International Journal of Educational Studies and Policy (IJESP)

Volume: 7, Issue: 2, May 2026

Artistic Thinking Disposition Scale for Middle School Students: A Validity and Reliability Study

Serkan Aslan[1], Mehmet Ali Gökdemir[2]

ABSTRACT

This study developed and validated a 19-item multidimensional scale to assess middle school students' (Grades 5–8) disposition towards artistic thinking in visual arts education within a developmentally appropriate and domain-specific framework. Addressing the limited availability of such instruments, the study aimed to provide a theoretically grounded and practically useful tool for research and instructional evaluation purposes. Nine experts reviewed the initial pool of 28 items, yielding strong content-related evidence (CVR = .667–1.000; CVI = .988). Evidence pertaining to the response processes of 50 students indicated that the items were generally comprehensible (item-level indices = .85–1.00; overall index = .92). To examine the internal structure, exploratory factor analysis using principal axis factoring with Promax rotation was conducted on data from 499 students. The findings supported a three-factor structure explaining 38.232% of the extracted variance, with factor loadings ranging from .35 to .87. Confirmatory factor analysis on an independent sample of 389 students yielded acceptable model fit ($\chi^2/df = 1.93$; CFI = .914; TLI = .902; RMSEA = .049; SRMR = .051). Reliability and model-based validity findings generally supported the psychometric adequacy of the scale, and multi-group CFA indicated acceptable scalar invariance across genders. Overall, the artistic thinking disposition scale offers a valid, reliable, and developmentally sensitive measure of artistic thinking disposition.

Keywords: Artistic thinking disposition, visual arts education, middle school students, scale development, psychometric properties.

DOI: https://doi.org/10.5281/ijesp.1903627

Article Info:

Received: 06.03.2026 **Accepted:** 14.04.2026 **Article Type:** Research Article

Cite as: Aslan. S. & Gökdemir, M. A. (2026). Artistic thinking disposition scale for middle school students: A validity and reliability study. *International Journal of Educational Studies and Policy, 7*(2), 193-226.

[1]Corresponding Author: Serkan Aslan, Suleyman Demirel University, Faculty of Education, serkanaslan@sdu.edu.tr, ORCID: 0000-0001-8515-4233

[2]Mehmet Ali Gökdemir, Dicle University, Faculty of Art and Design, mali.gokdemir@dicle.edu.tr, ORCID: 0000-0002-5100-5026

Introduction

Visual arts education is positioned as a multicomponent learning domain in which students shape their thinking through discriminating, making sense of, interpreting, and expressing visual stimuli, beyond acquiring technical skills such as line, form, and composition (Schneider & Rohmann, 2021; Winner et al., 2013). The open-ended task structure of art activities and the nature of products, which are often evaluated according to subjective criteria, direct students towards trial and error, generating alternatives, and working with uncertainty. For this reason, explaining artistic learning solely in terms of cognitive competences and technical performance remains insufficient; it is also necessary to consider students' thinking dispositions in relation to artistic engagement (Schneider & Rohmann, 2021; Winner et al., 2013).

In this context, artistic thinking disposition refers to a student's readiness to think in the visual arts and their inclination to engage with artistic tasks in interpretive, imaginative, and expressive ways. In the literature, thinking disposition is conceptualised not simply as the possession of ability, but as a broader tendency that shapes whether, when, and how individuals engage in thinking. Within this framework, disposition comprises three interrelated components: sensitivity, referring to noticing and picking up cues; inclination, referring to the willingness to engage in and sustain thinking; and ability, referring to the repertoire needed for successful performance (Perkins et al., 1993). In visual arts education, artistic performance is often interpreted in terms of talent, ability, or technical skill, such as drawing accurately, composing effectively, or producing visually compelling work. However, artistic thinking cannot be explained solely through such performance-based indicators. From a dispositional perspective, what matters is not only whether students can perform successfully, but also whether they are willing and ready to engage in, sustain, and deepen thinking throughout artistic activity. In this regard, the literature on thinking dispositions intersects meaningfully with the Studio Habits of Mind framework. Habits such as Envision, Express, Observe, Reflect, and Stretch and Explore describe the cognitive and affective processes through which students initiate, regulate, and sustain artistic thinking while making art (Hetland et al., 2013). Accordingly, artistic thinking disposition can be conceptualised as a domain-specific orientation that extends beyond talent or technical proficiency and refers to students' willingness and readiness to engage in artistic thinking across the observe–interpret–produce cycle (Hetland et al., 2013; Perkins et al., 1993).

Artistic thinking disposition refers to a student's willingness and readiness to engage in, sustain, and deepen artistic thinking throughout the artistic process. From the perspective of the Studio Habits of Mind framework, this disposition is reflected in students' capacity to tolerate uncertainty, explore alternative pathways, take creative risks, and persist in production-oriented activity, particularly through habits such as Stretch and Explore, Envision, Observe, and Reflect (Hetland et al., 2013). In this sense, artistic thinking disposition represents a dispositional ground that activates forms of thinking requiring flexibility, exploration, and sustained engagement during the observe–interpret–produce cycle. This construct becomes especially important during the middle school period (Grades 5–8), a developmental stage in which peer evaluation becomes more salient, self-perceptions may become more fragile, and performance anxiety may intensify. Under such conditions, some students may withdraw from creative and production-oriented tasks that involve the possibility of making mistakes. Therefore, making artistic thinking disposition visible at the middle school level provides a functional conceptual basis for explaining continuity in students' artistic participation and for structuring instruction in a developmentally and target-sensitive manner (Winner et al., 2013).

A distinctive component of artistic thinking disposition is an orientation towards aesthetic processing and judgement. Models that conceptualise aesthetic experience as a multi-stage process emphasise that steps of perceptual analysis, cognitive meaning-making, and evaluation explain the formation of aesthetic judgement in ways that go beyond 'liking' (Leder et al., 2004; Leder & Nadal, 2014). From this perspective, an artistic thinking disposition is not merely an orientation towards production; it also encompasses a disposition to evaluate art products in a reasoned manner, develop interpretive reasoning, and defend aesthetic decisions based on evidence/criteria (Leder et al., 2004; Leder & Nadal, 2014). Such a framework provides a theoretical basis for ensuring that students' evaluations of art products do not remain at the level of 'I liked it/did not like it', but instead orient towards patterns of thinking that discriminate visual qualities, construct context and meaning, and arrive at reasoned judgement.

Another component concerns processes of visual imagery and visual transformation. Visual imagery is a fundamental cognitive resource that enables one to construct a visual representation in the mind in a manner akin to perceptual experience without external visual stimuli, and to examine and reorganise this representation during the production process in order to evaluate alternatives; in this respect, it serves students' internal testing of design alternatives and the revision of formal decisions throughout the process (Berardi, 2025; Burnett Heyes et al., 2017). Considered together with the studio habits of mind approach, the capacity for imagery to support 'envisioning (design)' and the cycle of 'generating alternatives–reorganising' during production, strengthens the explanatory power regarding the cognitive infrastructure of artistic thinking disposition (Hetland et al., 2013; Marks, 1973).

This theoretical line indicates that artistic thinking disposition can be addressed as a construct integrating components such as (i) noticing artistic problems and orienting towards undertaking them, (ii) sustaining patterns of exploring–risk-taking–trying alternatives during the production process, (iii) activating aesthetic evaluation and interpretive reasoning, and (iv) mobilising imagery/transformation-based design decisions (Hetland et al., 2013; Leder et al., 2004; Perkins et al., 1993). For this reason, the development of measurement approaches capable of reliably assessing middle school students' artistic thinking dispositions can be considered not merely a psychometric preference but also a strategic necessity for monitoring the quality of instructional processes, diagnosing students' strengths, and enriching learning opportunities in an evidence-based manner (Schneider & Rohmann, 2021; Winner et al., 2013).

When measurement instruments in the literature are examined, psychometric initiatives in the field of visual arts appear to cluster largely around the construct of attitude. Such scales typically operationalise evaluative components such as liking, interest, valuing, and intention to participate in relation to art or the visual arts course; thus, they primarily render visible students' affective orientations towards art rather than the ways they engage in thinking within artistic processes (Aslantaş, 2014; Dede, 2016). Likewise, some instruments developed in the arts have addressed broader constructs such as art literacy; however, such measures do not directly aim to capture middle school students' domain-specific artistic thinking patterns as they emerge during classroom activity, such as generating creative alternatives, making aesthetic justifications, drawing interpretive inferences, or adapting solution pathways (Yücetoker, 2014). Accordingly, the main gap in the literature is not simply the limited number of instruments, but the lack of a developmentally sensitive, student-centred, and domain-specific measure designed to assess disposition towards artistic thinking. In this respect, the present study differs from previous research by focusing specifically on artistic thinking disposition and by developing a multidimensional scale tailored to middle school students in the context of visual arts education.

This picture makes two critical gaps visible in the literature. First, the relative scarcity of studies focusing directly on artistic thinking disposition has limited the development of a coherent evidence base for conceptualising and measuring this construct as distinct from attitude within visual arts education. Consequently, evidence derived mainly from attitude-based measures may not be sufficient to explain continuity, persistence, and strategic flexibility in students' artistic thinking processes (Aslantaş, 2014; Dede, 2016). Second, although attitude scales based on student samples exist, some focus on primary school level, while others address middle school and high school groups together, making it difficult to disentangle developmentally differentiated patterns of disposition (Aslantaş, 2014; Dede, 2016). Therefore, the development of a scale that represents students' disposition towards artistic thinking within the performance and interpretation dimensions of visual arts education is linguistically and cognitively appropriate for the middle school level, and is conceptually differentiated from attitude and art literacy measures, emerging as both a methodological and theoretical necessity. This need is especially pronounced in the middle school period (Grades 5–8), a developmental stage in which peer evaluation becomes more salient, self-perceptions may become more fragile, and performance anxiety may intensify, potentially leading students to withdraw from creative tasks involving uncertainty and risk. Accordingly, a developmentally sensitive, student-centred instrument specifically designed for middle school students is required to make artistic thinking disposition visible and to address a significant gap in the literature (Yücetoker, 2014).

Given the context-sensitive nature of artistic thinking disposition, the development of a valid and reliable measurement tool requires a clearly delimited conceptual framework and a construct-sensitive measurement approach. In line with this, the present study aims to develop a valid and reliable multidimensional measurement instrument for assessing middle school students' artistic thinking disposition in the context of visual arts education and to examine its psychometric properties comprehensively.

Research Questions

Accordingly, the study sought answers to the following research questions:

1. Is sufficient evidence obtained regarding the content validity of the developed Artistic Thinking Disposition Scale (ATDS)?
2. Do the items in the scale load onto meaningful and theoretically coherent factors as a result of exploratory factor analysis (EFA)?
3. Does the factor structure obtained via EFA yield acceptable fit indices as a result of confirmatory factor analysis (CFA), and is the structure confirmed?
4. Are the scale's sub-dimensions and total score sufficiently reliable in terms of internal consistency and composite reliability indicators?
5. Does the Artistic Thinking Disposition Scale (ATDS) demonstrate measurement invariance across gender (configural, metric, scalar)?

Method

Research Design

This study is scale development research designed to identify the artistic dispositions of middle school students (Grades 5, 6, 7, and 8). The scale development process was planned within the framework of a multi-stage design in which evidence for validity and reliability was generated simultaneously and holistically. The factor structure of the scale was determined through exploratory factor analysis (EFA), and the resulting structure was intended to be tested and confirmed via confirmatory factor analysis (CFA). The study was structured in line with foundational sources grounded in theoretical and methodological principles related to the scale development process; accordingly, a systematic measurement instrument development approach integrating theoretical foundations with empirical evidence was adopted (Creswell, 2018; DeVellis & Thorpe, 2021; Gay & Mills, 2018).

