ORIGINAL ARTICLE

# Development and Validation of the Cebeci Test of Creativity: A Computerized Test of Figural Creativity

Sukru Murat Cebeci[1] | Selcuk Acar[2]

[1]Renzulli Learning, New Haven, Connecticut, USA | [2]Department of Educational Psychology, University of North Texas, Denton, Texas, USA

**Correspondence:** Selcuk Acar (selcuk.acar@unt.edu)

**Received:** 29 February 2024 | **Revised:** 18 April 2025 | **Accepted:** 7 May 2025

**Keywords:** Cebeci Test of Creativity | creativity assessment | divergent thinking | scale development

## ABSTRACT

This study presents the Cebeci Test of Creativity (CTC), a novel computerized assessment tool designed to address the limitations of traditional open-ended paper-and-pencil creativity tests. The CTC is designed to overcome the challenges associated with the administration and manual scoring of traditional paper and pencil creativity tests. In this study, we present the first validation of CTC, demonstrating strong internal and external validity across two studies with a large sample size of over 14,000 students in grades 1–8. The results provide support for the proposed unidimensional factor structure of CTC, with robust reliability ($\omega = 0.833$ and 0.872). Analyses of measurement invariance showed that the unidimensional factor structure of CTC holds consistently across all grade levels, with factor loadings exhibiting notable similarity. Additionally, the item intercepts demonstrate considerable uniformity across grades 3–5. The composite CTC scores were positively correlated with creative self-efficacy but not with Standard Progressive Matrices. The outcomes of our study indicate that CTC is a valuable and efficient tool for assessing creativity in educational settings. Its scalability and comprehensive evaluation of four key dimensions of creative ideation (i.e., fluency, flexibility, originality, and elaboration) make it particularly advantageous for educators seeking to assess students' creative potential.

## 1 | Introduction

Many traditional creativity tests utilize an open-ended paper-and-pencil format, making scoring and administration time consuming and cumbersome and hindering their widespread adoption for large-scale school or district assessments (Acar, Berthiaume, et al. 2023). Despite the empirical value of classic tests, such as the Torrance Tests of Creative Thinking (TTCT, 1998) and the Test of Creative Thinking-Drawing Production (Urban and Jellen 1996), and more recent developments such as the Evaluation of Potential for Creativity (EPoC; Lubart et al. 2011), educators often seek instruments that are cost-effective, psychometrically sound, quick to score, and comprehensive in assessing various aspects of creativity. Efforts to automate the scoring of these tests, attempted in the past (Paulus and Renzuli 1968), did not gain traction. However, during the last 15 years, technology has played a significant role in enhancing the scalability of creativity assessment tools. Specifically, semantic networks, semantic distance (Acar and Runco 2014; Beaty et al. 2022; Dumas et al. 2021), large language models, and supervised learning (Organisciak et al. 2023) have been used to make scoring classic divergent thinking tasks more efficient and innovative. Importantly, these efforts focused on improving the scoring of already existing tasks rather than revolutionizing the tasks themselves. The Cebeci Test of Creativity (CTC), a new test of

creative potential that presents a different kind of challenge (a flower design task) compared to traditional divergent thinking tests and provides evidence of internal and external validity through two studies with large sample sizes.

### 1.1 | Figural Tests of Creative Thinking

Figural tests of creativity have been used for decades to measure aspects of creativity that are typically not measured by verbal tests of creativity (Kim 2017; Richardson 1986). Figural tests of creativity are known for greater potential for inducing original ideation compared to verbal tests (Runco and Albert 1985). Runco (1986) found that the originality-to-fluency ratio was higher in the Pattern Meanings test than in the Alternate Uses test. Although this feature does not necessarily make figural tests more useful assessment tools than verbal tests, it suggests that figural tasks are more open to original interpretation, as compared to the verbal tests. One way that figural tests can be particularly useful is that they can help avoid potential bias in verbal tests of creativity due to language and culture barriers (Erwin and Worrell 2012). This feature of the figural creativity tests can especially be important for high-stakes decisions including the identification of creatively gifted and talented students from culturally, linguistically, and economically diverse groups (Luria et al. 2016; Peters 2022).

Well-known tests of figural creativity exist, of which the most prominent is the Torrance Tests of Creative Thinking-Figural (Torrance 1966, 1998), with two forms (Form A and Form B), each consisting of three activities (Picture Construction, Picture Completion, and Circles or Lines, depending on the forms). TTCT-F is administered in 30 min (10 min for each activity) along with the instructions to think of a picture that no one else will think of. TTCT-F blends single and multiple-response test formats, as Activities 1 and 2 involve a single prompt and Activity 3 corresponds to a multiple-response structure through repeated presentation of the same prompt over 3 pages. TTCT-F is a paper-pencil test and is administered individually or in groups. The original version was scored for fluency, flexibility, originality, and elaboration, whereas the current streamlined version (Torrance and Ball 1984) used indices such as fluency, originality, elaboration, abstractness of titles, and resistance to premature closure. TTCT-F is a figural test because both stimuli and responses (drawings) are mainly figural, yet it still includes a verbal component through the titles given for the drawings, which are then scored for the level of abstractness. TTCT-F has been shown to have a two-factor structure with fluency and originality loading under the same factor and the other three scoring indices often loading under the second factor (Acar, Ogurlu, et al. 2023; Said-Metwaly et al. 2021). Factor-based reliability was shown to be higher for the first factor than the second, and a composite reliability estimate combining the two factors was also good (Acar 2023; Acar, Dumas, et al. 2024; Acar, Lee, et al. 2024; Acar, Organisciak, et al. 2024).

Wallach and Kogan (1965) also developed figural tests such as Line Meanings and Pattern Meanings, but different from TTCT-F, participants are not asked to draw a picture or object. Instead, they list what the presented figure may look like. Thus, it is a figural test at the stimuli level with an entirely verbal response. Another well-known test of figural creativity is the Test for Creative Thinking—Drawing Production (TCT-DP; Urban and Jellen 1996), which uses a single page with various shapes spread throughout the page, most of which are presented in a square box, and respondents are asked to turn it into a drawing. TCT-DP is scored for 14 evaluation criteria, which are then composited into a single general score. A different test, the Evaluation of Potential for Creativity (EPoC; Lubart et al. 2011) involves a graphic component that is measured with both abstract and concrete stimuli for both divergent and convergent (integrative) skills. Figural creativity tasks were also part of the Berlin Structure-of-Intelligence Test for Youth: Diagnosis of Talents and Giftedness (Jäger et al. 2005), in which respondents draw pictures from objects (symbol completion), combine geometric objects to create new figures (symbol combination), or convert them into real objects (object design), and design logotypes (layout).

Classic figural tests were originally designed as paper-pencil tests, and some have been moved to computerized environments for administration while maintaining the original structure of the test (Guo 2019; Kwon 1996; Lau and Cheung 2010; Palaniappan 2012; Zabramski 2014). However, it is still likely that testing experience and outcomes could vary due to differences in perceptual demands, necessary motor skills, mode of item presentation, and familiarity with the devices used (Schroeders and Wilhelm 2010). In an adult sample of respondents, Guo (2019) compared computerized and paper-and-pencil versions of verbal and figural creativity tests and found no difference in reliability and performance. Lau and Cheung (2010) conducted a similar investigation with fourth-grade students and found that reliability, inter-correlation coefficients, and performance in paper-and-pencil versus online versions of verbal and figural Wallach–Kogan Creativity Tests did not differ. Kwon et al. (1998), on the other hand, found that computerized and paper-and-pencil versions did not yield equivalent performance among fifth and sixth-grade students. These differences can be explained by factors such as using a mouse versus a pencil to draw (or design) or specific functions available in a computerized environment (quicker "undo" or "delete" versus manual erase) (Kwon et al. 1998).

Computerized assessment offers several clear advantages, such as providing additional data on respondents' processes, including latency (Acar et al. 2019), the number of iterations or corrections, and the specific tools used during task completion (Bump 1994; Shoemaker and Bolt 1992), as well as enabling quicker and easier scoring (Acar, Dumas, et al. 2024; Acar, Lee, et al. 2024; Acar, Organisciak, et al. 2024). On the other hand, computerized testing may result in different performances (Backes and Cowan 2019; Wollscheid et al. 2016) depending on familiarity with specific computer functions, previous experience with computers, and availability of assistance when working on the tasks (Mazzeo and Harvey 1988; Horkay et al. 2006). All these factors may become further pronounced in relation to developmental stages and socio-economic status. Concerning the latter, Pender (2020), for example, found that students in wealthier districts performed better in computer-based assessments than in paper-and-pencil methods. This difference may be related to the amount of exposure to and familiarity with the electronic devices used for testing, as low-income students

may face financial barriers to accessing these devices both at home and in their schools. Developmentally, motor and perceptual skills may affect performance because the extent to which a young respondent can effectively use electronic devices (as opposed to a pencil) and keep their focus on the task instructions presented on the computer screen (versus a human) is directly related to their maturation and experience or familiarity with the devices (Schroeders and Wilhelm 2010). Importantly, figural creativity tasks require more use of drawing features, and such factors may impact performance on these tests more than on other tests (e.g., multiple-choice).

Recently, figural tests have been scored using artificial intelligence methods. In one such study, Sung et al. (2024) used figural tasks that resemble Line and Pattern Meanings tests (Wallach and Kogan 1965) but added the feature of rotating the presented figure on the computer screen. Sung and colleagues used the Word2Vec algorithm to score the verbal responses resembling the figure. Artificial intelligence methods have also been applied to drawing-based tests of creativity. In one such study, Cropley and Marrone (2025) employed image classification, specifically convolutional neural networks (CNNs), to analyze creative outputs from the Test of Creative Thinking–Drawing Production (TCT-DP; Urban and Jellen 1996). Training CNN algorithms with classifications ranging from binary (low vs. high creative) to seven levels (1–7), they achieved high accuracy ($\kappa = 0.83–0.94$) compared to manual classification using five classification schemes. Patterson et al. (2022) extended these methods to the Multi-Trial Creative Ideation task (MTCI; Barbot 2018) using ResNet, a deep CNN. They trained the Automated Drawing Assessment (AuDrA) platform, showing a strong correlation with human raters (average $r = 0.76$). Acar, Dumas, et al. (2024); Acar, Lee, et al. (2024); Acar, Organisciak, et al. (2024) used vision transformer models (i.e., BEiT) with both MTCI and TTCT-F, and successfully predicted human ratings of originality in TTCT-F when using drawings, titles, or both in combination, and reached a high accuracy with MTCI ($r = 0.85$).

### 1.2 | The Cebeci Test of Creative Thinking (CTC)

As summarized above, computerized assessment of creativity with figural tasks has gained interest recently, with most studies using classic divergent thinking tasks that have been studied for decades (Acar, Dumas, et al. 2024; Acar, Lee, et al. 2024; Acar, Organisciak, et al. 2024; Patterson et al. 2022; Sung et al. 2024). In this study, we present the CTC, a computer-administered assessment of creative potential based on the divergent thinking framework (Runco and Acar 2012). The CTC was designed for artificial intelligence scoring and developed to address the need for a scalable and efficient method for assessing creativity in educational settings. The initial version of CTC, introduced in 2014, enabled the use of multiple prompt units, various tasks, and color selection for designs. However, feedback from field studies revealed that students spent excessive time choosing colors, therefore limiting the number of designs they could create. In response, the second iteration of CTC eliminated the color palette selection, opting for a single color for all designs. Additionally, the test duration was shortened to 30 min to fit within a single classroom period. The final version of CTC employs a single prompt and a single task: drawing flowers using petals as units. This simplified design enables a more focused assessment of creativity.

