2006

The latent structure of scores on the Torrance Tests of Creativity-Figural

K. H Kim
B. Cramond
Deborah L Bandalos, *James Madison University*

Educational and Psychological Measurement
Volume 66 Number 3
June 2006 459-477

10.1177/0013164405282456
http://epm.sagepub.com
hosted at
http://online.sagepub.com

The Latent Structure and Measurement Invariance of Scores on the Torrance Tests of Creative Thinking–Figural

Kyung Hee Kim
Eastern Michigan University
Bonnie Cramond
Deborah L. Bandalos
The University of Georgia

There is disagreement among researchers as to whether creativity is a unidimensional or multidimensional trait. Much of the debate centers around the most widely used measure of creativity, the Torrance Tests of Creative Thinking (TTCT). This study used data from 1,000 kindergartners (ages 5-7), 1,000 third graders (ages 7-11) and 1,000 sixth graders (ages 10-13). Confirmatory factor analyses were conducted for both the two-factor model and one-factor model to determine which fit the data better. Measurement invariance across genders and grade levels was assessed using multiple group analyses in which sets of parameters were freed sequentially in a series of hierarchically nested models. The findings indicate that the structure of TTCT scores is consistent with a two-factor theory. Also, the results of the multiple group analyses indicate that model parameters for gender groups are more invariant than for grade levels in determining the fit of the model.

Keywords: *creativity; measurement invariance; Torrance Tests of Creative Thinking; latent structure; factor analysis; structural equation modeling; dimensionality; validity; reliability*

As Simonton (2000) related, creativity is often portrayed as a mysterious and even mystical process, more akin to divine inspiration than to mundane thought. Thus, creativity is often viewed as one of the most enigmatic subjects in cognitive psychology. According to several researchers (Barron, 1961; MacKinnon, 1961; Walberg, 1988), creativity and intelligence are separate constructs; that is, a highly intelligent person may or may not be highly creative. A growing interest in creativity has been accompanied in past decades by a search for appropriate assessment tools,

Authors' Note: Please address correspondence to Kyung Hee Kim, Assistant Professor of Educational Psychology, Department of Teacher Education, 313K Porter College of Education Building, Eastern Michigan University, Ypsilanti, MI 48197; e-mail: kkim7@emich.edu.

which has led to the development of creativity tests, specifically those designed for measuring creative potential in children (Mouchiroud & Lubart, 2001). Torrance (1988) suggested that testing is a legitimate way of learning about the nature of creativity. Creativity tests are also able to ascertain where a child's strengths and weaknesses lie (Cramond, 1994; Torrance, 1974). Therefore, tests of creativity are central to measuring creative potential, to our understanding of creativity, and, perhaps, to contributing to individuals' cognitive development.

Development of the Torrance Tests of Creative Thinking (TTCT)

The TTCT was first published by E. Paul Torrance and his associates in 1966. It has been renormed four times, in 1974, 1984, 1990, and 1998. There are two forms (A and B) of the TTCT-Verbal and two forms (A and B) of the TTCT-Figural. The TTCT-Verbal consists of five different types of activities: Ask-and-Guess, Product Improvement, Unusual Uses, Unusual Questions, and Just Suppose. The stimulus for each task includes a picture to which people respond in writing. Fluency, Flexibility, and Originality are used to score all activities (Torrance, 1966, 1974). For the scope of this study, only the TTCT-Figural was examined.

The TTCT has much to endorse its use (e.g., Cropley, 2000). It has been translated into more than 35 languages (Millar, 2002). It has become highly recommended in the educational field and is even used in the corporate world. It is the most widely used test of creativity (Colangelo & Davis, 1997) and has been used in more research than any other creativity test (Lissitz & Willhoft, 1985). The standard administration and scoring procedures (Davis & Rimm, 1994) as well as the development and evaluation (Colangelo & Davis, 1997) have made the TTCT especially useful for identifying gifted and talented students. The TTCT-Figural has had 25 years of extensive development and evaluation (Millar, 2002). It has large norming samples, valuable longitudinal validations, and high predictive validity for a very wide age range (Cropley, 2000). The TTCT-Figural is purported to be fair in terms of gender and race and for persons who have various language, socioeconomic status, and cultural backgrounds (Cramond, 1993; Torrance, 1977b). The scores can also be useful for counseling purposes (e.g., Cropley & Cropley, 2000).