Study Group

In this study, four separate study groups were formed to gather evidence for validity and reliability across different stages of the scale development process. Each group was determined in alignment with the aims of the study by taking into account the type of data required and the analytical needs of the relevant stage; accordingly, multiple sampling strategies were utilised. In scale development studies, selecting a sampling approach that is fit for purpose is a critical factor that directly influences not only the validity and reliability of the findings but also the generalisability of the developed measurement instrument to different groups (Cohen et al., 2018). In this context, criterion sampling and convenience sampling methods were used together in the present study. The sampling approaches used to identify participants and the relationship between these approaches and the stages of the scale development process are presented in Figure 1.

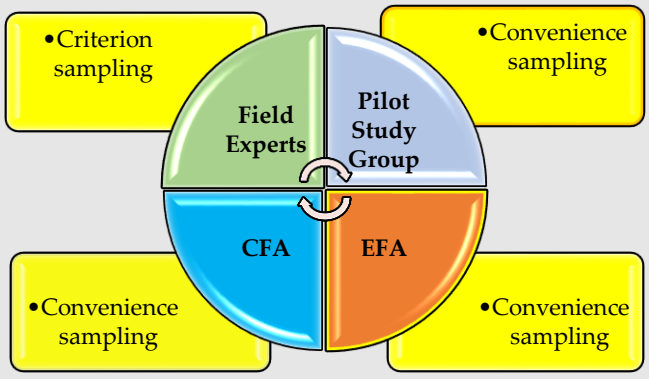

Figure 1. Sampling methods

Figure 1 presents the sampling approaches used in the study.

1. Field Experts (Content Validity Group): The first study group comprised field experts whose views were sought to gather evidence regarding the content validity of the developed Artistic Thinking Disposition Scale. At this stage, criterion sampling, which requires a specified level of expertise, was employed to identify participants. In criterion sampling, the researcher defines in advance the key characteristics required for inclusion in the sample; subsequently, individuals/cases meeting these characteristics are selected, and the rationale for why the criteria are critical is provided (Merriam & Tisdell, 2016). In this study, the criteria were defined as

holding a doctoral degree in visual arts education and having experience with the scale development process. In addition, methodological support was obtained from an associate professor specialising in educational measurement and evaluation to assess the scale's alignment with the principles of measurement instrument development. In this scope, feedback was obtained from a total of nine experts, comprising three professors, four associate professors, and two Dr Lecturers. The fact that the experts were employed at different universities introduced institutional diversity into the content validity evaluation and contributed to conducting the evaluations from a broader perspective.

2. Pilot Group: The second study group comprised middle school students who participated in the pilot administration of the scale. The pilot implementation was conducted with a total of 50 students enrolled in Grades 5, 6, 7, and 8. Of these students, 27 were female, and 23 were male. In determining the pilot study participants, convenience sampling, which allows data collection in a short time in line with the researcher's access opportunities, was preferred. Convenience sampling is frequently used in scale development processes because it enables pilot administrations to be conducted rapidly, at low cost, and in a manner suitable for testing implementation conditions (Gay & Mills, 2018). The primary aim at this stage was to evaluate the clarity of the items and the functionality of the instructions, to identify potential wording problems, and to obtain preliminary evidence to inform the necessary revisions prior to the main administration.

3. Exploratory Factor Analysis (EFA) Group: The third study group comprised a total of 499 students enrolled in middle schools in the city centre of Diyarbakır who took part in the EFA stage. This group included 240 female and 259 male students. When the distribution across grade levels was examined, it was determined that 103 students were in Grade 5, 85 students were in Grade 6, 142 students were in Grade 7, and 169 students were in Grade 8. Although criteria regarding sample size for EFA vary, it is stated that samples close to 500 can be considered at a "very good/excellent" level (Tabachnick & Fidell, 2019). Based on this rationale, it was judged that a sufficient sample size was achieved for the EFA. The EFA sample was determined using convenience sampling. In addition, particular attention was paid to ensuring that the EFA sample did not overlap with the pilot administration group in order to produce stronger validity and reliability evidence; thus, the consistency and trustworthiness of the findings were supported by using independent data sets in the analyses (Tabachnick & Fidell, 2019).

4. Confirmatory Factor Analysis (CFA) Group: The fourth study group comprised participants determined for CFA in order to test the factor structure obtained from the EFA on an independent sample. CFA is a technique that allows the structure revealed through exploratory analyses to be confirmed in a different sample and enables testing of the extent to which the proposed measurement model aligns with the theoretical construct (Blunch, 2013; Hair et al., 2022). Although different criteria exist in the literature regarding recommended sample size for CFA, it is emphasised that at least 200 participants may be sufficient in most cases, and that 250 and above may be preferred for more reliable estimates (Hu & Bentler, 1999; Tabachnick & Fidell, 2019). Accordingly, in this study, a sample consisting of a total of 389 students was formed for CFA, and it was judged that this number exceeded the recommended minimum sample sizes. This group included 201 female and 188 male students. The CFA sample was also determined using convenience sampling. However, care was taken to ensure that the CFA sample did not overlap with the pilot and EFA samples so that the model could be tested on an independent data set; it was accepted that this approach increases the strength of evidence

regarding construct validity and makes reliability evaluations more robust (Hair et al., 2022; Tabachnick & Fidell, 2019).

Data Collection Process

In this study, the data collection process was conducted incrementally and sequentially in a manner consistent with the standard stages adopted in scale development studies. In the first stage, a literature review lasting approximately six months was carried out, and a comprehensive item pool was generated based on this review. Following the literature review, a draft form comprising 28 items structured under four dimensions was prepared. After the draft form was created, expert opinions were obtained between 20 and 25 December 2025 in order to evaluate the items in terms of their scope and representational adequacy. After the necessary revisions were made in line with expert feedback, a pilot administration was conducted on 29–30 December 2025 to test the applicability of the scale and the clarity of the items. As no problems were encountered in terms of the administration flow and item wording in the pilot implementation, the study proceeded to data collection for the Exploratory Factor Analysis (EFA). EFA data were collected between 02 and 15 January 2026; based on the factor structure obtained through the EFA, Confirmatory Factor Analysis (CFA) was planned in order to test the scale structure on an independent sample. CFA data were collected between 17 and 27 February 2026.

Prior to data collection, ethical approval was obtained from the Social and Human Sciences Ethics Committee of Süleyman Demirel University (Decision Date: 11.12.2025; Decision No: E.1170302). Across all stages of the data collection process, participation was voluntary; participants were informed about the purpose and scope of the study, and the principle of confidentiality was carefully maintained. Data collection was conducted only after the necessary permissions had been obtained from the Provincial Directorate of National Education, relevant school administrations, and parents. Accordingly, the entire process was carried out in accordance with ethical principles and institutional authorisation procedures.

Scale Development Process

The development process of the Artistic Thinking Disposition Scale was planned and implemented step by step in line with widely accepted theoretical and psychometric guides in the field of scale development, based on the principle of defining the target construct clearly at a conceptual level and generating items capable of measuring this construct validly and reliably (Boateng et al., 2018; Clark & Watson, 1995; Worthington & Whittaker, 2006). Accordingly, the process comprised: (i) generating an item pool grounded in the theoretical framework, (ii) producing evidence regarding content validity through expert evaluations in order to strengthen the items' content representativeness, (iii) conducting content validity studies and a pilot administration to test the comprehensibility of items for the target age group and the implementation conditions, (iv) identifying the dimensional structure of the scale through Exploratory Factor Analysis (EFA) and testing the resulting structure on an independent sample via Confirmatory Factor Analysis (CFA), and (v) reporting reliability evidence through indicators such as internal consistency and composite reliability (Brown, 2015; Fabrigar et al., 1999; Kline, 2016; Raykov, 1997).

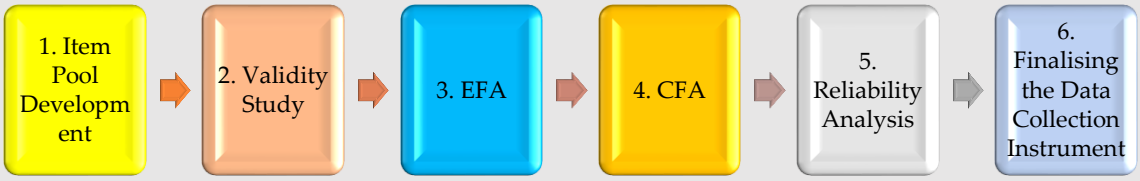

Figure 2. Scale development process

The scale development process presented in Figure 2 is explained in detail below.

1. Generating the Item Pool: During the item pool generation stage, artistic thinking disposition was conceptualised as a multidimensional construct that emerges in middle school students' engagement with the visual arts and jointly encompasses tendencies to (i) perceive and evaluate aesthetic stimuli in the environment, (ii) construct and transform mental images, (iii) remain open to trial-and-error and risk-taking in creative production, and (iv) express emotions and thoughts through visual means (Arnheim, 1969; Csikszentmihalyi, 1996; Dewey, 1934; Eisner, 2002; Leder & Nadal, 2014; Leder et al., 2004; Runco & Jaeger, 2012). In establishing this conceptual framework, the principles recommended in the scale development literature—defining the construct with high domain representativeness, keeping the indicator domain broad, and conducting item writing systematically—were followed; accordingly, the initial item pool was intentionally kept broader than the final form to allow for subsequent expert review and factor-analytic item elimination (Bandalos, 2018; Boateng et al., 2018; DeVellis & Thorpe, 2021). Items were therefore written to reflect situations that students could encounter directly in their daily lives and classroom experiences and were formulated as single-judgement, age-appropriate, and clearly worded statements grounded in observable experiences and indicators (Boateng et al., 2018; DeVellis & Thorpe, 2021). In item writing, prior measurement approaches relevant to the target domains, including visual aesthetic sensitivity tests and visual imagery measures, were also reviewed to support coverage of the indicator domain rather than direct scale adaptation (Götz et al., 1979; Marks, 1973; Zhang et al., 2023). For content representativeness, models conceptualising aesthetic perception and aesthetic judgement as staged cognitive-affective processes, together with empirical evidence on the development of visual aesthetic sensitivity during the school years, were used to justify item-level indicators such as noticing aesthetic details, attending to visual qualities, and becoming aware of one's aesthetic response (Leder & Nadal, 2014; Leder et al., 2004; Zhang et al., 2023). Likewise, in the domain of visual imagery, statements targeting indicators such as mental visualisation and visual transformation were developed based on theoretical accounts of image generation, manipulation, and reconstruction, alongside recent findings showing meaningful individual differences in imagery ability (Finke et al., 1992; Kosslyn, 1994; Wright et al., 2024). With respect to risk-taking and experimentation in creative production, integrative definitions of creativity emphasising novelty and appropriateness, together with recent school-based evidence on the relationship between creative performance and risk-taking behaviour, informed items targeting trying new ways, sustaining production despite the possibility of error, and seeking unusual solutions (Amabile, 1996; Graciano et al., 2023; Runco & Jaeger, 2012; Sternberg & Lubart, 1995; Woo & Lee, 2024). Finally, the domain of expression through art was grounded both in foundational educational approaches that position visual arts as a medium of meaning-making and self-expression (Dewey, 1934; Eisner, 2002) and in recent review evidence indicating that emotional expression and relief are frequently reported mechanisms in arts-based practices (Bosgraaf et al., 2020; Cheng, 2025). Accordingly, the item pool included statements designed to

capture students' tendency to convey and share their emotions and thoughts through visual means.

2. Validity Study: In this study, validity studies were conducted using a multidimensional approach to gather evidence regarding the validity of the developed Artistic Thinking Disposition Scale. First, content validity examinations were carried out to determine the extent to which the scale represents the content domain of the intended construct. Through expert reviews, items were examined in terms of their relationship with the construct, representativeness, and appropriateness of wording; revisions were made to items deemed necessary. Subsequently, evidence based on response processes was gathered to examine how students in the target age group interpreted the item statements and whether they experienced any difficulties during the response process. In addition, convergent and discriminant validity evidence was also evaluated in order to more robustly test whether the intended construct was represented as expected within the measurement model. For convergent validity, the primary criterion was that items assumed to represent the same construct generate a sufficient level of shared variance under the relevant factor; for discriminant validity, the primary criterion was that factors representing different constructs are distinguishable from one another (Brown, 2015; Hair et al., 2022).

