International Journal for Talent Development and Creativity

The Psychometric Properties of the Torrance Tests of Creative Thinking Figural Form A

Süreyya Yörük

Volume 11, Number 1-2, 2023

URI: https://id.erudit.org/iderudit/1115434ar
DOI: https://doi.org/10.7202/1115434ar

See table of contents

Publisher(s)

International Centre for Innovation in Education/Lost Prizes International

ISSN

2291-7179 (print)
2563-6871 (digital)

Explore this journal

Cite this article

Yörük, S. (2023). The Psychometric Properties of the Torrance Tests of Creative Thinking Figural Form A. *International Journal for Talent Development and Creativity*, *11*(1-2), 279–297. https://doi.org/10.7202/1115434ar

Article abstract

The Torrance Tests of Creative Thinking Figural Forms A and B are widely used to measure creative potential. Despite their common application in research, there has been a lack of focus on the psychometric properties of the tests. Thus, the scoring of the items is based on some unexamined hypotheses. The items are hypothesized to be equally difficult, and the response categories are hypothesized to be equally distributed. Additionally, it is hypothesized that the items measure a single cognitive factor for each creative thinking skill and that each creative thinking skill is a different cognitive factor. Given the impact of these hypotheses on the validity of the test scores, it is crucial to investigate these four hypotheses. In the present study, Rasch-based analyses and correlation analyses were conducted to examine all these hypotheses for Form A. The data were collected from 157 second-grade students in Turkey. The findings showed that a) the items were equally difficult for only elaboration, internal visualization, and humor; b) the response categories were equally distributed for only resistance to premature closure and elaboration; c) the items measured a single cognitive factor for each creative thinking skill; and d) certain creative thinking skills were highly correlated ($r \geqslant .90$). Overall, the items in Form A possess sufficient quality for assessing the majority of the creative thinking skills. Nevertheless, some revisions to the scoring of the items may be needed.

The Psychometric Properties of the Torrance Tests of Creative Thinking Figural Form A

Süreyya Yörük
Marmara University, Türkiye

Abstract

The Torrance Tests of Creative Thinking Figural Forms A and B are widely used to measure creative potential. Despite their common application in research, there has been a lack of focus on the psychometric properties of the tests. Thus, the scoring of the items is based on some unexamined hypotheses. The items are hypothesized to be equally difficult, and the response categories are hypothesized to be equally distributed. Additionally, it is hypothesized that the items measure a single cognitive factor for each creative thinking skill and that each creative thinking skill is a different cognitive factor. Given the impact of these hypotheses on the validity of the test scores, it is crucial to investigate these four hypotheses. In the present study, Rasch-based analyses and correlation analyses were conducted to examine all these hypotheses for Form A. The data were collected from 157 second-grade students in Turkey. The findings showed that a) the items were equally difficult for only elaboration, internal visualization, and humor; b) the response categories were equally distributed for only resistance to premature closure and elaboration; c) the items measured a single cognitive factor for each creative thinking skill; and d) certain creative thinking skills were highly correlated ($r \geq .90$). Overall, the items in Form A possess sufficient quality for assessing the majority of the creative thinking skills. Nevertheless, some revisions to the scoring of the items may be needed.

Keywords: Torrance Tests of Creative Thinking; Rasch measurement theory; dimensionality; item difficulty; response category difficulty.

Introduction

Assessment of the potential for creativity has been a primary research interest for years (Mednick, 1968; Runco et al., 2001; Torrance, 1984; Wilson et al., 1953). Several types of instruments, such as the Remote Associates Test (Mednick, 1968) and the Torrance Tests of Creative Thinking (Torrance, 1966, 1984), were developed for this purpose. However, the Torrance Tests of Creative Thinking (TTCT) Figural Forms A and B (Torrance, 1984) are possibly the most prominent (Kaufman et al., 2008). The TTCT-Figural Forms are commonly used for creativity research (Krumm et al., 2014; Palaniappan & Torrance, 2001) and for the identification of creatively gifted students (Hunsaker et al., 1991; Kaufman et al., 2012).

The TTCT-Figural Forms hold significant importance for educators, practitioners, and researchers in different fields, as the tests provide a valuable framework for assessing various creative thinking skills (Kaufman et al., 2008; Torrance, 1984). The TTCT-Figural Forms provide insights into students' creative thinking skills, allowing educators to monitor the development of their students' creative potential over time. Practitioners, such as psychologists and counselors, can use the tests to identify and support gifted students with distinctive creative abilities (Kaufman et al., 2012). Additionally, the tests can assist practitioners in designing personalized interventions that target specific aspects of creativity, fostering a more nuanced approach to talent development. Researchers benefit from the tests to explore the relationships between creativity and various factors, such as age, culture, and educational level (Cheng et al., 2010; Kim & VanTassel-Baska, 2010; Matud & Grande, 2007; Wechsler, 2006; Zhang et al., 2020). Overall, the TTCT-Figural Forms serve as useful tools in assessing, examining, nurturing, and understanding the creative potential levels of individuals in different populations.

The TTCT-Figural Forms were developed in 1966 but underwent updates to their current versions in 1984 (Torrance, 1984). Both forms consist of three subtests, each containing a certain number of visual stimuli (i.e., items). The first subtest has one item, the second subtest has 10 items, and the third subtest has 30 items. In each subtest, examinees are instructed to draw pictures using the items as the starting points and to give a title to each drawing (Torrance, 2006a, 2006b).

The current versions of the TTCT-Figural Forms measure 18 creative thinking skills. Five of the creative thinking skills are norm-referenced, and they are called fluency, originality, elaboration, abstractness of titles, and resistance to premature closure (Torrance et al., 1992). The other 13 creative thinking skills are criterion-referenced, and they are labeled as the creative strengths (Torrance et al., 1992). The creative strengths include emotional expressiveness, storytelling articulateness, movement or action, expressiveness of titles, synthesis of incomplete figures, synthesis of lines or circles, unusual visualization, internal visualization, extending or breaking boundaries, humor, richness of imagery, colorfulness of imagery, and fantasy.

When responses are scored for the TTCT-Figural Forms, a response to an item is interpreted in a different way for each creative thinking skill measured on that item (Torrance et al., 1992). For instance, responses to the items in the second subtest are evaluated in 16 different ways and scored for 16 different creative thinking skills. Therefore, the tests provide different subscores based on different interpretations of the same responses. Each form of the TTCT-Figural can be thought of as an instrument that can be used as 18 different tests, each of which is composed of the exact same items and is given the exact same responses.

Elaboration, abstractness of titles, and resistance to premature closure are manifested to varying degrees in a response (Torrance et al., 1992). Thus, responses are scored in a polytomous manner (i.e., scored with partial credit) for these three creative thinking skills, and they can be referred to as the polytomously scored creative thinking skills. The remaining creative thinking skills are either manifested in a response or not manifested and thus are scored in a dichotomous manner (i.e., 0 or 1). A score of 0 indicates that a creative thinking skill is not manifested, while a score of 1 indicates that the creative thinking skill is manifested. Those creative thinking skills can be referred to as the dichotomously scored creative thinking skills.

During scoring, all the items are treated to be equally difficult for each of the 18 creative thinking skills (Torrance et al., 1992). Thus, the manifestation of a creative thinking skill is given the same score across the items. Additionally, the response categories are treated to be equally distributed for abstractness of titles and resistance to premature closure (Torrance et al., 1992). Therefore, the distance between any two subsequent response categories is considered to be equal across the categories, and the response categories increase by one unit. For instance, 0-1-2 scoring is used for resistance to premature closure, and 0-1-2-3 scoring is used for abstractness of titles. Similarly, it is accepted that each additional requires the same amount of increase in ability to be added to a response for elaboration (Torrance et al., 1992).

Reliability of the test scores

The reliability of scores on the TTCT-Figural Forms ranges from acceptable to good. The internal consistency reliability coefficient over .90 was reported for the total score in the test manual (Torrance, 2008, 2017). Reliability coefficients over .70 were estimated in different studies, indicating that the tests provide scores with sufficient reliability (Ferrando, 2004; Ferrando et al., 2007; Krumm & Lemos, 2011; Liu, 2020, 2022; Liu et al., 2020; López, 2001; Prieto et al., 2003).

Validity of the test scores

Previous studies found weak correlations ($r < .20$) between scores on the TTCT-Figural Forms and scores on intelligence tests, providing evidence for discriminant validity (Cho et al., 2010; Palaniappan, 2008; Yong, 1994). As for predictive validity, the test manual (Torrance, 2008, 2017)

reported a significant correlation (r = .51) between the test scores and adult creative achievements. Moreover, predictive validity coefficients of .43 (Cramond et al., 2005) and .40 (Runco et al., 2010) were estimated 40 years and 50 years after administering the tests, respectively.

On the other hand, evidence for the internal structure of the tests is inconsistent. Some studies suggested that the tests are unidimensional (Aliotti et al., 1975; Clapham, 1998, 2004), whereas others indicated otherwise (Antunes & Almeida, 2007; Humble et al., 2018; Prieto et al., 2006). Two-factor (Bart et al., 2017; Humble et al., 2018; Kim, 2006; Kim et al., 2006; Krumm et al., 2014), three-factor (Antunes & Almeida, 2007; Auzmendi et al., 1996; Ferrando, 2006; Oliveira, 2007), and four-factor (Prieto et al., 2006) structures were identified in previous studies. A meta-analysis study (Said-Metwaly et al., 2018) provided evidence that supports a two-factor structure.

Fairness of the test scores

Fairness was examined in terms of gender (Awamleh et al., 2012; Buitink, 2017; Campos et al., 2000; Cheng et al., 2010; Kim et al., 2006; Kim & VanTassel-Baska, 2010; Matud & Grande, 2007; Zhang et al., 2020), ethnicity (Cheng et al., 2010; Palaniappan, 2008; Saeki et al., 2001; Tran, 2004), and socioeconomic status (Hermon et al., 2018; Johnson, 1974; Ogletree & Ujlaki, 1973; Voss, 1997). On no occasion did examinees in one group (females) consistently score significantly higher than did examinees in that group's counterpart (males). In some study settings, examinees in one group scored higher than examinees in its counterpart. However, in other study settings, the difference shifted or disappeared (Cheng et al., 2010; Kim & VanTassel-Baska, 2010; Matud & Grande, 2007).

