Title: Quantifying coloring skills in preschoolers

Abstract

Importance: Coloring is popular for preschool children, revealing their developmental states. However, interpreting coloring performances is challenging because commonly-used comments are descriptive and subjective with large variations.

Objective: To develop a scoring method to quantify children's coloring skills objectively.

Design: The colored blank train templates were analyzed using four indicators (entropy, complexity, coloring outside the lines, and unexpected blank areas) to form a summed score.

Setting: Community.

Participants: Data of 239 typically-developing children were used, aged 3 to 6 years.

Results: The summed score exhibited good internal consistency (Cronbach's $\alpha = 0.80$), discriminative validity ($p = 0.04$), convergent validity (rs = 0.66 and 0.59 with age and visual-motor integration), and acceptable factorial validity (Comparative fit index = 0.99, Standardized root-mean-square residual = 0.04, and Root-mean-square error of approximation = 0.13). Moreover, three coloring patterns (mature, transitional, and immature) were identified.

Conclusions and Relevance: The new method provides objective, reliable, and valid scores representing coloring skills in at least typically-developing children. In addition, the coloring patterns can be recognized. This method can be used to facilitate comparisons of children's coloring skills with peers, and provide valuable insight into children's development.

What This Article Adds: This study proposed a new method to objectively quantify children's coloring skills with sound reliability and validity in typically-developing children. The method can be used to evaluate children's coloring skills and patterns to shade the lights on their developmental stages.

Keywords: coloring skills, children, complexity, entropy, development

Introduction

Coloring is a popular and enjoyable activity among preschool children. This activity can be likened to a game due to its inherent characteristics of being fun, fostering creativity, and providing a sense of accomplishment. Thus, coloring skills is one of important play skills for children. Furthermore, coloring offers a multitude of developmental benefits. By engaging in coloring activities, children are able to channel their imagination and emotions through the vivid array of colors, simultaneously practicing patience and emotional regulation through the rhythmic and repetitive motions (Carsley et al., 2015; Turturro & Drake, 2022). Coloring exercises serve as a reflection of children's developmental progress, encompassing cognitive abilities (Holt et al., 2019), fine motor skills (Fitrianingsih & Sari, 2019; Oktavia et al., 2019; Priyantoro & Hasanah, 2023), and visual-motor integration (Martino & Lape, 2021). For instance, a child's aptitude for coloring can unveil their grasp of color recognition, arrangement of hues, attention span, dexterity, and the coordination between their hands and eyes. As coloring aptitude mirrors diverse developmental proficiencies, it can offer insights into a child’s overall growth, particularly concerning fine motor skills and cognition.

Coloring is commonly seen at home and in school, given the abundance of coloring materials and resources. Teachers have integrated coloring into early childhood education curricula, and parents have reported their children engaging in coloring activities at least once a week. Enhancing coloring skills becomes a meaningful educational objective for educators and parents alike, serving as a tool to enrich a child's developmental capabilities during their formative preschool years. Moreover, assessing coloring skills can serve as a mean for teachers, parents, and pediatric practitioners to gain a deeper understanding of a child's one of the play skills and the child’s comprehensive development, as discussed above. However, currently, the scoring of coloring skill is highly reliant on subjective descriptive comments, such as *good*, *beautiful*, or *not seriously coloring*. Although these subjective descriptions provide straightforward assessments of coloring performance, their utility is limited for two reasons. First, the descriptive comments are difficult to further interpret because the key indicators for determining coloring performance are largely unknown. Second, such subjective descriptions are difficult to analyze because of the lack of a consistent rating scale and the large variation in the descriptions. Moreover, the meanings of those descriptions could vary by person. Therefore, it is difficult to draw scientific conclusions from coloring tasks with these subjective comments. To address these issues, a quantitative scoring method with clear indicators for determining coloring skills in preschool children is needed.

Four indicators for determining children's coloring skills have been proposed from drawing theory and previous studies, including coloring outside the lines, unexpected blank areas, entropy, and complexity. The first two indicators, coloring outside the lines and unexpected blank areas, are proposed based on developmental milestones. It is assumed that the less frequently children color outside the lines and the fewer blank spaces remain, the better their coloring skills. These two indicators are commonly employed as rating criteria in assessments of fine motor skills. For example, the Peabody Developmental Motor Scales - Second Edition includes an item that assesses a child's ability to color within two lines (Foli & Fewell, 2000). Therefore, these two indicators can reasonably reflect children's levels of coloring skill.

