

A DEEPER DIVE INTO THE MEASUREMENT INVARIANCE
OF THE TORRANCE TEST OF CREATIVE THINKING:
CONFIRMATORY AND EXPLORATORY FACTOR ANALYSIS
ACROSS GENDER AND ETHNICITY

by

YOOJOONG KIM

(Under the Direction of Denis Dumas)

ABSTRACT

The Torrance Test of Creative Thinking-Figural (TTCT-F) is the most extensively used creativity measure in U.S. schools, especially for gifted identification. TTCT-F has been thought to be culturally fair across ethnicities and gender, although this belief has rarely been empirically tested. Our study examined the latent structure of the TTCT-F scales: Fluency, Originality, Elaboration, Abstractness of Titles (AT), and Resistance to Premature Closure (RPC), and then investigated measurement invariance across over- and under-represented ethnicities and across genders. Using data obtained from 379 U.S. elementary school students, we identified that the Innovative-Adaptive model, with an additional cross-loading for the RPC scale, was the best fit. We also found that this two-factor model was invariant across under- and over-represented ethnicities at configural, metric, and scalar levels. In contrast, even configural invariance did not hold across both boys and girls, suggesting that additional research is needed to understand creative thinking across genders.

INDEX WORDS: Torrance Test of Creative Thinking-Figural, Elementary students,
Innovative-Adaptive Model, Confirmatory Factor Analysis, Measurement
Invariance

A DEEPER DIVE INTO THE MEASUREMENT INVARIANCE
OF THE TORRANCE TEST OF CREATIVE THINKING:
CONFIRMATORY AND EXPLORATORY FACTOR ANALYSIS
ACROSS GENDER AND ETHNICITY

by

YOOJOONG KIM

M.A., Seoul National University, South Korea, 2021

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment
of the Requirements for the Degree

MASTER OF ART

ATHENS, GEORGIA

2025

© 2025

Yoojoong Kim

All Rights Reserved

A DEEPER DIVE INTO THE MEASUREMENT INVARIANCE
OF THE TORRANCE TEST OF CREATIVE THINKING:
CONFIRMATORY AND EXPLORATORY FACTOR ANALYSIS
ACROSS GENDER AND ETHNICITY

by

YOOJOONG KIM

Major Professor: Denis Dumas
Committee: George Engelhard Jr.
Matthew J. Madison

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2025

TABLE OF CONTENTS

	Page
LIST OF TABLES	iv
LIST OF FIGURES	iv
CHAPTER	
1 INTRODUCTION	1
2 LITERATURE REVIEW	3
2.1 TTCT-F and Gifted Identification	3
2.2 Gifted Identification Across Ethnicities and Genders	5
2.3 The Structure of the TTCT-F	7
2.4 Measurement Invariance in Educational Testing.....	9
2.5 Does the Structure of the TTCT-F Hold Across Ethnicities and Gender?	12
3 METHODOLOGY	14
3.1 Goals of Current Study	14
3.2 Participants.....	14
3.3 Focal Measure: TTCT-F	15
3.4 Data Collection Procedures.....	17
3.5 Data Analysis Overview	17
4 RESULTS AND ANALYSIS.....	19
4.1 Bivariate Correlations Among TTCT Scores	19
4.2 Fitting CFA Models to All Student Data	20
4.3 Multi-Group CFA Models for Measurement Invariance	24

4.4 Delving into Bivariate Correlation Across Gender.....	26
4.5 EFA by Gender	27
5 DISCUSSION.....	29
5.1 Key Findings.....	29
5.2 Practical Implications for Gifted Identification	33
5.3 Limitations and Future Directions	35
5.4 Conclusion	36
REFERENCES	37

LIST OF TABLES

	Page
Table 1: Bivariate Correlations among the TTCT-F Scales for the Entire Sample	54
Table 2: Descriptive Statistics of the TTCT-F Scales for the Entire Sample	55
Table 3: Correlations Between the TTCT-F Scales for Gender Groups.....	56
Table 4: Model Comparison via Fit Indices	57
Table 5: Measurement Invariance Fit Comparisons across Overrepresented and Underrepresented Ethnicities	58
Table 6: Factor Loadings of the TTCT-F Scales for Boys and Girls from the Exploratory Factor Analysis.....	59

LIST OF FIGURES

	Page
Figure 1: Theoretically Plausible Factor Models that were Fit	60
Figure 2: The Best Fitting Model for the Entire Elementary Student Sample, with Standardized Coefficients	61
Figure 3: Parallel Analysis Results for Boys	62
Figure 4: Parallel Analysis Results for Girls	63

CHAPTER 1

INTRODUCTION

For over five decades, the Torrance Test of Creative Thinking (TTCT; Torrance, 1966/2017) has been the most influential and extensively used creative thinking measure in U.S. schools, especially for gifted identification (Acar et al., 2021; Kaufman et al., 2012; Kim, 2011). Believed to be culturally fair across diverse ethnicities, genders, socioeconomic statuses, and languages, the TTCT is considered to play a vital role in promoting equitable access to gifted services, thereby including a wide range of students from diverse cultural backgrounds (Bart et al., 2017; Kim et al., 2006; Lee et al., 2024).

The TTCT is conceptually focused on tapping students' Divergent Thinking (DT; Guilford, 1967), referring to the ability to produce multiple original ideas from the same prompt. According to Guilford's Structure-of-Intellect Model (SOI; 1956, 1988), DT can be generally represented by ideational Fluency (quantity of ideas within a set amount of time), Originality (statistically infrequency of ideas), Flexibility (ideas from different categories), and Elaboration (detail of ideas). Drawing upon the conceptual framework of DT and DT tasks introduced by Guilford and his colleagues (1961, 1963), Paul Torrance created the TTCT which consists of two disparate modalities, namely Verbal and Figural. These tests have been re-normed six times and translated into more than 35 languages so far (Kim, 2010; Millar, 2002). The TTCT-Figural (TTCT-F) is generally regarded as a more reliable and valid option than the Verbal test, especially when assessing students from culturally diverse backgrounds (Cramond,

1993; Kim, 2011). While there is evidence of internal consistency (Acar et al., 2024), the actual psychometric evidence for the claims of cultural fairness of the TTCT-F is somewhat limited.

With the understanding that creative thinking should not be limited to divergent thinking, the earlier versions of the TTCT-F were thought to be insufficient in capturing the breadth of creative thinking, which led to the addition of additional scoring methods and scale: Abstractness of the Titles (AT), Resistance to Premature Closure (RPC), and the creative strengths checklist (Ball & Torrance, 1984; Hébert et al., 2002). Also, flexibility was removed due to its high overlap with fluency scores, which is known as the fluency confound (Clark & Mirels, 1970; Forthmann et al., 2020). At the same time, a criterion-referenced subscale that captures 13 creative strengths was added into the scoring system in the 1984 revision (Ball & Torrance, 1984; Kim, 2006). While the current scoring scales are considered to feature a broad scope of creative thinking and the research has addressed how the TTCT-F measures the latent constructs of creative thinking (Acar et al., 2023; Said-Metwaly et al., 2018), less is known about the correlates of the latent factors assessed by TTCT-F and how educators and school administrators can interpret these scores for educational purposes. Furthermore, beyond the previous findings that there tend to be no significant score differences in the TTCT-F scales across ethnicities and genders (Krumm et al., 2016; Torrance, 1971), little is known about whether the TTCT-F measures creative thinking in the same way across these groups.

CHAPTER 2

LITERATURE REVIEW

2.1 TTCT-F and Gifted Identification

Torrance advocated for including creativity tests as one of the multiple criteria in gifted identification (Torrance, 2004), believing that it could bring equity benefits for getting gifted and talented programs and services to students from culturally diverse backgrounds (Grantham, 2013; Lee et al., 2024; Torrance, 1971). He suggested that, as creativity is less influenced than other measures by past educational opportunity, students who may otherwise be missed can be identified using the TTCT-F (Torrance, 2004). Beyond merely using IQ tests and standardized academic achievement tests which can potentially put marginalized students at a disadvantage, the inclusion of the creativity assessments has been considered to improve underrepresentation of Culturally Linguistically, and Economically Diverse (CLED) students (Kaufman et al., 2012; Luria et al., 2016).

Among the creativity assessments used for gifted and talented programs including, divergent thinking tests (e.g., Wallach-Kogan Creativity Test; Wallach & Kogan, 1965; TTCT; Torrance, 2017), domain-specific creative products (e.g., CAT; Amabile, 1996), performance-based problem-solving assessments (e.g., DISCOVER; Maker, 2005), and creativity rating subscales (e.g., SRBCSS; Renzulli, 2010), the TTCT-F is the most widely used norm-referenced creative thinking test covering populations from preschool to adulthood (Kaufman et al., 2008, 2012; Lee et al., 2024). Given that the TTCT-F is one of the standardized creativity tests

possessing the grade and age norms representing U.S. populations, the TTCT-F could provide a good rationale for making high-stake decisions like determining eligibility for gifted programs (Kim, 2006). The TTCT-F consists of two parallel forms A and B, and includes activities of Picture Construction, Picture Completion, and Line/Circles, which might be accessible and engaging particularly for elementary-age children (Kim, 2017). Also, because domain-specific creativity emerges over the course of schooling within particular domains and is closely linked to domain knowledge and expertise (Baer, 2015; Dumas et al., 2024), a domain-general creative thinking assessment like the TTCT-F may be more suitable for younger students who have not sufficiently developed domain-specific knowledge yet.

Despite Torrance's warning against treating the TTCT scores as a single factor (1966, 1974), the single Creativity Index composite score has been frequently used for gifted identification in educational settings (Acar et al., 2023; Kim et al., 2006; Lee et al., 2024). The Creativity Index score is obtained by the average of the standardized scores for the five subscales (Fluency, Originality, Elaboration, AT, and RPC), where standardized means (*M*) of 100 and standard deviations (*SD*) of 20 were utilized (Torrance, 1966/2017). For instance, in the state of Georgia, the Creativity Index composite score at or above 90th percentile on the TTCT-F is considered as eligible for gifted education services (the cut-off score is around 125 depending on the distribution), within the students' grade level or age (Georgia Department of Education [GaDOE], 2022).

However, using the single Creativity Index scores for judging gifted eligibility may yield a limited understanding of creative thinking as expressed through the TTCT-F (Torrance, 1966, 1974; Kim, 2006). Solely depending on a composite score may not provide the granularity that is needed for an accurate assessment of an individual's creative thinking, as it oversimplifies the

complex structure of the construct. Once a student is identified as gifted, the Creativity Index fails to deliver any detailed information to support their individualized learning in developing creative thinking. Given that one of the main purposes of conducting the creativity tests is to identify a student's creative strengths and areas needing improvement (Cramond, 1994; Torrance, 1974), educators and school administrators might benefit from interpreting the scores in a way that fully illustrates creative thinking that is expressed through the TTCT-F.