Construct Validity Evidence of the TTCT

The extensive interest in tests of creativity, including the TTCT, has resulted in several validity studies concerned with the most accurate or appropriate way of measuring creativity. Validity is a critical issue for any test's survival. One useful way of conducting validity studies is analyzing the latent structure of scores from the instrument to provide a type of construct validity. In this study, we analyzed TTCT scores to understand their latent structure and to learn more about the cognitive function of creativity.

Since Guilford (1959, 1962) conceptualized divergent thinking as multidimensional, many researchers have concluded that creativity consisted of several independent psychological factors. The TTCT consists of five norm-referenced subscales: Flu-

ency, Originality, Abstractness of Titles, Elaboration, and Resistance to Premature Closure. Torrance (1966, 1974) has discouraged the use of composite scores for the TTCT. He warned that using a single score like a composite score may be misleading because each subscale score has an independent meaning.

However, studies on the TTCT have shown conflicting results regarding its dimensionality (Chase, 1985; Clapham, 1998; Dixon, 1979; Heausler & Thompson, 1988; Hocevar, 1979a, 1979b; Hocevar & Michael, 1979; Runco & Mraz, 1992; Treffinger, 1985). Hocevar (1979a, 1979b) concluded that the TTCT and Guilford's divergent thinking tests measured only Fluency rather than independent dimensions. In another study, Hocevar and Michael (1979) found that the heterotrait-monomethod coefficients were too high compared to monotrait-heteromethod coefficients based on multitrait-multimethod analyses using the TTCT and Guilford tests. However, only the TTCT-Verbal was used in their studies (Hocevar, 1979a, 1979b; Hocevar & Michael, 1979). Runco and Mraz (1992) also criticized the lack of discriminant validity of the TTCT dimensions in a study, including several other divergent thinking tests. Dixon's (1979) study also showed that in the scores of TTCT-Figural, Originality scores depended heavily upon Fluency scores. Abernathy Tannehill (1997) also indicated that the statistically significant correlation between Fluency and Originality suggested that the subscores of the TTCT may actually measure similar constructs. Likewise, Heausler and Thompson (1988) concluded that the correlations between the subscales were too high to provide meaningfully different information. After comparing the three subscores of Fluency, Flexibility, and Originality, Chase (1985) suggested that the correlation coefficients between them were so high (.74 to .80) that one single score could be appropriate for the three subscores. Treffinger (1985) also warned that interpretations of TTCT subscores as if they were independent should be avoided. Similarly, Hassan (1986) concluded that there was no justification for considering creativity as composed of the distinct traits recommended by Torrance (Fluency, Flexibility, Originality, and Elaboration).

There have only been a few published studies that analyzed the latent structure of TTCT scores. One of them was based on the TTCT-Figural Form A and modeled two components by a principal component analysis but concluded that the scores of the TTCT primarily reflected one general factor (Heausler & Thompson, 1988). Clapham's (1998) study had a similar conclusion that there was only one general factor; however, Resistance to Premature Closure explained the highest amount of the variance in the creativity index for both Forms A and B based on the result of a principal component analysis.