In this test, participants are presented with a single graphic element referred to as a "petal." Their task is to use this petal to design flowers within the empty space provided. Participants can move and rotate these petals with a computer mouse, adding and configuring them to bring their imaginative flower designs to life. After completing one flower, they have the option to begin anew, creating additional unique designs on separate pages. The goal is for respondents to craft as many distinct, interesting, surprising, and unexpected flowers as possible—ones that no one else would imagine. Upon finishing the test, each design is assessed using four key criteria: fluency, flexibility, originality, and elaboration. According to Guilford (1950), these creative abilities are tied to creative personality. For example, fluency shows a creative individual's ability to produce continuously and rapidly, including if a time pressure exists. Guilford believed fluency is important because "the person who is capable of producing a large number of ideas per unit of time, other things being equal, has a greater chance of having significant ideas" (p. 452). These "significant" ideas are often those that are novel and original. Guilford emphasized that novel and original ideas tend to be uncommon but acceptable, and benefit from making remote associations. Remoteness and uncommonness of the responses are greater in the succeeding responses compared to the early ones (Christensen et al. 1957), showing the critical role of fluency in arriving originality. This led some researchers to adopt the "extended effort principle" where fluency is the vehicle to produce original responses (Parnes 1961). The connection between fluency and originality can be viewed from the perspective of the Big-Five personality framework (Costa Jr. and McCrae 1992) where creativity is more strongly related to plasticity (versus stability) component. Grajzel et al. (2023) found that extraversion is more strongly related to fluency whereas openness is related with originality. Notably, plasticity consists of openness and extraversion (Silvia et al. 2009).

Guilford explained flexibility as the opposite of rigidity, which is a personality trait, and measured this ability in open-ended tasks because they enable tracking ideational trajectory in terms of respondent's ability to move away from fixation, to explore and demonstrate new ways of thinking. Flexibility is tied to fluency because ideational productivity benefits from one's ability to find a new way of thinking and depart from the old. Simultaneously, launching new ways of thinking could result in original ideation. Finally, Guilford's conceptualization of elaboration, which he defined as "the ability to work out the details of an idea or solution" (Guilford 1973, 2) is related to two other abilities (i.e., synthesis and complexity) that he mentioned in his seminal work.

Guilford's early conceptualization of the abilities underlying creativity highlights the interrelatedness of the concepts operationalized in CTC. However, Guilford's own assessments employed distinct tasks for scoring the various indices, avoiding stimulus dependency (Barbot 2018), and their factor-analytic studies (Kettner et al. 1959; Wilson et al. 1954) reported a multi-dimensional structure. In contrast, the structure of the CTC is more closely aligned with the Torrance Tests of Creative Thinking-Verbal (Torrance 1998) and Wallach and

Kogan (1965), both of which utilized the same set of stimuli to score different indices. While this approach introduces task dependency in the CTC, it enables an efficient collection of scores within a shorter timeframe compared to the time required to administer separate tasks for each scoring index. This efficiency is particularly crucial for CTC, which was designed for administration in school settings, where testing must be as minimally disruptive as possible.

Building on Guilford's conceptual framework as well as the structure of subsequently developed divergent thinking tasks (Torrance 1998; Wallach and Kogan 1965), the CTC conceptualizes creative potential as a single-factor assessment, encompassing these four indices that represent productivity (fluency), authenticity and novelty (originality), diversity and adaptability (flexibility), and elegance (elaboration), all qualities of human thought, respectively. This formulation is rooted in creative personality in which many, diverse, interesting, and complex ideas are generated as a reflection of open-mindedness and embracing change and progress (An et al. 2016; Grajzel et al. 2023; McCrae 1987).

### 1.3 | The Present Study

In this study, the first validation evidence of the CTC is summarized using a large sample. In two different studies with slightly different explicit instructions, we present findings related to the factor structure, model-based reliability, measurement invariance, and convergent and discriminant validity of CTC. Specifically, we test the unidimensional proposed factor structure of CTC and then assess the factor reliability. We also assess the measurement invariance by grade given the observed developmental fluctuations in creativity (Acar, Dumas, et al. 2024; Said-Metwaly et al. 2021; Torrance 1968). Next, we assess the nomological network of CTC through a self-report measure of creative self-efficacy and a test of general intelligence, namely Raven Standard Progressive Matrices Plus.

In the present study, we tested the hypothesis that CTC will show a unidimensional scale, addressing the following questions:

1. What is the factor reliability of CTC?
2. What is the extent of measurement invariance of CTC across different grade levels?
3. What is the relationship between CTC and self-reported creativity as well as general intelligence?

## 2 | Study 1

### 2.1 | Methods

#### 2.1.1 | Participants

In the first study, we administered the CTC to a total of 10,982 students in grades 1–8 across more than 40 schools in five different countries including Colombia, Italy, Sri Lanka, the United States, and Turkey. Of that number, 83 participants did not generate any meaningful responses, and thus were removed from the dataset. The analytical sample included 10,899 participants, of which the majority were from the United States, specifically in the following states: Connecticut, Florida, Hawaii, Maryland, and New Jersey. The distribution of students across grades was as follows: 1–1601, 2–544, 3–2045, 4–2284, 5–2162, 6–731, 7–466, and 8–1066.

#### 2.1.2 | Instruments

**2.1.2.1 | The Cebeci Test of Creativity.** CTC is a figural test of creative potential in which respondents are asked to develop flower designs by using petals that can be dragged and moved around with a mouse cursor. They receive the instructions on a computer screen and must complete a brief online training to learn how to use the functions of the activity to design their unique flowers (see Figure 1 for tutorial steps and the user interface of the CTC). Respondents can start the activity after they have demonstrated mastery in using the user interface. In accordance with the structure of classic divergent thinking tasks (Acar et al. 2020; Reiter-Palmon et al. 2019), respondents are asked to "design as many different flowers" emphasizing "the more flowers the better." They are also asked to "surprise" others, "try interesting designs" and "draw unique flowers that no one else thinks of" in a period of 30 min.

The CTC is scored for fluency, flexibility, originality, and elaboration. Fluency is scored as the number of meaningful flower designs. After counting meaningful and task relevant designs (all non-flower designs or ambiguous formations were omitted from scoring for fluency and other indices), final fluency scores are obtained by converting the total number of completed flowers on a 1–9 point scale by using the following ranges: one flower = 1 point, two flowers = 2 points, three flowers = 3 points, four flowers = 4 points, five or 6 flowers = 5 points, seven or eight flowers = 6 points, nine to 11 flowers = 7 points, 12 to 14 flowers = 8 points, 15 or more flowers = 9 points. This method improved normality and applied a ceiling to fluency scores in order to differentiate them from other divergent thinking scores, addressing the fluency confound problem. The fluency confound refers to the contamination of divergent thinking scores, such as originality and flexibility, by fluency (Clark and Mirels 1970; Hocevar 1979). Researchers have used a fixed fluency scores to mitigate the fluency confound on other divergent thinking measures, such as by asking participants to generate six responses (Guilford et al. 1960) or just one response (Acar, Dumas, et al. 2024; Acar, Lee, et al. 2024; Acar, Organisciak, et al. 2024; Zarnegar et al. 1988). Our approach is somewhat similar but differs in that we do not set an explicit productivity goal.

Flexibility was scored by counting the number of different flower design concepts used. Based on the sample of produced designs, we identified 21 different design concepts such as use of a stem, negative space, pot, fan, and compound. The total flexibility score is determined based on the number of distinct design concepts employed by a given respondent. When a design involves more than a single design concept, the number of

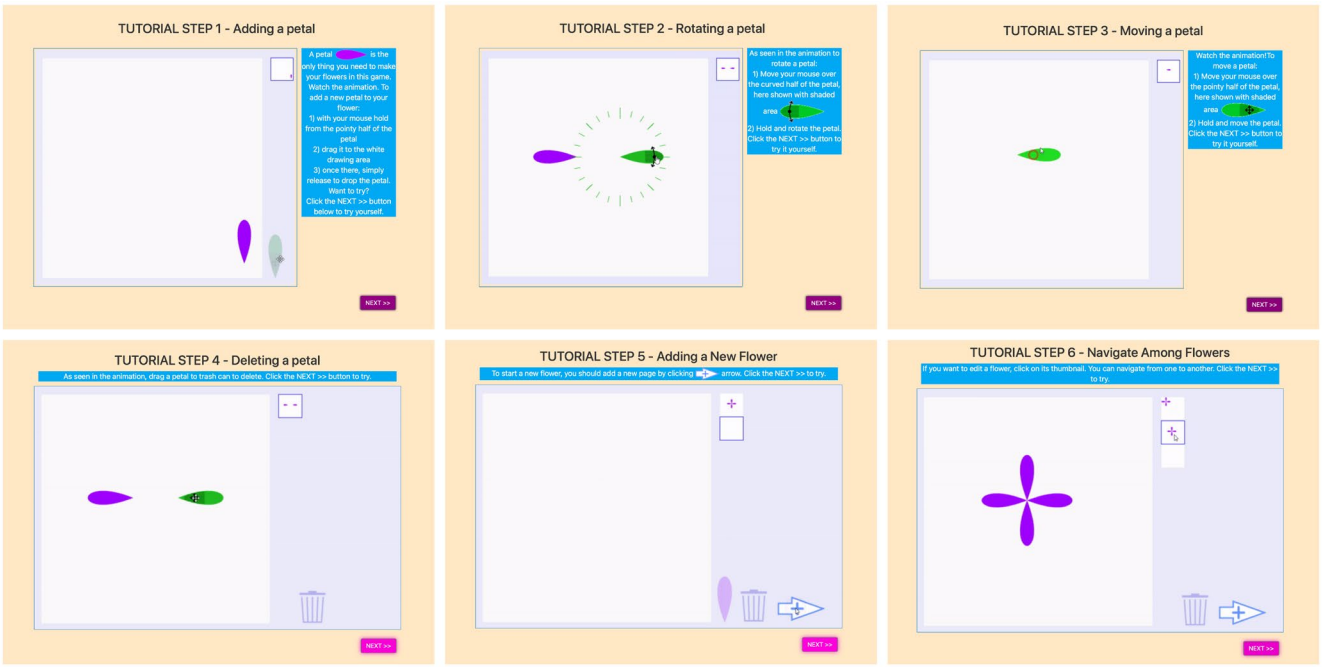



**FIGURE 1** | The user interface and tutorial steps of the CTC.

different design concepts within a given response is counted as if they belong to different responses. Originality focuses on the unusual and rarely produced designs based on past designs produced by other respondents. Here, each meaningful response could be scored as 0, 1, or 2 depending on statistical infrequency within the pool of 3000 participants' responses. The responses provided by more than 65% are scored as "0", those provided by between 64% and 10% received "1" and those provided by $<10\%$ received "2" points. Finally, elaboration was scored based on the number of petals used (each petal is awarded with 1 point), enhanced organization by positioning the petals in fine-tuned locations within the canvas (half a point per petal used), and the use of angular rotation of the petals (1 point added per rotation). Flexibility, originality, and elaboration scores were obtained at the individual level by a summative score across individual responses. Appendix A provides examples of high- and low-scoring hypothetical drawings generated in response to a different prompt.

