International Journal of Assessment Tools in Education
2024, Vol. 11, No. 3, 589–607
https://doi.org/10.21449/ijate.1475980
journal homepage: https://dergipark.org.tr/en/pub/ijate
Research Article

The mental imagery scale for art students: Building and validating a short form

Handan Narin Kızıltan[1*], Hatice Cigdem Bulut[2]

[1]Cukurova University, Faculty of Education, Department of Art Education, Adana Türkiye
[2]Northern Alberta Institute of Technology, Education Insights, Data & Research, Edmonton, AB Canada

ARTICLE HISTORY
Received: Apr. 30, 2024
Accepted: July 21, 2024

Keywords:
Mental imagery,
Scale shortening,
Ant colony optimization,
Item selection,
Psychometrics.

Abstract: Mental imagery is a vital cognitive skill that significantly influences how reality is perceived while creating art. Its multifaceted nature reveals various dimensions of creative expression, amplifying the inherent complexities of measuring it. This study aimed to shorten the Mental Imagery Scale in Artistic Creativity (MISAC) via the Ant Colony Optimization algorithm (ACO), a metaheuristic methodology for developing psychometrically robust brief scales. Answering 63 items in the original version of MISAC demands a higher cognitive load and, consequently, more time. Therefore, our goal was to shorten it while preserving its psychometric properties. In this study, responses to the MISAC were obtained from 500 undergraduate students enrolled in an art education program. The items on the short form of the MISAC were selected based on pre-specified validity criteria and content representability. The 28-item short form of MISAC demonstrated comparable performance to the original version regarding construct validity, criteria-related validity, and reliability coefficients. Moreover, strict invariance was attained across both gender groups in the validation process of the short form. These results highlight the utility of the shortened version of the MISAC as a valid measure with minimal loss of information of scores compared to the full version.

1. INTRODUCTION

Mental imagery, considered one of the critical cognitive skills for humans (Pérez-Fabello & Campos 2007), plays a crucial role in the perception of reality during the artistic production process (Ziss, 2011). Mental imagery occurs when perceptual information is accessed from memory and can be created by combining and manipulating stored perceptual information in new ways (Kosslyn et al., 2001). Therefore, mental images include both visual representations and various types of past mental encounters (Hilton, 2007).

Following the second half of the 20th century, interest in mental imagery has accelerated in fields such as behavioral and cognitive psychology, clinical psychology, neuroscience, marketing, and sport (Kosslyn et al., 2001; Park & Yoo, 2020; Pearson et al., 2015; Saulsman et al., 2019). In psychological research, mental imagery is utilized to prevent mental disorders and develop treatment methods (Saulsman et al., 2019; Schwarz et al., 2020). In addition, it has been used to

*CONTACT: Handan NARİN KIZILTAN ✉ handannarinn@gmail.com Cukurova University, Faculty of Education, Department of Art Education, Adana, Türkiye

e-ISSN: 2148-7456

change individuals' psychological attitudes, perceptions, and perspectives (Holmes & Mathew, 2010; Park & Yoo, 2020; Pearson, 2019; Saulsman et al., 2019). Mental imagery is also a crucial cognitive domain often highlighted in art education (Duncun, 2001; Heid et al., 2009) owing to its relation to creativity (Palmiero et al., 2016). Consequently, art and cognition are intricately linked, mutually reflecting and reinforcing each other (Bhattacharya & Petsche, 2002).

Artists create art by drawing on mental images developed through observing the world (Hetland et al., 2007). The personal records, diaries, and sketchbooks of well-known artists like Leonardo Da Vinci and Picasso, which reveal their internal worlds, demonstrate that they actively utilized imagery while creating their artwork. (Rosenberg, 1987; Vellera & Gavard-Perret, 2012). This same process applies to art students in visual arts classes. Art students learn to observe and use observation to generate mental images and plan ways to create their artwork (Hetland et al., 2007). Therefore, the power of imagination is an intrinsic and essential element for art students (Bhattacharya & Petsche, 2002; Chamberlain et al., 2019).

Imagination is a product of cognitive actions that facilitate the construction of new meanings (Efland, 2002). Metaphorical thinking is one of these cognitive actions utilized to imbue meaning in creating and evaluating artwork (Hetland et al., 2007; Serig, 2006). The tools artists employ during artwork production are integral to the cognitive process used by those interpreting the artwork to construct meanings (Efland, 2002). Metaphorical thinking involves expressing different concepts through a single, similar concept that can be represented in various ways (Deaver & Shiflett, 2011). It directs minds beyond existing similarities to new similarities it creates, leading to the discovery of a new dimension of meaning for the word (Lakoff & Johnson, 2010).

Consequently, previously undiscovered creative meanings are brought forth. When engaging in metaphorical thinking or drawing, an individual participates in the form of mental imagery (Dodson, 2013). Images also serve as metaphorical conceptualizations and a creative act of reinterpreting these concepts. Creative thinking involves the cognitive properties of metaphor capable of generating new meanings by establishing connections between different elements (Efland, 2002). Images are both a metaphorical conceptualization and a creative action. Therefore, when it comes to artistic creativity, a robust relationship exists between mental imagery, metaphorical thinking, and creative thinking. For this reason, the ability to form mental images can be associated with the data obtained from the metaphorical thinking ability test, which measures the ability to produce meaning, and the drawing test, which measures the ability to create creative images.

Given the ongoing importance and long history of mental imagery within art, there have been various studies focusing on mental imagery in art (Jankowska & Karwowski, 2015; Pérez-Fabello & Campos, 2007; Pérez-Fabello et al., 2016). Furthermore, the relationship between creativity in art and mental imagery is examined in several studies (Miller, 2014; Pérez-Fabello & Campos, 2007; Pérez-Fabello et al., 2016; Vellera & Gavard-Perret, 2012). A study by Drake et al. (2021) found that artists possess superior imagery skills compared to non-artists, as assessed by a self-report measure. Another study (Vellera & Gavard-Perret, 2012) found that an increase in mental imagery score corresponded with an increase in performance in creative tasks, as measured by two different tools. In another study by Jankowska and Karwowski (2020), the results from five separate studies, each employing various measurement tools, were combined. The study found that art students exhibited a higher level of mental imagery compared to the non-artist group. These studies provide evidence that mental imagery is considered an indicator of artistic creativity by using different measures.

There are primarily three ways to assess mental imagery (Ji et al., 2019): (a) Reporting naturally occurring mental imagery, (b) Laboratory assessments of mental imagery, and (c) Scales for mental imagery. Applying scales in the fields of art and creativity can be more convenient for researchers due to the focus in these fields not typically being placed on the neurocognitive basis of mental imagery. However, the construct of mental imagery has been a challenge to measure both validly and reliably due to its multidimensional nature (Cumming & Eaves, 2018). In the literature, a variety of measures are focused on different aspects of mental imagery (e.g., Betts' Questionnaire Upon Mental Imagery, [Betts' QMI; Betts, 1909]; Vividness of Visual Imagery Questionnaire

[VVIQ; Marks, 1973]; The Plymouth Sensory Imagery Questionnaire [Andrade et al., 2014]). However, the multidimensional structure of mental imagery requires the use of more than one measurement tool or longer measures, which include several factors (Calabrese & Marucci, 2006; Jankowska & Karwowski, 2015; Vellera & Gavard-Perret, 2012).

Given the drawbacks of longer measures, such as decreasing response rate and increasing response bias (e.g., careless responding [Niessen et al., 2016], exhibiting response styles [Weijters et al., 2010]), researchers conducting similar studies prefer using shorter measures or short versions of commonly utilized and adapted scales (e.g., short versions of Betts' QMI, Sheehan (1967), and VVIQ; Marks, 1995). As a result, scale-shortening procedures have recently gained popularity in psychological and cognitive assessments due to the development of automated methods (Basarkod et al., 2018; Schroeders et al., 2016).