The next two indicators were proposed by Sigaki and his colleagues (Sigaki et al., 2018). They utilized entropy and complexity to capture the local spatial patterns in the images, which might reflect coloring skill. Entropy refers the amount of disorder or chaos present in the images, with higher levels of disorder corresponding to higher entropy ratings. Thus, entropy captures the degree of organization or disorganization in coloring patterns. Complexity, on the other hand, represents the variability of patterns within an image, ranging from simplicity to complexity. Given the typical progression of children's motor and cognitive development, it is expected that children's coloring abilities will evolve from disordered (scribbles) to ordered patterns, and from simplicity (single color and simple patterns) to complexity (varied colors and intricate patterns).

The four aforementioned indicators (i.e., coloring outside the lines, unexpected blank areas, entropy, and complexity) demonstrate potential for scoring children's coloring skills. Thus, the aim of this study was to utilize these four indicators to comprehensively quantify children's coloring abilities from a coloring task. Moreover, the psychometric properties of this scoring method were examined. We expected that the newly developed scoring method would be reliable and valid; assist teachers, parents, and pediatric practitioners in better assessing children's play skills; and serve as an outcome measure for comparison with others.

Methods

Participants

This study was part of a large project utilizing artificial intelligence to assess children's motor skills, emotion states, and attention (Lin et al., 2023). The inclusion criteria were as follows: children who (1) were between the ages of 3 to 6 years, (2) could follow the researchers' instructions, and (3) agreed to be recorded on video. The data for this study were collected from a kindergarten in [masked for review]. A total of 239 children (mean age = 4.91 years, SD = 0.98) were recruited in our study, and around half of them (49.8%) were boys. Among the children, about 5 to 6% had a diagnosis of developmental delay/disability, and about 8% have/had received rehabilitation services (e.g., occupational therapy and physical therapy). Table 1 shows the demographics of the participants. Sample size over 200 was considered adequate for factor analysis (Comrey & Lee, 2013). This study was approved by the institutional review board of a medical center in [masked for review].*Measures*

1. **Coloring activity.** An A4-sized sheet of paper with a train template design was provided

for children to color. The train was composed of basic shapes of different sizes, including squares, triangles, circles, and rectangles, and it was printed on the top half of the paper (Appendix 1). The train was intentionally designed to include both large and small sizes, incorporating a variety of shapes suitable for preschool children in K1 to K3. To ensure its appropriateness, we sought validation after designing it. We invited two experienced senior pediatric occupational therapists to assess its suitability for preschool children. Both pediatric occupational therapists unanimously affirmed the train's suitability for this age group.

2. **The Beery-Buktenica Visual-Motor Integration, Chinese version (Berry-VMI-C).** In our study, we administered the Berry-VMI-C as a measure of convergent validity for coloring skill (Liu & Lu, 1999). The Berry-VMI-C comprises 27 items that necessitate children to imitate or copy geometric shapes. A higher sum of scores on this measure indicates a better ability of visual-motor integration. The Berry-VMI-C has demonstrated excellent psychometric properties (Liu & Lu, 1999).

Procedure

To recruit participants, we employed phone calls to establish communication with the school principal and provide an overview of our research project. Once the principal expressed their willingness to participate in our study, we disseminated the research cover letter and recruitment notice to parents through school teachers. Parents who indicated their agreement to partake in the research returned the Response Form to our team. Subsequently, researchers provided the parents with the informed consent documents. Upon receiving the informed consents from parents, the researchers then conducted group activities in the kindergarten, including coloring, origami, and copying a person, which were designed by the research team. The Berry-VMI-C was also administered to each child. Each group consisted of 3 to 6 children, and the entire administration time lasted approximately 40 to 50 minutes. After the group activities, the researchers scanned the colored pictures.

Data analyses

Four indicators were computed from the colored images, including entropy, complexity, coloring outside the lines, and the area of unexpected blank areas.

The entropy was calculated using the following formula:

$$\mathrm{H(P)} = \frac{1}{\ln(n)} \sum_{1}^{n} p_i \ln\left(\frac{1}{p_i}\right)$$

where P is the probability distribution = $\{p_i; i = 1, \ldots, n\}$; n is the number of possible permutations and $\ln(\mathrm{n})$ is the maximum value of the entropy (i.e., $\sum_{i=0}^{n} p_i \ln\left({}^{1}/{p_i}\right)$. The detailed explanation of the formula can be found in the article by Sigakia et al (Sigaki et al., 2018). The value of H quantifies the degree of "disorder" in the occurrence of the pixels of an image represented by a matrix. We have H ≈ 1 if the pixels appear in random order (i.e., disorganized), and H ≈ 0 if they always appear in the same order (i.e., organized).