3. Pilot Administration and Item Analysis: The pilot administration was conducted to test the administration process of the scale and the comprehensibility of the items for the target group. Findings obtained from the pilot study provided critical indications regarding whether the items were interpreted by participants as intended. Based on student feedback, the level of wording in some items was revised, potential ambiguities were removed, statements containing uncertainty were simplified, and the age-appropriateness of items was strengthened. The primary purpose of the revisions made at this stage was to increase the linguistic clarity and answerability of the instrument prior to the main administration and thereby reduce the risk of measurement error (DeVellis & Thorpe, 2021).

4. Exploratory Factor Analysis (EFA): Prior to conducting EFA to reveal the dimensional structure of the scale, the suitability of the data for factor analysis was evaluated using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity (Tabachnick & Fidell, 2019). After these prerequisites were met, factor loadings indicating the relationship between items and factors, eigenvalues used to evaluate the explanatory power of the structure, and the proportions of total variance explained were examined together during the EFA process. The primary decision criterion at this stage was that items cluster in a manner consistent with the theoretically expected structure and that each item represents the relevant construct at an adequate level (DeVellis & Thorpe, 2021; Tabachnick & Fidell, 2019).

5. Confirmatory Factor Analysis (CFA): CFA was conducted to test the structure obtained from the EFA on an independent sample. The primary aim in CFA was to test the fit of the proposed measurement model to the data and to determine whether the factor structure was confirmed in a manner consistent with theoretical expectations. In this context, model fit was interpreted by considering commonly used fit indices such as CFI, TLI, RMSEA, and SRMR (Brown, 2015; DeVellis & Thorpe, 2021; Hair et al., 2022). Considering fit indices jointly enabled a more balanced evaluation of model fit by reducing the limitations of interpretations based on a single criterion.

6. Reliability Analysis: To gather evidence regarding the reliability of the scale, indicators based on internal consistency and composite reliability were evaluated together. Accordingly, Cronbach's α and McDonald's ω coefficients were calculated; additionally, composite reliability

(CR) values were reported in a manner consistent with the measurement model approach. Corrected item–total correlations were also examined to evaluate item discrimination and consistency with the factor (DeVellis & Thorpe, 2021; Garson, 2013). This approach aims to provide more robust evidence regarding the internal consistency of the scale by reducing the limitations of evaluations based on a single reliability coefficient.

7. Finalising the Data Collection Instrument: Based on findings from the validity and reliability analyses, some items were removed from the scale, and revisions were made to items deemed necessary in order to enhance conceptual clarity and linguistic transparency. The primary principle here was to ensure that the scale both represents the intended construct more strongly and remains clear and applicable for students (DeVellis & Thorpe, 2021; Garson, 2022). After these revisions were completed, the Artistic Thinking Disposition Scale was finalised in a form with psychometric properties suitable for measuring middle school students' artistic dispositions.

Data Analysis

Data obtained from the study were analysed using IBM SPSS 30 and Mplus 8.11. A staged quantitative analysis strategy was adopted to evaluate the psychometric properties of the scale. First, evidence regarding content validity was obtained through expert review, and evidence based on response processes was gathered through student feedback to examine the comprehensibility and clarity of the items for the target age group. To examine the internal structure of the scale, exploratory factor analysis (EFA) was conducted first, followed by confirmatory factor analysis (CFA). Prior to EFA, factorability was assessed using the KMO coefficient and Bartlett's test. The emerging factor structure was evaluated based on factor loadings, eigenvalues, and explained variance (Tabachnick & Fidell, 2019). In the CFA stage, model fit was assessed using CFI, TLI, RMSEA, and SRMR, and these indices were interpreted jointly (Brown, 2015; Hair et al., 2022). In addition, convergent and discriminant validity were examined using model-based criteria (Hair et al., 2022). To obtain reliability evidence, Cronbach's α, McDonald's ω, composite reliability (CR), average variance extracted (AVE), and corrected item–total correlations were calculated (DeVellis & Thorpe, 2021; Garson, 2013). Finally, inter-factor relationships were examined using correlation coefficients.

Findings

Evidence on Content Validity

Expert evaluations showed that the items were generally consistent with the theoretical scope and intended dimensions of the scale. Most items were judged to be appropriate in terms of content relevance, dimension alignment, linguistic clarity, and suitability for the target age group. Qualitative feedback nevertheless indicated the need for revision in a limited number of items, particularly to improve wording clarity, reduce abstraction, strengthen the single-judgement structure, and minimise potential overlap between similarly worded items. Item-level CVR values ranged from .667 to 1.000. Twenty-seven items yielded a CVR of 1.000, indicating complete agreement among experts regarding their appropriateness, whereas only one item produced a lower CVR (.667). In line with the qualitative feedback, this item was revised to improve its concreteness and developmental appropriateness. The overall CVI was calculated as .988, and the mean proportion-based CVI was .994, both indicating a high level of content representativeness. Taken together, these findings suggest that the scale items adequately represented the construct of artistic thinking disposition. At the same time, expert feedback

contributed to refining a small number of items in order to improve linguistic precision, age-level appropriateness, and conceptual distinctiveness.

Response Process Evidence

Student feedback indicated that the draft items were generally understandable for the target age group. Across the 28 items, item-level comprehensibility indices ranged from .85 to 1.00, while the overall comprehensibility index was .92. These findings suggest that the wording of the items was largely clear and interpretable for middle school students. The qualitative feedback obtained from students nevertheless pointed to a limited number of items requiring revision. In particular, some students reported difficulty with expressions that were relatively abstract, potentially ambiguous, or not sufficiently aligned with their language level. In response to this feedback, several items were revised to improve clarity, simplify wording, and enhance developmental appropriateness. Taken together, the findings based on response processes indicate that the draft form was generally comprehensible to students and that the response process functioned as intended for most items. At the same time, student feedback provided an important basis for refining a small number of items in order to strengthen linguistic clarity and age-level appropriateness.

Findings Related to Exploratory Factor Analysis

Before proceeding to the exploratory factor analysis (EFA), preliminary analyses were conducted to determine whether the data set met the assumptions for factor analysis, including the evaluation of missing data and the examination of outliers. Such preliminary procedures are a fundamental requirement for the accuracy, stability, and interpretability of the factor solution to be obtained (Tabachnick & Fidell, 2019; Whittaker & Schumacker, 2022). In the first stage, the level of missingness in the data set was examined by considering the possibility that missing data arising from participants not responding to some items could bias the analysis results (Enders, 2022). Tabachnick and Fidell (2019) emphasise that, under certain conditions, a missing data rate exceeding 5% may meaningfully affect parameter estimates and standard errors. In this study, it was determined that item-level missing data rates remained below 5% (0.8%); therefore, analyses proceeded on the available data set without using any imputation method.

In the second stage, outliers that could distort analysis results and reduce representativeness of the sample distribution were evaluated using both univariate and multivariate approaches (Tabachnick & Fidell, 2019). Standardised Z scores were used to identify univariate outliers, and observations exceeding the ±3 threshold were considered potential outliers (Whittaker & Schumacker, 2022). To identify multivariate outliers, Mahalanobis distance was used; the distance of each observation from the multivariate mean was computed based on the χ^2 distribution, and observations with a significant distance at $p < .001$ were evaluated as outliers (Pallant, 2016; Tabachnick & Fidell, 2019). As a result of these analyses, no univariate or multivariate outliers were detected in the data set.

Following these preliminary procedures, the suitability of the data set for factor analysis was also tested prior to conducting EFA to determine the factor structure of the Artistic Thinking Disposition Scale. In this context, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were examined. A KMO value above .60 and a statistically significant Bartlett test indicate that the inter-variable correlation matrix contains a sufficient level of association for factor analysis (DeVellis & Thorpe, 2021; Field, 2024; Hair et

al., 2022). Accordingly, the KMO and Bartlett's test findings regarding the suitability of the sample for factor analysis are presented in Table 1.

Table 1. KMO and Bartlett's Test of Sphericity Results

Kaiser–Meyer–Olkin Measure of Sampling Adequacy	.914
Bartlett's Test of Sphericity Approx. Chi-Square	2828,051
df	171
Sig.	<.001

In this study, the KMO value was calculated as .914; considering that KMO values of .90 and above indicate an "excellent" level of sampling adequacy, the sample was judged to be highly appropriate for factor analysis (Field, 2024). This finding indicates that there is a pattern conducive to explaining shared variance among variables through a factor structure. To test whether the data set contained a sufficient level of inter-variable association required for factor analysis, Bartlett's test of sphericity was conducted, yielding approx. $\chi^2 = 2828.051$, df = 171, $p < .001$. This result indicates that the correlation matrix is not an identity matrix; therefore, there are significant relationships among items as required for factor analysis (DeVellis & Thorpe, 2021; Garson, 2021). When these findings from the KMO and Bartlett tests are considered together, it was concluded that the data set was suitable for factor analysis and that exploratory factor analysis could be applied to gather evidence regarding the scale's construct validity.

Examination of the correlation matrix showed that inter-item correlation coefficients ranged from .113 to .524. This distribution indicates that there are generally low-to-moderate relationships among items and that each item contributes to different aspects of the construct intended to be measured (Brown, 2015). The fact that correlation coefficients did not fall below .10 suggests that the items are related to a common construct, whereas the fact that coefficients did not exceed .80 indicates that there is no multicollinearity problem at a level that would distort factor analytic results (Akbulut, 2011; Tabachnick & Fidell, 2019). This evaluation was also supported by the determinant value of the correlation matrix being .003; a determinant greater than zero indicates that there is no singularity problem among variables at a level that would preclude factor analysis (Tabachnick & Fidell, 2019). Examination of the anti-image correlation matrix indicated that the diagonal measure of sampling adequacy (MSA) values ranged from .619 to .947. This finding suggests that each item has high sampling adequacy and that inclusion of the items in factor analysis is appropriate; in the literature, diagonal MSA values above .50 are considered sufficient for factorability (DeVellis & Thorpe, 2021). When communalities were examined, the explained variances of items ranged from .380 to .591. This result indicates that the items are represented at a moderate level by the factor structure and that the shared variance is generally at an acceptable level (Garson, 2022). Following these preliminary findings, an exploratory factor analysis was conducted to reveal the factor structure of the Artistic Thinking Disposition Scale. In the analysis, Principal Axis Factoring (PAF) was used as the extraction method, and Promax oblique rotation was applied in line with the assumption that factors may be correlated. The results indicated a three-factor structure with eigenvalues greater than 1; the eigenvalues of the factors, the proportions of variance explained, and their contributions to total variance are presented in Table 2.

Table 2. Variance Results of the Scale

Component	Initial eigenvalues (Total)	% of Variance	Cumulative %	Extraction sums of squared loadings (Total)	% of Variance	Cumulative %	Rotation sums of squared loadings (Total)
1	6,263	32,964	32,964	5,660	29,788	29,788	5,180
2	1,546	8,135	41,100	0,922	4,851	34,639	4,084
3	1,313	6,911	48,011	0,683	3,593	38,232	3,481
4	0,915	4,816	52,826	—	—	—	—
5	0,842	4,432	57,259	—	—	—	—
6	0,826	4,346	61,605	—	—	—	—
7	0,804	4,234	65,839	—	—	—	—
8	0,742	3,908	69,747	—	—	—	—
9	0,668	3,518	73,265	—	—	—	—
10	0,637	3,353	76,618	—	—	—	—
11	0,606	3,190	79,808	—	—	—	—
12	0,569	2,993	82,801	—	—	—	—
13	0,561	2,954	85,755	—	—	—	—
14	0,540	2,843	88,598	—	—	—	—
15	0,508	2,674	91,272	—	—	—	—
16	0,467	2,455	93,727	—	—	—	—
17	0,450	2,368	96,095	—	—	—	—
18	0,391	2,058	98,153	—	—	—	—
19	0,351	1,847	100,000	—	—	—	—

Findings presented in Table 2 demonstrate the three-factor structure of the Artistic Thinking Disposition Scale and provide evidence regarding the construct validity of this structure. As a result of the exploratory factor analysis, three factors with eigenvalues greater than 1 were identified. Following extraction, these three factors together explained 38.232% of the total variance. In multidimensional psychological and educational constructs, explained variance values do not need to be exceptionally high to support construct validity, particularly when the construct is theoretically complex and multifaceted. In this respect, the variance explained by the present factor structure may be considered to provide acceptable evidence for the construct validity of the scale (DeVellis & Thorpe, 2021; Kline, 2015).