The present dataset was scored by the developer of CTC, but to examine the inter-judge reliability, the developer of CTC has trained four other judges who scored designs from 87 participants following a 5-h long training. A two-way mixed effects intraclass correlation with five ratings showed a high level of consistency among the judges for fluency, ICC (3.5) = 0.96, 95% [0.94; 0.97]; flexibility, ICC (3.5) = 0.92, 95% [0.88; 0.94]; originality, ICC (3.5) = 0.84, 95% [0.78; 0.89]; and elaboration, ICC (3.5) = 0.94, 95% [0.92; 96].

**2.1.2.2 | Raven Standard Progressive Matrices-Plus.** Seventy-seven participants in the sample also completed Raven's Standard Progressive Matrices Plus (SPM; Raven 2000). The SPM is a measure of non-verbal, multiple-choice intelligence test aimed at measuring abstract reasoning and fluid intelligence. It uses 60 visual patterns with completion tasks presented in a $5 \times 5$ matrix format. Each matrix has a missing piece that the test-taker must identify from a set of options. The test presents items in five sets with 12 items in each with an increasing level of difficulty. Each correct answer scores 1 point. The test is used to assess general ability and as a screening tool for potential giftedness. SPM appears to have a one-factor instrument (Van der Ven and Ellis 2000), strong internal reliability (Burke and Bingham 1969) and a strong correlation with the Wechsler Adult Intelligence Scale (Burke 1972).

**2.1.2.3 | Creative Self-Efficacy.** To measure students' perception of their creative self-efficacy, we administered Creative Self-Efficacy (CSE; Tierney and Farmer 2002) scale to 8542 of the respondents. This tool includes three statements, and respondents rate their agreement on a 7-point scale ranging from 1 (indicating "very strongly disagree") to 7 (indicating "very strongly agree"). An example statement is, "I feel that I am good at generating novel ideas." In our sample, internal reliability was, $\omega = 0.893$ and 0.755 in Study 1 and Study 2 samples, respectively.

### 2.1.3 | Procedures

The CTC was administered following the receipt of parental consent forms. Before they started the test, students were told the following: "In this game, you will design flowers by simply using your computer mouse. It is fun and easy! Let's learn how to play this game. I will show you first and then you will try." Respondents completed the tests in supervised group settings in their respective schools, using personal computers provided by the school in a computer lab or a regular classroom environment with a laptop. All the designs were scored by the developer of CTC, and a brief training session was conducted

before other raters scored a subset of designs for inter-judge reliability.

#### 2.1.4 | Analytical Approach

Because we employed summative aggregation for flexibility, originality, and elaboration scores, we examined fluency confound on these scores using the delta method (Agresti 2002) as Forthmann et al. (2020) suggested. In this method, estimated correlation values of fluency with flexibility, originality, and elaboration scores are obtained based on the descriptive statistics of these four indices and then compared to the corresponding correlations of the observed scores. The magnitude of the difference between the observed and estimated scores is tested using a $z$-test. Fluency confound is evident when observed and estimated scores are not different from each other. We examined internal consistency based on McDonald's (1999) omega ($\omega$). We conducted confirmatory factor analyses using JASP 0.16.4. Model fit was assessed based on root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit indices (CFI), and the Tucker–Lewis index (TLI). To assess model fit, we used 0.10 as the cutoff index for the RMSEA and SRMR, and 0.90 for CFI and TLI (Browne and Cudeck 1993; Kline 2011). Regarding measurement invariance, we used 0.010 for ΔCFI, 0.015 for ΔRMSEA, and 0.030 for ΔSRMR (Chen 2007).

### 2.2 | Results

#### 2.2.1 | Preliminary Analyses

We assessed the fluency confound on flexibility, originality, and elaboration. The delta method analyses showed that estimated correlation values were significantly different from the observed correlation values for flexibility, $z = 4.53$, $p < 0.001$, originality, $z = 8.46$, $p < 0.001$, and elaboration, $z = -27.99$, $p < 0.001$ (See Table 1 for the descriptive statistics and the correlation matrix). These results indicate that flexibility, originality, and elaboration scores are not artifactual despite the strong correlations among them.

Due to the non-normal distribution of the data, we initially employed a log10 transformation. In Table 1, we present the descriptive statistics of the transformed data. While this procedure mitigated the issue, it did not fully resolve it. This consideration informed our choice of the model parameter estimator, discussed later. Pearson correlations (see Table 1) among the scoring indices range between 0.454 (fluency-originality) and 0.686 (fluency-flexibility).

#### 2.2.2 | Confirmatory Factor Analysis

Given that CTC is designed as a single-factor test, we conducted a confirmatory factor analysis in which Fluency, Flexibility, Originality, and Elaboration serve as predictors of the creative potential that CTC is intended to measure. Because of the skewed distribution of the scores, we opted for the Diagonally Weighted Least Squares (DWLS) estimator (Baghdarnia et al. 2014; Jöreskog 2001; Mindrila 2010). We also examined the squared multiple correlation ($R^2$) for potential multicollinearity. They ranged between 0.560 and 0.709.

Our single-factor model (see Figure 2) demonstrated an overall good fit [$\chi^2$ (2) = 182.41, $p < 0.001$; CFI = 0.990, TLI = 0.969; RMSEA = 0.091, SRMR = 0.049]. Factor loadings (see Figure 2) ranged between 0.71 and 84. Additionally, this single-factor model exhibited good internal consistency reliability ($\omega = 0.833$) and composite reliability (Raykov 1997) was also high: 0.858. The average variance extracted was 0.603. Table 2 presents the details of this model.

**2.2.2.1 | Measurement Invariance.** Subsequently, we examined measurement invariance across different grade levels. For configural measurement invariance, the model fit was strong [$\chi^2$ (16) = 166.55, $p < 0.001$; CFI = 0.991, TLI = 0.972, RMSEA = 0.083, SRMR = 0.048]. When assessing metric invariance, the model fit remained favorable [$\chi^2(37) = 332.59, p < 0.001$; CFI = 0.982, TLI = 0.977. RMSEA = 0.077, SRMR = 0.061]. However, scalar invariance [$\chi^2$ (58) = 937.54, $p < 0.001$; CFI = 0.946,

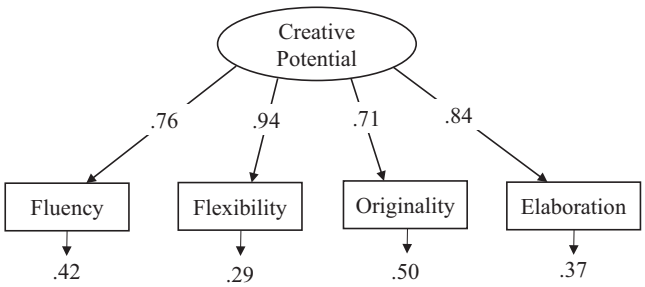



**FIGURE 2** | One factor model of Cebeci Test of Creativity (Study 1—grades 1–8).

**TABLE 1** | Descriptive statistics and bivariate correlations of the CTC scores ($N = 10{,}899$).

| | $M_{raw}$ | $SD_{raw}$ | 1 | 2 | 3 | 4 | $M_{log}$ | $SD_{log}$ |
|---|---|---|---|---|---|---|---|---|
| 1. Fluency | 5.31 | 2.08 | | 0.686 | 0.454 | 0.673 | 0.68 | 0.23 |
| 2. Flexibility | 3.44 | 1.84 | 0.660 | | 0.632 | 0.635 | 0.61 | 0.18 |
| 3. Originality | 3.23 | 3.20 | 0.450 | 0.635 | | 0.597 | 0.50 | 0.34 |
| 4. Elaboration | 78.52 | 47.54 | 0.561 | 0.554 | 0.616 | | 2.63 | 0.35 |

*Note:* $ps < 0.01$. Correlational analyses used log10 transformed indices. Lower diagonal for correlation matrix of raw scores and upper diagonal for transformed scores.

**TABLE 2** | Unstandardized parameter estimates of the unidimensional CTC model.

| | Indicator | | | Metric | | | Scalar | | | Strict | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | *Z* | Estimate | SE | *Z* | Estimate | SE | *Z* | Estimate | SE | *Z* |
| Fluency | 0.170 | 0.002 | 76.42 | 0.168 | 0.003 | 64.72 | 0.165 | 0.002 | 66.400 | −0.168 | 0.002 | −70.281 |
| Flexibility | 0.149 | 0.002 | 85.77 | 0.151 | 0.002 | 69.24 | 0.141 | 0.002 | 71.166 | −0.141 | 0.002 | −76.091 |
| Originality | 0.242 | 0.003 | 85.50 | 0.223 | 0.003 | 68.209 | 0.230 | 0.003 | 70.978 | −0.229 | 0.003 | −75.347 |
| Elaboration | 0.251 | 0.003 | 79.75 | 0.231 | 0.003 | 65.879 | 0.246 | 0.004 | 69.989 | −0.251 | 0.003 | −74.251 |

*Note*: *ps* < 0.001.

**TABLE 3** | Unstandardized parameter estimates of the unidimensional CTC model by grade (configural).

| Factor | Indicator | Estimate | SE | *z* |
|---|---|---|---|---|
| Grade 1 | Fluency | 0.219 | 0.006 | 37.376 |
| | Flexibility | 0.156 | 0.004 | 38.517 |
| | Originality | 0.194 | 0.006 | 34.866 |
| | Elaboration | 0.297 | 0.008 | 36.907 |
| Grade 2 | Fluency | 0.142 | 0.009 | 16.154 |
| | Flexibility | 0.147 | 0.008 | 18.990 |
| | Originality | 0.223 | 0.012 | 19.106 |
| | Elaboration | 0.258 | 0.015 | 17.631 |
| Grade 3 | Fluency | 0.179 | 0.005 | 34.946 |
| | Flexibility | 0.158 | 0.004 | 38.645 |
| | Originality | 0.234 | 0.006 | 38.097 |
| | Elaboration | 0.231 | 0.007 | 35.157 |
| Grade 4 | Fluency | 0.160 | 0.005 | 32.993 |
| | Flexibility | 0.153 | 0.004 | 37.312 |
| | Originality | 0.236 | 0.006 | 36.924 |
| | Elaboration | 0.222 | 0.006 | 34.152 |
| Grade 5 | Fluency | 0.143 | 0.005 | 29.707 |
| | Flexibility | 0.148 | 0.004 | 34.575 |
| | Originality | 0.234 | 0.007 | 34.119 |
| | Elaboration | 0.196 | 0.006 | 30.687 |
| Grade 6 | Fluency | 0.147 | 0.009 | 15.810 |
| | Flexibility | 0.133 | 0.007 | 18.355 |
| | Originality | 0.224 | 0.012 | 18.334 |
| | Elaboration | 0.218 | 0.013 | 16.587 |
| Grade 7 | Fluency | 0.159 | 0.01 | 15.589 |
| | Flexibility | 0.139 | 0.008 | 17.132 |
| | Originality | 0.230 | 0.014 | 17.055 |
| | Elaboration | 0.191 | 0.012 | 15.549 |
| Grade 8 | Fluency | 0.125 | 0.007 | 17.528 |
| | Flexibility | 0.132 | 0.006 | 20.957 |
| | Originality | 0.215 | 0.010 | 20.557 |
| | Elaboration | 0.187 | 0.010 | 18.939 |

TLI = 0.956, RMSEA = 0.106, SRMR = 0.079] and strict invariance [$\chi^2$ (86) = 1144.81, $p < 0.001$; CFI = 0.935, TLI = 0.964, RMSEA = 0.095, SRMR = 0.093] were violated. The results of the configural model are presented in Table 3. The unstandardized parameter estimates were the same for metric, scalar, and strict measurement invariance models, and they are presented in Table 2.