2.2 Gifted Identification Across Ethnicities and Genders

Underrepresentation of Black/African American, Hispanic/Latinx, and Native American students in gifted identification has been an important issue which needs to be addressed in the field of gifted education (Daniels, 1998; Ford et al., 2001; Ford & Harris, 1994; Yoon & Gentry, 2009). In contrast, Asian and White/European American students have been over-represented in gifted and talented education services, especially Asian students identified even at a higher rate than their White peers (Lee et al., 2024; Peters et al., 2019). Notably, the TTCT-F does seem to have a positive influence on this issue, because this identification gap tends to become larger when creativity assessment is not used, and only other identification methods such as IQ measures and standardized achievement tests are used (Hodges et al., 2018). This is because disadvantaged students may lack the opportunities to fully develop their potential or to develop the abilities assessed on identification tests, and therefore are not likely to be identified as gifted, despite having high learning potential (Dumas et al., 2020; Pfeiffer, 2012).

To address the unequal representation issue, a wider set of identification methods like nonverbal assessments, creative thinking tests, and recommendations from teachers or peers are used for identifying the talents and abilities which might otherwise be unrecognized through IQ measures and standardized achievement tests (Lohman & Foley-Nicpon, 2012). Because it is

simultaneously focused on creative thinking, and within a nonverbal modality, the TTCT-F is considered to be an especially useful measure for this purpose (Bart et al., 2017; Cramond, 1994; Torrance, 1971). For example, Black/African American students have tended to perform better than, or at least similar to, their White peers in classic studies using the earlier version of TTCT-F (Covington, 1969; Torrance, 1967, 1971). However, few studies have investigated whether the TTCT-F shows less bias in gifted identification, except one recent study revealed that the inclusion of the TTCT-F did not significantly improve the underrepresentation of Black/African American, Hispanic/Latinx, and Native American students (Lee et al., 2024). Hence, the field of gifted and talented education needs to further investigate whether the TTCT-F measures creative thinking similarly across different ethnicities.

On the other hand, in terms of gender, there tends to be no disproportional representation of males or females in gifted identification (Crombie et al., 1992; Petersen, 2018). However, boys were more likely to be identified than girls when only IQ measures and standardized achievement tests were used for determining eligibility, suggesting that this form of identification methods might possibly disadvantage girls (Petersen, 2018). This may indicate that including the TTCT-F in the identification processes could contribute to equal gender representation in gifted and talented services.

Concerning TTCT-F scores, it has been difficult to conclude about the presence of either male or female advantages on this test, with some samples showing higher scores for one group, and other samples showing the opposite (Baer & Kaufman, 2008; Saeki et al., 2001), which might contribute to proportional representation in gifted identification in terms of gender, if those sample-specific differences indicate equality in the population. Yet, males appear to be more likely to be represented in adulthood creative achievements than females. This might be due to

different socialization experiences and gender stereotypes (Bian et al., 2017; Cole & Zuckerman, 1987; Helson & Picano, 1990). Considering that school environments may be more supportive of creative expressions from boys than from girls (McVey, 2004), there may be a possibility that girls and boys respond to the creativity tests in a different way. At this point in time, as the student populations in the U.S. continues to grow more diverse (U.S. Department of Education, 2020), achieving proportional identification of different ethnicities and genders through unbiased assessments is becoming increasingly important.

2.3 The Structure of the TTCT-F

Examining the latent factor structure of a measure is one useful way to understand the underlying psychological constructs that the measure is assessing within the sample population (Kim et al., 2006; Wiebe et al., 2008). Although Torrance (1966, 1974) suggested that each subscale of the TTCT-F has an independent meaning, and therefore, the test scores should not be aggregated to a single score, researchers have been likely to believe that the structure of the TTCT-F, like the other divergent thinking tests, might comprise one overall creative thinking attribute (Abernathy-Tannehill, 1997; Chase, 1985). In this regard, some previous studies have shown that the TTCT-F scores consist of one general factor (Clapham, 1998; Heausler & Thompson, 1988), indicating that these scales may measure a similar construct. These findings were primarily due to high correlations between Fluency and Originality as well as Fluency and Flexibility (Clapham, 1998; Cramond et al., 2005), especially in the older versions of TTCT-F and TTCT-Verbal.

However, the current streamlined scoring system of the TTCT-F, which reflects the five norm-referenced scales (Fluency, Originality, Elaboration, AT, and RPC) along with 13 creative strengths, is distinct from the earlier version that was regarded as composed of a singular factor. With the current version of the scales, there has been a consensus that the TTCT-F consists of

two-factors rather than one general factor (Acar et al., 2023; Bart et al., 2017; Hahm et al., 2019; Humble et al., 2018; Kim et al., 2006; Krumm et al., 2014, 2016; Said-Metwaly et al., 2018; Xu et al., 2025; Yoon, 2017). These two latent factors were called “Innovative” and “Adaptive”, which were derived from Kirton’s Innovator and Adaptor distinctions (1976, 1978, 1989). The model suggests that the Innovative factor pertains to the ability to generate novel ideas quickly, whereas the Adaptive factor reflects the capacity for deep and detailed thinking (Kim et al., 2006). Relatedly, Fluency and Originality loaded onto the Innovative factor, and Elaboration and AT loaded onto the Adaptive factor consistently in the two-factor models (Acar et al., 2023; Said-Metwaly et al., 2018). These findings gave implications in terms of understanding and interpreting two disparate cognitive functions of creative thinking expressed through the TTCT-F (Alabbasi et al., 2022; Kim et al., 2006).

However, there has been an inconsistency regarding which factor RPC fell into, because RPC either loaded onto Innovative or Adaptive factors, or cross-loaded onto both (See Figure 1). This inconsistency might stem from the nature of RPC, which refers to the cognitive capability to keep one’s mind open and defer judgment while considering available alternatives to the solutions (Torrance, 1984, 1990, 1998). Given the meaning of RPC, it appears to be closely linked to both Innovative and Adaptive factors (Acar et al., 2023; Kim, 2006; Lubart & Zenasni, 2010). To elaborate, RPC might be a good mindset to reach originality by continuously producing multiple ideas (i.e., load on the Innovative factor), and it also might help in adding details and giving imaginative titles to those generated ideas through persistent commitment to the task (i.e., load on the Adaptive factor). Taking these past findings together, the theoretically and empirically plausible two-factor models as well as a general one-factor model are presented in Figure 1. Previous meta-CFA studies (Acar et al., 2023; Said-Metwaly et al., 2018)

demonstrated that the most optimal two-factor model involved RPC loading onto the Adaptive factor, while another model with the cross-loading of RPC was identified as the second most optimal. Considering that the best-fitting model manifested differently depending on the characteristics of the samples used in the previous studies (Bart et al., 2017; Humble et al., 2018; Kim et al., 2006; Krumm et al., 2014, 2016; Sahin, 2015; Xu et al., 2025), the factor structures of the TTCT-F scales could potentially vary across culturally different groups, particularly among younger students, and may not align with the most commonly accepted best-fitting models.

2.4 Measurement Invariance in Educational Testing

Measurement Invariance (MI) refers to the statistical concept that latent constructs, assessed through a measurement, are understood and interpreted in the same way across identifiable subgroups, such as different genders, race/ethnicities, and age/grades (Schmitt & Kulijanin, 2008). MI has been one fundamental goal in educational and psychological testing (AERA, APA, NCME, 2014; Dong & Dumas, 2020; Engelhard, 2008), because it ensures that a measure is unbiased and fair (Engelhard, 1989). Particularly, achieving MI is essential when comparing the latent means across different groups or time points, as failing to establish it may mislead the interpretation of such comparisons (Vanderberg & Lance, 2000). This is especially crucial in educational settings, where test scores impact high-stakes decisions that might have long-lasting impacts on students' lives (e.g., gifted identification or university admission; Millsap, 2011). Without MI, these decisions might disadvantage specific groups, leading to unintended negative outcomes (Millsap & Kwok, 2004). For example, lack of MI might result in over- or under-representation of particular ethnic or gender groups, thereby threatening fairness and consequential validity (Messick, 1989, 1994) of the test. Thus, the establishment of MI lays the

groundwork for making equitable and ethical decisions in educational contexts as well as strengthening the statistical robustness of measurements (Putnick & Bornstein, 2016).

The extent to which MI holds across diverse subgroups has been tested by using statistical techniques including confirmatory factor analysis (CFA) and item response theory (IRT; Stark et al., 2006). If the person-invariance, defined as an establishment of the stable item parameters (e.g., item difficulty) across subgroups (Engelhard, 2008), is of interest, Rasch models or IRT frameworks might be beneficial. However, the CFA-based analysis has stood out in MI studies due to its advantages over IRT when the MI study aims to examine item-invariance, which refers to whether the items function similarly across different groups (Engelhard, 2008). The CFA approach follows a step-by-step analysis to examine the extent to which MI holds, whereas IRT simultaneously tests the equivalence of item discrimination and item difficulty parameters, which are analogous to factor loadings and intercepts in CFA (Meade & Lautenschlager, 2004). The CFA model explores whether group differences in latent constructs stem from true differences or systematic measurement bias through a series of tests by using software like R packages Lavaan or OpenMx, LISREL, Amos, and Mplus. In particular, the factor models based on CFA analysis could be represented by the following regression equation where I indicates an observed variable or indicator, c implies an intercept, F represents the latent variable, f is a factor loading or regression coefficient, and u is uniqueness or residual.

$$I = c + fF + u$$

The steps of MI testing with CFA model are as follows. Before testing the invariance of the measurement, multivariate outliers for each group must be screened to ensure the dataset meets the assumptions for CFA. Then, multigroup CFA is carried out by sequentially fitting

configural, metric, scalar, and structural invariance models, with the parameters gradually constrained at each level. During these steps, if the model fit gets significantly worse than the previous invariance model, it could be interpreted as evidence that MI does not hold at that particular level (Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). In the first step, a test for configural invariance (Horn & McArdle, 1992) requires the same factor structure and patterns of factor loadings across different groups. Configural MI should be established to make the subsequent MI tests meaningful. The second step involves a test that the factor loadings are the same across groups, which is called metric invariance (Horn & McArdle, 1992). At this level of the MI tests, the factor loadings are constrained to equality for each group to compare the model fit with the configural model. If the model fit does not get worse and metric invariance is achieved, it suggests that individuals from different groups interpret the items in the same way. In the third step, a test of scalar invariance (Meredith, 1993), where the intercepts as well as factor loadings are constrained to be equal across groups, necessitates the intercepts of the regression equations for each observed variable are equivalent. Scalar invariance indicates that the observed scores would be the same when individuals from different groups have the same level of latent constructs. Once scalar invariance is achieved, latent factor means can be compared between groups. On the other hand, partial invariance might be tested without establishing scalar invariance, by freeing some item intercepts. The next step involves a test of strict invariance (Meredith, 1993), referred to as ensuring that measurement error is same across groups, by constraining the residuals of the regression equation for each observed variable to be equal. If strict invariance is not achieved, then reliability of each indicator could be interpreted as different across groups. However, strict invariance has been found to be difficult to achieve and not necessary to be tested, and therefore, not of interest to most researchers (Vandenberg, 2002).

The chi-square test is the most widely used test to examine the global model fit, however, this test is known to be sensitive to a large sample size given that it easily fails to reject the model even the differences between models are negligible (van de Schoot et al., 2012). In the case of multigroup CFA, there are mainly two types of fit indices that could be used for model comparisons (Kline, 2010). First, there are relative indices that compare a model fit to the baseline model, such as Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI). Second, absolute fit indices like Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) assess how well the model reproduces the observed data without comparing it to any other model. Some fit indices are likely to be sensitive to model complexity (e.g., RMSEA) and others sensitive to sample size (e.g., CFI, TLI). After testing the measurement invariance of the model, the model fit indices for each level, along with results (i.e., the numerical values of factor loadings and intercepts), are generally reported in a table (Putnick & Bornstein, 2016; van de Schoot et al., 2012).