Kirton (1976, 1978, 1989) proposed that creativity is composed of a single dimension ranging from an Innovative to an Adaptive orientation, on which an individual's positions are representative of a personal approach to creativity, problem solving, and decision making (Puccio, Treffinger, & Talbot, 1995). However, Innovative and Adaptive may be separate dimensions rather than opposite ends of the same continuum. The trainers of the scoring training program at the University of Georgia have suggested that the TTCT may have two factors based on years of experience in scoring the TTCT (B. Cramond, personal communication, January 13, 2003). Anecdotally,

trained scorers of the TTCT have noticed that there are two types of people. One type was people who produced quick and novel responses and did better on fluency and originality; the other was people who were detailed and deep thinkers and did better on elaboration and abstractness of titles. According to Kirton (1987), innovators prefer to create change by threatening the paradigm, whereas adaptors prefer to create change by working within the existing paradigm. However, "creative style" as proposed by Kirton (1976, 1978, 1989) has been conceptually distinct from degree of creativity and is concerned with the way an individual approaches problem solving rather than with individual's "creative ability" (Gelade, 2002). This indicates that a relationship does not necessarily exist between a person's style and his or her creative level. However, Isaksen and Puccio (1988) suggested that the distinction between creative style and creative level as measured by the TTCT might not be as clear as asserted by Kirton. Isaksen and Puccio, as well as Torrance and Horng (1980), found that innovators were statistically significantly more fluent and more original. Puccio and others (1995) also found that Innovators were highly original, transformational, and expressive; whereas Adaptors were more logical, adequate, and well crafted. Therefore, it was hypothesized that the Innovative and Adaptive types of creativity by Kirton might provide an explanation of the latent structure of TTCT scores.

Research Hypotheses and Purposes

The purpose of this study was to investigate the latent structure of TTCT scores as a model of creative functioning. The hypothesis was that there would be two factors inherent in the TTCT Figural: a factor labeled Innovative consisting of Fluency and Originality subscales, and a second factor labeled Adaptive consisting of Elaboration and Abstractness of Titles. It was further hypothesized that both factors would include the Resistance to Premature Closure subscale. The relationships between the factors and five subscales were based on Kirton's (1976, 1978, 1989) descriptions of Innovative and Adaptive, Puccio and others' (1995) findings, and years of experience in scoring of the TTCT. In addition, hypotheses regarding the subscales contributing to Innovative were based on the findings of Torrance and other researchers (Isaksen & Puccio, 1988; Torrance & Horng, 1980) about the relationship between Innovative style and measures of Fluency and Originality as reported above. The logic for the double contribution of Resistance to Premature Closure originated from Torrance's (1984, 1990, 1998) theory that creative persons would be able to keep their minds open and delay closure long enough to make mental leaps, whereas less creative individuals tend to leap to conclusions prematurely without considering the available information. Therefore, Resistance to Premature Closure may be a complement to either style of creativity.

A secondary purpose of this study was to compare the relative salience of grade in school to gender in the fit of the model. From the different results of several factor analytic studies on the TTCT, including the studies presented above, we wanted to examine whether the TTCT was measuring creativity differently for different groups. Thus, multiple group analyses were conducted to understand the latent structure of TTCT

scores across gender and grade in school. We hypothesized that grade group membership would be a more salient influence than gender; specifically, it was hypothesized that the factor model and parameter values would be more similar across boys and girls than across the three grade levels. This hypothesis was based on Torrance's conclusions that the TTCT-Figural was fair in terms of gender (Cramond, 1993; Torrance, 1977b). It should be noted here that invariance of the measurement model is often framed as an issue of validity. For example, differences in structures or patterns of factor loadings or factor correlations may result if subscales have differential relevance or meaning in different groups.

Method

Participants

The TTCT can be used with kindergartners through graduate students, and there are two parallel forms, A and B, of the TTCT-Figural, but in this study only the TTCT-Figural Form-A was studied using the results from 3,000 students. The data for this study contained the results of 1,000 kindergartners, third graders, and sixth graders, respectively. There was a mixture of boys and girls in the three different grade groups. These data were obtained through the Scholastic Testing Service Company. No information on relevant demographics was reported. In order to enhance anonymity, the company does not collect any ethnic or demographic data from examinees, thus no descriptions for those were available. Cases for which gender was not reported were excluded from the data analyses involving cross-gender comparisons. This resulted in totals of 1,459 boys and 1,538 girls for the latter analyses. For the grade level comparisons, gender identification was not needed, so the initial 1,000 cases for each of the grades were included.