Statement of the problem

Despite considerable research on the reliability, validity, and fairness of scores on the TTCT-Figural Forms, certain psychometric properties of the tests remain unexamined. There is a lack of evidence with regard to item difficulty for all the creative thinking skills and the distributions (difficulty levels) of the response categories for the polytomously scored creative thinking skills. Furthermore, evidence for the internal structure of the tests is insufficient and disputable (Almeida et al., 2008; Heausler & Thompson, 1988; Kaufman et al., 2008; Kazelskis, 1972; Zeng et al., 2011). There are no data based on item-level analyses that show whether a single cognitive factor loads on the items for each creative thinking skill and whether each creative thinking skill is a different cognitive factor.

The factor analysis studies addressed the internal structure of the TTCT-Figural Forms, but there are some issues concerning those studies. For instance, prior to factor analyzing the tests, each creative thinking skill was already treated to be a different cognitive factor. Moreover, it was accepted that a single cognitive factor loads on the items for a creative thinking skill. Additionally, subtest scores on the norm-referenced creative thinking skills were used for factor analyzing the tests, instead of item scores. Researchers omitted item scores, treated each creative thinking skill as one item, used the subscore on each creative thinking skill as the item score, and conducted the analyses at the subscore level (Auzmendi et al., 1996; Heausler & Thompson, 1988; Kim, 2006). As a result, this procedure resulted in factor reduction. It should also be noted that the creative strengths were not included in the analyses of those factor analysis studies.

Researchers did not conduct item-level analyses in previous studies, and therefore factors identified in the factor analysis studies did not correspond to any of the creative thinking skills measured on the tests. For example, Bart et al. (2017), Kim (2006), and Said-Metwaly et al. (2018) identified a two-factor model and treated those two factors as two types of creative personality. They argued that those factors were the innovative type of creative personality and the adaptive type of creative personality. Furthermore, they argued that the innovative factor loads on fluency and originality and that the adaptive factor loads on elaboration, abstractness of titles, and resistance to premature closure. However, without conducting item-level analyses and examining dimensionality for each creative thinking skill, one cannot be sure about these arguments.

The lack of evidence on item difficulty, response category difficulty, and the internal structure of the tests led researchers to accept certain hypotheses as true. These hypotheses are based on Torrance et al. (1992) and can be summarized as the following:

1. The items are at the same difficulty level for each creative thinking skill.
2. The response categories are equally distributed for abstractness of titles and resistance to premature closure. With respect to elaboration, each additional detail requires the same amount of increase in ability to be added to a response.
3. The items measure a single cognitive factor for each creative thinking skill.
4. Each creative thinking skill is a different cognitive factor.

Due to the absence of evidence, these hypotheses raise concerns about the scoring of the items and the validity of the test scores. If the items are not equally difficult and the response categories are not equally distributed, the validity of the test scores degrades because in this scenario, the ability levels of examinees could not be properly captured by the current scoring. Additionally, if the items do not measure a single cognitive factor for any of the creative thinking skills, the item scores based on a sole criterion become problematic. This is problematic because the item scores could not capture all the cognitive factors impacting responses for that creative thinking skill. Finally, if each creative thinking skill is not a different cognitive factor, total test scores are inflated for some examinees. The reason is that the same cognitive factor would be scored more than once under different creative thinking skills.

Purpose of the study

It is crucial to seek statistical evidence for the aforementioned four hypotheses, as they impact the item scoring and test validity. The primary objective of the present study is to conduct Rasch-based analyses and correlation analyses to examine those hypotheses for the TTCT-Figural Form A. The use of Rasch-based analyses makes the present study significant in comparison to other studies on the TTCT-Figural Forms because no previous study has applied Rasch measurement theory to the tests. With regard to its focus, the current study is the first of its kind. By addressing those four hypotheses, the present study will provide evidence for the internal structure of the test and show whether the scoring of the items are based on sound evidence.

Method

Participants

The participants were 157 second-grade students (83 girls; 52.87%) from northwestern Turkey. The students were enrolled in four different public schools. As a result of the researcher's personal connections with the principals of those schools, the sampling process was conducted using the convenience sampling method.

Due to the ages of the participants, the parents were informed about the study and were provided with a consent form. Two hundred and eight parents granted permission for their children, but 15 students did not participate due to absence or unwillingness. Thirty-six students were excluded from the analyses because those students did not respond to all the items analyzed in the study. There were no missing observations in the data.

Instrument

The TTCT-Figural Form A (Torrance, 1984) was administered. The item in Subtest 1, all 10 items in Subtest 2, and the first six items in Subtest 3 were analyzed. Given that each item in Subtest 3 is a pair of parallel lines, each item in the subtest was expected to work the same. However, item location may potentially affect the item parameters. Thus, to take the effect of item location into account, the first six items were analyzed.

Test administration

The subtests were administered sequentially. The examinees were able to respond to the item in Subtest 1 and to a sufficient number of items in Subtest 3 within the 10-minute time limit. Several examinees were unable to respond to all 10 items in Subtest 2 within the 10-minute time limit. Those examinees were provided with an additional 15 minutes of time so that each examinee could respond to all 10 items. This decision was based on Cohen and Swerdlik's (1999) suggestion in which they recommended that researchers provide extra time to all examinees if speed is not the object and if the purpose is to examine test properties.

Scoring process

As a certified scorer of the TTCT-Figural Forms, the author scored the tests. Another certified scorer scored 10% of the tests, and then the inter-rater reliability coefficients were estimated. The inter-rater reliability coefficients were used to check whether all the tests were scored accurately (Tinsley & Weiss, 1975). With the exception of elaboration, the items were scored following the guidelines provided by Torrance et al. (1992).

During the conventional scoring for elaboration, each item is not given an elaboration score. Rather, the scoring involves counting the number of details in a subtest. Accordingly, a subtest is treated as a single item, and a score based on the total number of details in the entire subtest is assigned as the elaboration score (Ball & Torrance, 1984; Torrance et al., 1992). For instance, in Subtest 1 in Form A, 1 point is given if there are 0-5 details, 2 points are given if there are 6-12 details, 3 points are given if there are 13-19 details, 4 points are given if there are 20-26 details, 5 points are given if there are 27-33 details, and 6 points are given if there are more than 33 details in total in the subtest (Torrance et al., 1992). Similar scoring is done for Subtests 2 and 3.

To examine the individual items for elaboration, a different type of scoring was employed. The following criteria were used. For up to five details, the number of details in an item was assigned as the elaboration score. For instance, if there were three details, the item was given 3 points. For six or more details, the item was given 6 points. This type of scoring made it possible to analyze the items using the rating scale model (Andrich, 1978).

Analyses

Prior to conducting the Rasch-based analyses and correlation analyses, means and standard deviations were estimated. Additionally, Cronbach's α (Cronbach, 1951) values were computed to check the reliability of the test scores.

Rasch analyses

The dichotomous Rasch model (Rasch, 1960) and the rating scale model (Andrich, 1978) were used for analyzing the first, second, and third hypotheses. The dichotomously scored creative thinking skills were analyzed through the dichotomous Rasch model, as the model is used for binary data. The polytomously scored creative thinking skills were analyzed through the rating scale model because the model was developed for analyzing Likert-type items (Likert, 1974). Separate analyses were conducted for each creative thinking skill. FACETS Version 3.65.0 (Linacre, 2009) was used for the analyses.

To investigate the first hypothesis, the difficulty levels of the items were estimated for each creative thinking skill, and the number of item groups regarding difficulty was checked using strata statistics (Linacre, 2017; Wright & Masters, 2002). Additionally, to determine whether the difficulty levels of the items will be stable across samples, item reliability indices were estimated (Bond & Fox, 2015). The item reliability index is recommended to be over .90 (Fox & Jones, 1998; Linacre, 2016).

To examine the second hypothesis, Rasch-Thrustone thresholds (Linacre, 2003; Thurstone, 1928) were estimated for elaboration, abstractness of titles, and resistance to premature closure. Rasch-Thrustone thresholds show the difficulty levels of the response categories for a polytomously scored item (Bond & Fox, 2015). Additionally, to check whether the response categories function properly, outfit mean-square values are considered. Linacre (2002b) suggested that the outfit mean-square value for each response category should be less than 2.0.

To address the third hypothesis, point-measure correlation coefficients, fit statistics, and eigenvalues obtained through the principal component analysis of the standardized Rasch residuals were considered in accordance with Linacre's (1998, 2017) suggestion. Point-measure correlation coefficients show the relationship between the examinee responses and Rasch measures (Bond & Fox, 2015). In a unidimensional data set, positive point-measure correlation coefficients should be estimated (Bond & Fox, 2015; Linacre, 2017).

Infit mean-square (IMNSQ), outfit mean-square (OMNSQ), and standardized fit statistics were estimated to examine item fit. The mean-square value shows the amount of variation involved in an item. An IMNSQ value and an OMNSQ value of 1.00 denotes 0% variation (Wright & Linacre, 1994). A mean-square value above 1.00 indicates more variation in the item measures, whereas a mean-square value below 1.00 implies less variation. Linacre (2002a, 2017) argued that up to 50% more variation is negligible and does not threaten unidimensionality. Therefore, IMNSQ and OMNSQ values under 1.50 are considered "productive for measurement" (Linacre, 2009, p. 192). IMNSQ values are sensitive to the responses of examinees whose ability levels are close to the difficulty levels of the items (Linacre, 2009).