According to Table 2, the first factor has an eigenvalue of 6.263 and explains 32.964% of the total variance at the initial stage; after extraction, this factor explains 29.788% of the variance. Because Items 11–19, which load on this factor, represent production- and expression-oriented dispositions—such as openness to trial-and-error in artistic production, developing original products, expressing oneself through visual means, and generating new solutions during the production process—this factor was labelled “Artistic Production and Visual Expression Disposition.” The second factor has an eigenvalue of 1.546 and explains 8.135% of the total variance initially; after extraction, it explains 4.851% of the variance. Items 6–10, which load on this factor, emphasise cognitive processes such as mental visualisation, transforming images, and establishing visual similarity; therefore, it was labelled “Imagery/Visual Thinking Disposition.” The third factor has an eigenvalue of 1.313 and explains 6.911% of the total variance initially; after extraction, it explains 3.593% of the variance. Items 1–5, which cluster under this factor, reflect students’ aesthetic attention to environmental and artistic stimuli, noticing visual qualities

such as form, colour, and texture, and the tendency to make sense of the affective and interpretive dimensions of an artwork; therefore, it was labelled “Aesthetic Sensitivity and Artistic Interpretation Disposition.” Taken together, these findings indicate that the scale measures artistic thinking disposition through three related yet distinguishable dimensions and that the factor structure is organised in a manner consistent with the intended theoretical framework.

To further support the number of factors in the Artistic Thinking Disposition Scale, the scree plot depicting the distribution of eigenvalues across factors was also examined within the EFA. The scree plot findings supported a three-factor solution for the scale. The plot showed a sharp decline from the first to the second factor, followed by a further decline from the second to the third factor at a more moderate rate; after the third factor, the curve levelled off noticeably. This pattern is consistent with Cattell’s (1966) elbow criterion, indicating that the factors explaining meaningful variance are concentrated in the first few dimensions, whereas the factors after the third contribute progressively less to the overall structure (Brown, 2015; Cattell, 1966). Indeed, when considered together with the eigenvalue table, the fact that only the first three factors had eigenvalues greater than 1 (Factor 1 = 6.263, Factor 2 = 1.546, Factor 3 = 1.313) confirms that the flattening in the scree plot occurred around the third factor and suggests that a three-factor structure constitutes an appropriate solution based on both statistical criteria and visual evidence (DeVellis & Thorpe, 2021; Tabachnick & Fidell, 2019).

This visual evidence supports the retention of a three-factor solution. Nevertheless, the scree plot does not provide information about the distribution of items across factors; rather, it is used to help determine the number of factors to be retained. The interpretation of item clustering was therefore based on the factor loading matrix obtained from the EFA. According to this pattern, Items 1–5 loaded on the third factor (Aesthetic Sensitivity and Artistic Interpretation Disposition), Items 6–10 loaded on the second factor (Imagery/Visual Thinking Disposition), and Items 11–19 loaded on the first factor (Artistic Production and Visual Expression Disposition). Taken together, these findings indicate that the scale has a theoretically meaningful and interpretable three-factor structure, and that the factor-retention decision was supported by both statistical and substantive considerations (Brown, 2015; DeVellis & Thorpe, 2021).

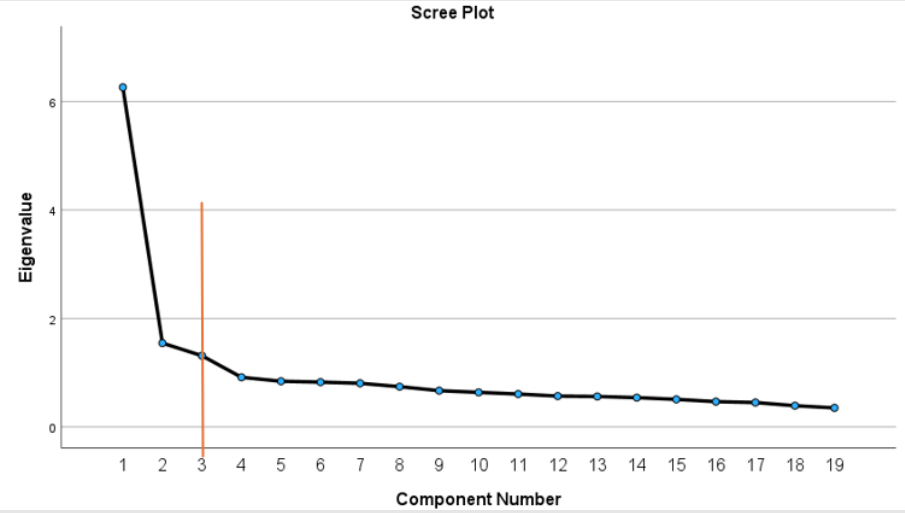

Figure 3. Line graph of eigenvalues for the scale items

The exploratory factor analysis conducted using the Promax rotation technique converged in six iterations, suggesting that the factor solution was not obtained immediately but stabilised after several optimisation steps. Such a pattern is common when some items display relatively low loadings, conceptual proximity to more than one factor, or minor cross-loading tendencies under oblique rotation. Despite this initial complexity, the final rotated solution demonstrated that the Artistic Thinking Disposition Scale was meaningfully organised under a three-factor structure. These results provided supportive evidence for the adequacy of the emerging three-factor structure. In the analysis process, the factor structure was evaluated by considering factor loadings, eigenvalues, proportions of total variance explained, corrected item–total score correlations, and Cronbach's alpha reliability coefficients together. Accordingly, the factor loadings, eigenvalues, percentages of variance explained, item–total correlations, and Cronbach's alpha coefficients for the scale are presented in detail in Table 3.

Table 3. Factor Loadings, Eigenvalues, Percentages of Variance Explained, and Item–Total Correlations

Item No	New item no	Factor 1	Factor 2	Factor 3	Communality	Item–Total Score Correlation	Cronbach Alfa
sde2	M1			,652	,324	,33	,89
sde3	M2			,612	,374	,43	,89
sde4	M3			,476	,297	,43	,89
sde6	M4			,461	,229	,36	,89
sde7	M5			,457	,308	,47	,89
sde8	M6		,733		,483	,50	,88
sde9	M7		,551		,382	,52	,88
sde10	M8		,373		,336	,53	,88
sde11	M9		,637		,389	,47	,88
sde12	M10		,622		,361	,45	,88
sde17	M11	,574			,399	,57	,88
sde18	M12	,529			,372	,54	,88
sde22	M13	,710			,485	,59	,88
sde23	M14	,702			,418	,54	,88
sde24	M15	,702			,437	,55	,88
sde25	M16	,572			,402	,58	,88
sde26	M17	,679			,456	,59	,88
sde27	M18	,663			,423	,56	,88
sde28	M19	,575			,388	,56	,88

When the findings presented in Table 3 are examined, it is evident that the items of the Artistic Thinking Disposition Scale generally cluster under three factors and that most items establish meaningful relationships with their corresponding factors. In this context, the factor loadings of the items under Factor 3 range from .457 to .652 (Items 1–5), the factor loadings of the items under Factor 2 range from .373 to .733 (Items 6–10), and the factor loadings of the items under Factor 1 range from .529 to .710 (Items 11–19). These findings indicate that the items display acceptable to strong associations with their respective dimensions and that the three-factor structure demonstrates a largely distinguishable pattern (DeVellis & Thorpe, 2021; Field, 2024). At the same time, it should be noted that Item 8 (sde10) showed a relatively low loading on Factor 2 (.373) and a weak secondary loading on Factor 1 (.194), suggesting limited overlap across these two dimensions for this item.

The item communalities (h^2) range from .229 to .485, indicating that the items explain varying but generally meaningful proportions of the variance related to the targeted construct. Although some items have relatively modest communality values, the overall pattern suggests that the items represent the measured attribute at an acceptable level within a multidimensional structure (DeVellis & Thorpe, 2021). Evidence regarding the discriminative power of the items is also supported by corrected item–total score correlations ranging from .33 to .59. More specifically, most of these correlations are .40 or above, indicating that the majority of items are consistently associated with the total score and that the scale generally possesses adequate capacity to reflect inter-individual differences (Field, 2024).

When the reliability findings are considered together, the Cronbach's alpha coefficients reported in Table 3 vary between .88 and .89, indicating high internal consistency across the factors (Kline, 2015a; Tabachnick & Fidell, 2019). Taken together, these indicators suggest that the scale's three-factor structure demonstrates a psychometrically adequate and generally robust profile, and that the items capture participants' artistic thinking dispositions in a reasonably consistent and structure-compatible manner. Within this framework, the factor loadings, communality values, and reliability coefficients obtained at this stage provided supportive evidence for the adequacy of the emerging three-factor structure.

Identification and Labelling of the Factors

The exploratory factor analysis results indicated that the Artistic Thinking Disposition Scale is organised under a three-factor structure. In the factor-labelling process, (i) the theoretical foundations of the scale, (ii) the semantic scope of the items, and (iii) the common themes shared by items clustering within each factor were considered jointly. This approach aims to ensure that factor labels not only reflect the statistical pattern but also explain, in an interpretable manner, the dimensions through which middle school students (Grades 5–8) express their perceptions and dispositions related to artistic thinking. Accordingly, each factor is presented below on the basis of the content coherence of its items.

Factor 1. Artistic Production and Visual Expression Disposition (Items 11–19): The first factor comprises Items 11–19 and represents students' willingness to produce in participation in visual arts activities, their creative trial behaviours, and their tendency to express their emotions and thoughts through visual means. The focus of this factor is that the student does not remain only in a position of perceiving/interpreting art, but also produces a visual product by using art as a language of expression. Items within this dimension, such as "I am ready to try new things in order to produce an original visual arts work" and "I can comfortably use the trial-and-error method", reflect characteristics such as risk-taking, flexibility, and openness to creativity in the production process. Items such as "I can express my emotions through visual arts works", "Sharing what I feel through visual arts makes me feel relieved", and "When doing a visual arts work, I can express my feelings and thoughts comfortably" indicate that visual arts constitute a space of emotional expression and self-narration for students. In addition, statements such as "I like making a visual work to explain an idea to others" and "I can express my thoughts more comfortably than explaining them verbally" emphasise the communicative aspect of visual production and reveal the place of visual language within the student's repertoire of self-expression. Accordingly, Factor 1 can be considered a strong dimension through which the student concretises artistic thinking at the behavioural level by integrating components of production, sharing, and visual narration.

Factor 2. Imagery/Visual Thinking Disposition (Items 6–10): The second factor comprises Items 6–10 and reflects imagery-based visual thinking dispositions involving processes such as generating mental images, transforming visual information, and establishing visual relationships between objects. The items in this dimension focus on the student's ability to visualise an image mentally, reconstruct it, and transform it in different ways to produce a creative mental representation. For example, items such as "When I close my eyes, I can imagine an object in different forms" and "When I look at a picture, new images form in my mind" reveal the student's capacity to generate mental imagery based on visual stimuli. Statements such as "I can change an image I have seen in my mind and imagine it again" and "I can imagine an object by transforming it into something else in my mind" emphasise the transformation and reconstruction of visual representations. The item "I like finding visual similarities between different objects" also reflects a tendency towards visual reasoning and analogy-making. In this respect, Factor 2 is a fundamental dimension that reveals, in a person-centred manner, the extent to which the student is predisposed to imagery and mental transformation processes, which constitute an important component of artistic thinking.