Measures

Although there have been several revisions of the TTCT-Figural manual, the test itself has remained unchanged. The TTCT-Figural consists of three activities: picture construction, picture completion, and repeated figures of lines or circles. These activities require 10 minutes each to complete (Torrance, 1966, 1974, 1984, 1990, 1998). All of the scores are continuous variables.

The original edition in 1966 measured Fluency, Flexibility, Originality, and Elaboration, which were taken from the divergent-thinking factors found in Guilford's Dimensions of Aptitude (Guilford, 1959; Torrance, 1966). The streamlined scoring system developed in 1979 results in 5 norm-referenced scores—Fluency, Originality, Elaboration, Abstractness of Titles, and Resistance to Premature Closure—and 13 criterion referenced scores, which make up the Creative Strength score (Torrance, 1984). These subscales are on somewhat different scales; thus, standard scores for those subscales were used for this study. The standard score ranges of each subscale are Fluency, 40-153; Originality, 40-154; Elaboration, 40-160; Abstractness of Titles, 40-

160; and Resistance to Premature Closure, 40-160. Because the scoring procedure of Creative Strengths is different from the other 5 subscales, it was not included in this study.

Because the TTCT is an open-ended instrument for which there is not a finite number of responses, the TTCT-Figural 1998 manual has provided the KR-21 internal consistency reliability estimates for the composite score using the 99th percentile scores for each grade or age as the estimate of the total number of items. These reliability estimates ranged between .89 and .94. According to the 1990 TTCT-Figural manual, the interrater reliability was above .90 (Torrance, 1990). According to the 1966 and 1974 TTCT-Figural manuals, the test-retest reliability coefficients (1-week, 2-week, 10-week, 6-month, and 3-year interval) were varied and not so high (from .50 to .93), but Torrance indicated that motivational conditions affect the reliability (Torrance, 1966, 1974). In the current study, the value of coefficient alpha was .79. Therefore, given the complexity of creative thinking, the TTCT-Figural can be seen as having reasonable reliability (Treffinger, 1985).

There have been many validity studies conducted on TTCT scores. In terms of predictive validity, TTCT scores have been statistically significantly correlated with creative achievement in 9-month, 7-year, 22-year, and 40-year longitudinal studies (Cramond, 1993; Millar, 2002; Torrance & Wu, 1981). Torrance's (1981) 22-year longitudinal study, Yamada and Tam's (1996) reanalysis, and Plucker's (1999) reanalysis of Torrance's data have concluded that the Creative Index was the best predictor for adult creative achievement. Plucker found that the path coefficient from the TTCT to adult creative achievement was .60, whereas the path coefficient from IQ score was .19. In terms of concurrent validity, Gonzales and Campos's (1997) study showed that the scores of the Spatial Test of Primary Mental Abilities (PMA) and the Gordon Test of Visual Imagery Control were statistically significantly correlated with the TTCT scores. This indicated that imagery was related to various aspects of creative thinking.

Data Analysis

Confirmatory factor analyses. Confirmatory factor analyses were conducted to test the fit of the proposed two-factor model for the entire sample and for each gender and grade level group, using the LISREL 8.53 program (Jöreskog & Sörbom, 2002a). We used covariance matrices generated by the PRELIS 2.51 program (Jöreskog & Sörbom, 2002b) as input to LISREL to analyze the confirmatory factor analysis model. Because the data were approximately normally distributed, maximum likelihood (ML) estimation was used for both confirmatory analyses and multiple group analyses.