On the other hand, OMNSQ values are sensitive to the responses of examinees whose ability levels are much higher or lower than the difficulty levels of the items (Linacre, 2009). Standardized fit statistics indicate how likely the misfit is for an item (Linacre, 2002a). An item with a good fit should have standardized infit (S-INFIT) and outfit (S-OFIT) statistics between -2.00 and 2.00 (Bond & Fox, 2015), as values out of this range denote misfit.

To estimate the variance in the item residuals, principal component analysis was conducted for each creative thinking skill. Both the dichotomous Rasch model and the rating scale model extract the primary cognitive factor from the data during the analyses (Bond & Fox, 2015). Therefore, if the data represent a unidimensional construct, random noise should be left in the residuals (Linacre & Tennant, 2009). In other words, there should not be a large eigenvalue that explains a considerable portion of the variance in the item residuals. According to Linacre (2017), an eigenvalue as large as 3.00 does not threaten unidimensionality as long as no single eigenvalue is considerably larger than the others.

Correlation analyses

To investigate the fourth hypothesis and detect the highly related creative thinking skills, item difficulties were correlated among the creative thinking skills. Because each creative thinking skill is hypothesized to be a different cognitive factor (Torrance, 1984), examinee performances on the items and the difficulty levels of the items are expected to vary across the creative thinking skills. In other words, the creative thinking skills are not expected to be highly correlated. A Pearson correlation (Pearson, 1909) coefficient value of .90 was determined as the cutoff for flagging highly related creative thinking skills because a coefficient value of .90 or above between two variables is an indication of collinearity and a large amount of shared variance (Field, 2009).

Results

Descriptive Statistics

The means (*M*), standard deviations (*SD*), item strata values, α values, inter-rater reliability coefficients, and item reliability coefficients are presented in Table 1. The reliability of scores on the TTCT-Figural Form A ranges from poor (α = .26 for unusual visualization, α = .36 for internal visualization) to good (α = .88 for expressiveness of titles, α = .90 for abstractness of titles) for the sample of the study (O'Rourke et al., 2005). The inter-rater reliability coefficients ranged from .84 to 1.00, indicating that there was a sufficient level of scoring reliability.

Table 1: Descriptive statistics and reliability coefficients.

	M	*SD*	Item Strata	Cronbach's α	Inter-rater reliability	Item reliability
AT	4.79	0.72	5.66	.90	.89	.94
CI	2.38	0.34	4.35	.63	.86	.90
El	22.92	1.69	2.50	.85	.84	.73
EE	1.07	0.24	3.12	.76	.89	.81
ET	2.10	0.32	4.11	.88	.90	.89
Ex	2.25	0.48	4.70	.65	.91	.91
Fa	1.35	0.27	3.18	.78	.91	.82
Fl	15.22	0.30	3.08	.71	.99	.81
Hu	0.64	0.19	1.16	.65	.85	.28
IV	0.99	0.28	1.20	.36	.90	.30
MA	1.85	0.35	4.07	.53	.89	.89
Or	9.38	0.49	6.15	.51	.93	.95
RC	10.57	0.89	5.87	.68	.89	.95
RI	2.74	0.36	4.07	.67	.87	.89
SA	1.41	0.27	3.88	.73	.90	.88
UV	4.40	0.43	5.73	.26	.87	.94

Notes. AT: Abstractness of titles. CI: Colorfulness of imagery. El: Elaboration. EE: Emotional expressiveness. ET: Expressiveness of titles. Ex: Extending boundaries. Fa: Fantasy. Fl: Fluency. Hu: Humor. IV: Internal visualization. MA: Movement or action. Or: Originality. RC: Resistance to closure. RI: Richness of imagery. SA: Storytelling articulateness. UV: Unusual visualization. The letter S in S1, S2, and S3 refers to subtest, as in Subtest 1. The letter I in I1 through I10 refers to item, as in Item 1. *NA* indicates that there is no infit mean-square value corresponding to that particular item.

Note that the analyses of synthesis of figures and synthesis of lines did not produce any interpretable results due to few manifestations. Overall, two and six manifestations were observed for synthesis of figures and synthesis of lines, respectively. Nevertheless, this situation was expected because Torrance (1979) already mentioned that these two creative thinking skills can be manifested by a very low percentage of examinees.

Item difficulty

The item difficulties ranged from -3.41 to 1.98 across the creative thinking skills (see Table 2). The item strata values for only internal visualization (1.20) and humor (1.57) indicated that there was one item group with regard to difficulty. The strata values for the other creative thinking skills were over 2.00, denoting more than two item groups.

Table 2: Item difficulties.

Items	AT	CI	El	EE	ET	Ex	Fa	Fl	Hu	IV	MA	Or	RC	RI	SA	UV
S1I1	-1.95	-2.68	*NA*	-2.55	-3.41	*NA*	-3.10	-2.01	-1.20	-0.94	-2.48	-1.22	*NA*	-2.36	-3.75	-2.05
S2I1	0.03	0.34	-0.07	-1.37	-0.98	*NA*	-0.36	-0.81	-1.07	-0.24	-1.36	-0.23	0.22	0.07	-0.29	0.94
S2I2	0.08	0.42	-0.02	0.91	0.34	*NA*	0.29	0.77	0,40	-0.24	0.63	0.66	-0.10	0.01	0.53	0.59
S2I3	0.25	-0.36	0.06	0.91	0.01	*NA*	0.45	-0.08	0.13	1.12	-0.81	-1.54	-0.13	-0.62	-0.85	-1.42
S2I4	0.19	-1.64	-0.12	-0.41	-0.39	*NA*	-0.94	0.64	-0,64	0.18	-0.88	-0.32	-0.83	-0.75	-0.05	-0.72
S2I5	0.22	0.17	-0.04	-1.88	-0.29	*NA*	0.45	-0.29	-0,48	-0.28	1.72	-1.16	0.75	-0.05	0.07	-0.06
S2I6	0.27	-0.36	0.13	0.04	0.01	*NA*	0.45	0.71	-0.10	0.36	-0.32	-0.11	-0.67	-0.27	0.21	-0.90
S2I7	0.39	0.09	-0.07	0.91	0.34	*NA*	0.00	0.50	0,40	0.02	-0.15	-1.28	0.08	-0.16	0.37	-0.54
S2I8	0.33	0.17	0.06	0.22	0.34	*NA*	0.45	0.57	0,73	0.56	-0.06	-0.03	0.16	-0.32	0.53	-0.44
S2I9	-0.14	0.10	0.12	0.04	-0.20	*NA*	-0.36	-0.81	-0.10	0.56	-0.81	0.04	0.30	0.35	0.53	0.59
S2I10	0.33	0.84	0.24	0.91	0.72	*NA*	0.84	0.28	0.73	0.18	1.72	0.51	0.21	0.58	1.98	1.08
S3I1	*NA*	0.84	-0.04	0.64	0.11	-0.55	0.45	-0.98	-0.40	-0.12	0.24	0.75	*NA*	0.97	0.21	0.81
S3I2	*NA*	0.42	-0.18	0.42	0.34	-0.67	0.14	-0.18	-0.30	-0.47	0.49	0.71	*NA*	0.42	0.21	0.75
S3I3	*NA*	0.10	-0.06	0.04	0.59	-0.76	0.14	0.50	0.13	-0.12	0.49	0.59	*NA*a	0.21	-0.39	0.35
S3I4	*NA*	0.52	-0.08	0.04	0.59	0.55	0.29	0.11	0.73	-0.24	0.13	0.75	*NA*	0.58	-0.39	0.35
S3I5	*NA*	0.62	-0.04	0.91	0.86	0.59	0.14	0.43	0.13	-0.36	0.49	0.79	*NA*	0.58	0.53	0.30
S3I6	*NA*	0.42	0.10	0.22	1.01	0.84	0.63	0.64	0.13	0.02	0.96	1.08	*NA*	0.76	0.53	0.39

Notes. AT: Abstractness of titles. CI: Colorfulness of imagery. El: Elaboration. EE: Emotional expressiveness. ET: Expressiveness of titles. Ex: Extending boundaries. Fa: Fantasy. Fl: Fluency. Hu: Humor. IV: Internal visualization. MA: Movement or action. Or: Originality. RC: Resistance to closure. RI: Richness of imagery. SA: Storytelling articulateness. UV: Unusual visualization. The letter S in S1, S2, and S3 refers to subtest, as in Subtest 1. The letter I in I1 through I10 refers to item, as in Item 1. *NA* indicates that there is no infit mean-square value corresponding to that particular item.

The item reliability indices were between .28 and .95. The two lowest reliability indices were estimated for internal visualization (.30) and humor (.28). The reliability indices were excellent (≥ .90) for originality, abstractness of titles, resistance to premature closure, unusual visualization, extending or breaking boundaries, and colorfulness of imagery. For the other creative thinking skills, the reliability indices were below the cutoff point, but they were still sufficiently large (≥ .80), except elaboration (.73).

Response category difficulty

The seven difficulty levels of the response categories for an item with a difficulty level of 0 logits for elaboration are as follows: 1) -1.68 (no detail), 2) -0.70 (adding one detail), 3) -0.25 (adding two details), 4) 0.04 (adding three details), 5) 0.32 (adding four details), 6) 0.71 (adding five details), and 7) 1.52 (adding six or more details). The outfit mean-square values for the response categories were 1.00, 1.10, 0.90, 1.10, 0.90, 1.30, and 0.90, respectively from the first category to the last.

The four difficulty levels of the response categories were 1) -3.35 (giving the lowest level of title), 2) -1.81 (giving the second level of title), 0.96 (giving the third level of title), and 3) 4.69 (giving the highest level of title) for an item with a difficulty level of 0 logits for abstractness of titles. The outfit mean-square values for the response categories were 1.10, 0.90, 0.90, and 1.80.

Finally, the three difficulty levels of the response categories for an item with a difficulty level of 0 logits for resistance to premature closure are as follows: 1) -1.32 (lowest level of resistance), 2) 0 (medium level of resistance), and 3) 1.34 (highest level of resistance). The outfit mean-square values were 1.00, 1.00, and 1.10 for the response categories.