#### 2.2.3 | External Validity

We tested the external validity of the CTC with SPM and CSE. When CSE was added to the model, model fit was good ([$\chi^2$ (13) = 213.948, $p < 0.001$; CFI = 0.994, TLI = 0.990, RMSEA = 0.043, SRMR = 0.033]). The regression path for CSE ➔ CTC was $b = 0.073$, SE = 0.005, $z = 15.570$, $p < 0.001$. When SPM was added into the model, model fit was poor ([$\chi^2$ (10) = 32.394, $p < 0.001$; CFI = 0.670, TLI = 0.505, RMSEA = 0.172, SRMR = 0.220]) and the regression path was not significant, $b = 0.082$, SE = 0.078, $z = 15.570$, $p < 0.001$.

## 3 | Study 2

In Study 2, we replicated the same analyses with a new sample of elementary school students. To mitigate the non-normal distribution, we presented the most common three responses and instructed participants to generate flower designs that were different from these. This strategy is consistent with the recent findings underlining the importance of explicit instructions in divergent thinking tasks (Acar et al. 2020; Said-Metwaly et al. 2021). Although Study 1 used explicit instructions, the presentation of the most common designs aimed to further emphasize originality. Specifically, we aimed to emphasize the objective of designing original flower designs by presenting unoriginal examples by going beyond the written-verbal instructions. By using these instructions, we attempted to divert students' thinking from the path of least resistance (Ward and Kolomyts 2019), which is often responsible for common, familiar, and easily accessible (and thus typically unoriginal) design ideas. Besides the goal of addressing non-normality, we expected this instructional change to increase original designs with potential decrements in fluency. Importantly, we kept our emphasis on fluency in our Study 2 instructions. In a similar manner to Study 1, students also completed the Creative Self-Efficacy survey (Tierney and Farmer 2002) as part of external validation. Additionally, we examined the factor structure of CTC when the fluency confound is totally removed by using average originality, flexibility, and elaboration scores (Runco and Acar 2012).

### 3.1 | Methods

#### 3.1.1 | Participants and Instruments

Our sample included 3967 third, fourth, and fifth graders. After removing the tests with no meaningful designs, the final analytical sample was 3923 with 1282 third graders, 1334 fourth graders, and 1307 fifth graders. Data were collected from 26 schools in 10 different US states: Colorado, Connecticut, District of Columbia, Florida, New Jersey, New York, Ohio, Pennsylvania, and Washington. The majority of the data were collected in New Jersey (96.1%). Data collection took place in groups. Besides the CTC, all participants also completed the Creative Self-Efficacy (Tierney and Farmer 2002).

### 3.2 | Results

#### 3.2.1 | Preliminary Analyses

Using similar techniques in Study 1, we assessed the fluency confound on flexibility, originality, and elaboration scores using the delta test (Agresti, 2002; Forthmann et al. 2020). The analyses showed that estimated correlation values were significantly different from the observed correlation values for flexibility, $z = 5.23$, $p < 0.001$, originality, $z = 3.81$, $p < 0.001$, and elaboration, $z = -16.94$, $p < 0.001$.

Like our Study 1 results, non-normality was still an issue. We again used log10 transformation, which improved the distribution but there was still non-normality. We conducted the analyses with these transformed scores. As presented in Table 4, the largest correlation was between fluency and flexibility ($r = 0.713$) and the smallest was between elaboration and originality ($r = 0.619$).

#### 3.2.2 | Confirmatory Factor Analysis

We tested the same one-factor model using the DWLS estimator (Baghdarnia et al. 2014; Jöreskog 2001; Mindrila 2010). The single-factor model (see Figure 3 and Table 5) demonstrated an overall good fit [$\chi^2$ (2) = 58.95, $p < 0.001$; CFI = 0.991, TLI = 0.974; RMSEA = 0.085, SRMR = 0.039]. Additionally, this single-factor model exhibited good reliability ($\omega = 0.861$). Construct reliability was 0.872, and the average variance extracted value was 0.632. $R^2$ ranged between 0.528 and 0.789.

**3.2.2.1 | Measurement Invariance.** We examined measurement invariance across different grade levels in this sample, too. The model fit was strong for configural [$\chi^2$ (6) = 60.502, $p < 0.001$; CFI = 0.991, TLI = 0.973, RMSEA = 0.083, SRMR = 0.048], metric [$\chi^2$ (12) = 77.231, $p < 0.001$; CFI = 0.989, TLI = 0.984, RMSEA = 0.064, SRMR = 0.053], and scalar invariance [$\chi^2$ (18) = 134.459, $p < 0.001$; CFI = 0.981, TLI = 0.981, RMSEA = 0.070, SRMR = 0.055] whereas strict measurement

**TABLE 4** | Descriptive statistics and bivariate correlations of the CTC indices with third, fourth, and fifth graders ($N = 3923$).

| | $M_{raw}$ | $SD_{raw}$ | 1 | 2 | 3 | 4 | $M_{log}$ | $SD_{log}$ |
|---|---|---|---|---|---|---|---|---|
| 1. Fluency | 5.01 | 1.94 | | 0.713 | 0.460 | 0.638 | 0.66 | 0.22 |
| 2. Flexibility | 3.35 | 1.65 | 0.683 | | 0.681 | 0.700 | 0.61 | 0.17 |
| 3. Originality | 3.14 | 2.95 | 0.465 | 0.677 | | 0.619 | 0.50 | 0.32 |
| 4. Elaboration | 70.75 | 41.66 | 0.542 | 0.621 | 0.699 | | 1.76 | 0.33 |

*Note:* $ps < 0.01$. Correlational analyses used log10 transformed indices. Lower diagonal for the correlation matrix of raw scores and the upper diagonal for transformed scores.

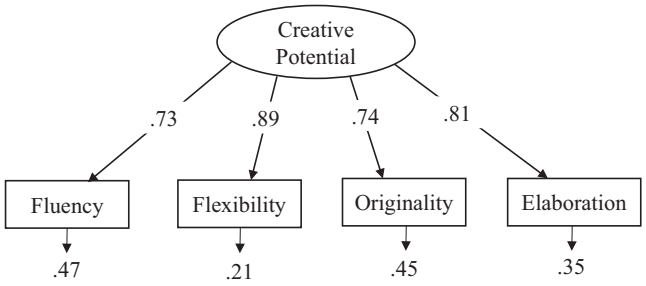



**FIGURE 3** | One factor model of Cebeci Test of Creativity (Study 2—grades 3–5).

invariance [$\chi^2(26) = 152.556$, $p < 0.001$; CFI = 0.979, TLI = 0.986, RMSEA = 0.061, SRMR = 0.063] was violated due to changes in the fit indices. The estimates for metric, scalar, and strict measurement invariance were presented in Table 5. Table 6 presents the unstandardized parameter estimates for the configural measurement invariance.

**3.2.2.2 | External Validity.** We tested the external validity of the CTC with CSE in Study 2 as well. The model fit was good ([$\chi^2$ (17) = 75.456, $p < 0.001$; CFI = 0.996, TLI = 0.995, RMSEA = 0.030, SRMR = 0.032]). The regression path for CSE ➔ CTC was $b = 0.298$, SE = 0.010, $z = 31.272$, $p < 0.001$. Data from the Raven test was not available in this sample.

**3.2.2.2.1 | Fluency Confound and the Factor Structure.** While CTC scores appear to be distinguished from fluency scores based on the delta test, fluency scores are still involved with originality, flexibility, and elaboration scores as summative aggregation is employed to individual responses to obtain scores for each participant. Thus, we examined the factor structure of CTC by using average scores for originality, flexibility, and elaboration. Fluency was negatively correlated with average originality ($r = -0.235$), average flexibility ($r = -0.667$), and elaboration ($r = -0.179$). As a result, the reliability of the composite scale with these average scores was low ($\alpha = 0.272$) and the CFA model with these four indices did not converge. Thus, we explored the possibility of a factor structure that is solely based on average scores (without fluency included in the exploratory factor analysis) for a factor that reflects "creative quality." The internal consistency reliability of the three average scores was good ($\alpha = 0.716$). Kaiser-Meyer-Olkin measure of sampling adequacy was 0.646 and Bartlett's test of specificity was significant ($\chi^2(3) = 2845.67$, $p < 0.001$), showing the correlations among the average scores were substantive: $r$s = 0.563, 0.533, and 0.371. Of these correlations, the largest value was between average flexibility and average originality and the lowest between average flexibility and average elaboration scores. The unidimensional factor explained 66% of the variance.

## 4 | Discussion

This is the first study to examine the psychometric properties of the new CTC. Across two different studies using a large sample size, a unidimensional factor structure was supported by the data. We observed measurement invariance across both studies.

**TABLE 5** | Unstandardized parameter estimates of the unidimensional CTC model in study 2 sample.

| | Indicator | | | Metric | | | Scalar | | | Strict | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | Z | Estimate | SE | Z | Estimate | SE | Z | Estimate | SE | Z |
| Fluency | 0.158 | 0.003 | 45.725 | 0.157 | 0.004 | 40.935 | 0.153 | 0.004 | 41.566 | 0.154 | 0.004 | 42.737 |
| Flexibility | 0.144 | 0.003 | 53.079 | 0.141 | 0.003 | 46.058 | 0.138 | 0.003 | 47.337 | −0.251 | 0.003 | 49.028 |
| Originality | 0.238 | 0.004 | 53.019 | 0.228 | 0.005 | 45.275 | 0.230 | 0.005 | 46.525 | 0.230 | 0.005 | 48.296 |
| Elaboration | 0.225 | 0.005 | 49.627 | 0.212 | 0.005 | 43.574 | 0.230 | 0.003 | 45.617 | −0.229 | 0.005 | 46.977 |

*Note:* $ps < 0.001$.

**TABLE 6** | Unstandardized parameter estimates of the unidimensional CTC model in study 2 sample by grade.

| | | Configural | | |
|---|---|---|---|---|
| **Factor** | **Indicator** | **Estimate** | **SE** | ***z*** |
| Grade 3 | Fluency | 0.171 | 0.006 | 27.597 |
| | Flexibility | 0.148 | 0.005 | 30.578 |
| | Originality | 0.215 | 0.007 | 29.611 |
| | Elaboration | 0.208 | 0.007 | 27.809 |
| Grade 4 | Fluency | 0.151 | 0.006 | 24.907 |
| | Flexibility | 0.139 | 0.005 | 29.191 |
| | Originality | 0.226 | 0.008 | 29.151 |
| | Elaboration | 0.221 | 0.008 | 27.268 |
| Grade 5 | Fluency | 0.149 | 0.006 | 24.254 |
| | Flexibility | 0.137 | 0.005 | 28.620 |
| | Originality | 0.248 | 0.009 | 28.946 |
| | Elaboration | 0.209 | 0.008 | 27.008 |

CTC had a strong factor-based reliability and correlated with creative self-efficacy, whereas the correlation with fluid intelligence was not significant.