This study aimed to shorten the Mental Imagery Scale in Artistic Creativity (MISAC), which was recently developed in art education. We employed methodological advances in scale-shortening techniques and utilized a metaheuristic approach (e.g., Ant Colony Optimization algorithm [ACO]) to shorten the MISAC. In addition, we gathered reliability and validity evidence for the shortened version of the MISAC. Also, we compared the psychometric features of the full version of the MISAC with that of the shortened version.

1.1. The MISAC

The MISAC measures the ability of individuals to recreate/remember objects, events, and phenomena based on their physical (movement, shape, color, place) and sensory modalities (e.g., sound, texture, and taste) (Narin, 2019). In this context, the scale measures the mental imagery ability of spatial, tactile, physical, kinesthetic, emotional, characteristic features, and affective experiences. While developing the MISAC, Mark's VVIQ (Mark, 1973) and Sheehan's Betts' QMI (1967) scales were considered. Mark's VVIQ scale includes four different contents (i.e., visualizing sentences about relatives or friends, the sunrise, a shop one often goes to, and the image of a country). Sheehan's Betts' QMI (1967) includes sensory modalities: visual, auditory, tactile (cutaneous), kinesthetic, gustatory, olfactory, and organic (whole body).

Unlike the scales mentioned above measuring the vividness of mental imagery, the MISAC measures the ability of mental imagery in terms of vividness, attention, and control. The most distinctive difference between the MISAC and other scales is its use in determining the mental imagery ability of a group within programs that require artistic creativity or in creative individuals such as those enrolled in art education programs. Notably, the MISAC can be utilized as a supplementary measurement tool for art and creativity research and for the selection procedures of students entering arts education or art-related programs. It can also be utilized to follow students' progress in different disciplines that require creative skills, such as visual communication design, art and design, and architecture.

Use of the MISAC not only considers the insights of mental imagery scales from working with participants with differing characteristics and creative individuals (Kozhevnikov et al., 2013; Miller, 2014; Pérez-Fabello et al., 2016; Vellera & Gavard-Peret, 2012) but also considers the limitations of current scales and attempts to overcome their shortcomings. For example, Sheehan's (1967) QMI contains smell as one of the sensory modalities; however, Arshamian and Larsson (2014) indicated that individuals, in most cases, cannot produce mental images based on the sense of smell. In addition, Kozhevnikov et al. (2013) noted that despite the importance of the ability to visualize and discriminate colors and textures of objects for artistic creativity, these aspects are often neglected in the current measures. Thus, the MISAC incorporates various conceptualizations regarding mental imagery within its list of items and factors.

Based on the original version of the MISAC, comprising seven factors and 63 items (Narin, 2019), the exploratory factor analysis (EFA) results revealed that the scale accounted for 49.6% of the total variance, with factor loadings values ranging from .45 to .74. The confirmatory factor analysis (CFA) supported the factor structure of the MISAC, as evidenced by good fit values (RMSEA = .05, NFI = .90, NNFI = .95, CFI = .95, SRMR = .06, IFI = .95) (χ^2 (1869) = 3525.56, p < .001). In the original version of the MISAC, each factor exhibited good internal consistency, with Cronbach's alpha values ranging between .82 and .89 (Narin, 2019).

1.2. Why Shorten The MISAC?

Long measures may cause fatigue, higher drop-out rates, and a lower response rate, as well as increase an unnecessary waste of time and energy, thereby reducing the quality of the gathered data (Basarkod et al., 2018; Olaru et al., 2015; Rammstedt & Beierlein, 2014). Also, if longer measures include items demanding a higher cognitive load, as seen in MISAC, then the undesired effects may be problematic regarding data quality. Responding to items regarding mental imagery might take longer than responding to items in other settings, as one must imagine the vividness of the object being questioned within an item. When items get more cognitively demanding, the respondents may likely adapt their response style as a shortcut (Krosnick et al., 2002). Thus, the length of scales and the level of cognitive load may be obstacles to obtaining the intended data quality.

Studies using mental imagery scales aim to determine the associations with other variables (Jankowska & Karwowski, 2015; Pérez-Fabello & Campos, 2007; Pérez-Fabello et al., 2016). Therefore, respondents may be required to respond to several questionnaires to provide scholars with a wide array of information regarding their visual and mental abilities. Due to assessment time and research funding sometimes being limited in designs that include multiple constructs, keeping the response rate and costs at a reasonable level is important, so shorter scales are more preferable (Rammstedt & Beierlein, 2014; Dogan & Bulut, 2024). Therefore, developing shorter versions of some scales has steadily increased over the past few years to eliminate these consequences.

In numerous higher education institutions, including those in Türkiye (e.g., O'Donoghue, 2011; Ozmutlu & Tomak, 2021; Taskesen, 2019; Tay, 2019; Yilmaz, 2016), scales or tests assessing creativity or related constructs hold significance in the selection process for art students, often complementing the evaluation of portfolios. However, the inclusion of multiple assessments, particularly longer ones, poses challenges for both candidates and the academic jury overseeing the selection process. This extended evaluation complicates the assessment for candidates and creates difficulties for the jury in making decisions based on these assessments. In response to these challenges, Turkish institutions frequently depend on evaluating drawing skills, including drawings of live models and imaginative design studies. (e.g., Dilmac & Kucuoglu, 2010; Taskesen, 2019); however, this approach introduces its own validity concerns. Including longer assessments, especially those focused on visualization skills, adds complexity to achieving thorough, reliable, and valid evaluations. Finding instruments that balance brevity with comprehensive assessment and validity poses a significant challenge. A shorter MISAC version can be a potential solution to bridge this evaluative gap.

1.2.1. *Ant-Colony optimization*

There has been a growing interest in the methods of automated approaches to scale shortening (Leite et al., 2008; Olaru et al., 2015; Schroeders et al., 2016; Yarkoni, 2010). The traditional approaches consist of examination of item-total correlations (Bowns et al., 2022; Carr et al., 2005), conducting EFA and choosing the highest factor loadings (Botes et al., 2021; Leite et al., 2008) researchers select items based on a reduction in the scale's Cronbach's alpha reliability coefficient if each item is removed (e.g., Bowns et al., 2022; Swindle et al., 2006). Inevitably, selecting the appropriate items via traditional approaches can take some time. More importantly, sequence effects or relying solely on one criterion within the abbreviating process result in unwanted biases. Furthermore, the required input from researchers is relatively high compared to automated approaches, depending on the number of items and factors on instruments and the number of criteria researchers consider (Yarkoni, 2010). Therefore, the workload can be a significant obstacle in this process. On the other hand, traditional scale shortening methods may miss the most optimal version, as researchers only consider limited alternate forms. As a result, not only can automation significantly reduce time spent on developing short measures, but also it allows researchers to achieve optimality or near-optimality (Jankowsky et al., 2020; Olaru et al., 2019; Raborn et al., 2019; Sandy et al., 2014).