The complexity was calculated using the following formula, which captures the degree of structural complexity present in the above-mentioned matrix:

$$\mathrm{C(P)} = \frac{D(P,U)H(P)}{D^*}$$

where $\mathrm{D(P,U)}$ is a relative entropic measure (the Jensen-Shannon divergence) between $\mathrm{P} = \{p_i; i = 1, \ldots, n\}$ and the uniform distribution $\mathrm{U} = \{u_i = {}^{1}/{n}\,; i = 1, \ldots, n\}$ defined as

$$\mathrm{D(P,U)} = \mathrm{S}\left(\frac{P+U}{2}\right) - \frac{S(P)}{2} - \frac{S(U)}{2}$$

where $\frac{P+U}{2} = \{p_i + {}^{1}/{n}\,; n, i = 1, \ldots, n\}$ and D^* is the maximum of $\mathrm{D(P,U)}$, and it is a normalization constant (obtained by calculating $\mathrm{D(P,U)}$ when just one component of P is equal to 1 and all others are zero). The quantity $\mathrm{D(P,U)}$ is zero when all permutations are equally likely to happen, and it is larger than zero if there are privileged permutations. Thus, C(P) is zero in both extremes of order (P $= \{p_i = \delta_{1,i}; i = 1, \ldots, n\}$) and disorder (P $= \{p_i = {}^{1}/_{n}; i = 1, \ldots, n\}$). A higher $\mathrm{C(P)}$ indicates more variability in the coloring pattern. A detailed explanation of the formula can be found in the article by Sigakia et al (Sigaki et al., 2018).

The coloring outside the lines was calculated by the following steps. First, the colored images were converted into grayscale images. Second, the contours of the colored train were drawn using the OpenCV programming function. Third, the contour image was converted into a binary image (Appendix 2a). Finally, the white regions of the binary image were calculated. The white regions included the whole train image and the areas that the coloring was outside the lines.

The area of unexpected blank areas was calculated using the following steps. First, we merged the black-and-white image obtained as described in the above paragraph with the grayscale image to create a new image (Appendix 2b). Next, we identified and measured the blank areas in the new image, which were then classified as unexpected blank areas.

Among the four indicators, smaller values of entropy, coloring outside the lines, and unexpected blank areas indicated better coloring skill, while larger values of complexity indicated better skill. To sum up the four indicators to represent coloring skill, and for easier understanding by the readers, we further added minus signs to the values of entropy, coloring outside the lines, and unexpected blank areas to transform the scale. The four indicators were transformed into z-scores and summed up to represent the coloring skill (i.e., coloring skill index).

The psychometric properties of the scoring method were examined, including the internal consistency, convergent validity, discriminant validity, and construct validity. To examine the internal consistency, Cronbach's α was used for the four indicators. Cronbach’s α higher than 0.8 and 0.7 indicates good and acceptable internal consistency, respectively (Ertaş et al., 2004).

To examine the convergent validity, Pearson's *r* was used to calculate the correlations between coloring skill index, the Berry-VMI-C scores, and the children's chronological ages. Pearson's *r* higher than 0.5 indicates good convergent validity.

To examine the discriminant validity, the children were categorized into two groups: children with typical development and children with developmental delay/disability. Children with developmental delay/disability were coded based on caregivers' reports. Children of caregivers reporting a diagnosis related to developmental delay/disability were categorized into the developmental delay/disability group. The two groups were first examined for differences in sex and age. If significant differences in sex or age existed between the two groups, analysis of covariance (ANCOVA) would be used to examine the differences in coloring skills between the two groups using sex and age as covariates. If there were no significant differences in sex and age between the two groups, independent *t* test would be used for comparing the two groups. Significant differences in coloring skill between the two groups indicated good discriminant validity.

Construct validity was examined using confirmatory factor analyses. Model fit was examined with the following fit indexes: the root-mean-square error of approximation (RMSEA), the comparative fit index (CFI), the goodness of fit (GFI), and the Standardized root-mean-square residual (SRMR). RMSEA $\leq .05$, CFI > 0.95, GFI > 0.95, and SRMR < 0.8 indicate good model fit.

Coloring patterns from the four indicators

Cluster analyses were then applied to the four indicators to classify children's coloring patterns. Two steps were used for the cluster analyses. First, a hierarchical method was used to explore the potential clusters. A tree diagram was used to help inspect the potential clustering. Second, K-means clustering was used to identify the final clusters. After the clusters were identified, analysis of variance (ANOVA) was used to investigate the differences in the four indicators, coloring skill index, VMI scores, and ages to explore the characteristics of the identified clusters.

Results

The demographics of the participants

A total of 239 children (mean age = 4.91 years, SD = 0.98) were recruited in our study, and around half of them (49.8%) were boys. Among the children, about 5 to 6% had a diagnosis of developmental delay/disability, and about 8% have/had received rehabilitation services (e.g., occupational therapy and physical therapy). Table 1 shows the demographics of the participants.