Factor 3. Aesthetic Sensitivity and Artistic Interpretation Disposition (Items 1–5): The third factor comprises Items 1–5 and reflects students' aesthetic awareness of objects and artworks in their environment, their tendency to attend to fine details, and their level of interpreting visual stimuli within an affective and meaning-making framework. Items in this factor involve cognitive processes such as attending to visual qualities of an object (e.g., colour, texture, and form), deriving pleasure from colour harmony in the environment, and being able to grasp the meaning/message conveyed by an artwork, while also making visible affective responses elicited by art products. For instance, statements such as "I pay attention to an object's visual characteristics such as colour, texture, and form" and "I like seeing the harmony of colours around me" represent the student's tendency to select aesthetic stimuli in everyday life and develop sensitivity to such stimuli. Similarly, the item "When I look at a visual arts product, I feel different emotions" reflects the affective dimension of encountering art, whereas the item "I can understand the thought or message it wants to convey" reflects the capacity to analyse and interpret the meaning of a visual product. In this respect, Factor 3 can be considered a dimension that holistically evaluates the process by which students do not merely "look at" artistic stimuli, but also notice them, are affected by them, and make sense of them.

When this three-factor structure is considered together, Factor 1 represents expressing emotions and thoughts through artistic production (visual expression–production), Factor 2 represents producing and transforming visual images in the mind (imagery/visual thinking), and Factor 3 represents noticing and making sense of artistic stimuli (aesthetic sensitivity–interpretation). These dimensions complement one another in a manner that encompasses the artistic thinking dispositions of middle school students (Grades 5–8) along a perceptual–cognitive–behavioural continuum; at the same time, because each factor offers a distinctive content domain, they function as distinguishable sub-structures within the scale.

Findings on the Reliability of the Scale

The findings in Table 4 include the reliability coefficients calculated for three factors and the overall scale:

Table 4. Reliability Findings (n=499)

Scale/Factor	M	SD	α	Split Half	Max	Min	Skewness	Kurtosis
1. Factor	3,64	,75	,87	.81	5,00	1,00	-,566	,335
2. Factor	3,60	,92	,75	.72	5,00	1,00	-,484	-,587
3. Factor	3,65	,89	,68	,60	5,00	1,00	-,471	-,706
General	3,63	,71	,89	,75	5,00	1,40	-,419	-,259

When the descriptive statistics and reliability indicators presented in Table 4 are considered together, it is evident that the Artistic Thinking Disposition Scale developed for middle school students in Grades 5, 6, 7, and 8 demonstrates generally adequate psychometric properties at both the subscale and total score levels. Subscale means ranging from M = 3.60 to M = 3.65, and an overall scale mean of M = 3.63, indicate that middle school students generally endorsed the scale items at a relatively high level and that their perceptions of artistic disposition tended to be positive. The standard deviations for the subscales ranged from .75 to .92, and .71 for the total scale, indicating that responses were not clustered in a single response option; thus, the scale produced sufficient variance to reflect individual differences among students.

With respect to reliability, Cronbach's alpha coefficients calculated for the subscales ranged from .68 to .87. At the same time, the total scale yielded α = .89. This finding supports that the scale—particularly at the total score level—has high internal consistency and that artistic disposition scores obtained from middle school students are generated reliably (DeVellis & Thorpe, 2021; Field, 2024; Tabachnick & Fidell, 2019). At the subscale level, Factor 1 demonstrated strong internal consistency (α = .87), Factor 2 yielded an acceptable level of reliability (α = .75), and Factor 3 produced an alpha value (α = .68) that falls within the borderline/acceptable range. Although this result may suggest that item homogeneity within the third factor is more limited than in the other factors, it can be considered an acceptable level of internal consistency in scale development studies, particularly for newly developed subscales (DeVellis & Thorpe, 2021).

Split-half reliability coefficients ranging from .60 to .81 across subscales and .75 for the total scale indicate that the scale's capacity to yield similar results across different item halves is generally adequate (Mertler & Vannatta Reinhart, 2017; Özdamar, 2016). In terms of distributional properties, skewness values ranged from −.419 to −.566, and kurtosis values ranged from −.706 to .335; because all values remained within the −1.5 to +1.5 thresholds, the score distributions can be considered approximately normal and generally suitable for parametric analyses (Tabachnick & Fidell, 2019).

Taken together, these findings indicate that the Artistic Thinking Disposition Scale developed for middle school students in Grades 5–8 yields consistent measurements at both subscale and total score levels and constitutes a usable measurement instrument for assessing students' artistic dispositions. Together with the high reliability evidence obtained at the total score level, the scale can be used with confidence in applied research.

Findings from Correlation Analyses of the Scale

The results of the correlation analysis conducted to examine the relationships between the total score and subscales of the Artistic Thinking Disposition Scale are presented in Table 5.

Table 5. Correlation matrix for the total score and subscales

Variables	**1**	**2**	**3**	**4**
1. General	1			
2. Aesthetic Sensitivity and Artistic Interpretation	,748**	1		
3. Imagery/Visual Thinking	,791**	,477**	1	
4. Artistic Production and Visual Expression	,900**	,490**	,563**	1

**p < .001*

Table 5 presents the Pearson correlation coefficients describing the relationships among the three sub-dimensions of the Artistic Thinking Disposition Scale developed for middle school students in Grades 5, 6, 7, and 8 and the overall total score (N = 499). The findings indicate that there are positive and statistically significant associations among the sub-dimensions ($p < .001$ for all). Examination of inter-subscale correlations shows that the association between Aesthetic Sensitivity and Artistic Interpretation Disposition and Imagery/Visual Thinking Disposition was $r = .477$, the association between Aesthetic Sensitivity and Artistic Interpretation Disposition and Artistic Production and Visual Expression Disposition was $r = .490$, and the association between Imagery/Visual Thinking Disposition and Artistic Production and Visual Expression Disposition was $r = .563$. This pattern suggests that the sub-dimensions are complementary components belonging to the same higher-order construct (artistic disposition); however, because the correlations are not extremely high/perfect, the dimensions appear to remain empirically distinguishable (DeVellis & Thorpe, 2021; Kline, 2015a).

When the relationships between subscale scores and the overall total score were examined, it was found that the Aesthetic Sensitivity and Artistic Interpretation Disposition factor was strongly associated with the total score ($r = .748$, $p < .001$), the Imagery/Visual Thinking Disposition factor was strongly associated with the total score ($r = .791$, $p < .001$), and the Artistic Production and Visual Expression Disposition factor was very strongly associated with the total score ($r = .900$, $p < .001$). Strong relationships between subscale scores and the total score are expected because the total score is a composite indicator derived from the subscales. Therefore, more discriminating evidence regarding the scale's structural integrity is derived from whether the inter-subscale associations are meaningful but not excessively high (Field, 2024). Taken together, these findings support that the sub-dimensions of the Artistic Thinking Disposition Scale are consistently related, operate within the same theoretical framework, and nevertheless allow each dimension to represent a distinct area of variance within the scale (DeVellis & Thorpe, 2021; Kline, 2015a).

Findings Related to Confirmatory Factor Analysis

To confirm the three-factor structure obtained from the exploratory factor analysis, a confirmatory factor analysis (CFA) was conducted on the item-level data of the Artistic Thinking Disposition Scale. Analyses were performed using Mplus 8.11, and the data set included 389 observations. The five-point Likert-type items were treated as continuous variables; MLR (robust maximum likelihood) estimation was used because it provides more robust estimation against departures from multivariate normality. Within the model, three latent variables were specified: Aesthetic Sensitivity and Artistic Interpretation Disposition (Items 1–5), Imagery/Visual

Thinking Disposition (Items 6–10), and Artistic Production and Visual Expression Disposition (Items 11–19); correlations among the latent variables were also freely estimated. Model estimation was completed without problems.

To examine distributional properties, univariate skewness and kurtosis values were evaluated. For the items, skewness values ranged from −1.363 to −0.245, and (excess) kurtosis values ranged from −0.991 to 1.064. These values fall within the ±2 (George & Mallery, 2020) and ±1.5 (Tabachnick & Fidell, 2019) ranges, indicating that univariate distributions were approximately normal. Multivariate normality was also examined. Mardia's multivariate normality statistics indicated significant departures from multivariate normality (p = 19): Mardia's skewness coefficient was $b_{1,p} = 44.560$, with the corresponding test statistic $\chi^2(1330) = 2888.953$, $p < .001$. Similarly, Mardia's kurtosis coefficient was $b_{2,p} = 477.663$, exceeding the expected value under multivariate normality $p(p + 2) = 399$; the deviation was significant ($z = 27.461$, $p < .001$). These findings indicate that multivariate normality was not satisfied. Consistent with this finding, the five-point Likert items were treated as approximately continuous, and CFA was estimated using MLR in order to obtain robust standard errors and fit statistics under departures from multivariate normality (Muthén & Muthén, 2017; Roos & Bauldry, 2021).

In the confirmatory factor analysis, several goodness-of-fit indices were reported jointly to evaluate the extent to which the proposed model fit the data, including χ^2/df, CFI, TLI, RMSEA, and SRMR. These indices are considered complementary criteria for assessing both relative and absolute model fit, and it is recommended that multiple fit indices be used when working with multidimensional structures (Brown, 2015; Byrne, 2012; Hu & Bentler, 1999; Schumacker & Lomax, 2016; Thompson, 2004). Table 6 presents the model fit indices obtained from the CFA.

Table 6. Fit Index Criteria and Confirmatory Factor Analysis Results

Fit index	Criteria for Excellent Fit	Criteria for Acceptable Fit	Obtained Value	Interpretation
χ^2/sd	$0 \leq \chi^2/sd \leq 3$	$3 < \chi^2/sd \leq 5$	287.941 / 149 = 1.93	Excellent fit
CFI	≥ .95	.90 ≤ CFI < .95	.914	Acceptable fit
TLI	≥ .95	.90 ≤ TLI < .95	.902	Acceptable fit
RMSEA	≤ .05	.05 < RMSEA ≤ .08	.049 (GA: .040–.057)	Excellent fit
SRMR	≤ .05	.05 < SRMR ≤ .10	.051	Acceptable fit

The confirmatory factor analysis findings presented in Table 6 indicate that the three-factor measurement model demonstrates an overall acceptable fit to the observed data, and—on some indices—an excellent level of fit. In evaluating model fit, reliance on the chi-square statistic alone was avoided due to its sensitivity to sample size; instead, the normed chi-square and both incremental and absolute fit indices were interpreted jointly (Brown, 2015; Byrne, 2012; Kline, 2016). The analysis yielded a normed chi-square value of $\chi^2/df = 1.93$ ($\chi^2(149) = 287.941$), which falls below 3 and thus supports the overall adequacy of model fit. Because the normed chi-square contributes to a more balanced interpretation of fit by taking model complexity into account, it is widely used as a criterion, particularly in structural modelling (Schermelleh-Engel et al., 2003; Tabachnick & Fidell, 2019).

Examination of incremental fit indices showed that CFI = .914 and TLI = .902 both exceeded the .90 threshold. This level suggests that the proposed measurement model yields a meaningful improvement over the independence model and that the factor structure exhibits an acceptable level of structural coherence with the observed data. However, because the CFI and

TLI values remain below .95, the fit should be interpreted as acceptable rather than excellent (Byrne, 2012; Hu & Bentler, 1999; Kline, 2016).

With respect to absolute fit indices, RMSEA = .049 was obtained, with a reported 90% confidence interval of [.040–.057]. An RMSEA value below .05 indicates that the model has the potential to exhibit a good fit in the population as well (Brown, 2015; Browne & Cudeck, 1993). In addition, the SRMR = .051 value is very close to the .05 cut-off, indicating that the discrepancy between the estimated and observed covariance/correlation structures remains low and that model fit is at least within acceptable bounds (Hu & Bentler, 1999). Taken together, these evaluations indicate that the CFA findings support a statistically coherent three-factor measurement model; in particular, the strong RMSEA value and the acceptable CFI/TLI values suggest that overall model fit is adequate. The path diagram of the model is presented in Figure 4.