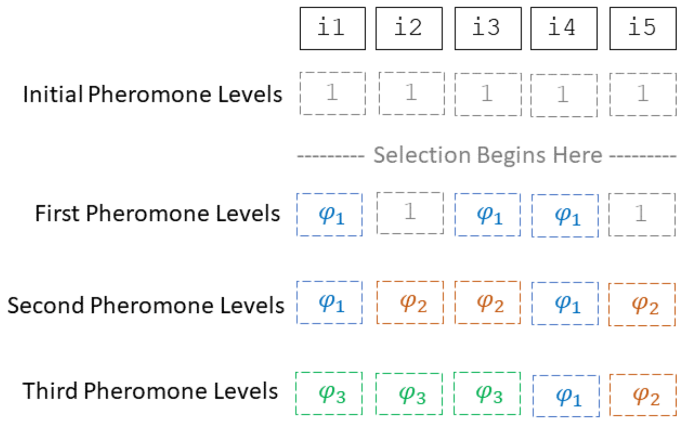
Automated approaches for scale abbreviation, such as Genetic Algorithms or Ant-Colony Optimization, make this process much faster and easier (Leite et al., 2008; Yarkoni, 2010). For instance, let us consider a situation where a researcher wants to shorten a 63-item with a seven-factor scale and use several criteria. If all 63 items of the long version are evenly distributed across the seven subscales (i.e., 9 items per subscale) and a short measure is constructed with 4 items per subscale, this would result in $\binom{9}{4}$^7= 504,189,521,813,376 possible combinations. After finding a suitable short form, the researcher may have to perform additional analysis for other criteria. However, using an automated approach, the process can be done more efficiently (Sandy et al., 2014; Yarkoni, 2010), and the researcher can use their time and expertise to evaluate the results instead of conducting multiple analyses. Research shows that automated approaches can provide better results than traditional approaches (Leite et al., 2008; Raborn et al., 2019; Sandy et al., 2014). For instance, Sandy et al. (2014) compared one rational approach and an automated approach (genetic algorithm approach) to develop a short scale. The validity and reliability properties of the scales developed separately by these approaches were similar. Similarly, Leite et al. (2008) showed that ACO excels at maximizing certain predefined qualities and outperforms methods for selecting items with traditional methods.

The current study used the Ant Colony Optimization (ACO) approach (Marcoulides & Drezner, 2003) to shorten the MISAC. The goal of using the ACO algorithm approach in this study was twofold. The first objective was to create a shortened version of MISAC to make it more practical for researchers to use in studies that do not have the capacity to use longer measures. Second, we wanted to maximize the model fit of a short form of MISAC in terms of converging on the previously validated mental imagery model. The ACO is one of the best-performing practices for producing short forms (Leite et al., 2008; Olaru et al., 2015; Raborn et al., 2019).

Interestingly, the ACO is a heuristic algorithm that incorporates the foraging behaviors of real ants to establish the shortest route to a food source in an automated model-fitting process (Marcoulides & Drezner, 2003). Deneubourg et al. (1983) found that ants produce pheromones while searching for a food source so that the ants that come after can utilize this chemical trail as feedback for determining the shortest path to the located food source. For example, ants will randomly try routes in the first step and produce pheromone chemicals during the search for routes to a food source. When a route is relatively long, its pheromone level will gradually dissipate, ultimately failing to attract other ants. Similarly, pheromone evaporation in the ACO algorithm can reduce the strength of pheromone routes over time. The evaporation rate can impact how well the ACO algorithm performs. This rate can encourage greater exploration of the solution space. However, it can also lead the algorithm to rapidly forget earlier successful solutions or prompt ants to follow existing routes more frequently, thus increasing the likelihood of the algorithm adhering to previously found shorter paths. At the end of this process, ants try to choose the shortest route over time.

In survey research, the ACO mimics those behaviors to generate short forms of scales by using the ‘pheromone’ levels of items (Olaru et al., 2015). For this, random models are generated through the ACO to determine the pheromone levels of items in the first iterations. Then, items that show the best fit in terms of specific criteria (i.e., model fit statistics) have higher probabilities of being selected in later iterations (Olaru et al., 2015). The process is complete once all the criteria are met by the number of items required for the short form. Figure 1 illustrates these steps in an example.

Figure 1. *An illustration of the item selection procedure using the ACO.*



In the initial stage, all items in this sample scale (i.e., i1, i2, i3, i4, and i5) have equal initial weights for the selection procedure. After selection begins in the ACO algorithm, a randomly short form is generated by selecting items 1, 3, and 4 and their pheromone levels (φ_1) are calculated for that short form. In the initial iteration, the ACO algorithm randomly selects items 1, 3, and 4, subsequently evaluating their suitability based on pheromone levels. These levels are calculated based on the criteria introduced to the algorithm (e.g., CFI > .95 and RMSEA < .06). These criteria can be various and are up to researchers and scale properties. Then, the pheromone levels influence and modify the weighting or significance of the selected items within the selection process (Leite et al., 2008). The algorithm integrates a pheromone evaporation mechanism, which reduces the current pheromone levels before adjusting them according to a pre-established rate determined by the researcher. This rate selection is pivotal, as it directs the algorithm's inclination towards favoring frequently selected items or encouraging greater exploration of potential item combinations in each iteration. Consequently, this step significantly contributes to fine-tuning the item selection process, emphasizing the influence or reliance on previously chosen items. In our five-item scenario, the process repeats for the second and third selections until the best items are chosen. If the third round marks the end, using the calculated pheromone levels helps identify the most suitable items based on how many the researcher aims to include in their shorter scale. As researchers can decide the criteria (e.g., model fit, number of items for each factor) and the parameters (i.e., number of ants and evaporation rate) to be introduced in the algorithm, the ACO provides flexibility and rapid solutions for the scale abbreviation process.

2. METHOD

2.1. Sample

The sample participants comprised 500 undergraduate students (29.2% males) aged 18-47 (*M*=22.3, *SD* = 3.86). The study recruited participants from five higher education institutions located in three different cities in Turkey, all of which specialize in providing education in the arts. The participants were drawn from the Fine Arts Education Department of the Education Faculty, as well as the Painting, Graphics, and Sculpture Departments of the Fine Arts Faculty. These departments were selected because they highly emphasize creativity and visual skills, which are essential for success and acceptance in the field. The number and percentage of first-year students, sophomores, juniors, and seniors were 119, 124, 125, 132 (23.8%, 24.8%, 25.0%, 26.4%), respectively. Before participating in the study, each student was given a detailed description of the research and asked to provide informed consent.

In this study, we present the results about the full scale and its properties (*N* = 420), obtained from a separate study (Narin, 2019). This sample shares resemblances with the sample characteristics employed in the current study. These undergraduates belong to the same programs, encompassing approximately 28% male students, with an approximate 27% distribution across each academic year, ranging from first-year students to senior student cohorts.

2.2. Instruments

The MISAC, consisting of 63 items, is utilized to assess the mental imagery of art education students and help to evaluate how clearly and vividly people remember various objects, situations, facts, and events, such as affective, tactile, and spatial experiences and actions experienced by the body. There are seven factors (spatial [10 items], tactile [10 items], physical [9 items], kinesthetic [9 items], emotional [8 items], characteristic feature [9 items], and affective experiences [8 items]) on the MISAC (Narin, 2019). The items on the MISAC are rated on a 7-point Likert-type scale (*1 = Very vivid and clear as in reality; 7 = no image appeared in my mind*). A high score on the MISAC indicates a high power of mental imagery. Notably, respondents require a maximum of 25 and an average of 15 minutes to answer the MISAC.

For example, an item from the spatial dimension can be given as "Imagine a café you often go to or your favorite café in your mind. How clear and vivid you can imagine these: (a) the location of tables, chairs, cash register…etc.''. Another item example regarding the physical dimension is as follows: "There are several actions/movements you experience using your body (e.g., arms, legs, and body). How clearly and vividly can you imagine when you think of yourself doing these movements? (a) Carrying a heavy load on your back".

TCIA (Test of Creative Imagery Abilities) is a test developed by Jankowska and Karwowski (2015) to measure creative imagery abilities. It was utilized by the authors of the current study after adapting it to Turkish (see Narin, 2019). The test consists of seven incomplete figures. Participants are asked to verbally produce and describe several images evoking these figures. Then, they are expected to select the most original image from those they produce, draw it, and title it. Next, the drawings are evaluated in three dimensions: vividness, originality, and convertibility. Also, the highest score that can be obtained on the TCIA test is 21. The test was utilized to establish criteria-related validity evidence in this study.