The unidimensional structure lies in the idea of creative potential, typically associated with divergent thinking (Runco and Acar 2012) and a creative personality (Davis 1989). Accordingly, individuals with a high creative potential are defined as those that produce many different unusual and elegant ideas, solutions, or problems, mirroring the concepts of fluency, flexibility, originality, and elaboration. The four concepts are tied to creative personality (Harrington 1975) in reflecting an inclination to keep an open mind (An et al. 2016; Grajzel et al. 2023; McCrae 1987), resist the temptation to reach quick and easily accessible solutions, autonomy, and independence (Runco 1992). The connection between creative personality and creative divergent thinking was theorized by the seminal work of Guilford (1950) who argued that creative abilities refer to the skills that creative individuals exhibit; thus, the abilities reflect the creative person's characteristics. Guilford then extended this argument to more specific skills such as sensitivity to problems, fluency, novelty, flexibility, synthesizing, reorganization, and redefinition, complexity, and evaluation. Divergent thinking tests have traditionally been used to operationalize these skills, and the CTC focuses on the most often used four of them. As predictors of the same latent construct, creative potential, we expected a unidimensional structure, and it was supported. On the other hand, the bivariate correlations of the individual indices were strong yet not confounded with each other. Correlations ranged between 0.454 and 0.686 in Study 1 and 0.619 and 0.713 in Study 2. Further, the delta method showed that summative originality, flexibility, and elaboration scores were not mere reflections of fluency scores. This is important because previous research has indicated that the divergent thinking indices may be confounded with each other (Clark and Mirels 1970; Hocevar 1979; Forthmann et al. 2020) and their correlations may be as high as 0.90 or higher. Such high correlations tend to indicate compromised discriminant validity (Clark and Watson 1995; Kline 2011; Gold et al. 2001). In this study, correlations were not excessively high (supporting discriminant validity) and yet observed correlations were different from estimated correlations (showing fluency confound is not overwhelming).

We also observed a strong factor-based reliability coefficients across the two studies ($\omega = 0.855$ and 0.865). High reliability coefficients are important for both research and practice, and especially in high-stakes uses such as gifted or creative identification. Such uses may sometimes take place by employing multiple tests or criteria such as academic achievement, creativity, and intelligence. These tests may sometimes be used in combination with a conjunctive, complementary, and compensatory models (McBee et al. 2014). The conjunctive use refers to the "and rule" where a high performance on two or more tests is expected to result in eligibility for gifted programs, whereas the "or rule" requires high performance on at least one of the measures used. The feasibility of the conjunctive rule requires a substantive correlation among the measures used to be able to identify a sufficiently large pool of students who performed high on both tests. This makes our path coefficients with SPM ($b = 0.08$) particularly important because a weaker correlation eliminates the option of applying the conjunctive rule (McBee et al. 2014). Based on the regression coefficient, which was not statistically significant, CTC should not be used along fluid intelligence tests because they seem to identify a different set of abilities.

CTC was significantly correlated with self-reported creative self-efficacy in both studies ($b$s = 0.073 and 0.298). Although significant, this small correlation is noteworthy and can be related to several factors. The first factor is that the self-reported measure of creative self-efficacy might be confounded by a general perception of self-efficacy that may have been projected onto creativity. Furthermore, factors such as meta-cognition, self-awareness, self-presentation bias, and an accurate sense of self may have been heavily involved in self-report measures, regardless of the target construct they are purported to measure (Paulus and Renzuli 1968). These limitations may be more pronounced in our study samples that involve elementary and middle school students, who may have a more transient and unstable view of themselves.

Both studies provided evidence of measurement invariance by grade, but this was stronger in Study 2 than in Study 1. In Study 1, we found evidence for configural and metric invariance, but in Study 2, we found evidence for configural, metric, and scalar (but not strict measurement invariance). This difference is probably due to the larger range of grades involved in Study 1 (i.e., grades 1–8) than in Study 2 (i.e., grades 3–5). This is understandable because the larger the grade difference, the more likely that developmental factors take part in creative performance and ideational skills. Considered together, these findings indicate that the unidimensional factor structure of CTC is applicable to all grade levels and factor loadings tend to be very similar. Furthermore, the item intercepts and residual variances are also quite similar across grades 3–5. Therefore, scores from CTC can be compared within and across grade levels in grades 3–5, but not across the grade levels beyond fifth grade. This finding

implies that generating grade-based norms and standardized scores is a proper and needed strategy to use CTC in a school context beyond sixth grade.

The intent behind the CTC is to provide a psychometrically robust test of creativity that is easy to administer in a school setting, either individually or in groups. Years of data collection have shown that the CTC has high usability, which is important because creativity is a highly desirable skill in the workforce, and educators need to consider how to develop it (Puccio 2017). Additionally, creativity holds a key place in the 21st Century Skills Framework (Partnership for 21st Century Skills 2008) and has been elevated to the highest level of thinking in Bloom's revised taxonomy (Anderson et al. 2001). However, educational practices are not on par with these developments, partly because creativity is not an explicit part of teacher education programs. Recently, The Organization for Economic Co-operation and Development's Programme for International Student Assessment (PISA; OECD 2023) has incorporated creativity into their assessment framework. This is a step forward toward a more creative education because one cannot control what they do not measure (Cruz-Cázares et al. 2013). The inclusion of creativity tests in educational assessment is key to measuring and developing creativity and incorporating it into instruction. The availability of tests such as the CTC fills this gap because large-scale assessment is a major challenge for open-ended tasks, and creativity tests are often open-ended. The ethical and responsible use of technology in the administration and scoring of these tasks can potentially transform education and reconsider educational priorities and practices.

### 4.1 | Limitations and Future Directions

In this first study on the validation of the CTC, there were time-related and logistical limitations to administer a larger number of instruments to expand the scope of validation. Therefore, future studies should aim to extend the validation evidence to other measures of creativity, particularly other divergent thinking tests, creative achievement measures, and product development tasks. Additionally, the current study used manual scoring of the responses, but efforts are underway to develop automated scoring based on artificial intelligence methods. Once implemented, this approach will enable the rapid and objective scoring of large volumes of data (e.g., flower designs) without the need for human judgment. Importantly, the continuing ratings of all existing data can be used to train the artificial intelligence models to increase their precision. Last, we do not have evidence on the relationship between CTC and academic achievement. Given that the CTC was primarily developed for K-12 settings and may be used for gifted identification, which may involve academic achievement scores, correlations of CTC with both grade point average and standardized test scores would provide a more complete picture of its nomological network.

The CTC applies a divergent thinking framework to graphic activities, but creativity with figural activities involves a broader set of abilities that can be considered in future research. Examples of these skills are utilized in other tests such as convergent and integrative abilities in Evaluation of Potential for Creativity (Lubart et al. 2011), resistance to premature closure, and extending and breaking the boundaries in Torrance's figural test (Torrance 1998), or perspective and humor in Urban's The Test for Creative Thinking—Drawing Production.

The interpretation of CTC should consider one key point: fluency remains a significant factor influencing the other three scores of the CTC, due to the summative aggregation of individual responses at the participant level. This is evident from the negative correlations between fluency and average originality, flexibility, and elaboration scores. These correlations help explain why the CFA model, which used these average scores, did not converge and why internal consistency reliability was low. As a result, the unidimensional CTC factor is primarily driven by fluency scores, while the originality, flexibility, and elaboration scores, though not artifactual as the delta test indicated, are still influenced by fluency. This finding has similarities to the structure of the TTCT Verbal, which also follows the same approach to score aggregation (Forthmann et al. 2020; Reiter-Palmon et al. 2019; Torrance 1998).

Although our scoring approach aligns with the notion that quantity breeds quality (Forthmann et al. 2025; Osborn 1963; Simonton 2004), the negative correlations between fluency and the average scores reflect the trade-off relationship, suggesting that a greater quantity of responses often results in lower quality for each individual response (Guilford 1968). This provides an important context for our findings on the factor structure of creative quality. When fluency bias is fully removed, average scores can still serve as a useful measure of creative quality. In relation to this, the CTC's test structure can be considered in future versions. CTC does not impose a fluency limit, and changes to the test structure by setting such a limit—like Activity 2 of the TTCT Figural—could impact the quality scores (Acar 2023; Acar, Dumas, et al. 2024; Acar, Lee, et al. 2024; Acar, Organisciak, et al. 2024; Zarnegar et al. 1988). Future research could investigate this alternative test structure for CTC.

## 5 | Conclusion

Overall, the findings across these two studies indicate that CTC is a reliable single-factor test. This single-factor structure seems to apply to all grade levels, satisfying configural measurement invariance and that factor loadings largely overlap across the grade levels (metric measurement invariance). Whereas there are differences in the intercepts of the observed variables in Study 1, we found support for scalar measurement variance, as well as in Study 2. Overall, our evidence has found that CTC is a valuable tool for assessing creativity in educational settings. It is efficient, scalable, and provides a comprehensive assessment of four key dimensions of creativity.

**Acknowledgments**

The authors gratefully acknowledge Dr. Sally Reis for her invaluable contributions in editing the manuscript and providing insightful feedback that significantly improved the quality of this work.

## Conflicts of Interest

The first author is also the developer of the Cebeci Test of Creativity, and the second author served as a consultant for this validation study.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References


Acar, S. 2023. "Does the Task Structure Impact the Fluency Confound in Divergent Thinking? An Investigation With TTCT-Figural." *Creativity Research Journal* 35, no. 1: 1–14. https://doi.org/10.1080/10400419.2022.2044656.

Acar, S., A. M. A. Alabbasi, M. A. Runco, and K. Beketayev. 2019. "Latency as a Predictor of Originality in Divergent Thinking." *Thinking Skills and Creativity* 33: 100574. https://doi.org/10.1016/j.tsc.2019.100574.

Acar, S., K. Berthiaume, K. Grajzel, D. Dumas, C. T. Flemister, and P. Organisciak. 2023. "Applying Automated Originality Scoring to the Verbal Form of Torrance Tests of Creative Thinking." *Gifted Child Quarterly* 67, no. 1: 3–17. https://doi.org/10.1177/00169862211061874.

Acar, S., D. Dumas, P. Organisciak, and K. Berthiaume. 2024. "Measuring Original Thinking in Elementary School: Development and Validation of a Computational Psychometric Approach." *Journal of Education & Psychology* 116: 953–981. https://doi.org/10.1037/edu0000844.

Acar, S., L. E. Lee, and R. Scherer. 2024. "A Reliability Generalization of the Torrance Tests of Creative Thinking-Figural." *European Journal of Psychological Assessment* 40, no. 5: 396–411. https://doi.org/10.1027/1015-5759/a000819.

Acar, S., U. Ogurlu, and A. Zorychta. 2023. "Exploration of Discriminant Validity in Divergent Thinking Tasks: A Meta-Analysis." *Psychology of Aesthetics, Creativity, and the Arts* 17, no. 6: 705–724. https://doi.org/10.1037/aca0000469.

Acar, S., P. Organisciak, and D. Dumas. 2024. "Automated Scoring of Figural Tests of Creativity With Computer Vision." *Journal of Creative Behavior* 59, no. 1: e677. https://doi.org/10.1002/jocb.677.

Acar, S., and M. A. Runco. 2014. "Assessing Associative Distance Among Ideas Elicited by Tests of Divergent Thinking." *Creativity Research Journal* 26, no. 2: 229–238. https://doi.org/10.1080/10400419.2014.901095.

Acar, S., M. A. Runco, and H. Park. 2020. "What Should People Be Told When They Take a Divergent Thinking Test? A Meta-Analytic Review of Explicit Instructions for Divergent Thinking." *Psychology of Aesthetics, Creativity, and the Arts* 14, no. 1: 39–49. https://doi.org/10.1037/aca0000256.