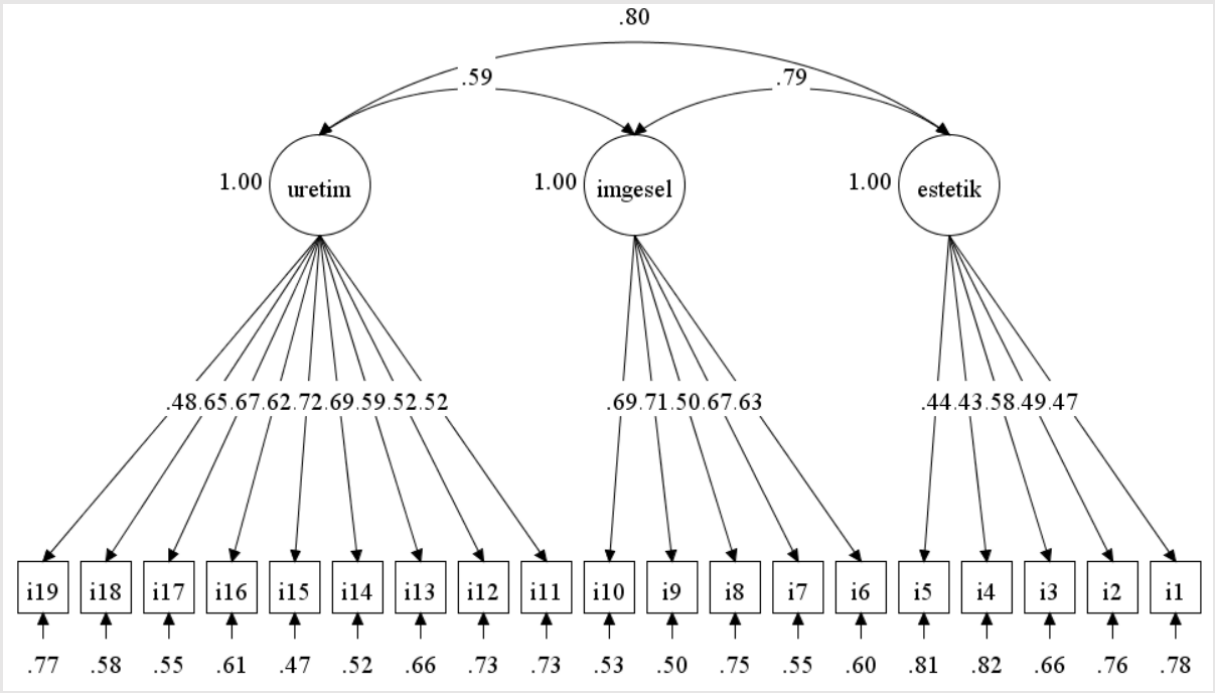

Figure 4. First-Order confirmatory factor analysis diagram of the ATDS Scale

Figure 4 presents the structural diagram for the first-order, three-factor CFA model specified for the Artistic Thinking Disposition Scale. In line with the theoretical structure, the model was defined through three latent variables: Artistic Production and Visual Expression Disposition (Factor 1), Imagery/Visual Thinking Disposition (Factor 2), and Aesthetic Sensitivity and Artistic Interpretation Disposition (Factor 3). Each item loaded only on its designated factor, and correlations among the factors were freely estimated on theoretical grounds (Brown, 2015; Kline, 2016). In the diagram, latent variances were fixed to 1.00. The inter-factor relationships were reported as Factor 1–Factor 2 r = .59, Factor 2–Factor 3 r = .79, and Factor 1–Factor 3 r = .80. The fact that all standardised loadings exceeded .40 indicates that the items represent their corresponding latent constructs at an adequate level and that the measurement model provides evidence supporting construct validity (Hair et al., 2019; Tabachnick & Fidell, 2019; Thompson, 2004). The standardised factor loadings reported in the

figure further indicate that, at the subscale level, the items generally demonstrate sufficient representational strength for their intended latent constructs. Loadings in the Aesthetic dimension ranged from .43 to .58, loadings in the Imagery/Visual Thinking dimension ranged from .50 to .71, and loadings in the Artistic Production and Visual Expression dimension ranged from .48 to .72. The clustering of most loadings above the moderate range strengthens the evidence for the measurement model's construct validity (Hair et al., 2022; Tabachnick & Fidell, 2019; Thompson, 2004).

The error (residual) variances shown in the diagram are interpreted based on the standardised solution logic and reflect the variance in each item that is not explained by its corresponding factor. Accordingly, residual variances ranged from .66 to .82 for items in the Aesthetic Sensitivity and Artistic Interpretation and Artistic Production and Visual Expression dimensions, from .50 to .75 for items in the Imagery/Visual Thinking dimension, and from .47 to .77 for items in the Artistic Production and Visual Expression dimension. This pattern indicates that, for some items, the explained variance may remain relatively limited.

Based on the CFA findings, McDonald's omega (ω) internal consistency coefficients were computed for the overall scale and for each subscale. The results indicated ω = .60 for the Aesthetic Sensitivity and Artistic Interpretation subscale, ω = .77 for the Imagery/Visual Thinking subscale, ω = .84 for the Artistic Production and Visual Expression subscale, and ω = .88 for the total scale score. These findings indicate that reliability is strong at the total score level, and that internal consistency is at a good level for the Imagery/Visual Thinking and Artistic Production and Visual Expression subscales. That reliability is relatively more limited for the Aesthetic Sensitivity and Artistic Interpretation subscale (Hair et al., 2022).

In this study, to evaluate convergent validity for three subscales of the Artistic Thinking Disposition Scale, average variance extracted (AVE) and composite reliability (CR) values were computed based on the standardised factor loadings obtained from the first-order CFA. As a result of these calculations, the following values were obtained: for Aesthetic Sensitivity and Artistic Interpretation, AVE = .24, CR = .60; for Imagery/Visual Thinking, AVE = .42, CR = .78; and for Artistic Production and Visual Expression, AVE = .37, CR = .84. While CR values of .60 and above indicate that composite reliability is adequate, the fact that AVE values fell below .50—particularly for the Aesthetic Sensitivity and Artistic Interpretation and Artistic Production and Visual Expression dimensions—suggests that the average variance explained by the indicators of the relevant factors may be relatively limited (Fornell & Larcker, 1981; Hair et al., 2022; Kline, 2016). When AVE and CR are considered jointly, it can be inferred that convergent validity is stronger for the Imagery/Visual Thinking dimension. In contrast, a more cautious interpretation is warranted for the Aesthetic Sensitivity and Artistic Interpretation dimension. At the same time, AVE values below .50 for both dimensions suggest that the average variance explained by the items through the relevant factor may be relatively limited (Hair et al., 2019; Kline, 2015b). However, Fornell and Larcker (1981) and Hair et al. (2019) note that convergent validity may still be considered acceptable when CR is .60 and above, even if AVE falls below .50. Within this framework, considering the AVE and CR findings together, convergent validity for both subscales can be considered supported.

Composite reliability (CR) is regarded as a more sensitive, model-based indicator of internal consistency than Cronbach's alpha, because CR provides a more realistic reliability estimate at the construct level by accounting for each item's factor loading and error variance (Hair et al., 2019; Raykov & Marcoulides, 2011). In the literature, CR values of .60 and above

are interpreted as indicating adequate composite reliability, while values of .70 and above indicate acceptable-to-good internal consistency (Fornell & Larcker, 1981; Kline, 2015b). In this study, based on calculations from the standardised factor loadings obtained in the CFA, CR values of .60 for Factor 1, .78 for the Imagery/Visual Thinking factor, and .84 for the Artistic Production and Visual Expression factor were obtained. These findings indicate that the subscales demonstrate adequate composite reliability and that the measurement model yields consistent parameters at the construct level.

The inter-factor correlations reported in Figure 4 (r = .59–.80), particularly the higher associations between Artistic Production and Visual Expression and Aesthetic Sensitivity and Artistic Interpretation (r = .80) and between Imagery/Visual Thinking and Aesthetic Sensitivity and Artistic Interpretation (r = .79), indicate that the sub-dimensions generate a meaningful amount of shared variance around a common higher-order construct. Although relationships at this level warrant monitoring of discriminant validity, they may also suggest a hierarchical organisation within the holistic nature of the "artistic disposition" construct (Brown, 2015; Kline, 2016). In the literature, it is emphasised that, for discriminant validity, inter-factor correlations below .85 indicate adequate statistical distinctiveness between constructs (Hair et al., 2019; Kline, 2015b). In this study, the correlations among the three factors remained below this threshold, supporting discriminant validity overall.

Findings on Measurement Invariance (Gender)

In this study, measurement invariance was tested sequentially at the configural, metric, and scalar levels to determine whether the Artistic Thinking Disposition Scale measures the same latent structure equivalently across female and male groups. Analyses were conducted using MLR estimation, which provides robust standard errors and robust fit statistics under departures from multivariate normality (Brown, 2015; Byrne, 2012; Kline, 2016). In comparing models, given the sensitivity of the chi-square difference test to sample size, the decision process was also supported based on changes in incremental fit indices (ΔCFI, ΔRMSEA, ΔSRMR) (Chen, 2007; Hu & Bentler, 1999). The findings are presented in Table 8.

Table 8. Multi-Group CFA results for the scale

Model	**χ^2**	**sd**	**RMSEA**	**%90 GA**	**CFI**	**TLI**	**SRMR**	**$\Delta\chi^2$**	**Δsd**	**p**	**ΔCFI**	**ΔRMSEA**	**ΔSRMR**
Configural	495.919	298	.058	.049–.067	.883	.865	.064	–	–	–	–	–	–
Metric	510.880	314	.057	.048–.066	.883	.873	.070	13.733	16	.6186	+.000	−.001	+.006
Scalar	535.091	330	.057	.048–.065	.878	.874	.073	23.914	16	.0914	−.005	+.000	+.003

Measurement invariance across gender was tested through configural, metric, and scalar models. The results supported configural, metric, and scalar invariance for the three-factor structure across female (n = 201) and male (n = 188) groups. At the configural invariance stage, the factor structure was specified identically in both groups and parameters were freely estimated without equality constraints across groups. The fit indices for the configural model were $\chi^2(298) = 495.919$, RMSEA = .058 (90% CI = .049–.067), CFI = .883, TLI = .865, and SRMR = .064.

This pattern indicates that the three-factor structure operates with the same configuration across the two gender groups; in particular, RMSEA and SRMR suggest approximate fit, whereas the CFI/TLI values indicate more limited incremental fit (Brown, 2015; Hu & Bentler, 1999).

At the metric invariance stage, factor loadings were constrained to be equal across the female and male groups. The fit indices for the metric model were $\chi^2(314) = 510.880$, RMSEA = .057 (90% CI = .048–.066), CFI = .883, TLI = .873, and SRMR = .070. The scaled chi-square difference test comparing the metric and configural models was non-significant ($\Delta\chi^2(16) = 13.733$, $p = .6186$). In addition, incremental fit changes were ΔCFI = +.000, ΔRMSEA = −.001, and ΔSRMR = +.006. These findings indicate that the strength of the relationships between items and their factors (i.e., factor loadings) is equivalent across gender groups, supporting metric invariance (Chen, 2007).

At the scalar invariance stage, item intercepts were constrained to be equal across groups in addition to factor loadings. The fit indices for the scalar model were $\chi^2(330) = 535.091$, RMSEA = .057 (90% CI = .048–.065), CFI = .878, TLI = .874, and SRMR = .073. The scaled chi-square difference test comparing the scalar and metric models was non-significant ($\Delta\chi^2(16) = 23.914$, $p = .0914$). Incremental fit changes also remained small (ΔCFI = −.005, ΔRMSEA = +.000, ΔSRMR = +.003). This pattern indicates that, in practical terms (ΔCFI/ΔRMSEA/ΔSRMR), the constraints did not meaningfully degrade model fit, and scalar invariance can be considered acceptable (Chen, 2007). Accordingly, the methodological basis for latent mean comparisons across gender groups is supported (Brown, 2015; Kline, 2016).