The Metaphoric Thinking Test (MTT) consists of 10 concepts and three initial sentences. The test aims to measure the participants' ability to make sense of an image, create conceptual images, and establish a similarity relationship (see Narin, 2019). The participants are expected to select only three of the ten concepts provided to them in the test, create sentences using the selected concepts in a new and different way, and complete the incomplete initial sentences in a way that creates new meaning and context. The associated concepts and sentences based on these concepts are then evaluated in the context of creative thinking with a rubric prepared by the researcher according to three levels: non-creative (0 points), partially creative (1 point), and high-level creative (2 points). The highest score that can be obtained on the MTT is 12. The test was utilized to establish criteria-related validity evidence in this study.

2.3. Procedures

First, the normality assumptions for each item were checked by using the criteria of ±2 for skewness and ±7 for kurtosis coefficients (West et al., 1995). All the items had low percentages (<5%) for the missing values, and all met the normality assumptions. To check whether the seven-factorial structure of the MISAC fits our data, we carried out CFA using the *lavaan* package (Rosseel, 2012) in R (R Core Team, 2022). All analyses were conducted using a diagonally weighted least squares estimator. As an indicator of a good fit, values below .05 for the root mean square error of approximation (RMSEA) and values above .95 for the Tucker-Lewis Index (TLI) and the Comparative Fit Index (CFI) were considered (Yu, 2002).

2.3.1. *Item Selection via ACO*

After the model fit was guaranteed, we ran the ACO algorithm to shorten the MISAC for our data using the *ShortForm* package (Raborn & Leite, 2018). The ACO algorithm mimics ants' behaviors to establish the shortest route to a food source as a model for searching the model fit processes of structural equation modeling (Marcoulides & Drezner, 2003). The goal of this approach was to reach an optimal or near-optimal model with a fewer number of items. For this, an iterative process is started with the ACO by using several parameters (i.e., ants, evaporation, and steps) and criteria (i.e., model fit indices) until a specified convergence criterion is met (i.e., the number of iterations) (see Leite et al., 2008).

In this current study, we also chose the same values of the model fit statistics mentioned earlier to evaluate the quality of the shortened scales generated by the ACO. As ACO follows a heuristic approach for calculating the probabilities of items to be selected for the short form, ACO may generate different short forms in each run (Leite et al., 2008). Thus, in the item selection process, we attempted to shorten the MISAC by selecting four or five items for each factor with minor modifications to the tuning parameters (i.e., the number of ants, evaporation rate, and steps) as follows (Raborn & Leite, 2018, p. 10):

i. ants = 120,
ii. evaporation = .95 (i.e., the percentage of the pheromone retained after evaporation between iterations), and
iii. steps = 20 (i.e., a numeric value that sets the rule for stopping, which is the number of ants in a row for which the model does not change).

The algorithm's computational process took approximately one hour to run with these parameters. The ACO algorithm was rerun 24 times to select optimal item candidates encompassing each factor's context and aligned with relevant theoretical representations. After each run, we identified frequently selected items for each factor.

Furthermore, after the 15th run, content experts identified 13 items to be excluded from the short form due to their content. Consequently, we omitted these 13 items from the algorithm for the remaining runs. Then, we thoroughly reviewed the top five item sets selected by the algorithm. Subsequently, we engaged in discussions regarding item coverage with two content experts. Finally, collaborating with these experts and authors, we collectively chose the most suitable version of the short form. The codes used in this study are available in Appendix A.

2.3.2. *Gathering validity and reliability evidence*

The means and standard deviations were also calculated for each factor and item. In addition, we calculated both Cronbach's α and McDonald's (1999) ω as reliability evidence by using the *psych* package (Revelle, 2019). Notably, we considered ω and α >.70 as a threshold for moderate reliability (Brunner et al., 2012; Nunnally & Bernstein, 1994), acknowledging the contextual considerations and potential trade-offs associated with reliability standards in research. The inter-correlations of the factors from the shortened MISAC with external criteria (i.e., metaphorical thinking and creativity imagery abilities) were calculated to gather concurrent validity evidence. The purpose of this analysis was to check whether the correlations obtained between the factors of the full scale and external variables were maintained within the shortened scale.

The ACO algorithm allows the selection of invariant items among specified groups, as demonstrated in various studies (Jankowsky et al., 2020; Olaru et al., 2019; Schroeders et al., 2016). However, due to our sample's gender imbalance (29.2% males) and relatively small male group size (n = 145), our initial analysis using modified functions from Jankowsky et al. (2020) and Olaru and Jankowsky (2022) showed consistent differences between groups in almost every selection. This finding indicated the necessity for freely estimating coefficients in each selection, undermining the ACO algorithm's optimization. Consequently, we could not employ

ACO for item selection based on measurement invariance. Therefore, we checked for measurement invariance across genders by utilizing the *lavaan* package as described in Bulut (2020) after the item selection process.

Regarding this validity evidence, we aimed to show that the shortened version of the MISAC was equally robust across gender groups. A stepwise procedure that started from the least restrictive model to the more restricted model (i.e., configural, metric, scalar, and strict invariance model, respectively) was adopted (see Van de Schoot et al., 2012). Furthermore, to test the measurement invariance, differences between the model fits previously evaluated with the same criteria and values of $\Delta\chi^2$ and ΔCFI were calculated. Chen's rule was followed (i.e., the ΔCFI is <.01) (Chen, 2007) to control whether both models fit equally well statistically.

3. RESULTS

In this study, the results of the analysis conducted in the prior research by Narin (2019) were shared to prove that the shortened scale has similar psychometric features to the full scale. Therefore, the information in Tables 1, 2, and 3 regarding the full scales was obtained from Narin's study (2019). Following the item selection process, the most optimal results were achieved by selecting four items from each factor, consistently chosen by ACO algorithms. The selected items from the shortened scale are provided in Appendix B. As shown in Table 1, the results of the CFA model of the shortened scale confirmed the hypothesized 7-factor model of the full scale and demonstrated a good fit. Furthermore, the model fit statistics were determined to be very similar to the full scale of the MISAC.

Table 1. *Model fit statistics of the full and shortened scales of the MISAC.*

Scales	χ^2	*df*	*p*	CFI	TLI	RMSEA	Lower	Upper
Full	9450.99	1953	< .001	0.98	0.98	0.01	0.00	0.02
Shortened	371.92	329	< .001	0.97	0.97	0.02	0.00	0.02

Means, standard deviations (SD), reliability coefficients, and zero-order correlations of the full and shortened scales of the MISAC are provided in Table 2. The reliabilities of the shortened scale were lower than the coefficients of the full scale; however, they were higher than .70 and ranged from .72 to .80 for all factors. Thus, the values were within an acceptable range.

Table 2. *Means, SDs, reliability coefficients, and zero-order correlations of the full and shortened scales of the MISAC.*

		M	*SD*	α	ω	1	2	3	4	5	6	7
Full scale	1. Spatial	5.60	1.09	.90	.92	1						
	2. Tactile	6.12	0.79	.87	.89	.42*	1					
	3. Physical	5.92	0.91	.88	.91	.49*	.53*	1				
	4. Kinesthetic	5.01	1.10	.86	.89	.40*	.36*	.47*	1			
	5. Emotional	5.43	1.18	.85	.89	.35*	.30*	.49*	.26*	1		
	6. Characteristic	5.60	0.98	.84	.88	.38*	.54*	.41*	.42*	.35*	1	
	7. Affective	5.78	0.96	.83	.88	.38*	.44*	.57*	.43*	.39*	.42*	1
Shortened	1. Spatial	5.79	1.14	.80	.80	1						
	2. Tactile	6.07	0.93	.72	.73	.44*	1					
	3. Physical	5.98	0.94	.74	.75	.46*	.50*	1				
	4. Kinesthetic	5.23	1.20	.78	.79	.36*	.39*	.38*	1			
	5. Emotional	5.40	1.28	.75	.76	.30*	.33*	.39*	.23*	1		
	6. Characteristic	5.58	1.12	.72	.72	.39*	.50*	.37*	.36*	.32*	1	
	7. Affective	5.74	1.14	.77	.78	.38*	.44*	.46*	.40*	.36*	.36*	1

Note: Inter-dimensional scale correlations within each form. * $p < .001$

As shown in Table 2, zero-order correlations between the factors of the shortened scale were similar to those between the factors of the full scale. To gather concurrent validity evidence, the correlation coefficients were calculated between the external variables (i.e., metaphorical thinking and creativity imagery abilities) and the factors of the shortened scale and compared with the result obtained from the full scale. As presented in Table 3, the direction and magnitude of these relationships in the full scale (computed using sum scores) were generally maintained within the shortened scale.