The internal consistency of the coloring skill index

The results of our study showed that Cronbach's α was 0.80, indicating good internal consistency.

Convergent validity of the coloring skill index

The coloring skill index was moderately correlated with age ($r = 0.59$, $p < .05$) and visual-motor integration ($r = 0.66$, $p < .05$).

Discriminant validity of the coloring skill index

Because no significant differences in age existed between the typically developing group and the developmental delay/disability group ($p < .05$), the t test was used to examine differences in coloring skill. The average scores of coloring skill were 0.10 (SD = 3.10) and -1.74 (SD = 3.78) for children without and with developmental delay/disability, respectively. A significant difference in coloring skill existed between the two groups (T = 2.05, $p = 0.04$).

Construct validity of the coloring skill index

Because the complexity was calculated from entropy, complexity and entropy had a high negative correlation. Therefore, we added a correlation between complexity and entropy in the model. The results of the confirmatory factor analyses found that the fit indexes of the one-factor model were acceptable (Chi-squared = 7.93, CFI = 0.99, GFI = 0.98, SRMR = 0.036, and RMSEA = 0.13). Figure 1 shows the factor structure and factor loadings of coloring skill.

Coloring patterns classified from the four indicators

Three potential clusters were first identified from the tree diagram of the hierarchical method (Appendix 3). Therefore, three clusters were applied to the K-means clustering analyses. The results showed three distinct clusters of children's colored pictures. Significant differences in the four indicators were found among the three clusters (Table 2). Significant differences among the three clusters were found in the four indicators, coloring skill, VMI scores and ages. Post hoc analyses showed that the four indicators, coloring skill index, VMI scores and ages were best/highest in cluster 1, followed by cluster 2 and then cluster 3. We further inspected a cross table of children's ages and the clusters (Appendix 4). In cluster 1, most of the children were between 6.0 and 6.4 years old (n=28 in 106; 26.4%); in cluster 2, most were between 4 and 5 years old (n=52 in 122; 42.6%); and in cluster 3, most were 3 years old (n=9 in 11). According to the characteristics of the pictures in the three clusters, they could be termed mature coloring (Figure 2a), transitional coloring (Figure b), and immature coloring (Figure 2c).

Discussion

This study aimed to develop a scoring method to quantify children's coloring skills through a coloring task. Four indicators were used to assess the coloring task, and an overall coloring score (i.e., coloring skill index) was calculated to represent each child’s coloring skill. This scoring method had acceptable internal consistency, good discriminant validity, and good convergent validity with the children’s chronological age and visual-motor integration. Therefore, the overall coloring scores seem to validly reflect how well children perform in coloring. Prospective users may access the authors to access the algorithm.

Overall, the scoring method had acceptable to good validity in our study. The coloring skill index showed adequate correlations to the children’s ages and VMI scores, and they were discriminative between children with and without developmental delay. Moreover, adequate-to-good data-model fits were found for the four indicators, suggesting that the four indicators assess the same latent trait and supporting its good factorial validity (Byrne, 1998). Therefore, the scores of the four indicators can be validly used to calculate a single score representing a child’s overall coloring skill. Based on these findings, this study developed a valid method for quantifying children’s coloring skills.

The internal consistency of the four indicators (0.8) exceeded the criterion for group-level comparisons (0.7) (Huang et al., 2016). This finding suggests that the overall coloring scores can reliably assess coloring performance in children. Although the Cronbach’s α of the coloring skill index was lower than 0.90, a criterion for individual-level comparisons, it is still a promising result because the value of Cronbach's α is influenced by the number of items included in the scale; typically, more items tend to elevate the value (Tavakol & Dennick, 2011). Considering that the coloring skill index was calculated based on only four indicators, the current reliability could be considered promising.

The four indicators match the graphic development theory proposed by Lowenfeld and Brittain (Lowenfeld & Brittain, 1987). According to the graphic developmental theory, the drawings of preschool and school-aged children can be categorized into four stages: scribbling (2 to 4 years), pre-schematic (4 to 6 years), schematic (7 to 9 years), and drawing realism (9 to 11 years). During the scribbling stage, children focus more on drawing lines, orientations, and forms than the use of colors, resulting in simple colors with rough scribbles. As children progress to the pre-schematic and schematic stages, children become more focused on the meanings of the colors, and their fine motor skills develop to allow them to color within the lines. Consequently, their coloring tends to be more complex and organized. Thus, the four indicators are consistent with the graphic developmental theory.