Scale Administration and Scoring

The Artistic Thinking Disposition Scale developed in this study is a measurement instrument designed to determine the artistic thinking and orientations towards artistic activities of middle school students in Grades 5, 6, 7, and 8. The scale is formatted as a 5-point Likert-type instrument, and students are asked to respond to each statement in a manner that reflects their own feelings and thoughts. Response options are: 1 = Strongly disagree, 2 = Disagree, 3 = Partly agree, 4 = Agree, and 5 = Strongly agree. The instructions clearly explain the purpose of the administration and how the items should be answered, and students are explicitly encouraged to read each item carefully and mark the option most appropriate for them. All items are positively worded; therefore, scoring is direct, and there are no reverse-coded items. The scale consists of 19 items, each scored from 1 to 5. Accordingly, the minimum possible total score is 19, and the maximum possible total score is 95. Higher total scores indicate higher levels of artistic disposition. Scores derived from the scale allow both an overall evaluation through the total score and a more detailed examination through subscale scores. According to the exploratory factor analysis findings, the scale comprises three sub-dimensions: Aesthetic Sensitivity and Artistic Interpretation Disposition (Items 1–5), Imagery/Visual Thinking Disposition (Items 6–10), and Artistic Production and Visual Expression Disposition (Items 11–19) (DeVellis & Thorpe, 2021; Field, 2024).

Scale Completion Time

Completion time is a functional indicator for evaluating the practicality of a measurement instrument and, particularly for middle school students, the process characteristics related to attention span, reading comprehension, and response accuracy (DeVellis & Thorpe, 2021). Accordingly, the 19-item form of the Artistic Thinking Disposition Scale was administered to

389 middle school students; during administration, each student's completion time was recorded, and descriptive indicators of completion time were calculated (Johnson & Morgan, 2016).

Table 9. Completion time of the scale

Completion Time	Min.	Max.	Mean
Artistic Thinking Disposition Scale	5	11	8

As shown in Table 7, students completed the scale in a minimum of 5 minutes and a maximum of 12 minutes, with an average completion time of 10 minutes. The observed variation in completion time is considered to be potentially related to individual characteristics such as students' reading speed, reading comprehension level, ways of interpreting the items, and sustaining attention. An average completion time of 10 minutes indicates that the scale can be completed within a practical, manageable timeframe that is suitable for field administrations with middle school students. Nevertheless, taking individual differences among students into account, it is recommended that participants be given sufficient time during administration and that careful reading of the items be emphasised, particularly at the instruction stage. This finding supports that the scale can be used confidently within a planable time window in both classroom-based implementations and online environments.

Discussion

The Artistic Thinking Disposition Scale (ATDS) developed within the scope of this study is a domain-specific measurement instrument designed to evaluate, within a multidimensional framework, the dispositions that emerge in middle school (Grades 5–8) students' engagement with visual arts. Within the study, *artistic disposition* was conceptualised as an integrated construct encompassing: (i) aesthetic sensitivity and artistic interpretation disposition, (ii) imagery/visual thinking disposition, and (iii) artistic production and visual expression disposition. This approach enabled the joint consideration of the perceptual–cognitive, affective, and behavioural aspects of artistic experience (Arnheim, 1969; Csikszentmihalyi, 1996; Dewey, 1934; Eisner, 2002; Leder & Nadal, 2014; Leder et al., 2004; Runco & Jaeger, 2012). The adoption of item-writing principles such as "single judgement, developmental appropriateness, clarity, and translation into observable experiential language" indicates that the deliberately broad item pool provided a functional basis for expert review and factor-analytic elimination (Bandalos, 2018; Boateng et al., 2018; DeVellis & Thorpe, 2021).

Evidence for content-related support was strengthened through both quantitative and qualitative findings based on expert judgement. The CVR values calculated using Lawshe's approach ranged from .667 to 1.000, and the CVI value based on the mean of the item-level indicators was .988, indicating that the items represented the theoretical content domain at a high level (Ayre & Scally, 2014; Lawshe, 1975; Polit & Beck, 2006). In addition, expert feedback suggested the need to reduce abstraction in some items, strengthen the single-judgement principle, and minimise potential overlap among conceptually similar statements. These revisions indicate that content-related evidence should be considered not only in terms of quantitative agreement among experts, but also as a process of improving conceptual precision and linguistic clarity through qualitative refinement (DeVellis & Thorpe, 2021). Evidence based on response processes was also supportive. Student feedback collected from 50 middle school students indicated that the items were generally clear and comprehensible for the target age group, with item-level comprehensibility indices ranging from .85 to 1.00 and an overall comprehensibility index of .92. Taken together, these findings suggest that the items were largely understandable and linguistically appropriate for the intended respondents, while also providing

a basis for further refinement of a limited number of items (DeVellis & Thorpe, 2021; Garson, 2022; Johnson & Morgan, 2016).

The confirmation of the EFA-derived pattern via CFA further strengthened the evidence regarding the internal structure of the ATDS. In the first-order CFA with three correlated factors, model fit was obtained as $\chi^2(149) = 287.941$, $\chi^2/df = 1.93$, CFI = .914, TLI = .902, RMSEA = .049 (90% CI = .040–.057), and SRMR = .051. These values indicate acceptable model fit, with particularly favourable evidence from RMSEA and the normed chi-square, and acceptable fit according to CFI, TLI, and SRMR. Consistent with the findings indicating departures from multivariate normality, the five-point Likert items were treated as approximately continuous, and CFA was estimated using MLR in order to obtain robust standard errors and fit statistics under non-normal conditions. However, because inter-factor correlations suggested the possibility of a shared higher-order structure, a second-order CFA was also tested. In that model, fit indices were weaker, $\chi^2(150) = 345.414$, CFI = .880, TLI = .863, RMSEA = .058 (90% CI = .050–.066), and SRMR = .114, indicating that the first-order correlated three-factor model provided a better representation of the data than the second-order solution. In this respect, the findings suggest that the subscale scores of the ATDS have stronger interpretive value than a single higher-order total score, particularly in practice-oriented uses.

Reliability findings also supported the psychometric adequacy of the scale, although the coefficients varied across subscales. Cronbach's α coefficients were .87 for Factor 1, .75 for Factor 2, and .68 for Factor 3. At the same time, the total scale yielded α = .89. These findings indicate high internal consistency for the total scale, strong reliability for Factor 1, acceptable reliability for Factor 2, and borderline but still usable reliability for Factor 3. Split-half coefficients ranged from .60 to .81 across the subscales and were .75 for the total scale, suggesting generally adequate consistency across item halves. In addition, CFA-based omega coefficients were .60 for Aesthetic Sensitivity and Artistic Interpretation, .77 for Imagery/Visual Thinking, .84 for Artistic Production and Visual Expression, and .88 for the total scale. Composite reliability values ranging from .60 to .84 further supported the adequacy of the subscales. Although AVE values were relatively low, especially for the Aesthetic Sensitivity and Artistic Interpretation dimension (AVE = .24), this pattern suggests the need for further item refinement rather than undermining the overall usability of the instrument.

The suitability of the scale for gender-based comparisons was also supported by measurement invariance findings. In multi-group CFA, the configural model yielded $\chi^2(298) = 495.919$, RMSEA = .058, CFI = .883, TLI = .865, and SRMR = .064; the metric model yielded $\chi^2(314) = 510.880$, RMSEA = .057, CFI = .883, TLI = .873, and SRMR = .070; and the scalar model yielded $\chi^2(330) = 535.091$, RMSEA = .057, CFI = .878, TLI = .874, and SRMR = .073. The chi-square difference tests were non-significant at both the metric stage, $\Delta\chi^2(16) = 13.733$, p = .6186, and the scalar stage, $\Delta\chi^2(16) = 23.914$, p = .0914. In addition, incremental fit changes remained within acceptable limits (ΔCFI = .000/.005, ΔRMSEA = -.001/+ .000, ΔSRMR = +.006/+ .003). Taken together, these findings indicate that scalar invariance across gender was supported at an acceptable level, providing a defensible basis for latent mean comparisons between female and male students.

Contributions to the Field

This study makes several contributions to the field of visual arts education and scale development. First, it introduces a developmentally sensitive and domain-specific measurement instrument focused directly on artistic thinking disposition, thereby addressing a conceptual and methodological gap in a literature that has more frequently concentrated on attitude, general orientation, or broader constructs such as art literacy. In this respect, the ATDS contributes to the literature by operationalising artistic thinking disposition as a multidimensional construct rather than reducing it to skill, performance, or affective preference alone. This feature gives the scale particular theoretical value.

Second, the study contributes methodologically by providing a psychometrically tested instrument based on multiple forms of evidence, including content-related evidence, response-process considerations, internal structure analyses, reliability indicators, and gender-based measurement invariance testing. The inclusion of scalar invariance evidence is especially important, as it supports the use of the ATDS in comparative studies and strengthens the scale's potential for broader empirical use. In this respect, the study offers not only a new instrument but also a robust example of scale development practice in the field of visual arts education.

Third, the scale contributes to educational practice by enabling researchers and practitioners to evaluate students' artistic thinking dispositions across distinct yet interrelated domains. This creates a useful basis for examining the effects of instructional interventions such as aesthetic literacy activities, imagery-based design tasks, and production-oriented visual arts workshops. Thus, the ATDS has the potential to support both research and classroom-based evaluation by making visible student tendencies that are often difficult to assess through performance products alone.

Limitations and Suggestions for Future Research

Despite these contributions, the study has several limitations that should be taken into account. First, the scale development process did not include criterion-related validity evidence based on correlations with an established external instrument. Although the internal structure and reliability findings were supportive, the absence of concurrent or predictive validity evidence limits the extent to which the ATDS can currently be evaluated in relation to conceptually adjacent constructs. Future studies should therefore examine the relationships of the ATDS with external measures such as creativity, visual arts attitude, or related cognitive-affective constructs.

Second, the scale was developed and validated using Turkish-speaking middle school students within a specific educational and cultural context. For this reason, the generalisability of the findings to different linguistic, cultural, and educational settings remains limited. Future research should test the scale in different samples, regions, and cultural contexts, and cross-cultural adaptation studies should be conducted before broader international use is assumed. In connection with this point, the provision of an English version of the items in the appendix would facilitate future adaptation studies.

Third, although the overall psychometric results were supportive, some findings suggest the need for further refinement. In particular, relatively low AVE values, especially in the Aesthetic Sensitivity and Artistic Interpretation dimension, indicate that some items may still contain substantial measurement error and that convergent validity could be strengthened through future item revision. Similarly, because the second-order model showed a weaker fit than the correlated first-order model, future research may further explore whether the construct is

best represented through correlated subdimensions rather than a total-score structure. Finally, because the scale is based on self-report data, future studies may benefit from combining it with performance-based, observational, or portfolio-based indicators in order to obtain a richer picture of students' artistic thinking dispositions.

Conclusion

In conclusion, the ATDS appears to be a valid and reliable measurement instrument for assessing middle school students' disposition towards artistic thinking in the context of visual arts education. The findings support a three-factor correlated structure and indicate that the scale is particularly informative when interpreted at the subscale level. By offering a developmentally appropriate, student-centred, and domain-specific framework, the ATDS contributes to the literature on visual arts education and provides a practical basis for research, instructional evaluation, and future scale adaptation studies. At the same time, further evidence from criterion-related studies, cross-cultural applications, and item refinement research will be important for strengthening the scale's broader applicability and interpretive power.

Disclosure statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

Conflicts of Interest

The authors declare no conflict of interest.

Ethics

This study was approved by the Ethics Committee of Süleyman Demirel University Beşeri ve Sosyal Bilimler. The ethics approval was granted on 11.12.2025 under the protocol number E.1170303. All procedures involving human participants were conducted in accordance with institutional and national ethical standards, the Helsinki Declaration, and applicable regulations. Informed consent was obtained from all participants.

CRediT authorship contribution statement

S. Aslan: Conceptualization, Methodology, Formal Analysis, Data Collection, Writing – Original Draft

M. A. Gökdemir: Visualization, Writing – Review & Editing, Data Collection

Declaration of Generative AI and AI-assisted Technologies Usage

The authors used ChatGPT (OpenAI) and Grammarly solely for language refinement, grammar correction, and supporting an appropriate academic tone during the revision process. Additionally, AI tools were used to assist in writing and refining the syntax for the analyses performed in Mplus 8.11. No AI tool was used to generate research ideas, theoretical content, data, analyses, or interpretations. All substantive intellectual contributions were made by the authors.