Table 3. *Correlations between external variables and factors of the full and shortened scales of the MISAC.*

Factors	Full scale (N = 420)		Shortened scale (N = 500)	
	MTT	TCIA	MTT	TCIA
Spatial	0.11^{*}	0.01	0.10^{*}	0.05
Tactile	0.18^{***}	0.03	0.14^{**}	0.09^{*}
Physical	0.02	0.05	0.01	0.12^{**}
Kinesthetic	0.11^{*}	0.15^{**}	0.06	0.12^{**}
Emotional	0.01	0.01	0.09^{*}	0.06
Characteristic	0.13^{*}	0.08	0.12^{**}	0.17^{***}
Affective	0.11^{*}	0.09	0.12^{**}	0.16^{***}

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

We tested the measurement invariance in the shortened scale to gather additional validity evidence and to examine whether this form of the scale maintained the same factorial structure across gender groups. Thus, the results of measurement invariance tests are presented in Table 4.

Table 4. *Measurement invariance tests across gender*.

Invariance test	χ^2	df	CFI	RMSEA	$\Delta\chi^2$	ΔCFI
Configural	350.68	658	.977	.015	-	-
Metric	433.49	679	.973	.015	24.212	.003
Scalar	455.67	700	.968	.017	34.301	.010
Partial scalar	433.05	699	.976	.014	5.4924	.003
Strict	461.25	727	.972	.015	38.095	.004

The first line of Table 4 shows the results of the baseline model (i.e., the model parameters are freely estimated across gender groups). These results were compared with later comparisons of more restrictive models. In the metric model, factor loadings were restricted to be equal for both genders. When the model fit values were compared, the chi-square difference test was not statistically significant ($\Delta\chi^2 = 24.212$, $df = 21$, $p = .28$), and ΔCFI was lower than 0.1, which indicated that the metric model fit the data equally across gender. As for the scalar model, the chi-square difference test was significant ($\Delta\chi^2 = 455.67$, $df = 21$, $p < .005$), and the change in CFI was also above an acceptable fit. Therefore, there was a lack of scalar invariance for the shortened scale. Thus, partial scalar invariance tests were established by freely estimating regression coefficients between the Physical Factor and item 24 (M24) for females and males. Then, it was indicated by the comparison of the adjusted scalar model and the metric model that partial scalar invariance was established for the scale ($\Delta\chi^2 = 5.4924$, $df = 20$, $p = .9$, ΔCFI < .1). Finally, strict invariance was checked by using the adjusted scalar model. However, the residuals were constrained to be equal for females and males. Thus, it was shown in the results that the chi-square difference test was not significant ($\Delta\chi^2 = 38.095$, $df = 28$, $p = .9$) and the ΔCFI was lower than 0.1. As a result, strict invariance across females and males was established. In sum, when these results were combined with moderate reliability and a good model fit, it was concluded that the shortened form had similar features to the full scale.

4. DISCUSSION and CONCLUSION

The primary objective of this current study was to develop a reliable and valid short form of the MISAC by utilizing the ACO algorithm. In addition, another aim was to gather validity and reliability evidence for the shortened version of the MISAC and test measurement invariance across gender groups. The ACO produced a 4-item per factor with a total of 28 items in the shortened scale (around 44% shorter). This finding suggests that responding to the shortened version of the MISAC can take approximately eight minutes on average. Hence, the shortened scale can allow researchers to collect data more flexibly and efficiently while reducing time, cost, and respondent burden (Basarkod et al., 2018).

Notably, the factorial structure and inter-correlations were maintained for the factors within the shortened scale. Furthermore, the shortened scale maintained the content representation across the seven factors underlying the MISAC. The item selection process inevitably involves a trade-off between their predictive strength and ensuring comprehensive content coverage (Leite et al., 2008; Raborn et al., 2019). Additionally, item sampling methods are closely connected to the specific elements within the construct being studied and the available item pool (Jankowsky et al., 2020). Given the relatively constrained size of the MISAC's original item pool, the ACO methodology adeptly extracted items that aligned statistically and conceptually with the intended content.

Our findings regarding the association between mental imagery and external variables (i.e., metaphorical thinking and creativity imagery abilities) were consistent with the results of the full scale. Obtaining the same results with the shortened version of the MISAC indicated that the short version has similar relationships with external variables, as seen in the full version. Research demonstrates that mental imagery serves as a foundational cognitive skill not only in creating mental representations of "images" but also in comprehending metaphors, thereby indicating its pivotal role in cognitive processes and creativity (Cornelissen & Clarke, 2010; Pérez-Fabello et al., 2016). Therefore, this finding holds crucial significance, indicating that scores derived from the short form effectively pinpoint the nuanced interplay between mental imagery and mentioned external variables. This validation solidifies the utility and applicability of the shortened MISAC in assessing and understanding the intricate cognitive mechanisms at play.

Smith et al. (2020) noted that shortening a scale brings several drawbacks. One such drawback related to reliability was evident in this study. With fewer items included in each factor, the shortened MISAC demonstrated only acceptable reliability. The measurement invariance results also revealed that strict invariance across females and males was attained using the adjusted scalar model. This conclusion stemmed from the observed disparity in the regression coefficient between item 24 (Sensing the texture of warm water) for females and males within the Physical Factor, suggesting varying interpretations of this item between genders. This discrepancy might be linked to gender's substantial influence on thermal perception (Schellen et al., 2013). The mental perception of warm water's temperature and texture may vary depending on gender. Hence, we recommend considering the shortened MISAC depending on the sample characteristics and research objectives. The original MISAC form might remain preferable when investigating gender differences.

The results showed that the ACO algorithm produced a shortened scale that satisfactorily showed good psychometric properties. Hence, the shortened scale can be considered a suitable alternative to the full scale in measuring mental imagery in the context of artistic creativity. The results of this current study are similar to previous studies that indicate that the ACO algorithm provides an effective procedure for shortening scales (Leite et al., 2008; Marcoulides & Drezner, 2003). Nevertheless, the ACO algorithm should be run multiple times to determine the appropriate items for content representability, as the item selection process should not be based solely on the algorithms (Kleka & Soroko, 2018). Therefore, these automated algorithms may guide researchers in efficiently selecting their items (Yarkoni, 2010). The manual selection of items does have disadvantages and does not always offer an optimal solution (Olaru et al., 2015; Sandy et al., 2014). Thus, running automated algorithms and examining results in terms of relevant theories may be preferable.

Overall, researchers aiming to collect data regarding mental imagery may utilize the shortened version of the MISAC to save time while maintaining a high level of reliability, validity, and similar features to the full scale. So, researchers aiming to use scales that include items with relatively demanding cognitive loads, such as the MISAC, can follow similar procedures to obtain psychometrically sound brief scales.