2.5 Does the Structure of the TTCT-F Hold Across Ethnicities and Gender?

The TTCT-F has been generally thought to measure creative thinking in the same way across gender (Kim 2006; Torrance, 1977) and ethnic groups (Cramond, 1993; Torrance & Torrance, 1972). Despite the importance of establishing the fairness of the TTCT-F across ethnicities in order to ensure equal representation in gifted identification (Kim et al., 2006; Lee et al., 2024), no single study exists which investigated whether the factor structure of the TTCT-F holds invariantly across over- and under-represented ethnicities. This might be due to limited access to sociodemographic data when scoring the TTCT-F, given that the testing company that scores the test does not collect private information, or probably due to the difficulty of collecting a sufficiently large number of TTCT-F responses from various ethnic groups.

On the other hand, prior studies were likely to report that the latent structure of the TTCT-F was invariant across genders (Krumm et al., 2014, 2016; Wu et al., 2025) but not across grades or ages (Kim et al., 2006), suggesting that more careful interpretation may be needed when comparing the TTCT-F scores from different age groups. Even among the studies which revealed that the TTCT-F was fair in terms of gender (Kim et al., 2007; Krumm et al., 2014, 2016; Wu et al., 2025), two studies have identified that there were significant mean differences in Innovative or Adaptive factors and some of the scales (Fluency, Originality, or Elaboration) between boys and girls, where girls' scores for these scales were consistently higher than boys (Krumm et al., 2014; Wu et al., 2025). This might suggest that there is a gender discrepancy on the TTCT-F scales, indicating that the expression of creative thinking could vary by gender, particularly in younger populations. Combining insights from the studies on ethnicity and gender related to the TTCT-F structures, it is necessary to investigate the structures of the TTCT-F across under- and over-represented ethnicities as well as genders.

CHAPTER 3

METHODOLOGY

3.1 Goals of Current Study

This study examines the factor structure of the TTCT-F scales: Fluency, Originality, Elaboration, Abstractness of Titles (AT), and Resistance to Premature Closure (RPC), and then investigates measurement invariance across ethnicities (White/European American and Asian students together as those who are over-represented in U.S. gifted and talented programs; and Black/African American and Hispanic/Latinx students as those who are under-represented) and genders (boys, girls). Specifically, we explored (a) four theoretically plausible models, including three different two-factor models and a one-factor model through Confirmatory Factor Analysis (CFA), and (b) the measurement invariance across under- and over-represented ethnicities as well as genders (boys and girls) by conducting Multigroup CFA.

In the present investigation, we addressed the following research questions:

Research Question 1: Which model is the best-fit for the entire sample composed of U.S. elementary students among the three different two-factor models and one-factor model?

Research Question 2: Does the factor structure of the TTCT-F scales hold invariantly across under- and over-represented ethnicities as well as across boys and girls?

3.2 Participants

This study was a part of a larger federally funded project which aimed to develop and validate cost-effective creativity assessments utilizing automated scoring methods. 379 elementary school

students from grades 3, 4, and 5 in the U.S. ($M_{\text{age}} = 9.33$, $SD = 0.97$) participated in this study. The sample was composed of 137 (36.1%) third grade, 130 (34.3%) fourth grade, and 112 (29.6%) fifth grade students. Of the total sample, there were 178 (47.0%) male and 201 (53.0%) female students. With respect to race and ethnicity, 125 (33.0%) were Hispanic or Latinx; 112 (29.6%) Black or African American; 68 (17.9%) Asian; 60 (15.8%) White or European American; 8 (2.1%) Multi-ethnic identities; and 3 (0.8%) American Indian or Alaska Native identity. Moreover, 131 (34.6%) were English Language Learners (ELLs), 49 (12.9%) receiving special education (SPED) services; and 60 (15.8%) identified as gifted and talented (GT).

3.3 Focal Measure: TTCT-F

The TTCT-F form A (Torrance, 1998/2017), which is composed of three activities, was administered in this study. The first activity, Picture Construction, required augmenting an abstract shape and titling the resulting drawing. The second activity, Picture Completion, consisted of ten incomplete figures and instructed participants to add additional detail in order to compose a drawing around those figures, and students also provided a title for their drawing. The third activity, Lines, required completing a drawing using the same pair of parallel lines as many times as possible as well as adding titles for each drawing (Torrance, 1966, 1974, 1990). Students were given 10 minutes per activity to both draw and title their drawings. Across these three activities, the TTCT-F quantifies creative thinking through five norm-referenced scores which are each explained below.

Fluency

Because all participants received the same amount of time on each task, Fluency was operationalized as the quantity of ideas posited during that time. If participants conveyed an idea through drawing using the provided figures in activity 2 and 3, they received 1 point for each

drawing. However, duplicates, unrecognizable, and abstract doodles were not counted toward fluency, resulting in 0 points. In the TTCT-F scoring procedure, fluency is scored first before the other dimensions, and drawings are only scored for the other dimensions if they counted toward the fluency score. Fluency was scored on activity 2 and 3 but not on activity 1 because there was only one item in this activity and therefore multiple ideas are not possible.

Originality

Originality indicates the number of unusual ideas, focusing on how the provided figures are used in each response. Originality was scored based on the zero-originality list for each activity which functioned as criteria for judging originality of the idea. The responses listed in the zero-originality list received 0 points, and those not listed were scored 1. Additionally, bonus points were granted when their drawing contained two or more figures across the items.

Originality was scored for all the activities, but bonus originality scores were granted just for activity 2 and 3.

Resistance to Premature Closure (RPC)

RPC measured the ability to delay closing the parallel lines in activity 2, implying how open-minded or curious they were. 0 points were awarded if the incomplete figure in each item was entirely closed or used as the response itself. 2 points were granted if the figure was never closed or closed incidentally to convey an idea. If the level of closure was somewhat between entirely closed and opened, they received 1 point following the scoring guidelines. RPC was only scored for activity 2.

Elaboration

Elaboration represents how many details were added to the basic idea expressed in the drawing. Additional ideas were counted by the number of shadings, decorations, etc. and

elaboration was scored on an interval scale ranging from 0 to 6 depending on how many details the responses had. All three activities were scored for Elaboration.

Abstractness of the Titles (AT)

AT is defined as making a title which goes beyond the obvious description of the idea in the drawing (Torrance et al., 1992). This represents finding and communicating the essential story in the drawing. AT was scored for activity 1 and 2 on a scale ranging from 0 to 3, according to the level of abstractness and imaginativeness.

3.4 Data Collection Procedures

TTCT-F Form A (Torrance, 1998) was given in groups (less than 20 children in each group) following the administration guidelines. The groups were formed depending on the timing of parental consent. Before conducting the study, we also obtained the participants' verbal assent. There were a few exceptions where some participants took the TTCT-F on another day, depending on what their school schedule required. Students received a toy gift as compensation for their participation. The data was collected during normal school hours from six different branches of the same public-school network located in the Southcentral region of the United States. The TTCT response booklets were sent to Scholastic Testing Service to obtain scores from a certified TTCT-F rater. Concerning the rater reliability for the TTCT-F, the Pearson correlation coefficients (r) for scoring Fluency fell between 0.96 and 0.99, and for scoring Originality, RPC, and Elaboration, they ranged from 0.90 to 0.99 on average, suggesting a high level of reliability (Torrance, 1966/2017).

3.5 Data Analysis Overview

Bivariate correlations between the TTCT scores were assessed to investigate how and to what extent they were related. Then, CFA was conducted for the entire sample to test which model

was the best-fit among the theoretically plausible two-factor models as well as a one-factor model of the TTCT-F scores. Then, multigroup CFA was conducted to compare the factor structure of the TTCT-F across ethnicities (under- and over-represented in U.S. gifted and talented programs) and gender (boys and girls) with the best-fitting model.

CHAPTER 4

RESULTS AND ANALYSIS

4.1 Bivariate Correlations Among TTCT Scales

Correlations were calculated to analyze the relations among the TTCT-F scales. Bivariate correlations are reported in Table 1. Notably, Fluency and Originality were highly correlated ($r = .77, p < .01$), indicating that those participants who generated a large number of ideas also produced more original ideas (sometimes referred to as a fluency confound; Forthmann et al., 2020; Hocevar, 1979). Fluency showed a moderate positive correlation with RPC ($r = .58, p < .01$), and a small correlation with Elaboration ($r = .28, p < .01$). Interestingly, Fluency showed the least positive association with AT ($r = .10, p < .05$). Originality was moderately correlated with RPC ($r = .57, p < .01$) and Elaboration ($r = .37, p < .01$), while showing a weaker positive correlation with AT ($r = .10, p < .05$). Elaboration was moderately associated with AT ($r = .48, p < .01$) and RPC ($r = .32, p < .01$). AT showed a small positive correlation with RPC ($r = .28, p < .01$). What stood out in the correlation table was that AT showed a relatively weaker correlation with the other scores except Elaboration. Also, Elaboration showed the strongest positive association with AT while being less strongly correlated with the other scores. This pattern demonstrates a close relationship between Fluency and Originality, just as AT and Elaboration are tightly intertwined. With these findings in mind, we carried out CFA.

4.2 Fitting CFA Models to All Student Data

Model Configuration

Four theoretically plausible and empirically supported two-factor models and the one-factor model were fit to the data from all participants. The compared models were: (a) Model 1- One-factor model; (b) Model 2- Innovative (Fluency, Originality, RPC) and Adaptive (Elaboration, RPC, AT); (c) Model 3- Innovative (Fluency, Originality, RPC) and Adaptive (Elaboration, AT); and (d) Model 4- Innovative (Fluency, Originality) and Adaptive (RPC, Elaboration, AT) (See Fig. 1).

CFA was carried out for the entire sample using Mplus 8.8 software to test which model was the best-fit among the four different factor models of the TTCT scales. As the skewness and kurtosis values for these scales were all within the range of normal distribution (See Table 2), maximum likelihood estimation was utilized for producing fit indices and model parameters. The variance of the Innovative and Adaptive factors was set to 1 for model identification and standardization. To determine the best-fitting model for the whole sample, the χ^2 test and fit indices were used. The fit indexes used for the CFA were the comparative fit index (CFI), the Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The cutoff criteria for indicating a good model-data fit were a value of .90 or above for CFI and TLI, .08 as an upper boundary for RMSEA, and .10 or less for SRMR (Cordon & Finney, 2008; Vandenberg & Lance, 2000). The holistic approach was adopted by combining fit indices, measurement invariance testing, and theoretical justifications, instead of solely depending on the cutoff criteria (Chen, 2007).

Modification indices were used to check if additional parameters were needed in the model in order to substantially improve model-data-fit. These modification indices suggested the

need to correlate some item error variances in Models 2, 3, and 4: the error covariances between Fluency and AT as well as between Originality and Elaboration. Adding the path of residual covariance between Fluency and AT was also justified by the well-established finding about a trade-off between idea quantity and quality in the field of creativity research (Forthmann et al., 2018), given that producing many drawings might have negatively influenced creating reflective titles in the limited time. The residuals of Originality and Elaboration were also allowed to covary in each of the models, which we believed made theoretical sense given the way that the scoring procedures for Elaboration may also have identified aspects of drawings that made them more Original.