Our study revealed that children's coloring skill can be classified into three distinct levels: mature, transitional, and immature coloring. Furthermore, upon examining their correlations with age, we observed that mature coloring typically emerges after the age of 5, transitional coloring tends to manifest between the ages of 4 and 5, and immature coloring is commonly observed at the age of 3. These findings can serve as valuable references for parents, clinicians, and school teachers, highlighting the importance of encouraging coloring activities for children in the middle preschool stage. For example, if a child shows immature coloring, filling blanks with varying shapes may be promising practice for them. Such coloring activities can provide the just-right challenge and can conceptually optimize their improvement in coloring.

There are two potential applications for the results of our study. First, an app-based assessment of coloring skill can be developed. Pediatric practitioners, parents, and school teachers can upload a photo of a colored image, and the assessment tool will provide immediate feedback on a child's coloring skill. Second, since a child's coloring skill can be quantified, it can be used to track their progress and compare with their peers. This quantification may serve as a screening tool for parents and caregivers to detect children at high risk of developmental delay early on. Additionally, it can be used as an evaluation tool for pediatric practitioners as a reference for interventions.

It is important to note that the coloring skill quantified in our study was a specific skill unrelated to the sense of use of colors (or aesthetic feeling). The entropy and complexity indicators measure children's understanding of coloring concepts, while coloring outside the lines and unexpected blank areas measure their motor skills. Our method of assessment does not evaluate whether a child uses appropriate colors or whether the colored image is aesthetically pleasing. Instead, our focus is on evaluating how well a child can execute a coloring activity.

Study Limitations

There were two limitations to our study that should be acknowledged. First, the sample used in our study was recruited from one kindergarten, which may limit the generalizability of our findings. Therefore, it is necessary for future studies to recruit more diverse samples. Second, some of the pictures analyzed in our study had issues with color pens running out of ink due to repeated use, which could have impacted our ability to accurately calculate complexity and unexpected blank areas. However, it should be noted that this issue only affected a small portion of the pictures (less than 15 pictures), and it is unlikely to have significantly impacted the overall trends observed in our study.

Implications for Occupational Therapy Practice

This study has the following implications for occupational therapy practice:

1. The new method can provide objective, reliable, and valid scores representing children’s

coloring skills. Therefore, children's coloring performances can be compared with peers, contributing valuable insights into their developmental stages.

2. The coloring patterns can be reorganized by the method. Thus, the targets of interventions may be optimized depending on their current patterns and those of the next level.

Conclusions

Our study aimed to develop a new scoring method to quantify coloring skill from a coloring task. Four indicators, namely, entropy, complexity, coloring outside the lines, and unexpected blank areas, were used to generate a summed score representing overall coloring skill. The new scoring method had good internal consistency, discriminative validity, convergent validity with ages and visual-motor integration, and adequate construct validity. Therefore, the quantification method proposed in this study can be used to identify children's coloring skill, compare their performances with their peers, and gain valuable insights into children's engagement with coloring tasks.

Acknowledgements

We appreciate all the caregivers and children who participated in our study. We also thank all the teachers in kindergartens who referred children.

This study was supported by [masked for review].

The authors have no conflicts of interest to disclose. All authors agree with the stated authorship and contributions of this article.

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

References

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.

Carsley, D., Heath, N. L., & Fajnerova, S. (2015). Effectiveness of a classroom mindfulness coloring activity for test anxiety in children. *Journal of Applied School Psychology*, *31*(3), 239-255. https://doi.org/10.1080/15377903.2015.1056925

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology press.

Ertaş, M., Siva, A., Dalkara, T., Uzuner, N., Dora, B., Inan, L., . . . Şirin, H. (2004). Validity and reliability of the Turkish Migraine Disability Assessment (MIDAS) questionnaire. *Headache: The Journal of Head and Face Pain*, *44*(8), 786-793. https://doi.org/10.1111/j.1526-4610.2004.04146.x

Fitrianingsih, N., & Sari, N. S. N. I. (2019). The influence of picture coloring on fine motor development In children aged 4-5 years. *Journal of Science Innovare*, *2*(1), 19-22. https://doi.org/10.33751/jsi.v2i01.1525

Foli, R., & Fewell, R. R. (2000). *Peabody Developmental Motor Scales (Examiner's Manual)*. Pro-Ed Inc.

Holt, N. J., Furbert, L., & Sweetingham, E. (2019). Cognitive and affective benefits of coloring: Two randomized controlled crossover studies. *Art Therapy*, *36*(4), 200-208. https://doi.org/10.1080/07421656.2019.1645498

Huang, C. Y., Lin, G. H., Huang, Y. J., Song, C. Y., Lee, Y. C., How, M. J., . . . Hsieh, C. L. (2016). Improving the utility of the Brunnstrom recovery stages in patients with stroke: Validation and quantification. *Medicine (Baltimore)*, *95*(31), e4508. https://doi.org/10.1097/MD.0000000000004508

Lin, G. H., Lee, S. C., Yu, Y. T., & Huang, C. Y. (2023). Machine learning-based brief version of the Caregiver-Teacher Report Form for preschoolers. *Research in Developmental*

Disabilities, *134*, 104437. https://doi.org/10.1016/j.ridd.2023.104437

Liu, H., & Lu, L. (1999). *The Beery–Buktenica Developmental Test of Visual–Motor Integration—Chinese version: Manual*. Psychological Publishing.