Point-measure correlation coefficients

A separate set of point-measure correlation coefficients was estimated for each creative thinking skill. All the coefficients were positive, ranging from .14 to .73 (see Table 1). The point-measure correlation coefficients showed that all the items worked in the same direction for each creative thinking skill.

Table 3: Point-measure correlation coefficients.

Items	AT	CI	El	EE	ET	Ex	Fa	Fl	Hu	IV	MA	Or	RC	RI	SA	UV
S1I1	.63	.54	*NA*	.60	.62	*NA*	.66	.16	.50	.40	.51	.30	*NA*	.62	.62	.35
S2I1	.71	.27	.50	.53	.73	*NA*	.42	.43	.55	.32	.38	.22	.46	.32	.41	.22
S2I2	.57	.40	.45	.32	.52	*NA*	.40	.50	.40	.46	.27	.32	.54	.43	.29	.12
S2I3	.64	.32	.52	.31	.61	*NA*	.41	.31	.39	.13	.48	.22	.49	.47	.53	.32
S2I4	.69	.43	.51	.53	.67	*NA*	.54	.55	.52	.38	.38	.33	.49	.44	.37	.33
S2I5	.61	.25	.40	.52	.62	*NA*	.36	.31	.43	.35	.21	.25	.32	.29	.37	.15
S2I6	.63	.40	.50	.53	.57	*NA*	.48	.55	.41	.27	.45	.30	.54	.47	.43	.41
S2I7	.63	.38	.52	.42	.59	*NA*	.46	.43	.44	.36	.41	.30	.47	.42	.32	.35
S2I8	.58	.25	.44	.40	.62	*NA*	.45	.38	.30	.20	.24	.31	.48	.40	.29	.40
S2I9	.69	.35	.47	.48	.61	*NA*	.45	.29	.28	.18	.42	.41	.54	.33	.33	.19
S2I10	.54	.41	.42	.33	.46	*NA*	.43	.35	.27	.40	.14	.35	.54	.34	.18	.23
S3I1	*NA*	.43	.45	.45	.56	.60	.52	.39	.25	.39	.37	.35	*NA*	.24	.37	.29
S3I2	*NA*	.42	.56	.42	.54	.59	.54	.42	.50	.39	.28	.30	*NA*	.39	.47	.15
S3I3	*NA*	.46	.49	.51	.51	.63	.41	.53	.35	.29	.38	.35	*NA*	.40	.57	.31
S3I4	*NA*	.25	.43	.53	.52	.60	.41	.51	.31	.27	.37	.38	*NA*	.21	.44	.29
S3I5	*NA*	.35	.45	.43	.50	.60	.38	.51	.41	.26	.40	.50	*NA*	.25	.36	.23
S3I6	*NA*	.34	.42	.38	.46	.56	.44	.42	.43	.32	.36	.45	*NA*	.23	.33	.22

Notes. AT: Abstractness of titles. CI: Colorfulness of imagery. El: Elaboration. EE: Emotional expressiveness. ET: Expressiveness of titles. Ex: Extending boundaries. Fa: Fantasy. Fl: Fluency. Hu: Humor. IV: Internal visualization. MA: Movement or action. Or: Originality. RC: Resistance to closure. RI: Richness of imagery. SA: Storytelling articulateness. UV: Unusual visualization. The letter S in S1, S2, and S3 refers to subtest, as in Subtest 1. The letter I in I1 through I10 refers to item, as in Item 1. *NA* indicates that there is no point-measure correlation coefficient to estimate for that particular item.

Mean-square values

All the IMNSQ values were below 1.50 for the dichotomously scored and polytomously scored creative thinking skills. The IMNSQ values showed that the amount of variation involved in each item was below the suggested cutoff point (1.50) for all the creative thinking skills. Table 4 presents the IMNSQ values.

Table 4: Infit mean-square values.

Items	AT	CI	El	EE	ET	Ex	Fa	Fl	Hu	IV	MA	Or	RC	RI	SA	UV
S1I1	1.02	1.04	NA	1.20	1.31	*NA*	1.13	1.20	1.14	1.07	1.09	0.99	*NA*	0.93	1.24	0.99
S2I1	0.89	1.10	1.06	1.18	0.78	*NA*	1.23	0.86	0.99	1.02	1.17	1.09	0.93	1.08	1.06	1.01
S2I2	1.09	0.93	1.06	1.13	1.16	*NA*	1.12	0.99	0.91	0.84	1.05	1.02	0.95	0.92	1.11	1.07
S2I3	0.90	1.07	0.87	1.11	0.83	*NA*	0.99	1.14	0.98	1.05	0.91	1.04	1.01	0.93	0.92	1.02
S2I4	0.90	1.14	0.88	0.86	0.81	*NA*	1.04	0.88	0.92	0.89	1.12	1.00	1.23	1.02	1.04	0.99
S2I5	1.15	1.14	1.07	1.27	0.99	*NA*	1.08	1.11	1.04	1.00	1.00	1.06	1.14	1.10	1.05	1.08
S2I6	0.97	1.03	1.05	0.72	1.05	*NA*	0.82	0.88	1.02	1.00	0.93	1.04	1.07	0.91	0.87	0.94
S2I7	0.89	0.94	1.04	0.81	0.85	*NA*	0.96	1.02	0.82	0.94	0.89	0.97	1.00	0.96	1.10	0.99
S2I8	1.19	1.16	0.90	1.02	0.81	*NA*	0.89	1.14	0.98	1.06	1.11	1.03	1.00	0.99	1.02	0.92
S2I9	0.87	1.04	0.98	0.95	0.97	*NA*	1.09	1.13	1.26	1.06	1.01	0.95	0.90	0.98	1.00	1.05
S2I10	1.17	0.82	1.09	1.09	1.09	*NA*	0.82	1.17	0.99	0.88	1.04	0.99	0.86	0.98	1.06	0.97
S3I1	*NA*	0.79	1.10	0.79	1.04	1.04	0.74	0.87	1.13	0.95	0.93	0.99	*NA*	1.06	0.99	0.95
S3I2	*NA*	0.85	0.79	0.91	1.00	1.09	0.75	0.98	0.89	0.98	1.05	1.04	*NA*	0.92	0.81	1.01
S3I3	*NA*	0.85	1.02	0.80	1.08	0.96	1.14	0.90	1.07	1.03	0.85	0.99	*NA*	0.89	0.71	0.95
S3I4	*NA*	1.08	1.04	0.76	0.98	0.94	0.95	0.84	0.94	1.09	0.91	0.98	*NA*	1.09	0.96	0.97
S3I5	*NA*	0.93	0.99	0.77	1.01	0.96	1.13	0.91	0.96	1.15	0.85	0.86	*NA*	1.08	1.01	0.99
S3I6	*NA*	0.98	1.20	1.19	0.99	1.01	0.83	1.10	0.93	0.99	0.88	0.90	*NA*	1.13	1.03	1.01

Notes. AT: Abstractness of titles. CI: Colorfulness of imagery. El: Elaboration. EE: Emotional expressiveness. ET: Expressiveness of titles. Ex: Extending boundaries. Fa: Fantasy. Fl: Fluency. Hu: Humor. IV: Internal visualization. MA: Movement or action. Or: Originality. RC: Resistance to closure. RI: Richness of imagery. SA: Storytelling articulateness. UV: Unusual visualization. The letter S in S1, S2, and S3 refers to subtest, as in Subtest 1. The letter I in I1 through I10 refers to item, as in Item 1. *NA* indicates that there is no infit mean-square value corresponding to that particular item.

The OMNSQ values showed that several items had misfit for the dichotomously scored creative thinking skills (see Table 5). OMNSQ values greater than 1.50 were estimated for some items for emotional expressiveness, storytelling articulateness, movement or action, expressiveness of titles, internal visualization, humor, and richness of imagery. However, the S-OFIT values showed that misfit was not likely for those items, except the item in Subtest 1 for expressiveness of titles (S-OFIT = 2.90) and storytelling articulateness (S-OFIT = 2.50). The S-OFIT statistics for expressiveness of titles and

storytelling articulateness indicated that the item in Subtest 1 did not work as intended for certain examinees whose ability levels were much higher or lower than the difficulty level of that item. On the other hand, both the OMNSQ and S-OFIT statistics were within the suggested range for good fit for all the items for the polytomously scored creative thinking skills.

Table 5: Outfit mean-square values.