Comparison of Model Fit

To determine which model obtained the best-fit, the χ^2 test and other fit indices were compared. The Chi-square values and fit indices for the models are provided in Table 4. Model 1 (One-factor) demonstrated a significant Chi-square value, $\chi^2(5) = 107.20, p < .001$, CFI and TLI values less than .90, and RMSEA value greater than .08, suggesting a poor-fit. The fit indices indicated that Model 2 consisting of Innovative (Fluency, Originality, and RPC) and Adaptive (RPC, Elaboration, and AT) with the residual covariances was the best-fitting model among the four. The Chi-square value of Model 2 was not significant, $\chi^2(1) = .04, p > .05$, and the fit indices were very good, as shown by the fact that the CFI and TLI were above .95, and RMSEA and SRMR were less than .05. In Model 2, the residual covariances of Fluency and AT as well as Originality and Elaboration were added. Although Model 3 with the residual covariances showed a significant Chi-Square value of 6.02 ($df = 2, p < .01$), it was necessary to pair with other fit indices given that Chi-Square value is sensitive to large sample size (Meade et al., 2008). Overall, the fit indices for Model 3 (CFI = .99, TLI = .97, RMSEA = .07, SRMR = .02) indicated

an acceptable fit. Model 4 showed a poor fit even with the residual covariances added, as shown by a significantly greater Chi-square value, $\chi^2(2) = 58.50, p < .001$, with the CFI and TLI below acceptable ranges and the RMSEA far above .08.

To determine whether Model 2 was more optimal than Model 3 since they both achieved acceptable fit, a Chi-square difference test was carried out to compare the model fit of Model 2 and 3. As a result, Model 2 showed a significantly better fit than Model 3, $\Delta\chi^2(1) = 5.98, p < .05$. Model 2 also demonstrated better CFI ($>.99$), TLI ($>.99$), and RMSEA ($<.001$), indicating a superior fit in comparison to Model 3. Besides, considering that the previous meta-CFA studies identified Models 2 and 4 as the competitive models, rather than Model 3 (Acar et al., 2023; Said-Metwaly et al., 2018), the fit Model 3 may be less likely to replicate in other datasets. Taking this evidence together, Model 2 was identified as optimal in our analysis.

Interpreting Model Coefficients

The standardized path coefficients in the CFA model were examined to compare the strength of the relations between the latent constructs and the indicator variables (see Figure 2). Innovative and Adaptive factors showed a moderately strong latent correlation ($r = .41, p < .01$), indicating a positive relationship while still being separable. Concerning the Innovative factor, Fluency loaded onto the factor with the highest coefficient ($\beta = .89, p < .01$), Originality the second highest ($\beta = .87, p < .01$), and RPC the lowest ($\beta = .57, p < .01$). Interestingly, RPC relatively weakly cross-loaded onto the Adaptive factor ($\beta = .20, p < .01$) in this study, while RPC tended to load onto the Adaptive factor higher than onto Innovative in prior studies (Krumm et al., 2014; Said-Metwaly et al., 2018). Adaptive was strongly loaded onto by Elaboration ($\beta = .75, p < .01$), and AT ($\beta = .64, p < .01$), suggesting that these indicators represent the Adaptive factor.

The residual correlations were as follows: Fluency with AT ($r = -.36, p < .001$), Originality with Elaboration ($r = .29, p < .01$). These factor coefficients indicated that there may be a negative association between Fluency and AT and a positive association between Originality and Elaboration after accounting for the shared variance explained by the two latent factors.

Reliability Coefficients of the Innovative-Adaptive Model

McDonald's ω (McDonald, 1970, 1999) and Coefficient H (Hancock & Mueller, 2001) were computed for the two latent factors (Innovative and Adaptive) to assess the internal reliability of the best-fitting Innovative-Adaptive model (Model 2). Cronbach's α has been the default estimate of internal consistency in psychology literature for many decades (Hogan et al., 2000). However, the α coefficient assumes tau-equivalence, which means that the standardized factor loadings for all items are equal and therefore is inappropriate in cases when the item loadings differ (Crutzen & Peters, 2017; Yang & Green, 2011). In addition, the best-fitting two-factor model did not meet the assumptions of Cronbach's α , specifically no cross loadings across factors and uncorrelated errors (McNeish, 2018), which led us to use alternative methods. Accordingly, ω and H coefficients, which do not have such strict assumptions were employed to obtain more accurate estimates of internal reliability. McDonald's Omega Total and Coefficient H were calculated using McNeish (2018)'s reliability excel sheet. The reliability coefficients for the Innovative factor were notably higher ($\omega = 0.84; H = 0.87$) than those observed for the Adaptive factor ($\omega = 0.61; H = 0.66$). Internal consistency was acceptable for the Innovative factor with both reliability coefficients exceeding 0.70, while the Adaptive factor barely passed or failed to meet the threshold criteria (Cohen & Swerdlik, 2018). These findings align with prior

work which reported the lower reliability estimates for the Adaptive factor compared to the Innovative factor (Acar, 2023; Acar et al., 2024).

4.3 Multi-Group CFA Models for Measurement Invariance

Based on the best fit model (Model 2) we identified, Multigroup CFA was conducted to compare the factor structure of the TTCT-F across ethnicities (under-represented, over-represented in U.S. gifted and talented programs) and genders (boys, girls).

Under- and Over-represented Ethnicities

Because the sample size within each ethnicity (i.e., African-American/Black, Asian, Hispanic, European-American/White, American Indian, and Multi-Ethnic) was not large enough to accommodate a full comparison of the measurement parameters across each individual group, they were combined in theoretically and practically-informed ways to accomplish this analysis. Specifically, the Hispanic and African-American/Black students were combined into the group *Under-represented*, because of their under-representation in U.S. gifted and talented programs. White and Asian students were also combined into the group *Over-represented*, because of their over-representation in U.S. gifted and talented programs (Erwin & Worrell, 2012; Lee et al., 2024; Peters et al., 2019; Yoon & Gentry, 2009).

In order to assess the measurement invariance of the Innovative-Adaptive model, we started by fitting a *configurally* invariant multigroup CFA model, in which the same two-factor configuration was present across the groups but no other parameters were constrained to be equal. Then, the loadings in that multigroup model were constrained to be equal, resulting in a *metric* invariance model. Third, the item intercepts were constrained to be equal across the groups, resulting in a *scalar* invariance model. Comparison of fit indices across these models

allowed for inferences about the invariance of the Innovative-Adaptive model across over- and under-represented groups (Cordon & Finney, 2008; Rodriguez & Shaffer, 2020)

The multigroup CFA showed that measurement invariance was established at configural, metric, and scalar levels across over- and under-represented ethnic groups (see Table 5), which implies that the Innovative-Adaptive model functioned in a similar way across the different ethnicity groups. In step 1, configural invariance was tested by fitting without any equality constraints. This model demonstrated a non-significant Chi-square value, $\chi^2(2) = 1.00, p > .05$, CFI and TLI values greater than .90, and RMSEA value almost near zero, indicating an excellent fit. The good fit suggested that the overall factor configuration held across groups. In step 2, metric invariance was assessed by placing equality constraints on the factor loadings. Metric invariance was supported, as indicated by the fit indices in Table 5. Though there was a noticeable increase in RMSEA from the configural model ($\Delta\text{RMSEA} = .052$), RMSEA was still acceptable (RMSEA = .052). Thus, the metric model did not practically worsen the fit statistics than the configural model, suggesting that Innovative and Adaptive factors were indicated similarly by the scales across the under- and over-represented ethnic groups. In step 3, scalar invariance was tested by constraining the intercepts to be equal across these groups. Also, scalar invariance was established as represented by the fit indices in Table 5. There were only slight changes in CFI, TLI, RMSEA, and SRMR from the metric to scalar model, suggesting that adding the constraints to the intercepts did not significantly worsen the model fit. This means that group mean differences in the scales are consistent with group mean differences in latent Innovative and Adaptive factors. As scalar invariance was achieved, comparing mean differences of the TTCT scales between the under- and overrepresented ethnicities would be considered valid.

Gender

Multigroup CFA across gender was carried out following the same procedures as with ethnicity above. However, in the case of Gender, the configurally invariant model failed to converge. We attempted many changes to estimation settings in Mplus, including constraining the residual variance-covariance matrix to be positive, removing outliers from the data, and increasing the number of model iterations, but none of these strategies were able to make the configurally invariant model converge. This failure to converge suggests that the latent structure of the TTCT-F implied by the Innovative-Adaptive model did not hold across both boys and girls in this dataset, and therefore the multigroup model was not able to be fit. In order to more fully examine this finding, we first closely investigated the differences in the inter-correlations among the TTCT scales across genders below and then conducted Exploratory Factor Analysis (EFA) for boys and girls separately.

4.4 Delving into Bivariate Correlation Across Gender

To further investigate how the scores were similarly or differently correlated in boys and girls, a Fisher's Z-test was used to examine if there were significant differences between the correlation coefficients across those groups. Table 3 presents bivariate correlations across genders. Notably, the correlation between Fluency and Originality was significantly larger in girls ($r = .86, p < .001$) than in boys ($r = .68, p < .001$): $z = -4.45, p < .001$. These findings suggest that producing more ideas very consistently led to greater originality for girls, while boys did not consistently show increased originality despite generating many ideas.

On the other hand, Fluency and AT showed a significantly weaker positive correlation in girls ($r < .01, p = .997$) than in boys ($r = .23, p < .001$), $z = 2.24, p < .05$, indicating that there was a relatively weak but positive association between these variables for boys, while no

relationship was found for girls. These correlations suggest that girls may be capable of generating abstract and imaginative titles without being influenced by the number of ideas, whereas boys may tend to create more abstract titles when they have more ideas. As seen by Table 2, boys also demonstrated a descriptively stronger relationship between AT and Originality, Elaboration, and RPC when compared to girls, but these differences were not statistically significant.

4.5 EFA by Gender

To identify the latent structure of the TTCT-F scales for boys and girls, a separate EFA was conducted for each gender group. The correlation matrices were analyzed with the estimation method of Maximum Likelihood (ML) and Promax rotation, based on the assumption of continuous data (Fabrigar et al., 1999) and potentially correlated latent factors (Costello & Osborne, 2005). The skewness and kurtosis values of the TTCT-F scales were within the range of normal distribution for both boys and girls (skewness and kurtosis between -1 and $+1$), supporting the use of the ML estimation. The use of Promax rotation was further justified by the presence of the positive correlation between the Innovative and Adaptive factors ($r = .41, p < .001$) in the best-fitting model (Model 2) for the entire sample. As suggested by EFA research practice (Thompson & Daniel, 1996), multiple criteria approaches were used in our analysis. Accordingly, we carried out parallel analysis in R version 4.3.0 using the *psych* package (Revelle, 2020), whose eigen values were extracted from the EFA run in Mplus 8.8.