Some methodological limitations in this study should be considered. First, there were limitations regarding the sample's representativeness, as it consisted solely of university students enrolled in undergraduate programs at art education institutions. Because of the gender imbalance in our sample and the limited number of male students, we were unable to utilize Jankowsky et al.'s (2020) and Olaru and Jankowsky's (2022) functions, which could have enabled us to select measurement invariance as a means to create a short form within the ACO algorithm. Therefore, for future research without these limitations, it is recommended that the functions be adapted to their specific dataset and the ACO algorithm employed accordingly.

Since this study was conducted within an art education group selected through a rigorous process, certain items might have been relatively effortless for participants to imagine mentally. To thoroughly investigate mental imagery within artistic work, other programs that require creative skills, such as design, architecture, and communication, should also be included in future research. Additionally, in this study, we could not consider students' academic year levels as a grouping variable due to the limited sample size in specific year cohorts. Future research could explore potential mean-level differences in students' abilities in mental imagery throughout their university education in the analysis.

In this study, there were no external variables that could reveal moderate or high correlations within our data set. Thus, future research can include additional variables to collect convergent or divergent validity. Exploring drawing skills, visual thinking abilities, and imaginative thinking skills through well-known assessments (e.g., The Torrance Tests of Creative Thinking) can be useful for gathering such evidence. Furthermore, latent group differences in mental imagery, in conjunction with these variables, can be examined while considering the previously mentioned grouping variables. Finally, since we aimed to shorten the scale, the reliability level decreased compared to the full scale. Therefore, using the shortened or full scale depends on the aim of future research. For example, suppose the plan is to utilize students' scores for individual-level decisions, such as selecting individuals for programs that require artistic skills or within the diagnostic processes for especially talented individuals. In that case, we recommend utilizing the full scale, as is strongly emphasized in other studies (e.g., Kruyen et al., 2014). However, the shortened scale is recommended if the aim is to analyze scores at a group level, such as modeling mental imagery or determining associations with other relevant variables. This approach saves time and reduces response bias during assessment sessions.

Acknowledgments

This paper was extracted from the first author's doctoral dissertation.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s). **Ethics Committee Number:** Cukurova University, Social Sciences and Humanities Research Ethics Committee, 03/04/2018-E.49944.

Contribution of Authors

Handan Narin Kızıltan: Investigation, Conception, Literature Review, Materials, Data Collection and Processing, Visualization, Writing. **Hatice Çiğdem Bulut**: Investigation, Conception, Literature Review, Design, Data Analysis, Visualization, Supervision, Writing.

Orcid

Handan Narin Kızıltan https://orcid.org/0000-0002-2164-8389
Hatice Çiğdem Bulut https://orcid.org/0000-0003-2585-3686

REFERENCES

Andrade, J., May, J., Deeprose, C., Baugh, S.J., and Ganis, G. (2014). Assessing vividness of mental imagery: The Plymouth Sensory Imagery Questionnaire. *British Journal of Psychology*, *105*, 547-563. https://doi.org/10.1111/bjop.12050

Arshamian, A., & Larsson, M. (2014). Same but different: The case of olfactory imagery. *Frontiers Psychology*, *5*, 34. https://doi.org/10.3389/fpsyg.2014.00034

Basarkod, G., Sahdra, B., & Ciarrochi, J. (2018). Body image-acceptance and action questionnaire-5: An abbreviation using genetic algorithms. *Behavior Therapy*, *49*(3), 388-402. https://doi.org/10.1016/j.beth.2017.09.006.

Betts, G.H. (1909). *The distribution and functions of mental imagery*. New York, Columbia University.

Bhattacharya, J., & Petsche, H. (2002). Shadows of artistry: cortical synchrony during perception and imagery of visual art. *Cognitive Brain Research*, *13*(2), 179-186. https://doi.org/10.1016/S0926-6410(01)00110-0

Botes, E., Dewaele, J.M., & Greiff, S. (2021). The development and validation of the short form of the foreign language enjoyment scale. *The Modern Language Journal*, *105*(4), 858-876. https://doi.org/10.1111/modl.12741

Bowns, R., Loeffelman, J.E., Steinley, D., & Sher, K.J. (2022). A brief young adult alcohol problems screening test: Short form development using combinatorics. *Journal of American College Health*, 1-7. https://doi.org/10.1080/07448481.2022.2095870

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*, 796-846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Bulut, O. (2020). *Testing for Measurement Invariance in R*. Retrieved from https://okan.cloud/posts/2020-12-21-testing-for-measurement-invariance-in-r/

Calabrese, L., & Marucci, F.S. (2006). The influence of expertise level on the visuo-spatial ability: Differences between experts and novices in imagery and drawing abilities. *Cognitive Processing*, *7*(1). 118-120. https://doi.org/10.1007/s10339-006-0094-2.

Carr, T., Moss, T., & Harry, D. (2005). The DAS24: A short form of the Derriford Appearance Scale DAS59 to measure individual responses to living with problems of appearance. *British Journal of Health Psychology*, *10*(2), 285-298. https://doi.org/10.1348/135910705X27613

Chamberlain, R., Drake, J.E., Kozbelt, A., Hickman, R., Siev, J., & Wagemans, J. (2019). Artists as experts in visual cognition: An update. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(1), 58. https://doi.org/10.1037/aca0000156

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

Cornelissen, J.P., & Clarke, J.S. (2010). Imagining and rationalizing opportunities: Inductive reasoning and the creation and justification of new ventures. *Academy of Management Review*, *35*(4), 539-557. https://doi.org/10.5465/amr.35.4.zok539

Cumming, J., & Eaves, D.L. (2018). The nature, measurement, and development of imagery ability. *Imagination, Cognition and Personality*, *37*(4), 375-393. https://doi.org/10.1177/0276236617752439

Deaver, S.P., & Shiflett, C. (2011). Art-based supervision techniques. *The Clinical Supervisor*, *30*(2), 257-276. https://doi.org/10.1080/07325223.2011.619456

Deneubourg, J.L., Pasteels, J.M., & Verhaeghe, J.C. (1983). Probabilistic behaviour in ants: A strategy of errors? *Journal of Theoretical Biology*, *105*, 259-271. https://doi.org/10.1016/S0022-5193(83)80007-1

Dilmac, O., & Kucuoglu, A. (2010). Güzel sanatlar eğitimi bölümleri resim-iş eğitimi anabilimdallarinda sunum dosyasina dayali özel yetenek sinav modeli [A sample of special ability test based on portfolio at the fine arts teaching departments]. *Journal of Institute of Fine Arts*, *18*, 63-77.

Dodson, B. (2013). *Keys to drawing with imagination: Strategies and exercises for gaining confidence and enhancing your creativity*. Pegasus Press.

Dogan, B.G., Bulut, H.C. (2024) Abbreviation of parenting behaviors and temperament in children scales using genetic algorithms. *Current Psychology*, 43, 7044-7058. https://doi.org/10.1007/s12144-023-04863-z

Drake, J.E., Simmons, S., Rouser, S., Poloes, I., & Winner, E. (2021). Artists excel on image activation but not image manipulation tasks. *Empirical Studies of the Arts*, *39*(1), 3–16. https://doi.org/10.1177/0276237419868941

Duncun, P. (2001). Visual culture: Developments, definitions, and directions for art education. *Studies in Art Education*, *42*(2), 101-112. https://doi.org/10.1080/00393541.2001.11651691

Efland, A.D. (2002). *Art and cognition: Integrating the visual arts in the curriculum.* Teachers College Press.

Hetland, L., Winner, E., Veenema, S., & Sheridan, K.M. (2007). *Studio thinking: The real benefits of visual arts education*. Teachers College Press.

Heid, K., Estabrook, M., & Nostrant, C. (2009). Dancing with line: Inquiry, democracy, and aesthetic development as an approach to art education. *International Journal of Education & the Arts*, *10*(3).

