y M. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine, 2016, 15(2): 155-163.
[12] Boos D D, Duan K, Liu X. Pairwise comparisons for Levene-style variability parameters. Communications in Statistics-Simulation and Computation, 2023, 52(4): 1562-1576.
[13] Mukherjee H, Bhonge P. Assessing Skew Normality in Marks Distribution, a Comparative Analysis of Shapiro Wilk Tests. arXiv preprint arXiv:2501.14845, 2025.
[14] Conti P L, De Giovanni L. A Kruskal-Wallis Type Test for Multi-treatment Effect//Scientific Meeting of the Italian Statistical Society. Cham: Springer Nature Switzerland, 2024: 463-468.
[15] Arboretti R, Barzizza E, Biasetton N, et al. A review of multivariate permutation tests: Findings and trends. Journal of Multivariate Analysis, 2025: 105421.
[16] Sodipo E O, Sodipo A A, Adepoju K A. STATISTICAL ANALYSIS OF STUDENTS'ACADEMIC PERFORMANCE IN NIGERIA UNIVERSITIES: A CASE STUDY OF THE UNIVERSITY OF IBADAN, NIGERIA, 2014.