Items	AT	CI	El	EE	ET	Ex	Fa	Fl	Hu	IV	MA	Or	RC	RI	SA	UV
S1I1	1.25	1.05	NA	1.21	2.62	*NA*	1.42	1.43	1.18	1.05	1.13	0.99	*NA*	0.93	1.72	0.99
S2I1	0.68	1.12	1.08	1.18	0.66	*NA*	1.06	0.52	0.98	1.14	1.20	1.13	0.94	1.07	0.93	0.86
S2I2	1.36	0.57	1.01	0.64	1.01	*NA*	0.97	0.95	0.57	0.65	0.90	1.03	0.91	0.78	1.14	1.18
S2I3	0.92	1.43	0.77	0.53	1.04	*NA*	0.99	1.44	0.89	1.57	0.91	1.04	0.95	0.98	0.86	1.04
S2I4	0.64	1.14	0.88	0.96	0.71	*NA*	1.00	0.74	0.91	0.72	1.08	1.04	1.21	0.99	1.21	1.01
S2I5	1.01	1.43	1.21	1.12	0.88	*NA*	1.25	1.43	1.14	0.98	0.53	1.06	1.36	1.38	0.96	1.23
S2I6	0.89	0.90	0.88	1.51	0.96	*NA*	0.52	0.77	0.93	1.06	0.74	1.04	0.98	0.76	0.57	0.89
S2I7	0.91	1.02	0.93	0.41	0.79	*NA*	1.05	1.21	0.53	0.80	0.90	1.01	1.03	0.92	0.86	0.91
S2I8	1.04	1.17	0.99	1.10	0.39	*NA*	0.68	1.33	1.10	1.24	1.63	1.06	0.94	1.04	1.58	0.95
S2I9	0.79	0.87	0.98	0.67	1.00	*NA*	1.19	1.08	1.42	1.43	1.05	0.92	0.90	1.15	1.03	0.97
S2I10	1.40	0.45	1.06	0.53	1.43	*NA*	0.53	1.24	1.34	0.56	1.14	1.01	0.90	0.81	0.50	0.91
S3I1	*NA*	0.37	1.18	0.52	0.97	1.02	0.36	0.65	1.55	0.72	0.78	0.98	*NA*	1.11	0.96	0.82
S3I2	*NA*	0.73	0.78	0.94	1.06	1.09	0.44	0.86	0.67	0.95	0.84	1.05	*NA*	0.75	0.39	1.33
S3I3	*NA*	0.59	1.02	0.77	0.78	1.04	0.83	0.72	1.05	1.14	0.74	1.02	*NA*	0.99	0.42	0.92
S3I4	*NA*	1.17	1.20	1.57	1.03	0.94	1.48	0.74	1.11	1.23	0.95	0.93	*NA*	1.69	1.03	0.94
S3I5	*NA*	0.70	0.99	1.39	0.63	0.86	1.33	0.79	0.85	1.27	0.51	0.83	*NA*	1.17	0.57	1.15
S3I6	*NA*	0.89	1.08	0.84	1.27	0.93	0.90	1.17	0.62	0.96	0.34	0.83	*NA*	1.07	0.73	1.05

Notes. AT: Abstractness of titles. CI: Colorfulness of imagery. El: Elaboration. EE: Emotional expressiveness. ET: Expressiveness of titles. Ex: Extending boundaries. Fa: Fantasy. Fl: Fluency. Hu: Humor. IV: Internal visualization. MA: Movement or action. Or: Originality. RC: Resistance to closure. RI: Richness of imagery. SA: Storytelling articulateness. UV: Unusual visualization. The letter S in S1, S2, and S3 refers to subtest, as in Subtest 1. The letter I in I1 through I10 refers to item, as in Item 1. *NA* indicates that there is no infit mean-square value corresponding to that particular item.

Principal components analysis of residuals

All the eigenvalues (λ) were below 3.00, and there were several principal components with eigenvalues over 1.00 for each creative thinking skill (see Table 6). A relatively large eigenvalue (λ > 2.00) was estimated for elaboration (λ = 2.09), emotional expressiveness (λ = 2.03), and humor (λ = 2.34). However, eigenvalues greater than 1.00 were estimated for several other principal components for these three creative thinking skills. These findings indicated that there was no dominant principal component in the residuals.

Table 6: Eigenvalues obtained with the principal component analysis of the standardized Rasch residuals.

PC	AT	CI	El	EE	ET	Ex	Fa	Fl	Hu	IV	MA	Or	RC	RI	SA	UV
1st	1.49	1.92	2.09	2.03	1.77	1.47	1.65	1.76	2.34	1.92	1.59	1.72	1.53	1.53	1.81	1.66
2nd	1.36	1.54	1.59	1.61	1.59	1.25	1.51	1.66	1.81	1.67	1.48	1.58	1.37	1.45	1.67	1.51
3rd	1.26	1.49	1.41	1.47	1.49	1.18	1.40	1.43	1.45	1.34	1.41	1.42	1.24	1.40	1.55	1.44
4th	1.18	1.33	1.34	1.36	1.32	1.09	1.29	1.26	1.37	1.30	1.22	1.27	1.13	1.35	1.37	1.34
5th	1.12	1.26	1.24	1.19	1.23	0.90	1.22	1.21	1.28	1.17	1.18	1.24	1.05	1.27	1.18	1.31
6th	1.02	1.22	1.14	1.13	1.13	0.09	1.17	1.17	1.23	1.13	1.18	1.16	1.01	1.24	1.13	1.30
7th	0.99	1.13	1.09	1.07	1.09	*NA*	1.15	1.12	1.10	1.06	1.14	1.15	0.97	1.15	1.05	1.19
8th	0.93	1.09	1.05	0.99	1.01	*NA*	1.03	1.08	0.99	1.01	1.08	1.06	0.85	1.11	1.02	1.01
9th	0.84	1.02	0.95	0.98	0.96	*NA*	1.03	1.04	0.86	0.98	1.02	1.03	0.83	0.99	0.96	0.94
10th	0.73	0.94	0.81	0.90	0.91	*NA*	1.00	0.95	0.81	0.97	0.98	0.96	0.03	0.95	0.92	0.89
11th	0.09	0.79	0.73	0.84	0.87	*NA*	0.91	0.91	0.78	0.91	0.89	0.93	*NA*	0.93	0.89	0.87
12th	*NA*	0.77	0.72	0.79	0.81	*NA*	0.88	0.82	0.73	0.87	0.83	0.86	*NA*	0.88	0.81	0.85
13th	*NA*	0.72	0.68	0.69	0.76	*NA*	0.79	0.70	0.68	0.84	0.80	0.82	*NA*	0.77	0.76	0.83
14th	*NA*	0.67	0.62	0.65	0.72	*NA*	0.70	0.66	0.60	0.79	0.75	0.70	*NA*	0.73	0.67	0.71
15th	*NA*	0.66	0.53	0.57	0.63	*NA*	0.62	0.61	0.54	0.49	0.72	0.63	*NA*	0.63	0.57	0.59
16th	*NA*	0.37	0.01	0.51	0.54	*NA*	0.49	0.49	0.22	0.39	0.65	0.45	*NA*	0.54	0.53	0.54
17th	*NA*	0.07	*NA*	0.18	0.17	*NA*	0.14	0.13	0.20	0.15	0.08	0.01	*NA*	0.07	0.12	0.01

Notes. PC: Principal components. AT: Abstractness of titles. CI: Colorfulness of imagery. El: Elaboration. EE: Emotional expressiveness. ET: Expressiveness of titles. Ex: Extending boundaries. Fa: Fantasy. Fl: Fluency. Hu: Humor. IV: Internal visualization. MA: Movement or action. Or: Originality. RC: Resistance to closure. RI: Richness of imagery. SA: Storytelling articulateness. UV: Unusual visualization. *NA* indicates that there is no item corresponding to that particular principal component.

Item correlations

The correlation coefficients ranged from -.62 to .94 (see Table 7). Similar patterns of item difficulty were obtained for certain creative thinking skills, and this situation resulted in a few high correlation coefficients. A coefficient value of .90 or above was estimated between the following pairs: 1) abstractness of titles and expressiveness of titles (r = .94, p < .001), 2) abstractness of titles and fantasy (r = .91, p < .001), 3) expressiveness of titles and fantasy (r = .90, p < .001), and 4) richness of imagery and colorfulness of imagery (r = .93, p < .001).

Table 7: Correlation coefficients of item difficulties among the creative thinking skills.

	AT	CI	El	EE	ET	Ex	Fa	Fl	Hu	IV	MA	Or	RC	RI	SA	UV
AT		.76*	.08	.67*	.94*	*NA*	.91*	.82*	.62*	.63*	.64*	.27	-.13	.81*	.86*	.48
CI			.17	.61*	.80*	.20	.86*	.37	.65*	.16	.72*	.59*	.73*	.93*	.78*	.82*
El				.23	.22	.61*	.51*	.11	.37	.58*	.20	.03	.11	.08	.56*	-.03
EE					.80*	.06	.68*	.61*	.80*	.47	.44	.35	-.29	.59*	.67*	.35
ET						.78*	.90*	.77*	.77*	.38	.76*	.56*	.01	.83*	.84*	.58*
Ex							.43	.51	.37	.22	.31	.81*	*NA*	.33	.38	-.62
Fa								.64*	.71*	.47	.78*	.43	.35	.80*	.83*	.56*
Fl									.56*	.37	.54*	.28	-.59	.37	.64*	.16
Hu										.39	.58*	.41	.08	.57*	.62*	.33
IV											.02	-.21	-.25	.12	.38	-.13
MA												.50*	.46	.70*	.73*	.59*
Or													-.03	.73*	.47	.73*
RC														.60	.20	.50
RI															.78*	.86*
SA																.69*
UV																

Notes. AT: Abstractness of titles. CI: Colorfulness of imagery. El: Elaboration. EE: Emotional expressiveness. ET: Expressiveness of titles. Ex: Extending boundaries. Fa: Fantasy. Fl: Fluency. Hu: Humor. IV: Internal visualization. MA: Movement or action. Or: Originality. RC: Resistance to closure. RI: Richness of imagery. SA: Storytelling articulateness. UV: Unusual visualization. *NA* indicates that the correlation coefficient was not estimated because there were not mutual items. * shows the significant coefficients at p < .001.

Discussion

Item difficulty

Item difficulties, strata values, and item reliability indices were estimated to examine the first hypothesis (the items are equally difficult). The strata values below 2.00 denote one item group for internal visualization and humor in terms of difficulty (Linacre, 2013; Wright & Masters, 2002). The strata value for elaboration is over 2.00, but the items are centered on 0 logits, and the difference between the easiest and most difficult items is 0.36 logits. In other words, the items are at the same difficulty level for elaboration, but the six response categories resulted in a strata value over 2.00 (Bond & Fox, 2015). The strata values are greater than 2.00 for the other creative thinking skills, and this finding implies that the items are not at the same difficulty level (Linacre, 2013; Wright & Masters, 2002). Note that the strata values for abstractness of titles and resistance to premature closure are 5.66 and 5.87, respectively, which are greater than the numbers of response categories.