Agresti, A. 2002. *Categorical Data Analysis*. Wiley.

An, D., Y. Song, and M. Carr. 2016. "A Comparison of Two Models of Creativity: Divergent Thinking and Creative Expert Performance." *Personality and Individual Differences* 90: 78–84. https://doi.org/10.1016/j.paid.2015.10.040.

Anderson, L. W., D. R. Krathwohl, P. W. Airasian, et al. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives (Complete Edition)*. Longman.

Backes, B., and J. Cowan. 2019. "Is the Pen Mightier Than the Keyboard? The Effect of Online Testing on Measured Student Achievement." *Economics of Education Review* 68: 89–103. https://doi.org/10.1016/j.econedurev.2018.12.007.

Baghdarnia, M., R. F. Soreh, and R. Gorji. 2014. "The Comparison of Two Methods of Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) in Testing Construct Validity of Achievement Goals." *Journal of Educational and Management Studies* 4, no. 1: 22–38.

Barbot, B. 2018. "The Dynamics of Creative Ideation: Introducing a New Assessment Paradigm." *Frontiers in Psychology* 9: 2529. https://doi.org/10.3389/fpsyg.2018.02529.

Beaty, R. E., D. R. Johnson, D. C. Zeitlen, and B. Forthmann. 2022. "Semantic Distance and the Alternate Uses Task: Recommendations for Reliable Automated Assessment of Originality." *Creativity Research Journal* 34, no. 3: 1–16. https://doi.org/10.1080/10400419.2022.2025720.

Browne, M. W., and R. Cudeck. 1993. "Alternative Ways of Assessing Model Fit." In *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long, 136–162. Sage.

Bump, W. M. 1994. "A Comparative Analysis of Spelling Skills Utilizing a Computer-Based Spelling Assessment Instrument." Unpublished doctoral dissertation, Texas A&M University.

Burke, H. R. 1972. "Raven's Progressive Matrices: Validity, Reliability, and Norms." *Journal of Psychology* 82, no. 2: 253–257. https://doi.org/10.1080/00223980.1972.9923815.

Burke, H. R., and W. C. Bingham. 1969. "Raven's Progressive Matrices: More on Construct Validity." *Journal of Psychology* 72, no. 2: 247–251.

Chen, F. F. 2007. "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Structural Equation Modeling: A Multidisciplinary Journal* 14, no. 3: 464–504. https://doi.org/10.1080/10705510701301834.

Christensen, P. R., J. P. Guilford, and R. C. Wilson. 1957. "Relations of Creative Responses to Working Time and Instructions." *Journal of Experimental Psychology* 53, no. 2: 82–88. https://doi.org/10.1037/h0045461.

Clark, L. A., and D. Watson. 1995. "Constructing Validity: Basic Issues in Objective Scale Development." *Psychological Assessment* 7, no. 3: 309–319. https://doi.org/10.1037/1040-3590.7.3.309.

Clark, P. M., and H. L. Mirels. 1970. "Fluency as a Pervasive Element in the Measurement of Creativity." *Journal of Educational Measurement* 7, no. 2: 83–86. https://doi.org/10.1111/j.1745-3984.1970.tb00699.x.

Costa, P. T., Jr., and R. R. McCrae. 1992. *NEO-PI–R Professional Manual*. Psychological Assessment Resources.

Cropley, D. H., and R. L. Marrone. 2025. "Automated Scoring of Figural Creativity Using a Convolutional Neural Network." *Psychology of Aesthetics, Creativity, and the Arts* 19, no. 1: 77–86. https://doi.org/10.1037/aca0000510.

Cruz-Cázares, C., C. Bayona-Sáez, and T. García-Marco. 2013. "You Can't Manage Right What You Can't Measure Well: Technological Innovation Efficiency." *Research Policy* 42, no. 6–7: 1239–1250. https://doi.org/10.1016/j.respol.2013.03.012.

Davis, G. A. 1989. "Testing for Creative Potential." *Contemporary Educational Psychology* 14, no. 3: 257–274. https://doi.org/10.1016/0361-476X(89)90014-3.

Dumas, D., P. Organisciak, and M. Doherty. 2021. "Measuring Divergent Thinking Originality With Human Raters and Text-Mining Models: A Psychometric Comparison of Methods." *Psychology of Aesthetics, Creativity, and the Arts* 15, no. 4: 645–663. https://doi.org/10.1037/aca0000319.

Erwin, J. O., and F. C. Worrell. 2012. "Assessment Practices and the Underrepresentation of Minority Students in Gifted and Talented Education." *Journal of Psychoeducational Assessment* 30, no. 1: 74–87. https://doi.org/10.1177/0734282911428197.

Forthmann, B., C. Szardenings, and H. Holling. 2020. "Understanding the Confounding Effect of Fluency in Divergent Thinking Scores: Revisiting Average Scores to Quantify Artifactual Correlation." *Psychology of Aesthetics, Creativity, and the Arts* 14, no. 1: 94–112. https://doi.org/10.1037/aca0000196.

Forthmann, B., S. Weiss, and B. Goecke. 2025. "A Cognitive Interpretation Is Not at Odds With Equal Odds: A Latent Variable Investigation of Creative Thinking." *Imagination, Cognition and Personality* 44, no. 4: 362–386. https://doi.org/10.1177/02762366241311561.

Gold, A. H., A. Malhotra, and A. H. Segars. 2001. "Knowledge Management: An Organizational Capabilities Perspective." *Journal of Management Information Systems* 18, no. 1: 185–214. https://doi.org/10.1080/07421222.2001.11045669.

Grajzel, K., S. Acar, and G. Singer. 2023. "The Big Five and Divergent Thinking: A Meta-Analysis." *Personality and Individual Differences* 214: 112338. https://doi.org/10.1016/j.paid.2023.112338.

Guilford, J. P. 1950. "Creativity." *American Psychologist* 5, no. 9: 444–454.

Guilford, J. P. 1968. *Intelligence, Creativity, and Their Educational Implications*. Robert R. Knapp.

Guilford, J. P. 1973. *Characteristics of Creativity*. Illinois State Office of the Superintendent of Public Instruction, Gifted Children Section.

Guilford, J. P., P. R. Christensen, P. R. Merrifield, and R. C. Wilson. 1960. *Alternative Uses Manual*. Sheridan Supply Co.

Guo, J. 2019. "Web-Based Creativity Assessment System That Collects Both Verbal and Figural Responses: Its Problems and Potentials." *International Journal of Information and Education Technology* 9, no. 1: 27–34.

Harrington, D. M. 1975. "Effects of Explicit Instructions to "Be Creative" on the Psychological Meaning of Divergent Thinking Test Scores." *Journal of Personality* 43, no. 3: 434–454. https://doi.org/10.1111/j.1467-6494.1975.tb00715.x.

Hocevar, D. 1979. "Ideational Fluency as a Confounding Factor in the Measurement of Originality." *Journal of Educational Psychology* 71, no. 2: 191–196. https://doi.org/10.1037/0022-0663.71.2.191.

Horkay, N., R. E. Bennett, N. Allen, B. Kaplan, and F. Yan. 2006. "Does It Matter if I Take My Writing Test on Computer? An Empirical Study of Mode Effects in NAEP." *Journal of Technology, Learning, and Assessment* 5, no. 2: n2. https://doi.org/10.1002/j.2162-6057.1998.tb00809.x.

Jäger, A. O., H. Holling, F. Preckel, et al. 2005. "Berliner Intelligenzstruktur-Test Für Jugendliche: Begabungs- Und Hochbegabungsdiagnostik (BIS-HB) [Berlin Structure-of-Intelligence Test for Youth: Diagnosis of Talents and Giftedness]." Hogrefe.

Jöreskog, K. 2001. "Analysis of Ordinal Variables 2: Cross-Sectional Data." Unpublished Manuscript. http://www.ssicentral.com/lisrel/ordinal.htm.

Kettner, N. W., J. P. Guilford, and P. R. Christensen. 1959. "A Factor-Analytic Study Across the Domains of Reasoning, Creativity, and Evaluation." *Psychological Monographs: General and Applied* 73, no. 9: 1–31. https://doi.org/10.1037/h0093745.

Kim, K. H. 2017. "The Torrance Tests of Creative Thinking-Figural or Verbal: Which One Should We Use?" *Creativity. Theories–Research-Applications* 4, no. 2: 302–321. https://doi.org/10.1515/ctra-2017-0015.

Kline, R. B. 2011. *Principles and Practice of Structural Equation Modeling*. Guilford Press.

Kwon, M., E. T. Goetz, and R. D. Zellner. 1998. "Developing a Computer-Based TTCT: Promises and Problems." *Journal of Creative Behavior* 32, no. 2: 96–106.

Kwon, M. C. 1996. *An Exploratory Study of a Computerized Creativity Test: Comparing Paper-Pencil and Computer-Based Versions of the Torrance Tests of Creative Thinking*. Texas A&M University.

Lau, S., and P. C. Cheung. 2010. "Creativity Assessment: Comparability of the Electronic and Paper-And-Pencil Versions of the Wallach–Kogan Creativity Tests." *Thinking Skills and Creativity* 5, no. 3: 101–107. https://doi.org/10.1016/j.tsc.2010.09.004.

Lubart, T. I., M. Besançon, and B. Barbot. 2011. *Evaluation Du Potential Créatif (EPoC)*. Editions Hogrefe France.

Luria, S. R., R. L. O'Brien, and J. C. Kaufman. 2016. "Creativity in Gifted Identification: Increasing Accuracy and Diversity." *Annals of the New York Academy of Sciences* 1377, no. 1: 44–52. https://doi.org/10.1111/nyas.13136.

Mazzeo, J., and A. L. Harvey. 1988. "The Equivalence of Scores From Automated and Conventional Educational and Psychological Tests: A Review of the Literature." *ETS Research Report Series* 1988, no. 1: i–27.

McBee, M. T., S. J. Peters, and C. Waterman. 2014. "Combining Scores in Multiple-Criteria Assessment Systems: The Impact of Combination Rule." *Gifted Child Quarterly* 58, no. 1: 69–89. https://doi.org/10.1177/0016986213513794.

McCrae, R. R. 1987. "Creativity, Divergent Thinking, and Openness to Experience." *Journal of Personality and Social Psychology* 52, no. 6: 1258–1265. https://doi.org/10.1037/0022-3514.52.6.1258.

McDonald, R. P. 1999. *Test Theory: A Unified Treatment*. Erlbaum.

Mindrila, D. 2010. "Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) Estimation Procedures: A Comparison of Estimation Bias With Ordinal and Multivariate Non-Normal Data." *International Journal of Digital Society* 1, no. 1: 60–66.

OECD. 2023. "PISA 2022 Assessment and Analytical Framework." OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en.

Organisciak, P., S. Acar, D. Dumas, and K. Berthiaume. 2023. "Beyond Semantic Distance: Automated Scoring of Divergent Thinking Greatly Improves With Large Language Models." *Thinking Skills and Creativity* 49: 101356. https://doi.org/10.1016/j.tsc.2023.101356.

Osborn, A. F. 1963. *Applied Imagination: Principles and Procedures of Creative Problem-Solving*. Scribner.

Palaniappan, A. K. 2012. "Web-Based Creativity Assessment System." *International Journal of Information and Education Technology* 2, no. 3: 255–258. https://doi.org/10.7763/IJIET.2012.V2.123.