For boys, the eigenvalues were 2.85 for the 1st factor and 0.97 for the 2nd factor, indicating only 1 factor exceeded Kaiser's criterion (eigen value > 1.0), and the observed eigenvalue for only 1 factor exceeded the 95th percentile of eigenvalues derived from the parallel samples but not 2 factors in parallel analysis. However, the two-factor model showed the better

fit than the one-factor model ($\Delta\chi^2 = 38.52, p < .001$), and an excellent absolute fit [$\chi^2(1) = 0.17, p = .68, \text{RMSEA} < .001$]. These conflicting results may arise as parallel analysis tends to underestimate factors when their correlations are high and the first factor is large (Hayton et al., 2004; Turner, 1998), and therefore we retained the two-factor solution. For girls, the eigen values were 2.50 for the 1st factor, and 1.33 for the 2nd factor, which are both above Kaiser's criterion (eigen value > 1.0). Parallel analysis results also indicated that the original eigenvalues of the 1st factor and the 2nd factor exceeded the 95th percentile of eigenvalues derived from the parallel samples (see Figure 4). The two-factor model—with Fluency, Originality, and RPC loading more strongly on the Innovative factor, and Elaboration and AT on the Adaptive factor (See Table 6)—was significantly better than the one factor model for girls ($\Delta\chi^2 = 70.83, p < .001$), and suggested an excellent absolute fit [$\chi^2(1) = .98, p = .32, \text{RMSEA} < .001$]. Overall, all the EFA analysis results for girls supported the two-factor structure.

In the EFA, boys exhibited stronger factor correlation ($r = .54, p < .001$) than girls ($r = .31, p < .001$), supporting greater interdependence between these latent traits for girls. Relatedly, RPC appeared to strongly load onto the Innovative factor ($\lambda = .55$) and weakly cross-load onto the Adaptive factor ($\lambda = .30$) in boys, whereas loading primarily onto the Innovative factor ($\lambda = .55$) but negligible loading on the Adaptive factor ($\lambda = .11$) in girls. Taken together, these results suggest Innovative and Adaptive latent traits tend to be more distinct in girls than in boys. Notably, a Heywood case (a negative residual variance of Fluency) was found in the girls' model, suggesting the instability of the Innovative-Adaptive model while Fluency served as a singular distinguishing feature. These findings also align with the Multigroup CFA results, where the Innovative-Adaptive model did not hold across both genders.

CHAPTER 5

DISCUSSION AND CONCLUSION

In this Discussion, we begin with Key Findings from the empirical investigation, before moving to Practical Implications for Gifted Identification, and then positing Future Directions in this line of work.

5.1 Key Findings

An Innovative-Adaptive Two-Factor Model Was the Best Fitting Model

The Innovative-Adaptive two-factor model with a cross-loading for RPC (Model 2), was the best fitting to the overall sample of elementary student TTCT-F responses. Our findings with the overall sample are therefore entirely aligned with past empirical work that established the Innovative-Adaptive two-factor model of the TTCT scales (i.e., Kim et al., 2006), which revealed that the latent structure of creative thinking was most accurately represented by the two-factor model with a cross loading for RPC. This model suggests that keeping open-minded and delaying judgment (i.e., RPC) might be a foundational psychological attribute for producing multiple original ideas quickly (i.e., Innovative factor) as well as adding details and giving imaginative but relevant titles to each drawing (i.e., Adaptive factor) in elementary students. Given that the Innovative-Adaptive two-factor model showed a considerably better fit compared to the one factor model, the TTCT-F might provide richer information about students' creative strengths and areas needing improvement if the scores are interpreted based on these Innovative and Adaptive constructs, rather than a single overall score, as might be used in school-based

practice. For instance, an elementary student might be relatively proficient in quickly coming up with novel ideas (Innovative factor) yet need to work on refining those ideas with additional details and distinctive titles (Adaptive factor): nuance that would be lost if a single overall score were used.

In contrast to previous studies, residual covariances between Fluency and AT as well as Originality and Elaboration were needed to make the model fit for the entire sample in this study. The negative residual correlation between Fluency and AT might imply that, after the variance for the latent Innovative and Adaptive factors is accounted for, these two scales might show a trade-off. Existing work on a trade-off between quantity and quality of ideas has been fruitful in creativity research (Forthmann et al., 2020; Gonthier & Besançon, 2024), and perhaps this line of inquiry could benefit from a closer look at the TTCT-F in the future, especially because quantity-quality relations in creativity could be different in samples of children than they are in adults. In addition, the positive residual correlation between Originality and Elaboration represents that the scoring procedures for Elaboration might also tap cognitive process involved in Originality. For instance, the students might need to add details to their drawings in order to effectively communicate an unusual or unconventional idea to the judges. This relation was not clearly explored or understood in earlier TTCT-F factor structure studies, making it interesting for future investigation.

Notably, the reliability coefficients were higher in the Innovative factor ($\omega = 0.84$; $H = 0.87$) than in the Adaptive factor ($\omega = 0.61$; $H = 0.66$). Although the Innovative factor showed acceptable internal consistency, the Adaptive factor appeared to be less reliable. These findings align with previous work that confirmed the acceptable reliability estimates for the Innovative

factor, but not for the Adaptive factor (Acar, 2023; Acar et al., 2024). Therefore, caution should be needed to interpret elementary students' scores based on the Adaptive factor.

The Two-Factor Model was Invariant Across Under- and Over-represented Ethnicities

The Innovative-Adaptive two-factor structure of creative thinking in the TTCT-F held invariantly across under- and over-represented ethnicities. These results show that the TTCT-F scales appear to be accurately characterized by the Innovative-Adaptive two-factor structure, and this latent structure is consistently represented across both under- and over-represented ethnic groups. More practically, it means that scores on the Innovative or Adaptive factors of the TTCT-F could be interpreted the same way—and therefore compared directly— across these ethnic groups. So, based on these results, utilizing the two latent mean scores of Innovative and Adaptive factors for gifted identification, or for understanding creative strengths and weaknesses, might not be biased against under-represented ethnicities such as Black/African American and Hispanic/Latinx students. This finding is in line with Torrance and his colleagues' arguments (Cramond, 1994; Torrance, 1971) which emphasized the fairness of the TTCT-F across ethnicities in terms of equal representation in gifted and talented services. It also adds nuance to the equity benefits of the TTCT-F by highlighting the importance of using the Innovative-Adaptive two-factor scores in educational settings.

The Two Factor Model was Not Invariant Across Genders

Interestingly, our findings did not align with earlier empirical work which revealed that the two-factor structure of the TTCT-F across genders was invariant (Kim et al., 2006; Krumm et al., 2014; 2016). A significantly weaker correlation between Fluency and Originality in boys compared to girls implied that generating many ideas may not be as effective in ensuring high idea quality for boys as it does for girls. Put another way, it might be that the quantity of boys'

ideas showed more of a trade-off with the originality of those ideas. This finding suggests that elementary students of different genders might use different cognitive or metacognitive strategies when engaging with creative thinking assessments. Considering that the detailed instructions for each activity spanned from 5 to 8 lines, which appear to be possibly too long and wordy for younger students, it might be hypothesized that boys may have been less attentive to the instructions and prioritized completing more items, therefore producing a more discernible trade-off between quantity and quality of ideas. On the other hand, we might hypothesize that girls perhaps followed the directions more carefully and therefore were still able to produce original and interesting ideas while coming up with multiple ideas. These hypotheses align with existing sustained attention studies which have shown that boys tend to aim to be as fast as possible whereas girls tend to focus more on avoiding errors (Efrat & Orna, 2022; Venker et al., 2007). In addition, a significantly larger correlation between Fluency and AT in boys than in girls was a new finding in studies of the TTCT. In particular, the zero correlation between these scales in girls suggests that creating imaginative and meaningful titles is an entirely discrete capability from generating many ideas, at least when there is a time-constraint. However, a positive correlation between Fluency and AT in boys implies that boys might be more capable of making abstract titles when they also have more ideas. This could possibly be due to boys' higher risk-taking tendency compared to girls on average (Harris & Jenkins, 2023), which might explain why boys are more willing to share the abstract titles of their drawings, as they appear less concerned about facing criticism, whereas girls might be more reluctant on average and worried about negative outcomes. Taken together, the Innovative-Adaptive two-factor model does not appear to ensure the invariance of the TTCT-F factors across genders in elementary students.

The Innovative and Adaptive Factors were Closely Related in Boys but Separable in Girls

Our EFA results (see Table 6) revealed that the Innovative and Adaptive latent traits tend to be more distinct in girls, implying that the quick generation of original ideas was a separable ability from deep and detailed thinking in girls, but these abilities were more interconnected with one another in boys. Relatedly, the factor cross-loadings of RPC on the Innovative and Adaptive factors differed by gender, suggesting RPC played a meaningful role in the Adaptive factor for boys but showed minimal cross-loading for girls. The tendency to remain open-minded and delay judgment (i.e., RPC) might be a key attribute that distinguishes the underlying mechanism of creative thinking in girls from those in boys. Moreover, given that EFA results for boys revealed the inconsistent factor structure solution, and the Heywood case of Fluency in the girls' EFA, the two-factor model might not fully represent the latent structure of creative thinking reflecting these gender differences in elementary school populations. With a larger sample, it may be possible to uncover a gender-specific model through separate EFA.

5.2 Practical Implications for Gifted Identification

Although the TTCT-F has been generally believed to provide equal affordances to students of diverse ethnicities and genders (Cramond et al., 2005; Kim, 2006; Lee et al., 2024), the way in which this test should be utilized for proportional representation in gifted identification remains uncertain. Relatedly, our findings can give practical guidance for gifted administrators or school psychologists who implement the TTCT-F in order to assess the creative thinking of elementary school students from different ethnicities and genders for academic purposes. First, the Innovative-Adaptive two-factor scores seem to accurately represent the domain-general creativity of elementary students from diverse backgrounds, and therefore, it is recommended that we shift towards using two-factor scores for gifted identification, rather than sticking with

the single overall score that has been the typical practice. Torrance's suggestion was to not entirely rely on a singular score on the TTCT, as it might fail to capture essential educational information (Torrance, 1966, 1974). Thus, gifted administrators should consider focusing on how to use the two-factor scores in more informed ways. For instance, a student who demonstrates high potential in producing multiple original ideas (i.e., high on Innovative) but who is relatively weaker on the elaboration of ideas and creation of imaginative titles (i.e. relatively low on Adaptive), or vice versa, might still be considered for gifted identification. Furthermore, the Innovative-Adaptive two-factor scores could deliver detailed information to support gifted students' individualized learning for developing creative ideas in school settings. Even after the gifted identification process is completed, gifted educators might interpret these scores in a way that could provide information about students' creative strengths and areas needing improvement (Cramond, 1994; Torrance, 1974).

Another practical implication is that the application of the Innovative-Adaptive two-factor scores appears to facilitate the proportional representation of the ethnicities that are under-represented in U.S. gifted and talented programs, through invariant psychometric properties. Given that the simple addition of the TTCT-F in the identification process did not guarantee equal representation of ethnicities in recent work (i.e., Lee et al., 2024), using the two-factor scores might give more access to gifted and talented services for Black/African American and Hispanic/Latinx students. However, before applying the Innovative-Adaptive two-factor model for gifted identification, it is important to carefully examine and consider how creative thinking is exhibited differently by elementary students of different genders. Therefore, gifted administrators and educators should be aware of this and be flexible in their interpretation of the TTCT-F results when evaluating boys and girls together. This study did not show measurement

invariance of the Innovative-Adaptive factors across genders, and specifically the model was not able to be fit to girls' responses. This finding calls into question whether the Innovative-Adaptive scoring model can be validly applied to girls' responses to the TTCT-F and suggests a need for future work specifically with girls' data, in order to improve their measurement.