The item reliability indices indicate that the difficulty levels of the items were estimated with a high precision for all the creative thinking skills, except elaboration, internal visualization, and humor (Bond & Fox, 2015; Fox & Jones, 1998). Therefore, it is highly probable that the items vary in difficulty for the overwhelming majority of the creative thinking skills. The item order for elaboration, internal visualization, and humor may change in other samples (Bond & Fox, 2015). However, this change may not produce more than one item group with regard to difficulty. Thus, based solely on the

findings of the present study, it can be argued that the first hypothesis is supported for only elaboration, internal visualization, and humor.

These results raise significant concerns about the validity of the test scores. This is because the TTCT-Figural Form A seems to produce unfair scores on certain creative thinking skills for some examinees who respond to more challenging items but receive the same score as other examinees who respond to less challenging items. To overcome this issue, the scoring of the items may need to be revised based on item difficulty for several creative thinking skills (Bond & Fox, 2015). In this way, each examinee receives a score that precisely reflects her or his performance on an item.

Response category difficulty

Rasch-Thurstone thresholds (Linacre, 2003; Thurstone, 1928) were estimated to address the second hypothesis (the response categories are equally distributed). The findings indicate that each additional detail does not require a considerably different increase in ability level to be added to a response because the difference between any two subsequent categories is much lower than 1.40 logits for elaboration (Linacre, 2002b). Given that the items on Form A are at the same difficulty level for elaboration in the present study, the findings validate the current interpretation of responses for elaboration.

The second hypothesis is also supported for resistance to premature closure, as the difference between any two response categories is almost the same across the response categories. The Rasch-Thurstone thresholds (Linacre, 2003; Thurstone, 1928) are below 1.40, in contrast to Linacre's (2002b) suggestion, but this can be neglected because the response category difficulties are very close to the cutoff point. It appears that each response category measures a narrow section of resistance to premature closure (Bond & Fox, 2015). Therefore, researchers and practitioners can continue using the equally distributed 3-point scale for scoring. However, the scoring may need to be revised based on item difficulty because there are multiple item groups for resistance to premature closure. In this way, raw scores could reflect more precise estimations of the examinees' performances.

The second hypothesis is not supported for abstractness of titles because the Rasch-Thrustone thresholds (Linacre, 2003; Thurstone, 1928) indicate that the response categories are not distributed evenly. Therefore, the current scoring method is not verified for abstractness of titles. The findings suggest that the scoring needs to be changed based on item difficulty and response category difficulty (Bond & Fox, 2015). In this way, the revised scoring can capture the ability level differences across the response categories for second-grade students.

It appears that the overwhelming majority of the second-grade students are not sufficiently developed to produce titles that convey abstract thought. Only four examinees produced the highest level of titles in the present study. This finding was expected and is aligned with Torrance's (1979) results. Nevertheless, some second graders may generate titles that correspond to the highest response category. Therefore, researchers and practitioners can continue using the 4-point scale scoring after revising it.

Dimensionality of the creative thinking skills

Point-measure correlation coefficients, fit statistics, and eigenvalues were estimated to investigate the third hypothesis (the items measure a single cognitive factor for each individual creative thinking skill). The positive point-measure correlation coefficients estimated in the present study suggest that the items operate uniformly in the same direction. In other words, the items measure a single cognitive factor for each creative thinking skill (Bond and Fox, 2015; Linacre, 2017).

The infit statistics are aligned with the point-measure correlations. The IMNSQ and S-INFIT values imply that each interpretation of responses provides information about a single cognitive factor (Linacre, 2009). Furthermore, the findings show that the items work productively for examinees whose ability levels are close to the difficulty levels of the items (Linacre, 2009). The outfit statistics also

support the unidimensional structure for the items for each creative thinking skill, except the item in Subtest 1 for expressiveness of titles and storytelling articulateness (Linacre, 2002a). Moreover, the outfit statistics suggest that the overwhelming majority of the items work productively for examinees whose ability levels are much higher or lower than the difficulty levels of the items (Linacre, 2009).

Estimating high OMNSQ and S-OFIT values for the item in Subtest 1 for expressiveness of titles and storytelling articulateness was not anticipated. This is because the item is hypothesized to measure a single cognitive factor (Torrance et al., 1992). The data of the present study show that some examinees whose ability levels are higher than the difficulty level of the item did not manifest those two creative thinking skills in their responses. The findings imply that additional factors (e.g., fatigue, boredom, and another thinking skill) seem to impact responses for that particular item. Nevertheless, it is also reasonable to infer that the high OMNSQ and S-OFIT values may be due to a few random responses given to the item in Subtest 1, as point-measure correlation coefficients and eigenvalues denote a single cognitive factor.

Finally, all the eigenvalues are below 3.00, and no single eigenvalue is considerably larger than the others. These findings suggest that the cognitive factor extracted through the Rasch-based analyses for each creative thinking skill is the major construct impacting examinee responses for that creative thinking skill (Linacre, 2017).

In conclusion, the scores on each creative thinking skill reflect potential on a single cognitive factor, and the third hypothesis is supported by the Rasch-based analyses for all the creative thinking skills. Therefore, using a scoring method based on a sole criterion for a creative thinking skill is sufficient for capturing examinee performances on a particular item.

Dimensionality of the TTCT-Figural

Pearson correlation (Pearson, 1909) coefficients were estimated to examine the fourth hypothesis (each creative thinking skill is a different cognitive factor). The correlation coefficients over .90 indicate that the examinees exhibited nearly identical performances on the items for particular creative thinking skills, including abstractness of titles, expressiveness of titles, fantasy, richness of imagery, and colorfulness of imagery. In other words, some creative thinking skills seem to require almost the same trait level to be manifested on the same item. One may argue that this situation occurred by chance and that each creative thinking skill is a different cognitive factor. However, because all the thinking skills on the TTCT-Figural Forms are related to creativity (Torrance, 1984), it is more reasonable to infer that the same underlying cognitive factor impacts examinee responses for pairs of thinking skills with high correlations.

It is not surprising to obtain a high correlation coefficient between certain pairs, including abstractness of titles and expressiveness of titles, as well as colorfulness of imagery and richness of imagery. The reason is that the items are scored on similar criteria for these pairs (Torrance et al., 1992). The scoring criteria used for fantasy are not similar to the criteria used for abstractness of titles and expressiveness of titles. Hence, estimating a high correlation coefficient between fantasy and each of the other creative thinking skills is not anticipated. However, the findings indicate that the same underlying cognitive factor impacts examinee responses for these three creative thinking skills. That cognitive factor probably enables an examinee to create drawings that manifest fantasy (e.g., characters from movies and books) and induces a certain level of tension in the examinee to express her or his fantasy-driven ideas through more abstract and expressive titles. If future studies support these findings, some creative thinking skills will need to be removed from the test to avoid inflating test scores.

At this point, the relation between fluency and originality should not be overlooked. Previous studies contended that fluency and originality are highly correlated (Chase, 1985; Ferrándiz et al., 2017; Heausler & Thompson, 1988; Runco & Mraz, 1992; Treffinger, 1985). The magnitude of the

coefficient is expected to increase up to .80 when total scores on each creative thinking skill are considered (Acar, 2023; Forthmann et al., 2020).

Despite previous studies, the correlation coefficient between fluency and originality is .28 in the present study. This finding suggests that the difficulty level of an item for fluency does not affect the originality of responses given to that item. Rather, the originality of responses is impacted by the shape of the stimulus, and this argument was supported in a recent study (Acar, 2023). In his study, Acar (2023) worked with 477 adults and found that the correlation between fluency and originality dropped when scores on Subtest 3 were excluded. In other words, using the same stimulus inflated the correlation coefficient between fluency and originality. Note that each item in Subtest 3 is a pair of parallel lines and that the items in Subtests 1 and 2 are different stimuli.

Overall, the fourth hypothesis is partially validated. The correlation analyses suggest that each creative thinking skill is not a different cognitive factor and that the TTCT-Figural Forms do not measure 18 creative thinking skills. This finding poses a problem with regard to the test scores because certain cognitive factors seem to be scored multiple times under different creative thinking skills, and thus the test scores for some examinees are inflated. This argument is particularly relevant to abstractness of titles and expressiveness of titles, and colorfulness of imagery and richness of imagery due to the similarities in the scoring criteria. However, attention needs to be paid to fantasy as well. The findings indicate that each of the other creative thinking skills is presumably a different cognitive factor.

The findings of the present study complemented the findings of the factor analysis studies on the TTCT-Figural Forms (Antunes & Almeida, 2007; Auzmendi et al., 1996; Bart et al., 2017; Clapham, 1998, 2004; Ferrando, 2006; Humble et al., 2018; Kim, 2006; Kim et al., 2006; Krumm et al., 2014; Oliveira, 2007; Prieto et al., 2006; Said-Metwaly et al., 2018). Previous studies were based on the hypotheses that the items measure a single cognitive factor for each creative thinking skill and that each norm-referenced creative thinking skill was a different cognitive factor. Because both of these hypotheses find support in the present study, one could argue that conducting a subscore-level factor analysis on the TTCT-Figural Forms is appropriate for identifying overarching factors (e.g., personality types), which encompass various cognitive factors (e.g., fluency, originality, and abstractness of titles) related to creativity.

Implications

The findings of the study have significant implications that concern practitioners administering the TTCT-Figural Forms and the Scholastic Testing Service, the company that holds the rights to the tests. It is crucial to consider the varying difficulty levels of the items and response categories to revise the scoring. Adapting the scoring method to account the variations in item and response category difficulty can enhance the validity and fairness of the test scores. On other hand, the current scoring criteria can be applied consistently across the items, as a single cognitive factor impacts responses for each creative thinking skill. However, a noteworthy consideration arises from the high correlation coefficients estimated among certain creative thinking skills. This finding suggests a careful reevaluation of the test content. It is prudent to revise the test by removing redundant creative thinking skills to ensure more valid and fair assessments.