Parnes, S. J. 1961. "Effects of Extended Effort in Creative Problem Solving." *Journal of Educational Psychology* 52, no. 3: 117–122. https://doi.org/10.1037/h0044650.

Partnership for 21st Century Skills. 2008. "21st Century Skills, Education & Competitiveness: A Resource and Policy Guide." https://files.eric.ed.gov/fulltext/ED519337.pdf.

Patterson, J. D., B. Barbot, J. Lloyd-Cox, and R. Beaty. 2022. "AuDrA: An Automated Drawing Assessment Platform for Evaluating Creativity." https://psyarxiv.com/t63dm.

Paulus, D. H., and J. S. Renzuli. 1968. "Scoring Creativity Tests by Computer." *Gifted Child Quarterly* 12, no. 2: 79–83. https://doi.org/10.1177/001698626801200202.

Pender, W. C. 2020. "The Impact of Computer Based Versus Paper Pencil on PARCC Test Results for Illinois Public Elementary Schools (Doctoral Dissertation, University of St. Francis)."

Peters, S. J. 2022. "The Challenges of Achieving Equity Within Public School Gifted and Talented Programs." *Gifted Child Quarterly* 66, no. 2: 82–94. https://doi.org/10.1177/00169862211002535.

Puccio, G. J. 2017. "From the Dawn of Humanity to the 21st Century: Creativity as an Enduring Survival Skill." *Journal of Creative Behavior* 51, no. 4: 330–334. https://doi.org/10.1002/jocb.203.

Raven, J. 2000. "The Raven's Progressive Matrices: Change and Stability Over Culture and Time." *Cognitive Psychology* 41, no. 1: 1–48. https://doi.org/10.1006/cogp.1999.0735.

Raykov, T. 1997. "Estimation of Composite Reliability for Congeneric Measures." *Applied Psychological Measurement* 21, no. 2: 173–184.

Reiter-Palmon, R., B. Forthmann, and B. Barbot. 2019. "Scoring Divergent Thinking Tests: A Review and Systematic Framework."

*Psychology of Aesthetics, Creativity, and the Arts* 13, no. 2: 144–152. https://doi.org/10.1037/aca0000227.

Richardson, A. G. 1986. "Two Factors of Creativity." *Perceptual and Motor Skills* 63, no. 2: 379–384. https://doi.org/10.2466/pms.1986.63.2.379.

Runco, M. A. 1986. "Flexibility and Originality in Children's Divergent Thinking." *Journal of Psychology* 120, no. 4: 345–352. https://doi.org/10.1080/00223980.1986.9712632.

Runco, M. A. 1992. "Children's Divergent Thinking and Creative Ideation." *Developmental Review* 12, no. 3: 233–264. https://doi.org/10.1016/0273-2297(92)90010-Y.

Runco, M. A., and S. Acar. 2012. "Divergent Thinking as an Indicator of Creative Potential." *Creativity Research Journal* 24, no. 1: 66–75. https://doi.org/10.1080/10400419.2012.652929.

Runco, M. A., and R. S. Albert. 1985. "The Reliability and Validity of Ideational Originality in the Divergent Thinking of Academically Gifted and Nongifted Children." *Educational and Psychological Measurement* 45, no. 3: 483–501. https://doi.org/10.1177/001316448504500306.

Said-Metwaly, S., B. Fernández-Castilla, E. Kyndt, W. Van den Noortgate, and B. Barbot. 2021. "Does the Fourth-Grade Slump in Creativity Actually Exist? A Meta-Analysis of the Development of Divergent Thinking in School-Age Children and Adolescents." *Educational Psychology Review* 33: 275–298. https://doi.org/10.1007/s10648-020-09547-9.

Schroeders, U., and O. Wilhelm. 2010. "Testing Reasoning Ability With Handheld Computers, Notebooks, and Paper and Pencil." *European Journal of Psychological Assessment* 26, no. 4: 284–292. https://doi.org/10.1027/1015-5759/a000038.

Shoemaker, A. L., and D. M. Bolt. 1992. "Computer Measurement of the Autokinetic Effect." *Perceptual and Motor Skills* 75, no. 3: 771–777. https://doi.org/10.2466/pms.1992.75.3.771.

Silvia, P. J., E. C. Nusbaum, C. Berg, et al. 2009. "Openness to Experience, Plasticity, and Creativity: Exploring Lower-Order, High-Order, and Interactive Effects." *Journal of Research in Personality* 43, no. 6: 1087–1090.

Simonton, D. K. 2004. *Creativity in Science: Chance, Logic, Genius, and Zeitgeist*. Cambridge University Press. https://doi.org/10.1017/CBO9781139165358.

Sung, Y.-T., H.-H. Cheng, H.-C. Tseng, et al. 2024. "Construction and Validation of a Computerized Creativity Assessment Tool With Automated Scoring Based on Deep-Learning Techniques." *Psychology of Aesthetics, Creativity, and the Arts* 18, no. 4: 493–509. https://doi.org/10.1037/aca0000450.

Tierney, P., and S. M. Farmer. 2002. "Creative Self-Efficacy: Its Potential Antecedents and Relationship to Creative Performance." *Academy of Management Journal* 45, no. 6: 1137–1148. https://doi.org/10.5465/3069429.

Torrance, E. P. 1966. *The Torrance Tests of Creative Thinking-Norms-Technical Manual Research Edition-Verbal Tests, Forms A and B – Figural Tests, Forms A and B*. Personnel Press.

Torrance, E. P. 1968. "A Longitudinal Examination of the Fourth Grade Slump in Creativity." *Gifted Child Quarterly* 12, no. 4: 195–199.

Torrance, E. P. 1998. *Torrance Test of Creative Thinking: Manual for Scoring and Interpreting Results, Verbal Forms A&B*. Scholastic Testing Service.

Torrance, E. P., and O. E. Ball. 1984. *The Torrance Tests of Creative Thinking: Streamlined Scoring Guide Figural A and B*. Scholastic Testing Service.

Urban, K. K., and H. G. Jellen. 1996. *Test for Creative Thinking - Drawing Production (TCT-DP)*. Swets and Zeitlinger.

Van der Ven, A. H. G. S., and J. L. Ellis. 2000. "A Rasch Analysis of Raven's Standard Progressive Matrices." *Personality and Individual Differences* 29, no. 1: 45–64. https://doi.org/10.1016/S0191-8869(99)00177-4.

Wallach, M. A., and N. Kogan. 1965. *Modes of Thinking in Young Children: A Study of the Creativity-Intelligence Distinction*. Holt, Rinehart & Winston.

Ward, T. B., and Y. Kolomyts. 2019. "Creative Cognition." In *The Cambridge Handbook of Creativity*, edited by J. C. Kaufman and R. J. Sternberg, 175–199. Cambridge University Press. https://doi.org/10.1017/9781316979839.011.

Wilson, R. C., J. P. Guilford, P. R. Christensen, and D. J. Lewis. 1954. "A Factor-Analytic Study of Creative-Thinking Abilities." *Psychometrika* 19: 297–311. https://doi.org/10.1007/BF02289230.

Wollscheid, S., J. Sjaastad, C. Tømte, and N. Løver. 2016. "The Effect of Pen and Paper or Tablet Computer on Early Writing–A Pilot Study." *Computers & Education* 98: 70–80.

Zabramski, S. 2014. "Creating Digital Traces of Ideas: Evaluation of Computer Input Methods in Creative and Non-Creative Drawing (Doctoral Dissertation, Acta Universitatis Upsaliensis)."

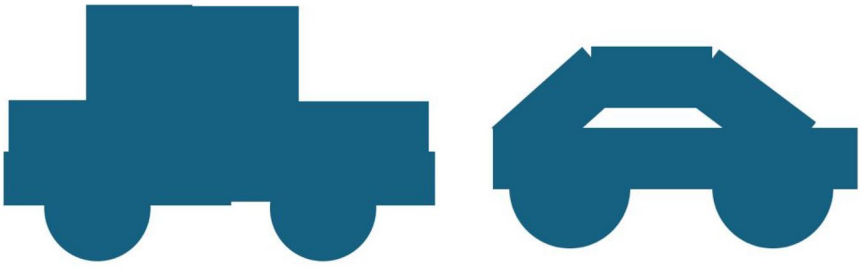
Zarnegar, Z., D. Hocevar, and W. B. Michael. 1988. "Components of Original Thinking in Gifted Children." *Educational and Psychological Measurement* 48, no. 1: 5–16. https://doi.org/10.1177/001316448804800103.

**Appendix A**

**Scoring Procedures With Example Responses From a Hypothetical Task**

1. Design vehicles with as many rectangles and circles as you wish.
2. Design AS MANY DIFFERENT VEHICLES AS YOU CAN. The more vehicles the better.

3. Surprise me! Try interesting designs. Draw unique vehicles that no one else thinks of.
4. Look at the examples below and DRAW SOMETHING ELSE.

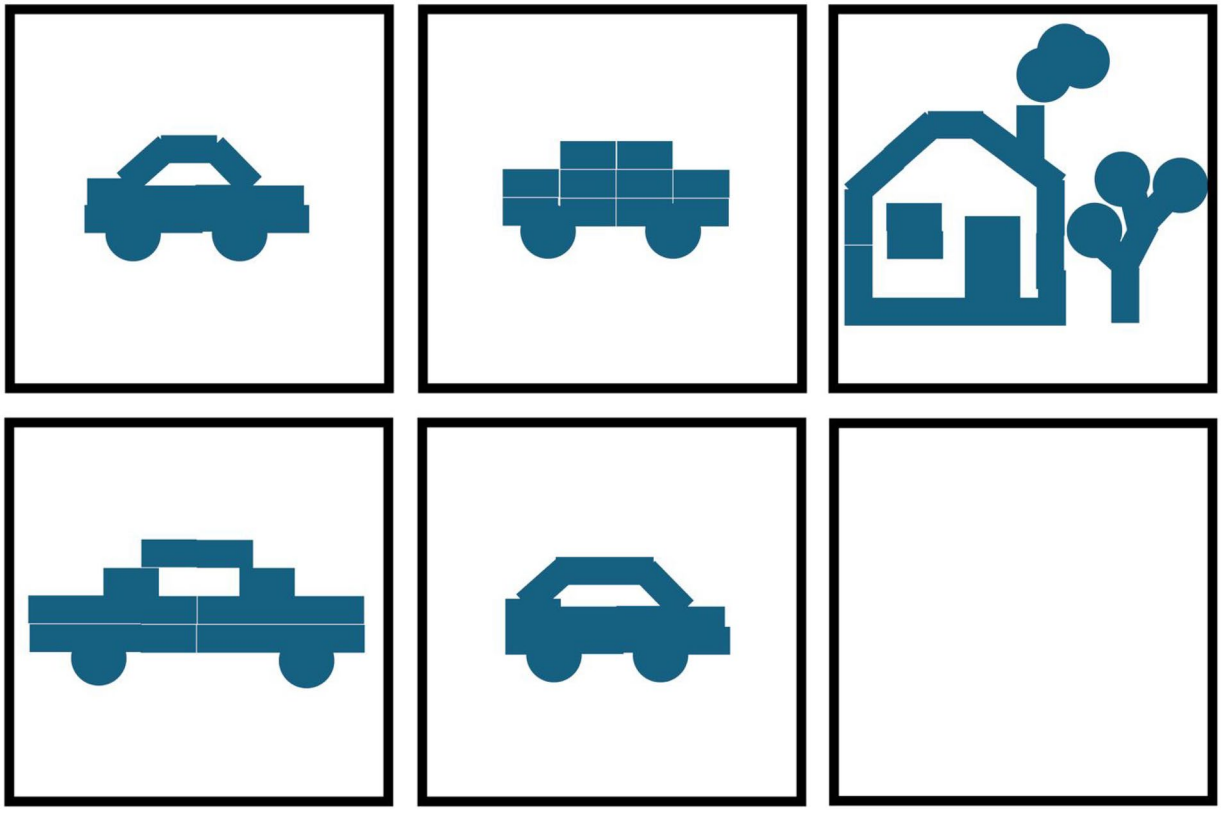


5. No questions or discussions with others are permitted.

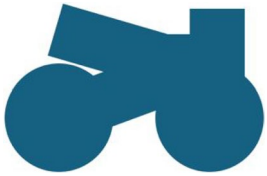
Please note that during the test the students can rotate the rectangle. They can arrange the shapes but can not change the size of the rectangle or the circle.