Holmes, E.A. & Mathew, A. (2010). Mental imagery in emotion and emotional disorders. *Clinical Psychology Review*, *30*(3), 349-362. https://doi.org/10.1016/j.cpr.2010.01.001

Jankowska, D.M., & Karwowski, M. (2015). Measuring creative imagery abilities. *Frontiers in Psychology*, *6*, 1591. https://doi.org/10.3389/fpsyg.2015.01591

Jankowska, D.M. & Karwowski, M. (2020). Mental imagery and creativity. https://doi.org/10.31234/osf.io/eyfxr

Jankowsky, K., Olaru, G., & Schroeders, U. (2020). Compiling measurement invariant short scales in cross–cultural personality assessment using ant colony optimization. *European Journal of Personality*, *34*(3), 470-485. https://doi.org/10.1002/per.2260

Ji, J.L., Kavanagh, D.J., Holmes, E.A., MacLeod, C., & Di Simplicio, M. (2019). Mental imagery in psychiatry: Conceptual & clinical implications. *CNS spectrums*, *24*(1), 114-126. https://doi.org/10.1017/S1092852918001487

Kleka, P., & Soroko, E. (2018). How to avoid the sins of questionnaires abridgement? *Survey Research Methods*, *12*(2), 147-160. https://doi.org/10.18148/srm/2018.v12i2.7224

Kosslyn, S.M., Ganis, G., & Thompson, W.L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*, 635-642. https://doi.org/10.1038/35090055

Kozhevnikov, M., Kozhevnikov, M., Yu, C.J., & Blazhenkova, O. (2013). Creativity, visualization abilities, and visual cognitive style. *British Journal of Educational Psychology*, *83*, 196-209. https://doi.org/10.1111/bjep.12013.

Krosnick, J.A., Holbrook, A.L., Berent, M.K., Carson, R.T., Hanemann, W.M., Kopp, R.J., ..., Conaway, M. (2002). The impact of "no opinion" response options on data quality: Nonattitude reduction or an invitation to satisfice? *The Public Opinion Quarterly*, *66*(3), 371-403. https://doi.org/10.1.1.141.7834

Kruyen, P.M., Emons, W.H., & Sijtsma, K. (2014). Assessing individual change using short tests and questionnaires. *Applied Psychological Measurement*, *38*, 201-216. https://doi.org/10.1177/0146621613510061

Lakoff, G., & Johnson, M. (2010). *Metaphors we live by* (G.Y. Demir, Trans.). (Rev. ed. 2nd ed.). Paradigma Press.

Leite, W.L., Huang, I.C., & Marcoulides, G.A. (2008). Item selection for the development of short forms of scales using an ant colony optimization algorithm. *Multivariate Behavioral Research*, *43*(3), 411-431. https://doi.org/10.1080/00273170802285743

Marcoulides, G.A., & Drezner, Z. (2003). Model specification searches using ant colony optimization algorithms. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(1), 154-164. https://doi.org/10.1207/S15328007SEM1001_8

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

Marks, D.F. (1995). New directions for mental imagery research. *Journal of Mental Imagery*, *19*(3-4), 153-167.

McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

Miller, L.A. (2014). A self-report measure of cognitive processes associated with creativity. *Creativity Research Journal*, *26*(2), 203-218. https://doi.org/10.1080/10400419.2014.901088.

Narin, H. (2019). Sanat eğitimi öğrencilerinin zihinsel imgeleme kapasitesini belirlemeye yönelik bir ölçme aracı geliştirme çalışması [*A study on developing a measurement tool for determining mental imagery capacity of art education students]*. [Unpublished doctoral thesis]. Cukurova University.

Niessen, A.S.M., Meijer, R.R., & Tendeiro, J.N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use?. *Journal of Research in Personality*, *63*, 1-11. https://doi.org/10.1016/j.jrp.2016.04.010

Nunnally, J.C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed). McGraw-Hill.

O'Donoghue, D. (2011). Has the art college entry portfolio outlived its usefulness as a method of selecting students in an age of relational, collective and collaborative art practice? *International Journal of Education & the Arts*, *12*(3), 1-27.

Olaru, G., Witthöft, M., & Wilhelm, O. (2015). Methods matter: Testing competing models for designing short-scale Big-Five assessments. *Journal of Research in Personality*, *59*, 56-68. https://doi.org/10.1016/j.jrp.2015.09.001

Olaru, G., & Jankowsky, K. (2022). The HEX-ACO-18: Developing an age-invariant HEXACO short scale using ant colony optimization. *Journal of Personality Assessment*, *104*(4), 435-446. https://doi.org/10.1080/00223891.2021.1934480

Ozmutlu, A., & Tomak, A. (2021). Examination of plastic arts field special ability exams in Turkey on the basis of their characteristics and risks. *International Journal of Interdisciplinary and Intercultural Art*, *6*(12), 39-56.

Palmiero, M., Piccardi, L., Nori, R., Palermo, L., Salvi, C., & Guariglia, C. (2016) Editorial: Creativity and Mental Imagery. *Frontiers Psychology*, *7*(1280). https://doi.org/10.3389/fpsyg.2016.01280

Park, M., & Yoo, J. (2020). Effects of perceived interactivity of augmented reality on consumer responses: A mental imagery perspective. *Journal of Retailing and Consumer Services*, *52*, 101912. https://doi.org/10.1016/j.jretconser.2019.101912

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), 624-634. https://doi.org/10.1038/s41583-019-0202-9

Pearson, J., Naselaris, T., Holmes, E.A., & Kosslyn, S.M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences*, *19*(10), 590-602. https://doi.org/10.1016/j.tics.2015.08.003.

Pérez-Fabello, M.J., & Campos, A. (2007). Influence of training in artistic skills on mental imaging capacity, *Creativity Research Journal*, *19*(2-3), 227-232.

Pérez-Fabello, M.J., Campos, A., & Campos-Juanatey, D. (2016). Is object imagery central to artistic performance?. *Thinking Skills and Creativity*, *21*, 67-74. https://doi.org/10.1016/j.tsc.2016.05.006.

R Core Team (2022). *R: A Language and environment for statistical computing.* (Version 4.0) [Computer software]. https://cran.r-project.org

Raborn, A.W., & Leite, W.L. (2018). ShortForm: An R package to select scale short forms with the ant colony optimization algorithm. *Applied Psychological Measurement*, *42*(6), 516-517. https://doi.org/10.1177/0146621617752993

Raborn, A.W., Leite, W.L., & Marcoulides, K.M. (2020). A comparison of metaheuristic optimization algorithms for scale short-form development. *Educational and Psychological Measurement*, *80*(5), 1-22. https://doi.org/10.1177/0013164420906600

Rammstedt, B., & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and ways to overcome them. *Journal of Individual Differences*, *35*(4), 212-220. https://doi.org/10.1027/1614-0001/a000141

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. [R package]. Retrieved from https://cran.r-project.org/package=psych.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Rosenberg, H.S. (1987). Visual artists and imagery. *Imagination, Cognition and Personality*, *7*(1), 77-93. https://doi.org/10.2190/AVJ5-N24B-P7MC-HR4R.