5.3 Limitations and Future Directions

Recruiting more diverse samples is one important mission in the field of educational psychology research in order to better understand the underlying mechanisms which manifest across different cultures and contexts (Matthews & López, 2020). Future work can extend our study, which relied on demographic details provided by schools, to include increasingly disaggregated categories of ethnicities (e.g., different subgroups within the Asian category, such as Chinese, Koreans, and Indians), socioeconomic status, and disability status, which were not part of our analysis because the demographic details provided by the schools, and the sample size, were not sufficient for this purpose. Future research could replicate the current CFA and multigroup CFAs across a broader range of these demographic subgroups. Furthermore, the intersectionality of the demographic variables (e.g., Hispanic girls, Black males) was unable to be examined in this study, as the limited sample size did not allow for a multigroup CFA among the specific intersecting groups. The results might have been different, had we been able to consider the disaggregated categories of race/ethnicities or the intersectionality of the ethnicities and genders.

Another potential limitation is that the Checklist of Creative Strengths scale was not included in this study, because this scale is based on the 13 criterion-referenced measures and uses different scoring methods (Kim et al., 2006). Qualitative studies (e.g., case studies, content analysis of TTCT responses) could also be conducted in the future to explore how creative

thinking is expressed differently across the Checklist of Creative Strengths scale among the different groups.

Although our study revealed gender differences in the strength of the associations among the TTCT-F scales (e.g., larger positive correlation between Fluency and Originality in girls than in boys), we were not able to fully investigate the underlying psychological reasons behind these differences. Given that the TTCT-F Form A scores were obtained as a part of the larger federally funded project, we were incapable of conducting the follow-up studies that could closely examine intra-individual creative processes while engaging in the TTCT-F test. Previous interview studies with children clearly have been helpful in revealing the psychological processes behind their idea generation (e.g., Berthiaume et al., 2024), and therefore conducting interviews with these children might yield a deeper understanding of how students of different genders use cognitive and metacognitive strategies in a different way. Lastly, given that only the TTCT-F form A was employed in this study, the results might have been different if the form B had been used. To broaden our findings, replicating studies with the TTCT-F form B is essential, in order to determine whether our findings would hold across different TTCT stimuli.

5.4 Conclusion

In our TTCT-F dataset, which was obtained from 379 U.S. elementary students of diverse cultural backgrounds, the Innovative-Adaptive two-factor model did appear to be the best latent structure to explain the TTCT-F scales. Although this model did hold invariantly across over- and under-represented ethnicities, it did not hold across genders. Our findings run counter to past empirical work where the two-factor model functioned similarly across boys and girls (Kim et al., 2016; Krumm et al., 2014, 2016). Differences in the way that creative thinking might

manifest across elementary students of different gender therefore requires additional research in order to be better understood.

References

- Abernathy Tannehill, R. L. (1997). *An analysis of selected creativity tests administered to students affiliated with the Cherokee tribe* (Order No. 9801472). Available from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (304363591). <https://www.proquest.com/dissertations-theses/analysis-selected-creativity-tests-adminstered/docview/304363591/se-2>
- Acar, S. (2023). Does the task structure impact the fluency confound in divergent thinking? An investigation with TTCTFigural. *Creativity Research Journal*, 35(1), 1–14. <https://doi.org/10.1080/10400419.2022.2044656>
- Acar, S., Dumas, D., Organisciak, P., & Berthiaume, K. (2024). Measuring original thinking in elementary school: Development and validation of a computational psychometric approach. *Journal of Educational Psychology*, 116(6), 953–981. <https://doi.org/10.1037/edu0000844>
- Acar, S., Lee, L. E., & Hodges, J. (2023). Assessing the Robustness of the Factor Structure of TTCT-Figural: A Meta-CFA Replication-Extension. *Creativity Research Journal*, 35(4), 547–567. <https://doi.org/10.1080/10400419.2023.2209393>
- Acar, S., Lee, L. E., & Scherer, R. (2024). A reliability generalization of the Torrance Tests of Creative Thinking-Figural. *European Journal of Psychological Assessment*. <https://doi.org/10.1027/1015-5759/a000819>
- AERA, APA, NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

- Alabbasi, A. M. A., Paek, S. H., Kim, D., & Cramond, B. (2022). What do educators need to know about the Torrance Tests of Creative Thinking: A comprehensive review. *Frontiers in psychology, 13*, 1000385. <https://doi.org/10.3389/fpsyg.2022.1000385>
- Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Boulder, CO: Westview
- Ball, O. E., & Torrance, E. P. (1984). Streamlines scoring workbook: Figural A. *Bensenville, IL: Scholastic Testing Service*.
- Baer, J. (2015). The Importance of Domain-Specific Expertise in Creativity. *Roeper Review, 37*(3), 165–178. <https://doi.org/10.1080/02783193.2015.1047480>
- Bart W. M., Hokanson B., & Can I. (2017). An investigation of the factor structure of the Torrance Test of Creative Thinking. *Educational Sciences: Theory & Practice, 17*(2), 515–528. <https://doi.org/10.12738/estp.2017.2.0051>
- Berthiaume, K. C., Acar, S., & Dumas, D. (2024). “My Brain Said That ... ”: A Qualitative Study of Sources of Children’s Creative Ideas. *Creativity Research Journal, 1*–25. <https://doi.org/10.1080/10400419.2024.2336739>
- Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children’s interests. *Science, 355*(6323), 389-391. <https://doi.org/10.1126/science.aah6524>
- Chase, C. I. (1985). Review of the Torrance tests of creative thinking. *The ninth mental measurements yearbook, 2*, 1630-1632.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal, 14*(3), 464-504. <https://doi.org/10.1080/10705510701301834>

- 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>
- Clark, P. M., & Mirels, H. L. (1970). Fluency as a pervasive element in the measurement of creativity. *Journal of Educational Measurement, 7*(2), 83-86.
<https://doi.org/10.1111/j.1745-3984.1970.tb00699.x>
- Cohen, R. J., & Swerdlik, M. E. (2009). *Psychological Testing and Assessment: An Introduction to Tests and Measurement* (7th Ed.). McGraw-Hill.
- Cole, J. R., & Zuckerman, H. (1987). Marriage, motherhood and research performance in science. *Scientific American, 256*(2), 119-125. <https://doi.org/10.7312/cole21260-017>
- Cordon, S. L., & Finney, S. J. (2008). Measurement Invariance of the Mindful Attention Awareness Scale Across Adult Attachment Style. *Measurement and Evaluation in Counseling and Development, 40*(4), 228-245.
<https://doi.org/10.1080/07481756.2008.11909817>
- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation, 10*(1). <https://doi.org/10.7275/jyj1-4868>
- Covington, N. R. (1968). *Creativity in culturally deprived adolescent boys*. The Florida State University.

- Cramond, B. (1994). The Torrance Tests of Creative Thinking: From design through establishment of predictive validity. In R. F. Subotnik & K. D. Arnold (Eds.), *Beyond Terman: Contemporary longitudinal studies of giftedness and talent* (pp. 229-254). Ablex.
- Cramond, B., 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>
- Crombie, G., Bouffard-Bouchard, T., & Schneider, B. H. (1992). Gifted programs: Gender differences in referral and enrollment. *Gifted Child Quarterly*, 36(4), 213-214.
<https://doi.org/10.1177/001698629203600408>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3), 297-334. <https://doi.org/10.1007/BF02310555>
- Crutzen, R., & Peters, G. J. Y. (2017). Scale quality: Alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review*, 11(3), 242–247.
<https://doi.org/10.1080/17437199.2015.1124240>
- Daniels, V. I. (1998). Minority students in gifted and special education programs: The case for educational equity. *The Journal of Special Education*, 32(1), 41-43.
<https://doi.org/10.1177/002246699803200107>
- Dong, Y., & Dumas, D. (2020). Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age. *Personality and Individual Differences*, 160, 109956. <https://doi.org/10.1016/j.paid.2020.109956>

- Dumas, D., Dong, Y., Grajzel, K., Forthmann, B., & Doherty, M. (2022). Understanding ideational fluency as a survival process. *British Journal of Educational Psychology*, 92(2), 645-666. <https://doi.org/10.1111/bjep.12469>
- Dumas, D., Forthmann, B., & Alexander, P. (2024). Using a model of domain learning to understand the development of creativity. *Educational Psychologist*, 1-16. <https://doi.org/10.1080/00461520.2023.2291577>
- Dumas, D., McNeish, D., & Greene, J. A. (2020). Dynamic measurement: A theoretical–psychometric paradigm for modern educational psychology. *Educational Psychologist*, 55(2), 88-105. <https://doi.org/10.1080/00461520.2020.1744150>
- Engelhard Jr, G. (1989). Historical Views of the Concept of Invariance and Measurement Theory in the Behavioral Sciences.
- Engelhard Jr, G. (2008). Historical perspectives on invariant measurement: Guttman, Rasch, and Mokken. *Measurement*, 6(3), 155-189. <https://doi.org/10.1080/15366360802197792>
- Efrat, B., & Orna, T. (2022). Sex differences in the sustained attention of elementary school children. *BMC psychology*, 10(1), 307. <https://doi.org/10.1186/s40359-022-01007-z>
- Erwin, J. O., & Worrell, F. C. (2012). Assessment practices and the underrepresentation of minority students in gifted and talented education. *Journal of Psychoeducational Assessment*, 30(1), 74-87. <https://doi.org/10.1177/0734282911428197>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, 4(3), 272-299. <https://doi.org/10.1037/1082-989X.4.3.272>

- Ford, D. Y. (2014). Segregation and the Underrepresentation of Blacks and Hispanics in Gifted Education: Social Inequality and Deficit Paradigms. *Roeper Review*, 36(3), 143–154.
<https://doi.org/10.1080/02783193.2014.919563>
- Ford, D. Y., & Harris III, J. J. (1994). Promoting achievement among gifted Black students: The efficacy of new definitions and identification practices. *Urban Education*, 29(2), 202-229. <https://doi.org/10.1177/0042085994029002006>
- Ford, D. Y., Harris III, J. J., Tyson, C. A., & Trotman, M. F. (2001). Beyond deficit thinking: Providing access for gifted African American students. *Roeper Review*, 24(2), 52-58.
<https://doi.org/10.1080/02783190209554129>
- Forthmann, B., Oyebade, O., Ojo, A., Günther, F., & Holling, H. (2018). Application of latent semantic analysis to divergent thinking is biased by elaboration. *The Journal of Creative Behavior*, 53(4), 559–575. <https://doi.org/10.1002/jocb.240>
- Forthmann, B., Szardenings, C., & Dumas, D. (2021). On the conceptual overlap between the fluency contamination effect in divergent thinking scores and the chance view on scientific creativity. *The Journal of Creative Behavior*, 55(1), 268-275.
<https://doi.org/10.1002/jocb.445>
- 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.
<https://doi.org/10.1037/aca0000196>
- Georgia Department of Education. (2024). *Resource manual for gifted education services*.
https://lor2.gadoe.org/gadoe/file/1dacad8d-a918-4dcc-afae-4ab1172eda82/1/_2024-2025%20Gifted%20Resource%20Manual_.pdf