Tables 1 and 2 contain correlation matrices and standard deviations for each gender and grade level group. All of the correlation coefficients between the variables were statistically significant at the .01 level except one ($p < .05$). Particularly, the correlation coefficients between Fluency and Originality (.812 for boys, .844 for girls, .836 for kindergartners, .785 for third graders, and .853 for sixth graders) were very high, as some researchers (Abernathy Tannehill, 1997; Chase, 1985; Dixon, 1979; Heausler &

Table 1
Correlations Between Variables for the Grade Levels With Standard Deviations (N = 3,000)

	Kindergartner (n = 1,000)					Third Grader (n = 1,000)					Sixth Grader (n = 1,000)				
	F	O	E	T	C	F	O	E	T	C	F	O	E	T	C
F															
O	.84					.79					.86				
E	.39	.38				.14	.15				.22	.25			
T	.45	.49	.49			.08	.15	.30			.31	.32	.41		
C	.65	.65	.34	.51		.47	.48	.17	.25		.66	.57	.20	.40	
SD	28.83	30.39	17.05	43.11	31.43	20.62	21.62	17.55	25.12	19.94	18.33	18.67	17.09	21.36	18.85

Note: F = Fluency; O = Originality; E = Elaboration; T = Abstractness of Titles; C = Resistance to Premature Closure. All correlations are statistically significant at $p < .01$ (two-tailed) except for r F-T = .08 in the third grade group, which is statistically significant at $p < .05$ (two-tailed).

Table 2
Correlations Between Variables for the Gender Groups With Standard Deviations (N = 2,997)

	F	O	E	T	C	*SD*
F		.84	.25	.29	.58	22.75
O	.81		.26	.33	.58	23.26
E	.27	.28		.39	.22	17.67
T	.39	.45	.35		.44	33.39
C	.63	.62	.27	.48		23.24
SD	23.55	25.47	16.64	34.74	25.74	

Note: Correlations for girls (n = 1,538) are above the diagonal and correlations for boys (n = 1,459) are below the diagonal. F = Fluency; O = Originality; E = Elaboration; T = Abstractness of Titles; C = Resistance to Premature Closure. All correlations are statistically significant at $p < .01$.

Thompson, 1988) had found before. Because several researchers have concluded that the TTCT is composed of only one general latent factor, we also conducted analyses specifying one general factor.

Multiple group analyses. The second part of this study was designed to examine whether the same factor model and parameter values would hold across both boys and girls and across the three grade levels. The factor loadings from Innovative to Originality and from Adaptive to Abstractness of Titles were set to 1.0 to scale the latent variable (Bollen, 1989). The analyses were accomplished through the multiple group procedures available in the LISREL 8.53 program. Measurement invariance across gender and grade was assessed using multiple-group procedures in which sets of parameters were freed sequentially in a series of hierarchically nested models. In addition to these multiple group analyses across gender and grade, three grade level invariance tests within each gender were conducted to examine possible interaction effects between grade and gender.

Results

One- and Two-Factor Models

Fit indexes for both the one- and two-factor models are shown in Table 3. In assessing model fit we followed the two-index strategy and indexes of fit suggested by Hu and Bentler (1998, 1999). This included reporting root mean square error of approximation (RMSEA) or standardized root mean square residual (SRMR) and supplementing it with one of the following: nonnormed fit index (NNFI [Tucker-Lewis index; TLI]), incremental fix index (IFI), comparative fit index (CFI), or relative noncentrality index (RNI). Chi-square differences were computed to test the difference in fit between the one- and two-factor models. For the NNFI, IFI, and CFI, values vary between 0 and 1.0, and values of .95 and above are considered to indicate a good model fit (Hu & Bentler, 1995, 1999). For the RMSEA (Steiger, 1990), values of about .05 are conventionally considered to indicate a close fit, and values up to about .08 are considered reasonable, whereas Hu and Bentler (1999) recommended a cutoff close to .06. For the SRMR (Jöreskog & Sörbom, 1986) a cutoff of .08 or less is recommended by Hu and Bentler (1999).