*Fluency*: The number of meaningful and relevant vehicle designs.

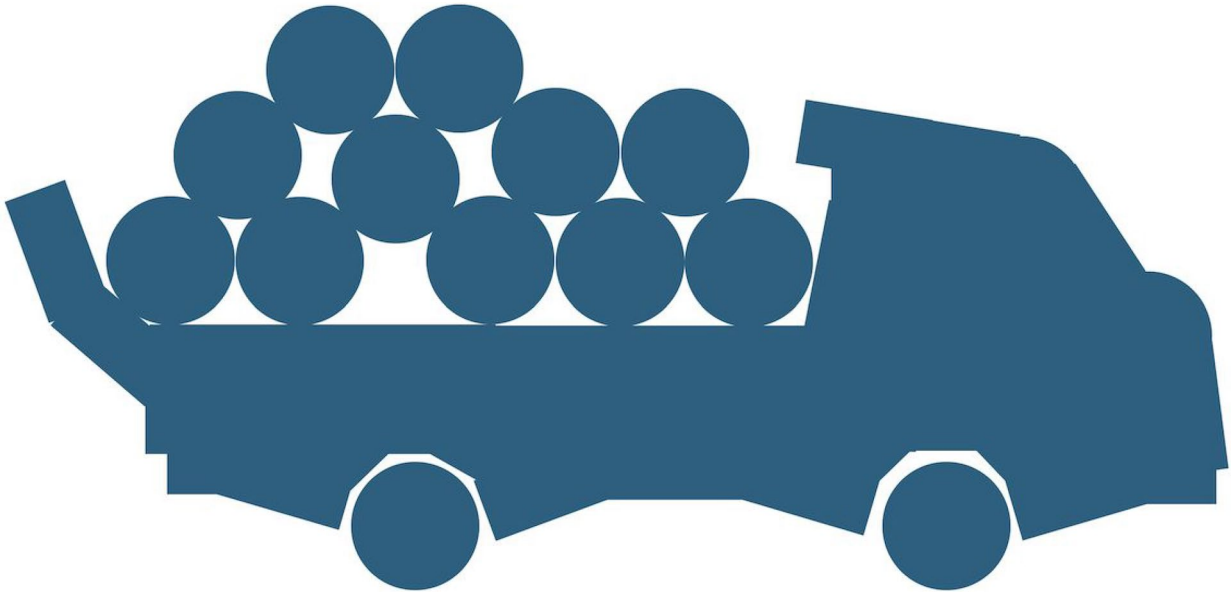
For the following sample with five responses, four of them are meaningful vehicles and one response, the house, is not relevant (invalid) because it is not a vehicle.



*Elaboration*: The number of stimuli used, the positioning of the stimuli, and use of angular rotation.



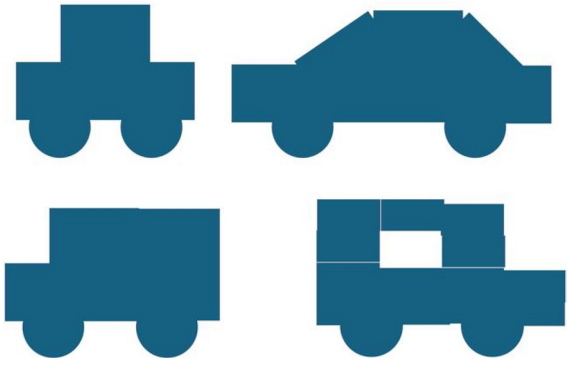
The bike design above uses four rectangles and two circles for a total of just six shapes. Only two rectangles are rotated. It is an example of low elaboration.



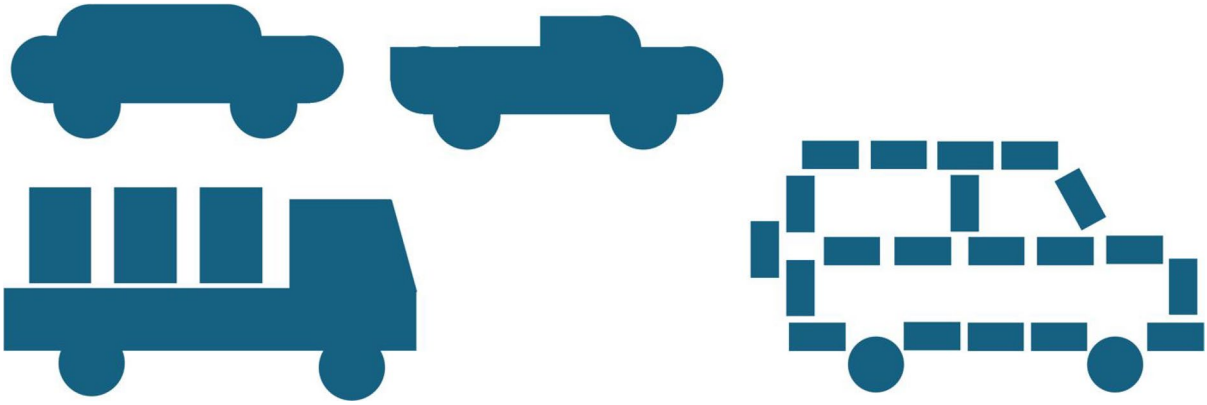
This truck design uses 42 rectangles and 15 circles for a total of 57 shapes. In total 15 rectangles are rotated. This is an example of high elaboration.

*Originality*: Statistically unusual relevant designs based on an existing pool of designs.

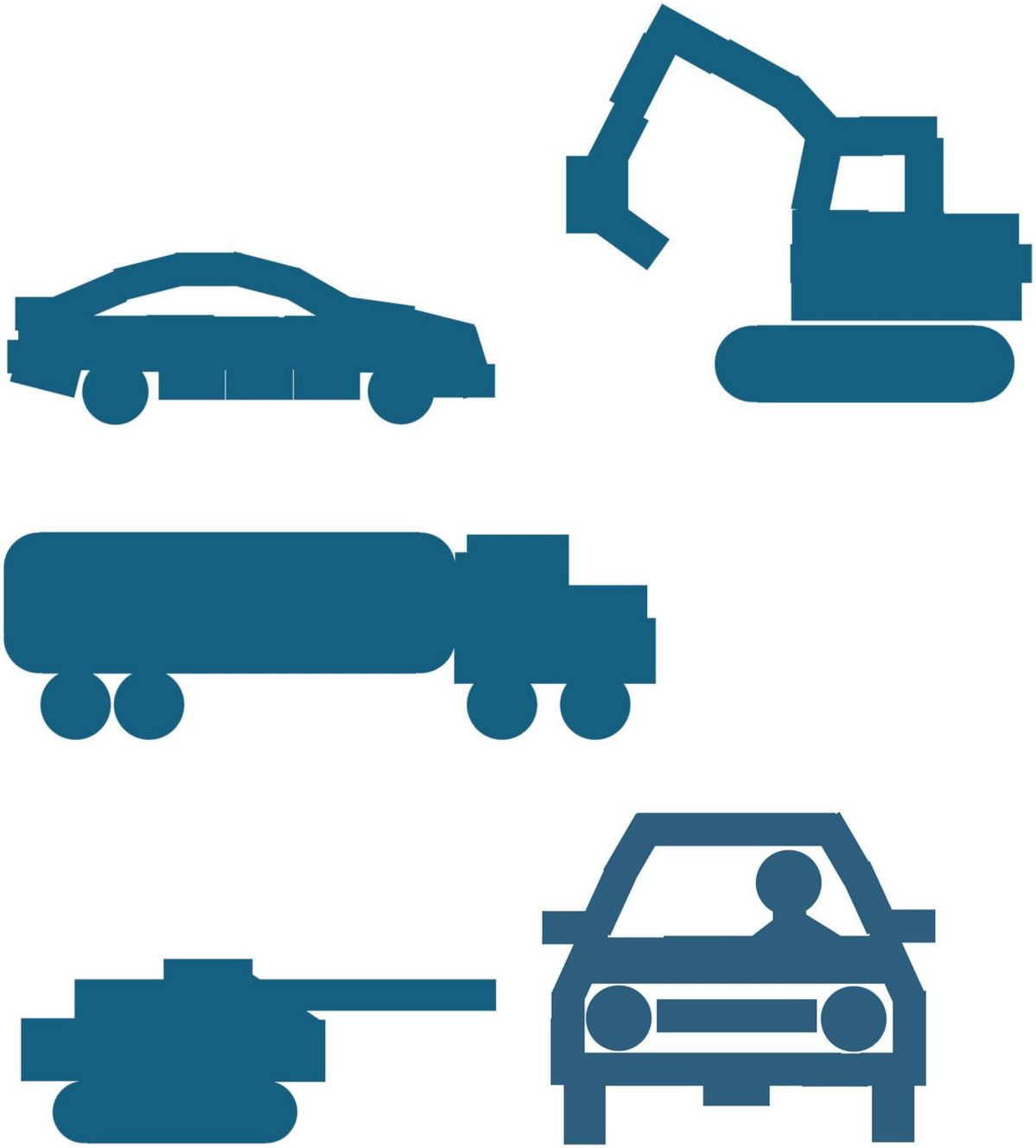
Originality = 0: The responses were similar to the sample given in the instructions and the most common responses from the test population.



Originality = 1: Relevant responses that are less common but not rare enough to have a full originality score.

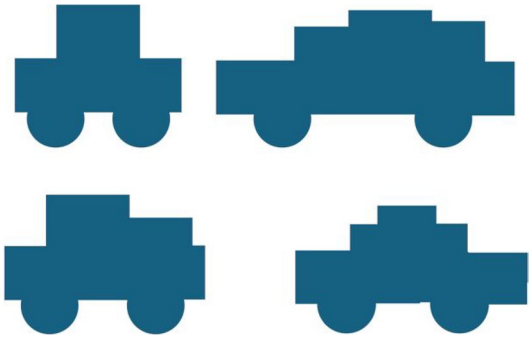


Originality = 2: Relevant responses that are unusual. We see rare vehicle types (e.g., excavator, oil tanker truck, and tank) or drawing stimulus used in unusual ways (e.g., the curves of the sports car) or we see rare viewing angles (e.g., front view).



*Flexibility*: The number of different design concepts.

The response set below shows a flexibility score of 1 because they are all similar.



The response set below shows a flexibility score of 4 because they are all coming from different vehicle types.