Lowenfeld, V., & Brittain, W. (1987). *Creative and mental growth*. Prentice-Hall.

Martino, E. M., & Lape, J. E. (2021). Occupational therapy in the preschool classroom-Promoting fine motor and visual motor skills for kindergarten readiness. *Journal of Occupational Therapy, Schools, & Early Intervention*, *14*(2), 134-152. https://doi.org/10.1080/19411243.2020.1822261

Oktavia, D., Bali, M., Rahman, H., Umar, U., Syakroni, A., & Widat, F. (2019). Exploration of fine motor skills through the application of paint. Proceedings of 1st Workshop on Environmental Science, Society, and Technology, WESTECH 2018, December 8th, 2018, Medan, Indonesia,

Priyantoro, D. E., & Hasanah, U. (2023). Implementation of coloring activities early childhood in developing fine motor skills. *Journal of Childhood Development*, *3*(1), 1-12. https://doi.org/10.25217/jcd |

Sigaki, H. Y., Perc, M., & Ribeiro, H. V. (2018). History of art paintings through the lens of entropy and complexity. *Proceedings of the National Academy of Sciences*, *115*(37), E8585-E8594. https://doi.org/10.1073/pnas.1800083115

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53-55. https://doi.org/10.5116/ijme.4dfb.8dfd

Turturro, N., & Drake, J. E. (2022). Does coloring reduce anxiety? Comparing the psychological and psychophysiological benefits of coloring versus drawing. *Empirical Studies of the Arts*, *40*(1), 3-20. https://doi.org/10.1177/0276237420923290

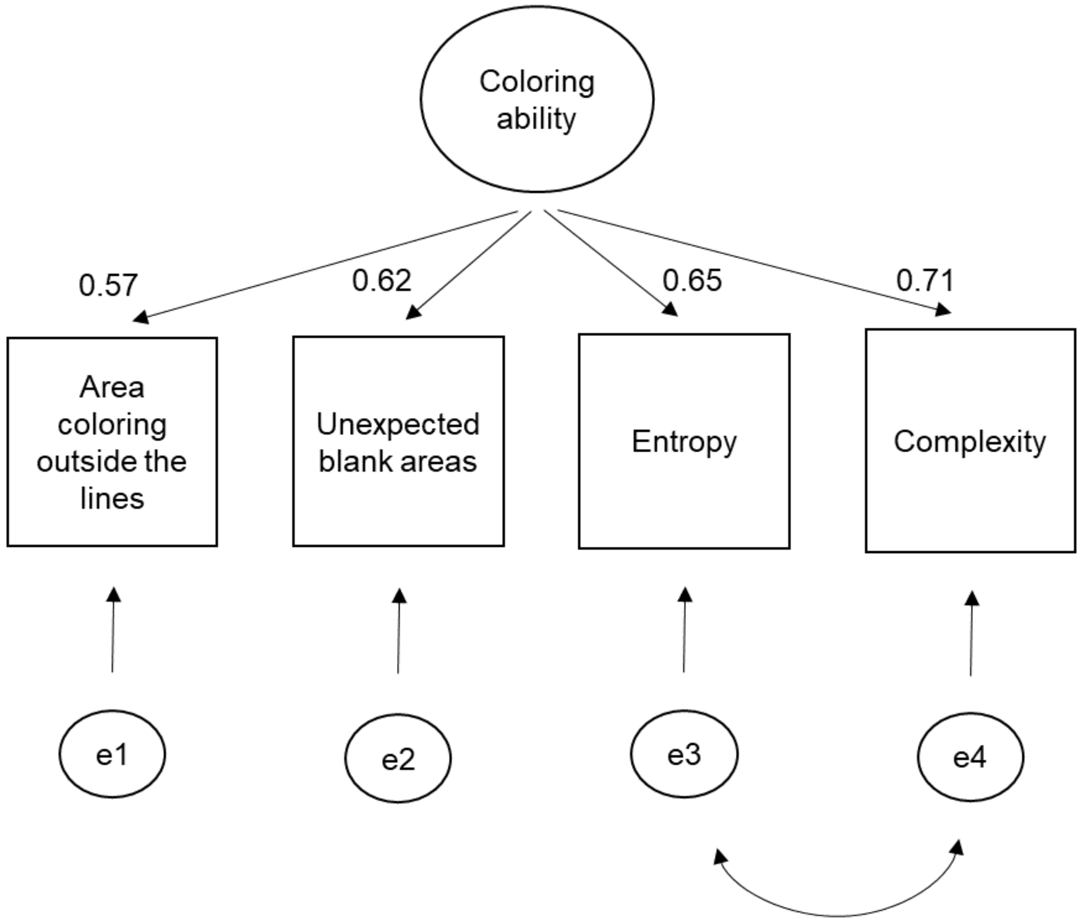


Figure 1. Factor structure and factor loadings of the coloring skill.