- Gonthier, C., & Besançon, M. (2024). It is not always better to have more ideas: Serial order and the trade-off between fluency and elaboration in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, 18(4), 480–492. <https://doi.org/10.1037/aca0000485>
- Grantham, T. C. (2013). Creativity and equity: The legacy of E. Paul Torrance as an upstander for gifted Black males. *The Urban Review*, 45, 518-538. <https://doi.org/10.1007/s11256-013-0257-2>
- Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53(4), 267–293. <https://doi.org/10.1037/h0040755>
- Guilford, J. P. (1959). Three faces of intellect. *American Psychologist*, 14(8), 469–479. <https://doi.org/10.1037/h0046827>
- Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1), 3-14. <https://doi.org/10.1002/j.2162-6057.1967.tb00002.x>
- Hahm, J., Kim, K. K., & Park, S. H. (2019). Cortical correlates of creative thinking assessed by the figural Torrance Test of Creative Thinking. *NeuroReport*, 30(18), 1289-1293. <https://doi.org/10.1097/WNR.0000000000001358>
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. duToit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future – A festschrift in honor of Karl Jöreskog* (pp. 195–216). Scientific Software International.
- Harris, C. R., & Jenkins, M. (2006). Gender Differences in Risk Assessment: Why do Women Take Fewer Risks than Men? *Judgment and Decision Making*, 1(1), 48–63. <https://doi:10.1017/S1930297500000346>

- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods*, 7(2), 191-205. <https://doi.org/10.1177/1094428104263675>
- 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>
- Helson, R., & Picano, J. (1990). Is the traditional role bad for women?. *Journal of personality and social psychology*, 59(2), 311.
- Hocevar, D. (1979). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology*, 71(2), 191–196. <https://doi.org/10.1037/0022-0663.71.2.191>
- Hodges, J., Tay, J., Maeda, Y., & Gentry, M. (2018). A meta-analysis of gifted and talented identification practices. *Gifted child quarterly*, 62(2), 147-174. <https://doi.org/10.1177/0016986217752107>
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, 60(4), 523–531. <https://doi.org/10.1177/00131640021970691>
- Horn, J. L., & Mcardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117–144. <https://doi.org/10.1080/03610739208253916>
- 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>
- 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.
<https://doi.org/10.1177/0734282911428196>
- Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity research journal*, 18(1), 3-14.
https://doi.org/10.1207/s15326934crj1801_2
- Kim, K. H. (2010). Measurements, causes, and effects of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 4(3), 131–135. <https://doi.org/10.1037/a0018964>
- Kim, K. H. (2011a). The creativity crisis: The decrease in creative thinking scores on the Torrance Tests of Creative Thinking. *Creativity research journal*, 23(4), 285-295.
<https://doi.org/10.1080/10400419.2011.627805>
- Kim, K. H. (2011b). The APA 2009 Division 10 debate: Are the Torrance Tests of Creative Thinking still relevant in the 21st century? *Psychology of Aesthetics, Creativity, and the Arts*, 5(4), 302–308. <https://doi.org/10.1037/a0021917>
- 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>
- Kirton, M. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology*, 61(5), 622–629. <https://doi.org/10.1037/0021-9010.61.5.622>

- Kirton, M. J. (1978). Adaptors and innovators in culture clash. *Current Anthropology*, 19(3), 611-612. <https://doi.org/10.1086/202148>
- Kirton, M. J. (1989). Adaptors and innovators atwork. In M. J. Kirton (Ed.), *Adaptors and innovators: Styles of creativity and problem-solving* (pp. 56-78). Routledge.
- Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.
- Krumm, G., Filippetti, V. A., Lemos, V., Koval, J., & Balabanian, C. (2016). Construct validity and factorial invariance across sex of the Torrance Test of Creative Thinking–Figural Form A in Spanish-speaking children. *Thinking Skills and Creativity*, 22, 180-189. <https://doi.org/10.1016/j.tsc.2016.10.003>
- 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>
- Lee, L. E., Rinn, A. N., & Rambo-Hernandez, K. E. (2024). What happens after nomination? Evaluating the probability of gifted identification with the Torrance Test of Creative Thinking. *Gifted Child Quarterly*, 68(2), 119-136. <https://doi.org/10.1177/00169862231222886>
- Lohman D. L., & Foley Nicpon M. (2012). Ability testing and talent identification. In S. L. Hunsaker (Ed.), *Identification: The theory and practice of identifying students for gifted and talented education services* (pp. 283-335). Creative Learning Press.
- Lubart, T., & Zenasni, F. (2010). A new look at creative giftedness. *Gifted and Talented International*, 25(1), 53-57. <https://doi.org/10.1080/15332276.2010.11673549>

- Luria, S. R., O'Brien, R. L., & Kaufman, J. C. (2016). Creativity in gifted identification: Increasing accuracy and diversity. *Annals of the New York Academy of Sciences*, 1377(1), 44-52.
- Maker, C. J. (2005). *The DISCOVER Project: Improving assessment and curriculum for diverse gifted learners* (RM05206). University of Connecticut, The National Research Center on the Gifted and Talented.
- Matthews, J. S., & López, F. (2020). Race-reimagining educational psychology research: Investigating constructs through the lens of race and culture. *Contemporary Educational Psychology*, 61, 101878. <https://doi.org/10.1016/j.cedpsych.2020.101878>
- McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, 23(1), 1-21. <https://doi.org/10.1111/j.2044-8317.1970.tb00432.x>
- McDonald, R. P. (1999). *Test theory: A unified approach*. Erlbaum.
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- McVey, L. A. (2003). *Single -sex schooling and girls' gender -role identity and creativity* (Order No. 3101155). Available from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (305327723). <https://www.proquest.com/dissertations-theses/single-sex-schooling-girls-gender-role-identity/docview/305327723/se-2>
- Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo Study of Confirmatory Factor Analytic Tests of Measurement Equivalence/Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(1), 60–72
https://doi.org/10.1207/S15328007SEM1101_5

- Meredith, W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543 (1993). <https://doi.org/10.1007/BF02294825>
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11. <https://doi.org/10.3102/0013189X018002005>
- Messick, S. (1994). Foundations of validity: Meaning and consequences in psychological assessment. *European Journal of Psychological Assessment*, 10(1), 1–9.
- Millar, G. W. (2002). *The Torrance kids at mid-life: Selected case studies of creative behavior*. Ablex Publishing.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.
- Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological methods*, 9(1), 93–115. <https://doi.org/10.1037/1082-989X.9.1.93>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Peters S. J., Gentry M., Whiting G. W., McBee M. T. (2019). Who gets served in gifted education? Demographic representation and a call for action. *Gifted Child Quarterly*, 63(4), 273–287. <https://doi.org/10.1177/0016986219833738>
- Pfeiffer, S. I. (2012). Current Perspectives on the Identification and Assessment of Gifted Students. *Journal of Psychoeducational Assessment*, 30(1), 3-9. <https://doi.org/10.1177/073428291142819>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review*, 41, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>

- Renzulli, J. (2010). Scales for Rating the Behavioral Characteristics of Superior Students: Technical and Administration Manual (3rd ed.). Routledge.
<https://doi.org/10.4324/9781003237808>
- Revelle, W. (2020). psych: Procedures for psychological, psychometric, and personality research. *R package version*, 2(5).
- Rodriguez, V. J., & Shaffer, A. (2020). Validation of the Regulating Emotions in Parenting Scale (REPS): Factor Structure and Measurement Invariance. *Journal of Family Psychology*. Advance online publication. <http://dx.doi.org/10.1037/fam0000808>
- 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.
<https://doi.org/10.1080/10400419.2018.1530534>
- Saeki, N., Fan, X., & Van Dusen, L. (2001). A comparative study of creative thinking of American and Japanese college students. *The Journal of Creative Behavior*, 35(1), 24-36.
<https://doi.org/10.1002/j.2162-6057.2001.tb01219.x>
- Şahin, F. (2015). A Research on the structure of intelligence and creativity, and creativity style. *Talent*, 5(1), 2-20.
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human resource management review*, 18(4), 210-222.
<https://doi.org/10.1016/j.hrmr.2008.03.003>
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified

strategy. *Journal of applied psychology*, 91(6), 1292-1306. <https://doi.org/10.1037/0021-9010.91.6.1292>

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and psychological measurement*, 56(2), 197-208.

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking: Norms technical manual (Research Edition)*. Princeton: Personnel Press.

Torrance, E. P. (1967). *Understanding the fourth grade slump in creative thinking*. University of Minnesota.

Torrance, E. P. (1971). Are the Torrance tests of creative thinking biased against or in favor of “disadvantaged” groups?. *Gifted child quarterly*, 15(2), 75-80.

<https://doi.org/10.1177/001698627101500201>

Torrance, E. P. (1974). *Norms-technical manual: Torrance Tests of Creative Thinking*. Bensenville: Scholastic Testing Service.

Torrance, E. P. (1977). *Discovery and nurturance of giftedness in the culturally different*. Council on Exceptional Children.

Torrance, E. P. (1984). *The Torrance Tests of Creative Thinking norms-technical manual for figural Forms A and B*. Bensenville: Scholastic Testing Service.

Torrance, E. P. (1990). *The Torrance Tests of Creative Thinking Norms—Technical Manual Figural (streamlined) Forms A & B*. Bensenville, IL: Scholastic Testing Service.

Torrance, E. P. (1998). *The Torrance Tests of Creative Thinking Norms—Technical Manual Figural (streamlined) Forms A & B*. Bensenville, IL: Scholastic Testing Service

- Torrance, E. P. (2004). The role of creativity in identification of the gifted and talented. U: DJ Treffinger (Ed.) *Creativity and giftedness* (pp. 79-87).
- Torrance, E. P. (2017). *The Torrance Tests of Creative Thinking: Norms-technical manual, figural streamlined forms A and B*. Bensenville: Scholastic Testing Service.
- Torrance, E. P., & Torrance, P. (1972). Combining creative problem-solving with creative expressive activities in the education of disadvantaged young people. *The Journal of Creative Behavior*, 6(1), 1–10. <https://doi.org/10.1002/j.2162-6057.1972.tb00902.x>
- Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and psychological measurement*, 58(4), 541-568. <https://doi.org/10.1177/001316449805800>
- U. S. Department of Education. (2020). *Civil rights data collection: 2017-18 state and national estimations*. <https://ocrdata.ed.gov/estimations/2017-2018>
- Vandenberg, R. J. (2002). Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures. *Organizational Research Methods*, 5(2), 139-158. <https://doi.org/10.1177/1094428102005002001>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, 3(1), 4-70. <https://doi.org/10.1177/109442810031002>
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492. <https://doi.org/10.1080/17405629.2012.686740>

- Venker, C. C., Goodwin, J. L., Roe, D. J., Kaemingk, K. L., Mulvaney, S., & Quan, S. F. (2007). Normative psychomotor vigilance task performance in children ages 6 to 11—the Tucson Children’s Assessment of Sleep Apnea (TuCASA). *Sleep and Breathing, 11*, 217-224. <https://doi.org/10.1007/s11325-007-0103-4>
- Wallach, M. A., & Kogan, N. (1965). A new look at the creativity-intelligence distinction 1. *Journal of personality, 33*(3), 348-369. <https://doi.org/10.1111/j.1467-6494.1965.tb01391.x>
- Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*(2), 575–587. <https://doi.org/10.1037/0012-1649.44.2.575>
- Xu, X., Chen, Y., Ye, X., Zhang, J., Luan, J., & Li, Y. Creativity: Exploring Factor Structure of the Torrance Tests of Creative Thinking in Chinese Preschoolers. *The Journal of Creative Behavior. <https://doi.org/10.1002/jocb.1529>*
- Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment, 29*(4), 377–392. <https://doi.org/10.1177/0734282911406668>
- Yoon, C. H. (2017). A validation study of the Torrance Tests of Creative Thinking with a sample of Korean elementary school students. *Thinking Skills and Creativity, 26*, 38-50. <https://doi.org/10.1016/j.tsc.2017.05.004>
- Yoon, S. Y., & Gentry, M. (2009). Racial and ethnic representation in gifted programs: Current status of and implications for gifted Asian American students. *Gifted Child Quarterly, 53*(2), 121-136. <https://doi.org/10.1177/0016986208330564>

Table 1*Bivariate Correlations among the TTCT-F Scales for the Entire Sample*

	F	O	E	AT	RPC
F	1.00				
O	.77**	1.00			
E	.28**	.37**	1.00		
AT	.10*	.23**	.48**	1.00	
RPC	.58**	.57**	.32**	.28**	1.00
SD	21.98	21.09	16.49	22.71	20.73

Note. F = Fluency; O = Originality; E = Elaboration; AT = Abstractness of Titles; RPC = Resistance to Premature Closure.