References

Akbulut, Y. (2011). *Sosyal bilimlerde SPSS uygulamaları*. İdeal.

Amabile, T. M. (1996). *Creativity in context*. Westview Press.

Arnheim, R. (1969). *Visual thinking*. University of California Press.

Aslantaş, S. (2014). Görsel sanatlar dersine yönelik tutum ölçeği geliştirilmesi. *International Journal of Turkish Education Sciences, 2014*(2), 185–196.

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counselling and Development, 47*(1), 79–86. https://doi.org/10.1177/0748175613513808

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Berardi, A. M. (2025). Varieties of imagery and perception: The structure of task differences. *Frontiers in Psychology, 16*, 1568148. https://doi.org/10.3389/fpsyg.2025.1568148

Blunch, N. J. (2013). *Introduction to structural equation modelling using IBM SPSS Statistics* (2nd ed.). SAGE Publications.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioural research: A primer. *Frontiers in Public Health, 6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Bosgraaf, L., Spreen, M., Pattiselanno, K., & van Hooren, S. (2020). Art therapy for psychosocial problems in children and adolescents: A systematic narrative review on art therapeutic means and forms of expression, therapist behaviour, supposed mechanisms of change, and effects. *Frontiers in Psychology, 11*, 584685. https://doi.org/10.3389/fpsyg.2020.584685

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Burnett Heyes, S., Pictet, A., Mitchell, H., Raeder, S. M., Lau, J. Y. F., Holmes, E. A., & Blackwell, S. E. (2017). Mental imagery-based training to modify mood and cognitive bias in adolescents: Effects of valence and perspective. *Cognitive Therapy and Research, 41*(1), 73–88. https://doi.org/10.1007/s10608-016-9795-8

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Chen, F. F. (2007). Sensitivity of goodness-of-fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheng, Y. (2025). From brushstrokes to brighter minds: A systematic review of the influence of out-of-school visual art activities on adolescents. *Psychology Research and Behavior Management, 18*, 2401–2416. https://doi.org/10.2147/PRBM.S566460

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education*. Routledge.

Creswell, J. W. (2018). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. HarperCollins.

Dede, H. (2016). Öğrencilerin sanata karşı bakış açılarını ortaya koymaya yönelik bir tutum ölçeği. *İdil, 5*(25), 1559–1576. https://doi.org/10.7816/idil-05-25-13

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications.

Dewey, J. (1934). *Art as experience*. Minton, Balch & Company.

Eisner, E. W. (2002). *The arts and the creation of mind*. Yale University Press.

Enders, C. K. (2022). *Applied missing data analysis*. Guilford Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th ed.). SAGE Publications Ltd.

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and applications*. MIT Press.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Garson, G. D. (2013). *Validity and reliability*. Statistical Associates Publishing.

Garson, G. D. (2021). *Data analytics for the social sciences: Applications in R*. Routledge.

Garson, G. D. (2022). *Factor analysis and dimension reduction in R: A social scientist's toolkit*. Routledge.

Gay, L. R., & Mills, G. E. (2018). *Educational research: Competencies for analysis and applications*. Pearson.

George, D., & Mallery, P. (2020). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge.

Götz, K. O., Borisy, A. R., Lynn, R., & Eysenck, H. J. (1979). A new visual aesthetic sensitivity test: I. Construction and psychometric properties. *Perceptual and Motor Skills, 49*(3), 795–802. https://doi.org/10.2466/pms.1979.49.3.795

Graciano, P., Lermen, F. H., Reichert, F. M., & Padula, A. D. (2023). The impact of risk-taking and creativity stimuli in education towards innovation: A systematic review and research agenda. *Thinking Skills and Creativity, 47*, 101220. https://doi.org/10.1016/j.tsc.2022.101220

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). *Multivariate data analysis*. Pearson.

Hetland, L., Winner, E., Veenema, S., & Sheridan, K. M. (2013). *Studio thinking 2: The real benefits of visual arts education*. Teachers College Press.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Irwing, F. P., Booth, T., & Hughes, D. J. (2018). *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale, and test development*. Wiley.

Johnson, R. L., & Morgan, G. B. (2016). *Survey scales: A guide to development, analysis, and reporting*. Guilford Press.

Kline, P. (2015). *A handbook of test construction: Introduction to psychometric design*. Routledge.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. MIT Press.

Lane, S., Raymond, M. R., & Haladyna, T. M. (2015). *Handbook of test development*. Routledge.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode—Developments and challenges in empirical aesthetics. *British Journal of Psychology, 105*(4), 443–464. https://doi.org/10.1111/bjop.12084

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology, 95*(4), 489–508. https://doi.org/10.1348/0007126042369811

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Mertler, C. A., & Vannatta Reinhart, R. (2017). *Advanced and multivariate statistical methods: Practical application and interpretation*. Routledge.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Özdamar, K. (2016). *Ölçek ve test geliştirme yapısal eşitlik modellemesi*. Nisan Kitapevi.

Pallant, J. (2016). *SPSS survival manual*. McGraw-Hill Education.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly, 39*(1), 1–21.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173–184. https://doi.org/10.1177/01466216970212006

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.

Roos, J. M., & Bauldry, S. (2021). *Confirmatory factor analysis* (Quantitative Applications in the Social Sciences, Vol. 189). SAGE Publications, Inc.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*(2), 23–74.

Schneider, V., & Rohmann, A. (2021). Arts in education: A systematic review of competency outcomes in quasi-experimental and experimental studies. *Frontiers in Psychology, 12*, 623935. https://doi.org/10.3389/fpsyg.2021.623935

Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to structural equation modeling*. Routledge.

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

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics*. Pearson.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.

Whittaker, T. A., & Schumacker, R. E. (2022). *A beginner's guide to structural equation modelling*. Routledge.

Winner, E., Goldstein, T. R., & Vincent-Lancrin, S. (2013). *Art for art's sake? The impact of arts education*. OECD Publishing. https://doi.org/10.1787/9789264180789-en

Woo, J., & Lee, S.-Y. (2024). The moderating effect of creative metacognition on adolescents' risk-taking in creative performance. *Asia Pacific Education Review, 25*(1), 99–115. https://doi.org/10.1007/s12564-024-09927-8

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Wright, D. J., Scott, M. W., Kraeutner, S. N., Barhoun, P., Bertollo, M., Campbell, M. J., Waltzing, B. M., Dahm, S. F., Esselaar, M., Frank, C., Hardwick, R. M., Fuelscher, I., Marshall, B., Hodges, N. J., Hyde, C., & Holmes, P. S. (2024). An international estimate of the prevalence of differing visual imagery abilities. *Frontiers in Psychology, 15*, 1454107. https://doi.org/10.3389/fpsyg.2024.1454107

Yücetoker, İ. (2014). Sanat okuryazarlığı ölçeğinin hazırlanması ve geliştirilmesi. *Sanat Eğitimi Dergisi, 2*(1), 112–126. https://doi.org/10.7816/sed-02-01-06

Zhang, J., Du, X., Zhang, X., & Bai, X. (2023). The development of visual aesthetic sensitivity in students in China. *Frontiers in Psychology, 14*, 1071487. https://doi.org/10.3389/fpsyg.2023.1071487

Ek: Ölçek maddeleri (Türkçe versiyonu)

Not: Yazarlardan izin alınmadan ölçek bilimsel çalışmalarda kaynak gösterilmek kaydıyla kullanılabilir.

1-5. Maddeler Estetik Duyarlık ve Sanatsal Yorumlama Eğilimi, 6-10. Maddeler İmgesel / Görsel Düşünme Eğilimi, 11-19. maddeler Sanatsal Üretim ve Görsel İfade Eğilimi

Madde No	SANATSAL DÜŞÜNME EĞİLİMİ ÖLÇEĞİ	Hiç katılmıyorum	Katılmıyorum	Kısmen katılıyorum	Katılıyorum	Tamamen katılıyorum
1.	Bir nesnenin renk, doku ve biçim gibi görsel özelliklerine dikkat ederim.	1	2	3	4	5
2.	Bir görsel sanat ürününe (resim, heykel, görsel vb.) baktığımda değişik duygular hissederim.	1	2	3	4	5
3.	Bir görsel sanat ürününe (resim, heykel, görsel vb.) baktığımda, anlatmak istediği düşünceyi veya mesajı anlayabilirim.	1	2	3	4	5
4.	Çevremdeki renklerin birbirine uyumunu görmek hoşuma gider.	1	2	3	4	5
5.	Bir görsel sanat eserindeki güzel ayrıntılar ilgimi çeker.	1	2	3	4	5
6.	Bir nesneyi gözlerimi kapatınca farklı biçimlerde hayal edebilirim.	1	2	3	4	5
7.	Bir resme baktığımda zihnimde yeni görüntüler oluşur.	1	2	3	4	5
8.	Farklı nesneler arasında görsel benzerlikler bulmayı severim.	1	2	3	4	5
9.	Gördüğüm bir görüntüyü zihnimde değiştirerek yeniden hayal edebilirim.	1	2	3	4	5
10.	Bir nesneyi zihnimde başka bir şeye dönüştürerek hayal edebilirim.	1	2	3	4	5
11.	Özgün bir görsel sanat çalışması ortaya koymak için yeni şeyler denemeye hazırım.	1	2	3	4	5
12.	Görsel sanat çalışmalarında deneme-yanılma yöntemini rahatça kullanırım.	1	2	3	4	5
13.	Duygularımı görsel sanat çalışmalarıyla (resim, üç boyutlu çalışma vb.) anlatabilirim.	1	2	3	4	5
14.	Bir fikrimi başkalarına anlatmak için görsel bir çalışma (resim, üç boyutlu çalışma vb.) yapmayı severim.	1	2	3	4	5
15.	Kendi düşüncelerimi ifade etmek için görsel bir çalışma (resim, üç boyutlu çalışma vb.) yapmayı severim.	1	2	3	4	5
16.	Hissettiklerimi görsel sanatlar yoluyla paylaşmak beni rahatlatır.	1	2	3	4	5
17.	Bir olayı görsel sanat çalışmalarıyla (resim, üç boyutlu çalışma vb.) anlatmak benim için kolaydır.	1	2	3	4	5
18.	Görsel sanat çalışması yaparken duygu ve düşüncelerimi rahatça ifade edebilirim.	1	2	3	4	5
19.	Görsel sanat çalışmaları yaparken düşüncelerimi sözle anlatmaktan daha rahat ifade ederim.	1	2	3	4	5

Appendix: Scale Items (English Version)

Item No.	Artistic Thinking Disposition Scale
1	I pay attention to the visual qualities of an object, such as its colour, texture, and form.
2	When I look at a work of visual art (e.g. a painting, sculpture, or image), I experience different emotions.
3	When I look at a work of visual art (e.g. a painting, sculpture, or image), I can understand the idea or message it is intended to convey.
4	I enjoy noticing the harmony between colours in my surroundings.
5	Beautiful details in a work of visual art capture my interest.
6	When I close my eyes, I can imagine an object in different forms.
7	When I look at a picture, new images form in my mind.
8	I enjoy finding visual similarities between different objects.
9	I can mentally change and reimagine an image I have seen.
10	I can imagine an object by transforming it into something else in my mind.
11	I am willing to try new things in order to create an original visual artwork.
12	I can use trial and error comfortably in visual art activities.
13	I can express my feelings through visual art activities (e.g. painting or three-dimensional work).
14	I enjoy creating a visual work (e.g. a drawing or three-dimensional work) to communicate an idea to others.
15	I enjoy creating a visual work (e.g. a drawing or three-dimensional work) to express my own ideas.
16	Sharing what I feel through visual art makes me feel at ease.
17	I find it easy to represent an event through visual art activities (e.g. painting or three-dimensional work).
18	I can express my feelings and thoughts comfortably when creating visual art.
19	By creating visual art, I can express my thoughts more comfortably than I can with words.