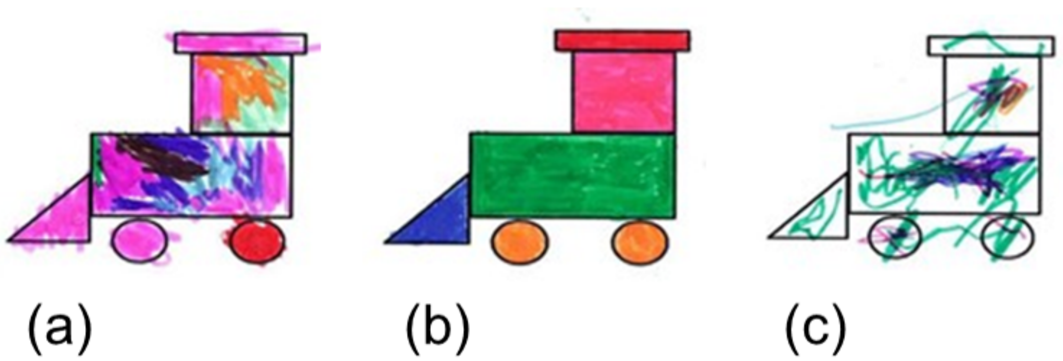


Figure 2. The examples of the pictures in the three clusters. (a) is in cluster 1, which could be defined as mature coloring, (b) is in cluster 2, which could be defined as transitional coloring, and (c) is in cluster 3, which could be defined as immature coloring.

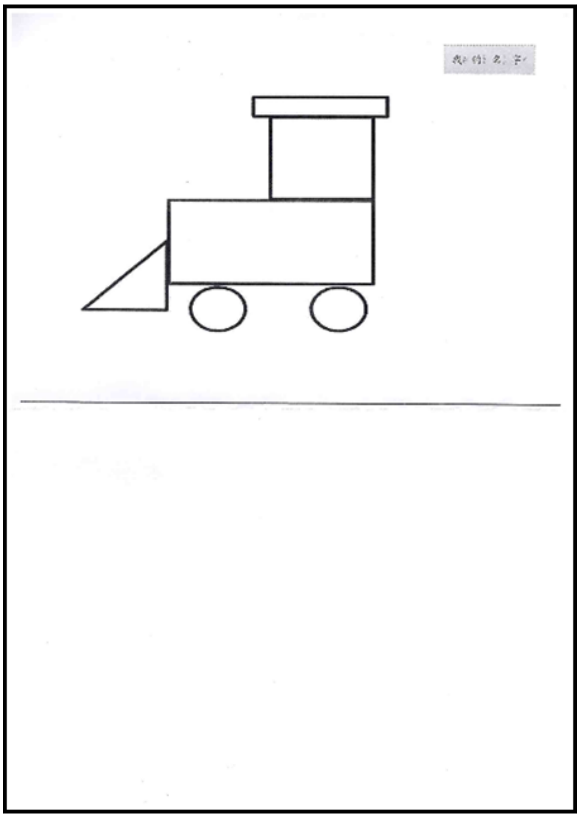
Table 1. Demographics of the participants (N=239)

Variables	Statistics
Sex (male): n (%)	119 (49.8)
Age (years): n (%)	
3	46 (19.2)
4	82 (34.3)
5	66 (27.6)
6	45 (18.8)
Having a diagnosis of developmental delay/disability: n (%)	
Developmental delay	12 (5%)
Hyper activity and inattention	4 (1.7%)
Autism spectrum disorder	1 (0.4%)
Rehabilitation received: n (%)	21 (8.8)
Mother's educational level	
Family income	