* $p < .05$. ** $p < .01$.

Table 2*Descriptive Statistics of the TTCT-F Scales for the Entire Sample*

	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
F	103.99	21.98	-.36	-.11
O	104.39	21.09	-.37	.03
E	76.76	16.49	.32	-.52
AT	92.82	22.71	.05	-.45
RPC	108.04	20.73	-.62	.13

Note. F = Fluency; O = Originality; E = Elaboration; AT = Abstractness of Titles; RPC = Resistance to Premature Closure.

Table 3*Correlations between the TTCT-F Scales for Gender Groups*

	Boys (<i>n</i> = 176)					Girls (<i>n</i> = 199)				
	F	O	E	AT	RPC	F	O	E	AT	RPC
F	1.00					1.00				
O	.68***	1.00				.86***	1.00			
E	.26***	.43***	1.00			.28***	.30***	1.00		
AT	.23**	.40***	.51***	1.00		.00	.07	.48***	1.00	
RPC	.56***	.65***	.42***	.42***	1.00	.61***	.48***	.23***	.13	1.00
SD	22.48	21.79	16.46	22.05	23.10	21.30	20.37	16.04	23.15	18.33

Note. F = Fluency; O = Originality; E = Elaboration; AT = Abstractness of Titles; RPC = Resistance to Premature Closure.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4*Model Comparison via Fit Indices*

Models	Chi-Square Test			Fit Indices			
	χ^2	<i>df</i>	χ^2/df	CFI	TLI	RMSEA	SRMR
Model 1	107.20***	5	21.44	.85	.70	.23	.09
Model 2	.04	1	.04	>.99	>.99	>.99	<.001
Model 3	6.02**	2	3.01	.99	.97	.07	.02
Model 4	58.50***	2	29.25	.92	.59	.27	.07

Note. The values of the best fit model were bolded.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5*Measurement Invariance Fit Comparisons across Overrepresented and Underrepresented Ethnicities*

Invariance Level	χ^2	df	$\Delta\chi^2$	CFI	Δ CFI	TLI	Δ TLI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
M1. Configural Invariance	1.00	2		>.999		>.999		<.001		.007	
M2. Metric Invariance	8.93	6	7.94	.996	.005	.985	.030	.052	.052	.061	.054
M3. Scalar Invariance	14.13	9	5.19	.992	.004	.983	.002	.056	.004	.060	.001

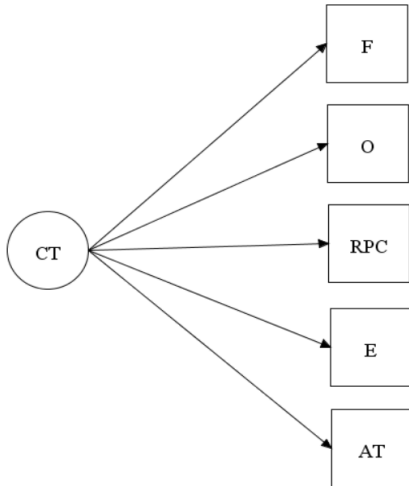
Table 6*Factor Loadings of the TTCT-F Scales for Boys and Girls from the Exploratory Factor Analysis*

Scale	Boys (<i>n</i> = 176)		Girls (<i>n</i> = 199)	
	Factor 1: Innovative	Factor 2: Adaptive	Factor 1: Innovative	Factor 2: Adaptive
Fluency	0.886	-0.121	1.087	-0.100
Originality	0.754	0.183	0.801	0.050
Elaboration	0.022	0.687	0.109	0.663
AT	-0.043	0.749	-0.177	0.771
RPC	0.549	0.302	0.551	0.105
Factor correlation	0.538	-	0.314	-

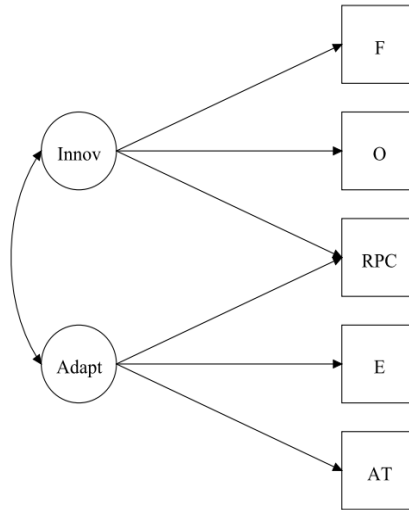
Note. AT = Abstractness of Titles; RPC = Resistance to Premature Closure. Pattern coefficients equal to or higher than .40 are bolded.

Figure 1

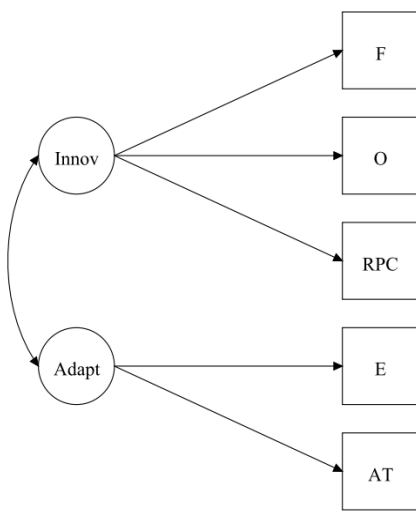
Theoretically Plausible Factor Models that were Fit



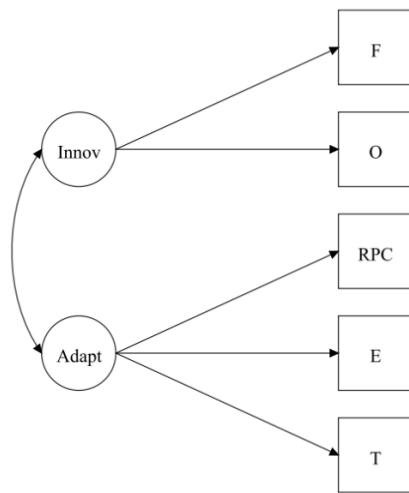
Model 1



Model 2



Model 3

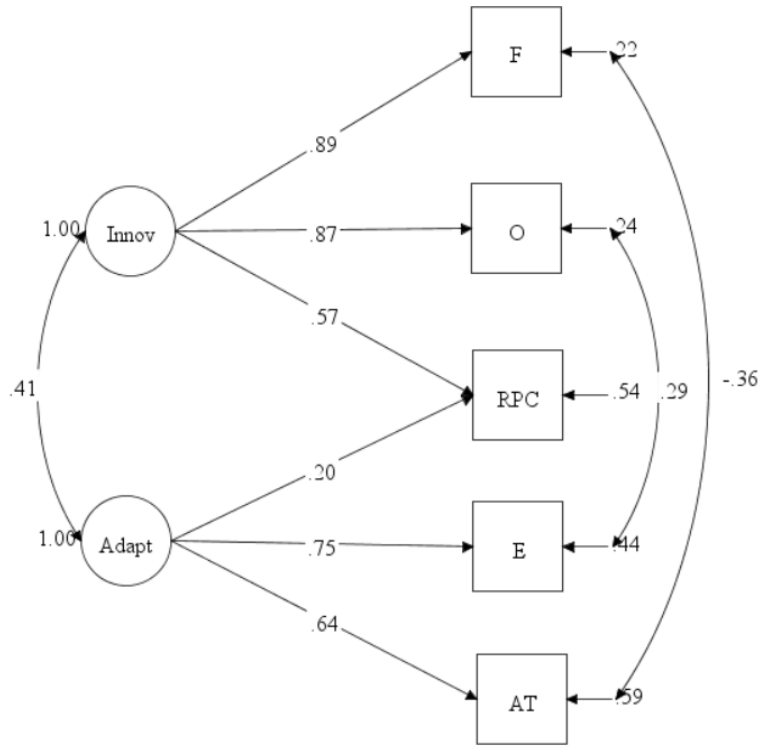


Model 4

Note. CT = Creative Thinking factor; Innov = Innovative factor; Adapt = Adaptive factor; F = Fluency; O = Originality; RPC = Resistance to Premature Closure; E = Elaboration; AT = Abstractness of Titles.

Figure 2

The Best-fitting Model for the Entire Elementary Student Sample, with Standardized Coefficients



Note. Innov = Innovative factor; Adapt = Adaptive factor; F = Fluency; O = Originality; RPC = Resistance to Premature Closure; E = Elaboration; AT = Abstractness of Titles.

Figure 3

Parallel Analysis Results for Boys

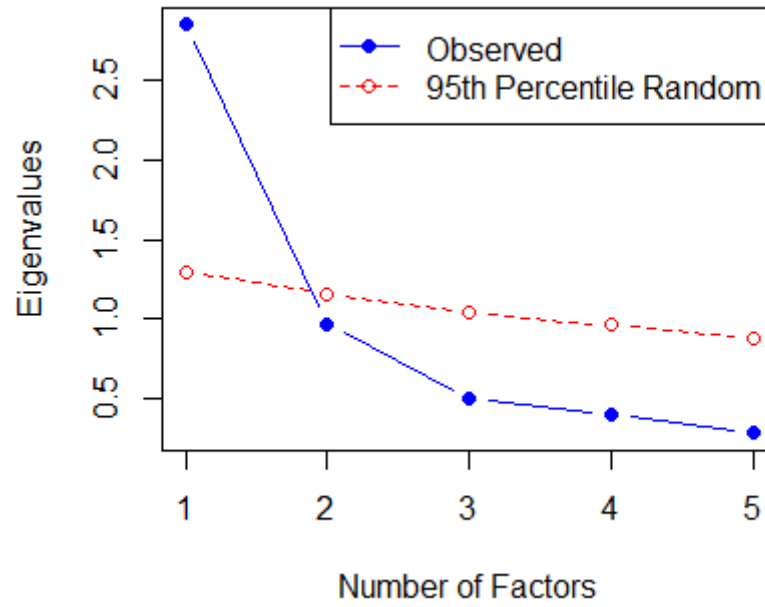


Figure 4

Parallel Analysis Results for Girls