Limitations

Linacre (1994) stated that a sample size as low as 150 is large enough to perform Rasch analyses. According to Arrindell and van der Ende (1985), the minimum sample size required for principal component analysis is 100. Thus, the sample size was large enough for both Rasch-based and principal component analyses. However, future research should be conducted with larger samples so that synthesis of incomplete figures and synthesis of lines can also be investigated. Additionally, larger samples can produce more precise item statistics.

The items were scored based on the scoring manual published in the United States (Torrance, 2017), and this situation possibly impacted the scoring for originality. Because there was no zero-originality list for a Turkish sample, whether a response was original was determined using the US manual (Torrance, 2017). It is likely that a different zero-originality list would have been used if there were a Turkish manual and that this situation may have potentially impacted originality analyses.

Directions for future research

This study is the first one to examine the items on the TTCT-Figural Form A using Rasch measurement theory. More research that uses Rasch measurement theory to examine the items on each form of the TTCT-Figural is needed. Future research should involve samples comprising adults and students from different grade levels to test the hypotheses examined in the present study. Moreover, differential item functioning should be investigated for gender, ethnicity, and socioeconomic status. Finally, the reliability of the test scores and factors impacting reliability should be examined in more detail.

Conclusion

The present study addresses four hypotheses concerning the items on the TTCT-Figural Form A. The insights gained from this study shed light on critical aspects of the test's structure and scoring methods. The observed variability in item and response category difficulties underscores the necessity for a nuanced approach to scoring. Additionally, the identification of a single cognitive factor for each creative thinking skill provides strong evidence for the test's construct validity. The low and moderate correlations among certain creative thinking skills offer further evidence for construct validity. However, considerably high correlations among certain creative thinking skills warrant a reevaluation of the test's content. Collectively, the findings of the present study contribute to a more comprehensive understanding of the TTCT-Figural Forms' properties and offer valuable directions for optimizing the scoring procedures.

References

Acar, S. (2023). Does the task structure impact the fluency confound in divergent thinking? An investigation with TTCT-figural. *Creativity Research Journal, 35*(1), 1-14. https://doi.org/10.1080/10400419.2022.2044656

Aliotti, N. Britt, M., & Haskins, G. (1975). Relationships among creativity, intelligence, and achievement measures in upward bound students. *Psychology in the Schools, 12*(4), 423-427.

Almeida, L. S., Prieto, L. P., Ferrando, M., Oliveira, E., & Ferrándiz, C. (2008). Torrance test of creative thinking: The question of its construct validity. *Thinking Skills and Creativity, 3*(1), 53–58. https://doi.org/10.1016/j.tsc.2008.03.003

Andrich, D. (1978). Application of a psychometric model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, *2*(4), 581-594. https://doi.org/10.1177/014662167800200413

Antunes, A. P., & Almeida, L. (2007). Avaliar a criatividad: Contibutos para validade de alguns subtestes do TPCT (Torrance pensamento creativity test). *Revista Psicologia e Educaçao, 6*(1), 37–53.

Arrindell, W. A., & van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement, 9*(2), 165-178. https://doi.org/10.1177/014662168500900205

Auzmendi, E., Villa, A., & Abedi, J. (1996). Reliability and validity of a newly constructed multiple-choice creativity instrument. *Creativity Research Journal, 9*(1), 89-95. https://doi.org/10.1207/s15326934crj0901_8

Awamleh, H., Al Farah, Y., & El-Zraigat, I., (2012). The level of creative abilities dimensions according to Torrance formal test (B) and their relationship with some variables (sex, age, GPA). *International Education Studies, 5*(6), 138-147.

Ball, O. E. & Torrance, E. P. (1984). *Torrance tests of creative thinking: Streamlined scoring workbook: Figural A*. Bensenville, IL: Scholastic Testing Service.

Bart, W. M., Hokanson, B., & Can, I. (2017). An investigation of the factor structure of the Torrance tests of creative thinking. *Educational Sciences: Theory & Practice, 17*(2), 515-528. https://doi.org/10.12738/estp.2017.2.0051

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Buitink, M. (2017). *A gender-comparison between verbal and figural divergent thinking in 4-year-old children, using the Torrance test of creative thinking and the alternative uses task* [Unpublished master's thesis]. Utrecht University. Retrieved from: https://studenttheses.uu.nl/bitstream/handle/20.500.12932/26834/Masterthesis%20Buitink%2c%20M-3762688.pdf?sequence=1&isAllowed=y

Campos, A., Lopez, A., Gonzalez, M. A., & Perez-Fabello, M. J. (2000). Aspects of creativity affected by imaging capacity. *North American Journal of Psychology, 2*(2), 313-322.

Chase, C. I. (1985). Review of the Torrance tests of creative thinking. In J. V. Mitchell Jr. (Ed.), *The ninth mental measurements yearbook* (pp. 1631-1632). Buros Institute of Mental Measurement. University of Nebraska

Cheng, Y., Kim, K. H., & Hull, M. F. (2010). Comparisons of creative styles and personality types between American and Taiwanese college students and the relationship between creative potential and personality types. *Psychology of Aesthetics, Creativity, and the Arts, 4*(2), 103–112. https://doi.org/10.1037/a0017430

Cho, S. H., Nijenhuis, J. T., van Vianen, A. E. M., Kim, H. B., & Lee, K. H. (2010). The relationship between diverse components of intelligence and creativity. *The Journal of Creative Behavior, 44*(2), 125-137. https://doi.org/10.1002/j.2162-6057.2010.tb01329.x

Clapham, M. M. (1998). Structure of figural forms A and B of the Torrance tests of creative thinking. Educational and Psychological Measurement, 58(2), 275-283. https://doi.org/10.1177/0013164498058002010

Clapham, M. M. (2004). The convergent validity of the Torrance tests of creative thinking and creativity interest inventories. *Educational and Psychological Measurement, 64*(5), 828-841. https://doi.org/10.1177/0013164404263883

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurement* (4th ed.). Mayfield.

Cramond, C., Matthews-Morgan, J., Bandalos, D., & Zuo, L. (2005). A report on the 40-year follow-up of the Torrance tests of creative thinking: Alive and well in the new millennium. *Gifted Child Quarterly, 49*(4), 283-291. https://doi.org/10.1177/001698620504900402

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. doi:10.1007/bf02310555

Ferrándiz, C., Ferrando, M., Soto, G.M., Sainz, M., & Prieto, M.D. (2017). Divergent thinking and its dimensions: What we talk about and what we evaluate? *Anales De Psicologia, 33*(1), 40-47.

Ferrando, M. (2004). *Creatividad e inteligencias multiples* [Creativity and multiple intelligences] [Unpublished doctoral dissertation]. Universidad de Murcia.

Ferrando, M. (2006). *Creatividad e inteligencia emocional: Un estudio empırico en alumnos con altas habilidades* [Unpublished doctoral dissertation]. Universidad de Murcia.

Ferrando, M., Ferrándiz, C., Bermejo, M., Sánchez, C., Parra, J., & Prieto, M. (2007). Estructura interna y baremación del test de pensamiento creativo de Torrance. *Psicothema, 19*(3), 489-496.

Field, A. (2009). *Discovering statistics using IBM SPSS statistics* (3rd ed.). Sage Publications.

Forthmann, B., Szardenings, C., & Holling, H. (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*(1), 94–112. doi:10.1037/aca0000196

Heausler, N. L., & Thompson, B. (1988). Structure of the Torrance tests of creative thinking. *Educational and Psychological Measurement, 48*(2), 463-468. https://doi.org/10.1177/0013164488482021

Hermon, A., Palomera, B., & Hokanson, B. (2018, October 25). *Creativity and family income: Comparing creativity in eighth graders* [Conference session]. The Association for Educational Communications and Technology (AECT) 2018 Convention, Kansas City, MO, United States.

Humble, S., Dixon, P., & Mpofu, E. (2018). Factor structure of the Torrance tests of creative thinking Figural form A in Kiswahili speaking children: Multidimensionality and influences on creative behavior. *Thinking Skills and Creativity,* 27, 33-44. https://doi.org/10.1016/j.tsc.2017.11.005

Hunsaker, S. L., Abeel, L. B., & Callahan, C. M. (1991, June). *Instrument use in the identification of gifted and talented children*. Paper presented at the Meeting of the Jacob K. Javits Gifted and Talented Education Program Grant Recipients, Washington, DC.

Johnson, R. A. (1974). Differential effects of reward versus no-reward instructions on the creative thinking of two economic levels of elementary school children. *Journal of Educational Psychology, 66*(4), 530-633.