As Table 3 shows, the statistically significant chi-square difference tests for each gender and grade level group and for the total group indicate that the two-factor model had a much better fit. This was supported by the values of the fit indexes. We further analyzed the two-factor model by examining parameter estimates for the total (N = 3,000) group. Parameter estimates, standard errors, and values of R^2 for the total group are shown in Table 4. The large values of the factor loadings and large R^2 values indicated that the subscales were strongly related to their factors. However, the low R^2 value for Elaboration suggests that this subscale is not as highly related to its factor as were the other subscales.

Table 3
Results of Model Comparison of One- and Two-Factor Models

Group	Number of Factors	χ^2	*df*	NNFI	IFI	CFI	RMSEA	SRMR	$\Delta\chi^{2a}$
Boys (*n* = 1,459)	One	194.80*	5	.89	.94	.94	.16	.062	
	Two	33.94*	3	.97	.99	.99	.083	.015	160.86*
Girls (*n* = 1,538)	One	322.74*	5	.80	.90	.90	.20	.091	
	Two	41.32*	3	.96	.99	.99	.090	.028	281.42*
Kindergarten (*n* = 1,000)	One	177.51*	5	.88	.94	.94	.18	.072	
	Two	14.69*	3	.99	1.00	1.00	.062	.014	162.82*
Third grade (*n* = 1,000)	One	138.63*	5	.79	.90	.90	.16	.085	
	Two	19.35*	3	.96	.99	.99	.087	.029	119.28*
Sixth grade (*n* = 1,000)	One	228.58*	5	.79	.90	.90	.21	.096	
	Two	25.87*	3	.96	.99	.99	.087	.029	202.71*
Total group (*N* = 3,000)	One	517.10*	5	.85	.92	.92	.18	.077	
	Two	69.37*	3	.97	.99	.99	.085	.020	447.73*

Note: NNFI = nonnormed fit index; IFI = incremental fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.
a. Indicates comparisons are to the previous model, 1 with 2, 2 with 3, and so forth.
*$p < .01$.

Table 4
Parameter Values and Standard Errors for the Two-Factor Model (N = 3,000)

Path	Factor Loading	Measurement Error Variance	R^2	Factor Correlation
F-Innovative	20.98 (0.35)	96.45 (6.38)	.82	.52 (.02)
O-Innovative	22.25 (0.37)	98.65 (7.08)	.83	
C-Innovative	12.38 (0.53)	294.67 (9.04)	.51	
E-Adaptive	8.19 (0.29)	231.50 (6.55)	.22	
T-Adaptive	27.15 (0.93)	428.08 (41.73)	.63	
C-Adaptive	7.45 (0.58)	294.67 (9.04)	.51	

Note: Parameter values are unstandardized; standard errors are in parentheses. F = Fluency; O = Originality; E = Elaboration; T = Abstractness of Titles; C =Resistance to Premature Closure.

Multiple Group Analyses Across Gender

Separate covariance matrices for boys (n = 1,459) and girls (n = 1,538) were used as input for the multiple group analyses across gender. These analyses involved testing to determine whether the two-factor model would fit equally well for boys and girls. A series of nested models were analyzed and compared by examining the change in model χ^2 values. The first model in this sequence is one in which all model parameters (factor loadings, measurement error variances, and factor variances and covariance) were invariant. In the second model, factor loadings were freed to vary across groups, measurement error variances were freed in the third model, and factor variances and covariance were freed in the final model. A statistically significant decrease in χ^2 values between adjacent models indicates that the release of cross-group invariance constraints results in a statistically significantly better fit. This is taken as an indication that the constrained parameters are not invariant. Chi-square values and difference tests, and values of other fit indexes for the series of analyses, are shown in Table 5. The comparisons of models resulted in statistically significant χ^2 differences for the analyses in which the factor loadings, error variances, and factor variances and covariance were freed across groups, indicating that the factor loadings ($\Delta\chi^2(4)$ = 16.64, $p < .01$), error variances ($\Delta\chi^2(5) = 42.38$, $p < .01$), and factor variances and covariance ($\Delta\chi^2(3) = 18.89$, $p < .01$) differed statistically significantly across gender.