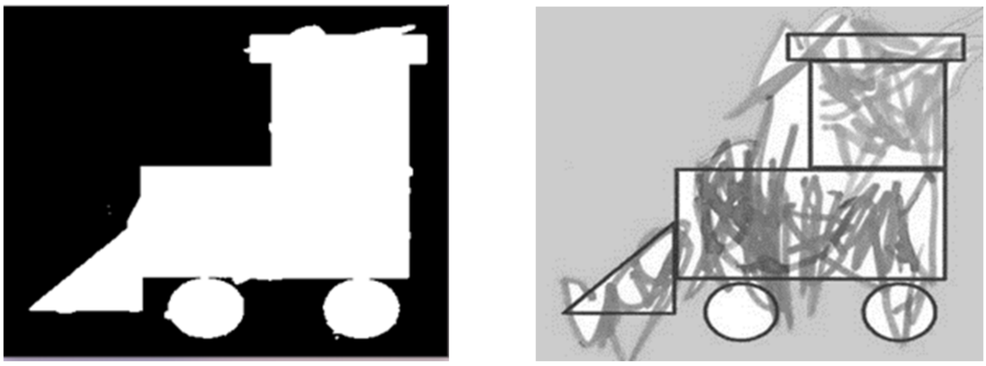
Table 2. The coloring patterns identified from the four indicators

	Cluster groups			
	1 (n=106): Mean (SD)	2 (n=122): Mean (SD)	3 (n=11): Mean (SD)	Post-hoc analyses
Coloring outside the lines	0.32	-0.01	-3.01	1 > 2 > 3
	(0.23)	(0.47)	(3.08)	
Unexpected blank areas	0.31	0.01	-3.2	1 > 2 > 3
	(0.30)	(0.60)	(2.5)	
Entropy	0.89	-0.64	-1.46	1 > 2 > 3
	(0.64)	(0.51)	(0.88)	
Complexity	0.87	-0.61	-1.67	1 > 2 > 3
	(0.58)	(0.55)	(0.93)	
Coloring skill index	2.40	-1.24	-9.34	1 > 2 > 3
	(1.27)	(1.54)	(3.98)	
Age	65.68	54.33	41.00	1 > 2 > 3
	(10.15)	(9.77)	(4.15)	
VMI scores	14.02	10.07	4.00	1 > 2 > 3
	(3.08)	(3.52)	(2.41)	

Appendix 1. The blank train for children to color

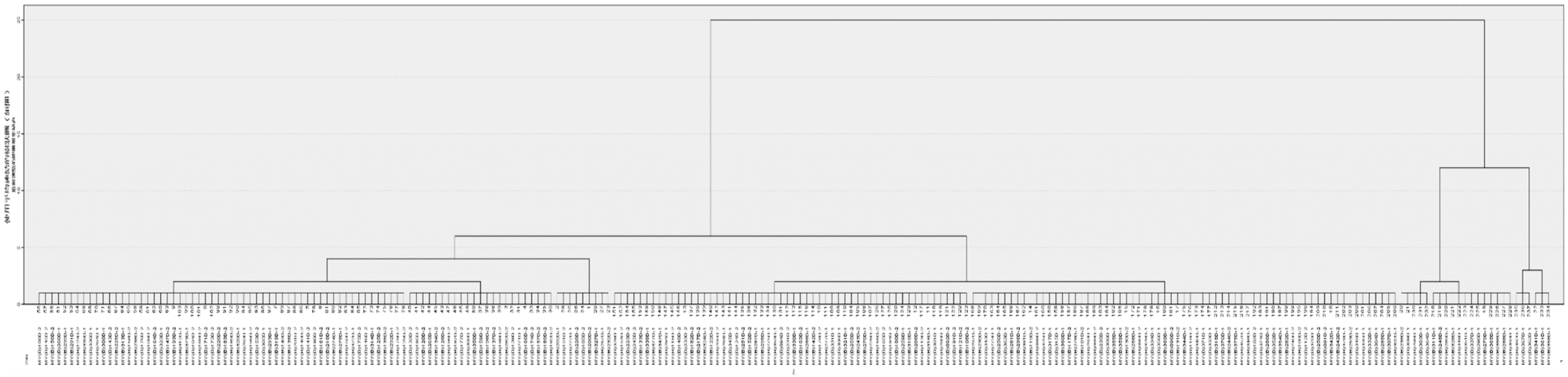


Appendix 2. (a) An example of areas colored outside the lines (and the train template area), and (b) an example of unexpected blank areas. The two examples are not from the same pictures.


(a) (b)

404

Appendix 3. The tree diagram of the hierarchical methods. Three potential clusters were identified.



Appendix 4. The cross-table between children's ages and the clusters

	ages								
Cluster	3.0–3.4	3.5–3.9	4.0–4.4	4.5–4.9	5.0–5.4	5.5–5.9	6.0–6.4	6.5–6.9	Total
1	2	2	12	16	18	15	28	13	106
2	13	20	28	24	19	13	4	1	122
3	9	0	2	0	0	0	0	0	11
Total	24	22	42	40	37	28	32	14	239