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment.* John Wiley & Sons.
Kaufman, J. C., Plucker, J. A., & Russell, C. M. (2012). Identifying and assessing creativity as a component of giftedness. *Journal of Psychoeducational Assessment, 30*(1), 60-73.
Kazelskis, R. (1972). The convergent, divergent and factorial validity of the Torrance figural tests of creativity. *Southern Journal of Educational Research, 6*(3), 123-129.
Kim, K. H. (2006). Is creativity unidimensional or multidimensional? Analyses of the Torrance tests of creative thinking. *Creativity Research Journal, 18*(3), 251–259. https://doi.org/10.1207/s15326934crj1803_2
Kim, K. H., Cramond, B., & Bandalos, D. L. (2006). The latent structure and measurement invariance of scores on the Torrance tests of creative thinking-Figural. *Educational and Psychological Measurement, 66*(3), 459-477. https://doi.org/10.1177/0013164405282456
Kim, K. H., & VanTassel-Baska, J. (2010). The relationship between creativity and behavior problems among underachieving elementary and high school students. *Creativity Research Journal, 22*(2), 185–193.
Krumm, G., & Lemos, V. (2011). Análisis de las propiedades psicométricas de la prueba de figuras del test de pensamiento creativo de Torrance (TTCT). Forma B, en la provincia de Entre Ríos, Argentina [Analysis of the psychometric properties of the Figural Torrance Test of Creative Thinking (TTCT) Form B in the province of Entre Rios, Argentina]. In M. C. Richaud de Minzi & V. Lemos (Eds.), *Psicología y otras ciencias del comportamiento. Compendio de investigaciones actuales* (pp. 731–748). Universidad Adventista del Plata.
Krumm, G., Lemos, V., & Filippetti, V. A. (2014). Factor structure of the Torrance tests of creative thinking figural form b in Spanish-speaking children: Measurement invariance across gender. *Creativity Research Journal, 26*(1), 72–81. https://doi.org/10.1080/10400419.2013.843908
Likert, R. (1974). *The method of constructing an attitude scale*. Routledge.
Linacre, J. M. (1994). Sample size and item calibration [or Person Measure] stability. *Rasch Measurement Transactions, 7*(4), 328. Retrieved from https://www.rasch.org/rmt/rmt74m.htm
Linacre, J. M. (1998) Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement, 2*(3), 266-283.
Linacre, J. M. (2002a). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), *878. Retrieved from https://www.rasch.org/rmt/rmt162f.htm*
Linacre, J. M. (2002b). Understanding Rasch measurement: Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.
Linacre, J. M. (2003). Estimating 50% cumulative probability (Rasch-Thurstone) thresholds. *Rasch Measurement Transactions, 16*(4), 901. Retrieved from https://www.rasch.org/rmt/rmt164e.htm
Linacre, J. M. (2009). *Facets* (Version 3.65.0). MyCommerce-ShareIt
Linacre, J. M. (2016). *A user's guide to Winsteps Rasch-model computer programs*. MESA Press.
Linacre, J. M. (2017). *A user's guide to Winsteps Rasch-model computer programs*. MESA Press.
Linacre, J. M., & Tennant, A. (2009). More about critical eigenvalue sizes (variances) in standardized-residual principal components analysis (PCA). *Rasch Measurement Transactions, 23*(3), 1228. Retrieved from https://www.rasch.org/rmt/rmt233f.htm
Liu, H. Y. (2020). Factors affecting nursing students' creativity in Taiwan: Exploring the moderating role of creative personality. *Nurse education today, 88*, 104367. https://doi.org/10.1016/j.nedt.2020.104367
Liu, H. Y. (2022). Promoting creativity of nursing students in different teaching and learning settings: A quasi-experimental study. *Nurse Education Today, 108*, 105216. https://doi.org/10.1016/j.nedt.2021.105216
Liu, H. Y., Chang, C. C., Wang, I. T., & Chao, S. Y. (2020). The association between creativity, creative components of personality, and innovation among Taiwanese nursing students. *Thinking Skills and Creativity, 35*, 100629. https://doi.org/10.1016/j.tsc.2020.100629
López, O. (2001). *Evaluación y desarrollo de la creatividad* [Unpublished doctoral dissertation]. Universidad de Murcia.
Matud, M. P., & Grande, C. R. J. (2007). Gender differences in creative thinking. *Personality and Individual Differences, 43*(5), 1137–1147.
Mednick, S. A. (1968). The remote associates test. *The Journal of Creative Behavior, 2*(3), 213–214.
Ogletree, E. J., & Ujlaki, W. (1973). Effects of social class status on tests of creative behavior. *The Journal of Educational Research, 67*(4), 149-152.
Oliveira, E. P. L. (2007). *Alunos sobredotados: A aceleraçao escolar como resposta educativa* [Unpublished doctoral dissertation]. Universidade do Minho.
O'Rourke, N., Hatcher, L., & Stepanski, E. J. (2005). *A step-by- step approach to using SAS for univariate & multivariate statistics*. SAS Institute.
Palaniappan, A. K. (2008). Influence of intelligence on the relationship between creativity and academic achievement: A comparative study. *The International Journal of Learning, 15*(7), 267-277.
Palaniappan, A. K., & Torrance, E. P. (2001). Comparison between regular and streamlined versions of scoring of Torrance tests of creative thinking. *The Korean Journal of Thinking & Problem Solving, 11*(2), 5-7.

Pearson. K. (1909). Determination of the coefficient of correlation. *Science, 30*(757), 23-25. Doi: 10.1126/science.30.757.23

Prieto, M. D., López, O., Ferrándiz, C., & Bermejo, M. R. (2003). Adaptación de la prueba figurativa del test de pensamiento creativo de Torrance en una muestra de los primeros niveles educativos [Adaptation of the figurative test of Creative Thinking Torrance Test in a sample of the first educational levels]. *Revista de Investigación Educativa, 21*(1), 201– 213.

Prieto, M. D., Parra, J., Ferrándo, M., Ferrándiz, C., Bermejo, M. R., & Sánchez, C. (2006). Creative abilities in early childhood. *Journal of Early Childhood Research, 4*(3), 277-290. https://doi.org/10.1177/1476718X06067580

Rasch, G. (1960). P*robabilistic models for some intelligence and attainment tests.* Danmarks Paedagogiske Institut.

Runco, M. A., Plucker, J. A., & Lim, W. (2001). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal, 13*(3-4), 393-400.

Runco, M. A., & Mraz, W. (1992). Scoring divergent thinking tests using total ideational output and a creativity index. *Educational and Psychological Measurement, 52*(1)*,* 213-221. DOI: 10.1177/001316449205200126

Runco, M. A., Millar, G. Acar, A., & Cramond, B. (2010). Torrance tests of creative thinking as predictors of personal and public achievement: A fifty-year follow-up. *Creativity Research Journal, 22*(4), 361–368.

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., & Van Den Noortgate, W. (2018). The factor structure of the figural Torrance tests of creative thinking: A meta-confirmatory factor analysis. *Creativity Research Journal, 30*(4), 352-360. doi: 10.1080/10400419.2018.1530534

Saeki, N., Fan, X., & Van Dusen, L. V. (2001). A comparative study of creative thinking of American and Japanese college students. *The Journal of Creative Behavior, 35*(1), 24-38. https://doi.org/10.1002/j.2162-6057.2001.tb01219.x

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology 33*(4), 529-554. https://doi.org/10.1086/214483

Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology, 22*(4), 358–376. https://doi.org/10.1037/h0076640

Torrance, E. P. (1979). *The search for satori and creativity.* Creative Education Foundation.

Torrance, E. P. (1966). *Torrance tests of creative thinking: Norms technical manual (research edition)*. Personnel Press.

Torrance, E. P. (1984). *Torrance test of creative thinking: Norms technical manual, Figural (streamlined) forms A & B.* Scholastic Testing Service.

Torrance, E. P. (1984). *Torrance test of creative thinking: Norms technical manual, Figural (streamlined) forms A & B.* Scholastic Testing Service.

Torrance, E. P. (2000). *Research review for the Torrance tests of creative thinking Figural and verbal forms A and B.* Scholastic Testing Service.

Torrance, E. P. (2006a). *Thinking creatively with pictures: Figural booklet A.* Scholastic Testing Service.

Torrance, E. P. (2006b). *Thinking creatively with pictures: Figural booklet B.* Scholastic Testing Service.

Torrance, E. P. (2008). *Torrance test of creative thinking: Norms technical manual, Figural (streamlined) forms A & B.* Scholastic Testing Service.

Torrance, E. P. (2017). *Torrance test of creative thinking: Norms technical manual, Figural (streamlined) forms A & B.* Scholastic Testing Service.

Torrance, E. P., Ball, O. E., & Safter, H. T. (1992). *Torrance tests of creative thinking: Streamlined scoring guide for Figural forms A & B.* Scholastic Testing Service.

Tran, N. (2004). Cultural dimensions in creativity: A preliminary study about creativity among the Vietnamese people in America. *Education and Society, 22*(2), 71-81.

Treffinger, D. J. (1985). Review of the Torrance tests of creative thinking. In J. V. Mitchell Jr. (Ed.), *The ninth mental measurements yearbook* (pp. 1632–1634). Buros Institute of Mental Measurements

Voss, D. H. (1997). *Determining test fairness and differential validity of scores for the Torrance tests of creative thinking for kindergarten students* [Unpublished doctoral dissertation]. Texas Tech University.

Wechsler, S. (2006). Validity of the Torrance tests of creative thinking to the Brazilian culture. *Creativity research journal, 18*(1), 15-25. https://doi.org/10.1207/s15326934crj1801_3

Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin, 50*(5), 362-370.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370. Retrieved from https://www.rasch.org/rmt/rmt83b.htm#:~:text=These%20statistics%20have%20expectation%201.0,predictable%20than%20the%20model%20expects.

Wright, B. D., & Masters, G. N. (2002). Number of person or item strata: (4*Separation + 1)/3. *Rasch Measurement Transactions, 16*(3), 888. Retrieved from http://www.rasch.org/rmt/rmt163f.htm

Yong, L. M. S. (1994). Relations between creativity and intelligence among Malaysian pupils. *Perceptual and Motor Skills, 79*(2), 739-742.

Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal, 23*(1), 24–37. http://dx.doi.org/10.1080/10400419.2011.545713

Zhang, W., Ren, P., & Deng, L. (2020). Gender differences in the creativity–academic achievement relationship: A study from China. *The Journal of Creative Behavior, 54*(3), 725-732. doi: 10.1002/jocb.387

Acknowledgment

I express my deepest gratitude to Bonnie Cramond for her guidance and support for this study, as it is based on my doctoral research. I also thank Burak Türkman for his support in estimating inter-rater reliability.

About the Author

Dr. Süreyya Yörük is from Türkiye. He is an assistant professor at the Department of Special Education at Marmara University. He received his bachelor's degree from Istanbul University with an emphasis on gifted education. Later, he worked for two years as a gifted teacher in a private institution in Istanbul. Afterwards, he moved to the United States for his graduate endeavors. Süreyya Yörük completed his master's and doctorate at the University of Georgia with an emphasis on gifted education and creativity. Süreyya Yörük is interested in talent development, identification of gifted students, assessment of creativity, and educational statistics. He currently teaches undergraduate and graduate courses at Marmara University.

e-Mail: yoruk.sureyya@gmail.com