However, although statistically significant decreases in χ^2 were found, the various fit indexes still showed remarkably good fits to the model with the constraint of invariant error variances and/or factor variances and covariance. Furthermore, these differences did not appear to be substantial when we examined the actual parameter values. The measurement error variances were slightly larger for boys than girls except that of Elaboration-Adaptive although the factor covariance was notably larger for boys (377.83) than girls (253.30). Because the χ^2 difference tests could be influenced by the large sample sizes (Bentler, 1993; Jöreskog & Sörbom, 1989), we also considered Cheung and Rensvold's suggestion (2002) that a difference of CFI of less than or equal to .01 is an indication that the constrained parameters are invariant. Differences in the CFI values were 0 for all model comparisons except that between Models 2 and 3 as shown in Table 5. We examined the modification indexes (MIs) to determine

Table 5
Model Comparisons Across Gender (n = 1,459 for Boys and n = 1,538 for Girls)

Model	χ^2	df	NNFI	IFI	CFI	RMSEA	SRMR	$\Delta\chi^{2a}$	Δdf^a	ΔCFI[a]
1. All invariant	153.17*	18	.98	.98	.98	.071	.062			
2. Factor loadings free	136.53*	14	.97	.98	.98	.077	.058	16.64*	4	0
3. Model 2 + Error Variances Free	94.15*	9	.97	.99	.99	.079	.048	42.38*	5	.01
4. Models 2-3 + Factor Variances and Covariance Free	75.26*	6	.97	.99	.99	.087	.028	18.89*	3	0

Note: NNFI = nonnormed fit index; IFI = incremental fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.
a. Indicates comparisons are to the previous model, 1 with 2, 2 with 3, and so forth.
*$p < .01$.

whether parameter values differed across two groups. In multiple group analyses, values of the MIs for parameters held invariant represent the amount that the overall χ^2 value would decrease if the parameter values were allowed to differ. These examinations revealed that the MIs for the parameters were small and negligible. Thus, it can be concluded that at a practical level, the two-factor model may be sufficiently useful to describe the data for both boys and girls. Therefore no post hoc modifications of the model for these groups were undertaken.

Multiple Group Analyses Across Grades

Separate covariance matrices for kindergartners (n = 1,000), third graders (n = 1,000), and sixth graders (n = 1,000) were used as input for this set of multiple group analyses. As Table 6 shows, a lack of measurement equivalence across the kindergartners, third graders, and sixth graders was found; that is, error variances ($\Delta\chi^2[10]$ = 1,048.78, $p < .01$) and factor variances and factor covariance ($\Delta\chi^2[6] = 759.75$, $p < .01$) were statistically significantly different across the three groups. Furthermore, the differences of the CFI values (Δ Model 2 vs. Model 3: .16 for the measurement error variances; Δ Model 3 vs. Model 4: .12 for the factor variances and covariance) were greater than .01. Thus, it can be concluded that at both a statistical and a practical level, neither the same factor variances and covariance nor the same measurement error variances held for the three grade groups. As Model 1 in Table 6 shows, all of the fit indexes for the assessment of measurement equivalence in which the factor loadings, error variances, and factor variances and covariance were fixed as equal across groups showed a poor fit for the model.

We also looked at the actual parameter values for the three grade level groups. The values of the measurement error for Adaptive (1,002.37 for kindergarteners, 312.84 for third graders, and 89.99 for sixth graders) and R^2 (.67 for kindergarteners, .41 for third graders, and .75 for sixth graders) were quite different across the three groups. The values of the factor variances of Innovative (631.01 for kindergarteners, 371.35 for third graders, and 309.82 for sixth graders) and Adaptive (820.48 for kindergarteners, 393.80 for third graders, and 385.99 for sixth graders) and those of the factor covariance (593.90 for kindergarteners, 61.87 for third graders, and 123.17 for sixth graders) were also quite different across the three groups. We examined the MIs to determine which parameter values differed across the three groups. These examinations revealed measurement error variances and factor variances and covariance that were not invariant.