Sandy, C.J., Gosling, S.D., & Koelkebeck, T. (2014). Psychometric comparison of automated versus rational methods of scale abbreviation: An illustration using a brief measure of values. *Journal of Individual Differences*, *35*(4), 221-235. https://doi.org/10.1027/1614-0001/a000144

Saulsman, L.M., Ji, J.L., & McEvoy, P.M. (2019). The essential role of mental imagery in cognitive behaviour therapy: What is old is new again. *Australian Psychologist*, *54*(4), 237-244. https://doi.org/10.1111/ap.12406

Schellen, L., Loomans, M., de Wit, M., & van Marken Lichtenbelt, W. (2013). The influence of different cooling techniques and gender on thermal perception. *Building Research & Information*, *41*(3), 330-341. https://doi.org/10.1080/09613218.2013.772002

Schroeders, U., Wilhelm, O., & Olaru, G. (2016). Meta-heuristics in short scale construction: Ant colony optimization and genetic algorithm. *PloS One*, *11*(11), e0167110. https://doi.org/10.1371/journal.pone.0167110

Schwarz, S., Grasmann, D., Schreiber, F., & Stangier, U. (2020). Mental imagery and its relevance for psychopathology and psychological treatment in children and adolescents: A systematic review. *International Journal of Cognitive Therapy*, *13*, 303-327. https://doi.org/10.1007/s41811-020-00092-5

Serig, D. (2006). A conceptual structure of visual metaphor. *Studies in Art Education*, *47*(3), 229-247. https://doi.org/10.1080/00393541.2006.11650084

Sheehan, P.W. (1967). A shortened form of Betts' Questionnaire Upon Mental Imagery. *Journal of Clinical Psychology*, *23*(3), 386-389. https://doi.org/10.1002/1097-4679(196707)23:3<386::AID-JCLP2270230328>3.0.CO;2-S

Smith, G.T., McCarthy, D.M., & Anderson, K.G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102. https://doi.org/10.1037/1040-3590.12.1.102

Swindle, R., Cameron, A., & Rosen, R. (2006). A 15-item short form of the psychological and interpersonal relationship scales. *International Journal of Impotence Research*, *18*(1), 82–88. https://doi.org/10.1038/sj.ijir.3901381

Taskesen, S. (2019). An investigation on special talent exams in the division of art teaching. *Journal of Education and Training Studies, 7*(10), 86-97. https://doi.org/10.11114/jets.v7i10S.4554

Tay, J. (2019). *Art teachers' perceptions about visual arts giftedness: Content and construct validation of perceptions about art giftedness* [Doctoral dissertation, Purdue University]. ProQuest Dissertations Publishing.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-92. https://doi.org/10.1080/17405629.2012.686740

Vellera, C., & Gavard-Perret, M. L. (2012). *Is mental imagery ability an element for identifying creative consumers?*. https://halshs.archives-ouvertes.fr/halshs-00851322

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363-373. https://doi.org/10.1016/j.jrp.2010.04.001

Yilmaz, S. (2016). The relation between visual perception tests organized for special talent exams and academic achievement in art teaching program. *The Black Sea Journal of Social Sciences*, *8*(15), 55-74.

Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* [Doctoral dissertation, University of California]. ProQuest Dissertations Publishing.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236-247. https://doi.org/10.1016/j.ijresmar.2010.02.004

West, S.G., Finch, J.F., Curran, P.J., & Hoyle, R.H. (1995). Structural equation modeling: Concepts, issues, and applications. In Hoyle, R.H (Ed.). *Structural equation models with nonnormal variables: Problems and remedies*, (p. 55-75). Sage

Ziss, A. (2011). *The science of artistic assimilation of aesthetic reality* (2nd ed.). (Y. Şahan, Trans.). Hayalbaz Books.

6. APPENDIX

6.1. Appendix A. Utilized Code for Running the ACO Algorithm

```
# Load packages
library(ShortForm)
library(lavaan)
# Load data
misac <- read_excel("C:/.../Research/SA/misac.xlsx")
misac_v1 <- data.matrix(misac[,6:68])
# Run the ACO logarithm
misac_short <- antcolony.lavaan(data = misac_v1,
                    ants = 120,
                    evaporation = 0.95,
                    antModel = 'char =~ M6+ M2+ M8+ M7+ M10+ M3+ M1
                          kine =~ M16+ M20+ M18+ M14+ M15+ M12
                          tact =~ M22+ M26+ M21+ M29+ M23+ M27+ M24+ M25+ M30+ M28
                          spat =~ M43+ M45+ M48+ M41+ M44
                          phys =~ M51+ M54+ M52+ M55+ M56+ M59+ M53+ M58
                          emot =~ M69+ M67+ M66+ M65+ M63+ M70+ M64
                          affe =~ M75+ M72+ M76+ M80+ M73+ M74+ M79',
                    list.items = list(c('M6', 'M2', 'M8', 'M7', 'M10', 'M3', 'M1'),
                          c('M16', 'M20', 'M18', 'M14', 'M15', 'M12'),
                          c('M22', 'M26', 'M21', 'M29', 'M23', 'M27', 'M24', 'M25', 'M30', 'M28'),
                          c('M43', 'M45', 'M48', 'M41', 'M44'),
                          c('M51', 'M54', 'M52', 'M55', 'M56', 'M59', 'M53', 'M58'),
                          c('M69', 'M67', 'M66', 'M65', 'M63', 'M70', 'M64'),
                          c('M75', 'M72', 'M76', 'M80', 'M73', 'M74', 'M79')),
                    full = 50, i.per.f = c(4,4,4,4,4,4,4),
                    factors = c('char','kine','tact','spat', 'phys','emot','affe'),
                    steps = 20,
                    fit.indices = c('cfi','rmsea'),
                    fit.statistics.test = "(cfi > 0.95)&(rmsea < 0.05)",
                    summaryfile ='summary.txt',
                    feedbackfile ='iteration.html',
                    max.run = 1000)
# print selected items
misac_short$best.syntax
```

6.1. Appendix B. Results of confirmatory factor analysis of the shortened scale

Factor	Item	Rephrased Item Labels and prompts	Estimate (SE)
Spatial		How vividly can you imagine your favorite café in your mind?	
	M43	Visualizing the interior dimensions of the cafe	.67 (.06)***
	M45	Visualizing the placement of tables, chairs, cash register, etc. in the cafe	.80 (.06)***
	M44	Visualizing the height of the cafe's ceiling	.58 (.07)***
	M41	Visualizing the color and shape of the cafe's signboard	.77 (.06)***
Tactile		How clearly can you imagine the sensations you feel with your hands?	
	M22	Sensing the texture of cotton	.64 (.05)***
	M23	Sensing the texture of a thorn	.62 (.06)***
	M24	Sensing the texture of warm water	.57 (.07)***
	M28	Sensing the texture of silk fabric	.69 (.05)***
Physical		How vividly can you imagine these movements?	
	M51	Walking uphill	.66 (.04)***
	M56	Carrying a heavy load on your back	.71 (.05)***
	M59	Throwing a basketball	.65 (.06)***
	M53	Climbing a tree	.59 (.07)***
Kinesthetic		How clearly can you see various movements and situations related to a motorcycle and its actions?	
	M16	Overcoming a bump/obstacle on a motorcycle	.58 (.06)***
	M20	Dragging a fallen motorcycle on the ground	.75 (.07)***
	M18	Motorcycle colliding rapidly with a vehicle	.79 (.06)***
	M14	Motorcycle swiftly passing by	.66 (.07)***
Emotional		How vividly can you imagine a feeling or emotion?	
	M69	Feeling guilt	.69 (.08)***
	M67	Experiencing panic/shock	.69 (.07)***
	M66	Feeling doubt	.74 (.08)***
	M70	Expressing astonishment	.53 (.09)***
Characteristic		How vividly can you see a familiar friend in your mind?	
	M7	Appearance while expressing joy	.60 (.07)***
	M8	Appearance when angered	.52 (.07)***
	M10	Notable behavior while eating (e.g., eating habits)	.72 (.07)***
	M3	Notable behavior while walking/stepping (e.g., stride length)	.67 (.07)***
Affective		How clearly can you imagine the expressions or emotions?	
	M75	A cat with a full stomach	.68 (.06)***
	M80	A dog growling upon seeing a stranger	.70 (.07)***
	M73	Eating situation of a child with a sore throat	.77 (.06)***
	M74	Body of a sleep-deprived person	.59 (.06)***

Note. Total explained variance (R^2 = 59%), $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$