Multiple Group Analyses Across Grades Within Each Gender

We conducted further tests of grade level invariance within each gender to examine possible interactions between genders and grade levels. For both boys and girls, all of the fit indexes for the assessment of measurement equivalence in which the factor loadings, error variances, and factor variances and covariance were fixed as equal across groups showed the same poor fit for the model (NNFI = .80, IFI = .78, CFI = .78, and RMSEA = .20 for both boys and girls, and SRMR = .29 for boys and .24 for

Table 6
Model Comparisons Across Grades (n = 1,000 for the Three Grade Levels, Respectively)

Model	χ^2	*df*	NNFI	IFI	CFI	RMSEA	SRMR	$\Delta\chi^{2a}$	Δdf^a	ΔCFI[a]
1. All invariant	1,430.42*	33	.80	.78	.78	.20	.27			
2. Factor loadings free	1,868.45*	25	.65	.71	.71	.28	.35	–438.03*	8	–.07
3. Model 2 + Error Variances Free	819.67*	15	.75	.87	.87	.23	.26	1,048.78*	10	.16
4. Models 2-3 + Factor Variances and Covariance Free	59.92*	9	.97	.99	.99	.075	.029	759.75*	6	.12

Note: NNFI = nonnormed fit index; IFI = incremental fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.
a. Indicates comparisons are to the previous model, 1 with 2, 2 with 3, and so forth.
*$p < .01$.

girls; χ^2 716.88 for boys and 732.32 for girls). These results were similar to those from the invariance tests across grades for the total group as shown in Table 6. The values of difference χ^2 tests and fit indexes (Δ Model 1 vs. Model 2, Δ Model 2 vs. Model 3, and Δ Model 3 vs. Model 4) were also similar to those from the grade level invariance tests reported in Table 6. These results suggested that there were no grade level by gender interactions with regard to invariance of parameter values but that parameter invariance was primarily due to grade level differences.

Discussion

This study examined the possibility of a two-factor model based on Kirton's (1976, 1978, 1989) Adaptor-Innovator (A-I) Theory. We are aware of no previously published studies on complete relationships between A-I Theory and the latent structure of TTCT scores. We hypothesized that the proposed model with Factor Innovative and Factor Adaptive would be a good fit for the entire sample. According to this model, Innovative consisted of Fluency and Originality; Adaptive consisted of Elaboration and Abstractness of Titles; and both factors were related to Resistance to Premature Closure. Based on the results of this study, the proposed two-factor structure of TTCT scores fits well. We also conducted confirmatory factor analyses with one general factor to compare the results with one factor to those with two hypothesized factors in this study. The χ^2 values and different fit indexes suggested that the proposed two-factor model in this study was a much better fit.

These findings are inconsistent with the empirical and theoretical literature on the TTCT in that Torrance (1966, 1974) suggested six different factors and in that several factor analytic studies related to the TTCT concluded that the TTCT had only one factor. However, the findings in this study were consistent with the descriptions of Innovative and Adaptive of Kirton's A-I Theory and with other researchers' findings (Puccio et al., 1995). In addition, we feel the results have certain face validity because we have found the same pattern of results from a long experience in scoring of the TTCT.

Other findings were more consistent with previous research. One was that the mean standard scores for all of the subscales were greater for the third graders than sixth graders. This might be related to Torrance's (1968, 1977a; see also Davis, 1992; Kang, 1989; Marcon, 1995; Nash, 1974; Timmel, 2001; Torrance & Gupta, 1964; Walker, 1995; Williams, 1976) famous observation of the "fourth-grade slump," which is a large drop in creativity at the fourth grade associated with the imposition of social demands. The finding that all of the correlation coefficients between the variables were high, particularly the correlation coefficients between Fluency and Originality, was consistent with the work of Torrance and Safter (1999), who reasoned that the person who produces a large number of alternatives is more likely to produce original ones. The large measurement error covariances between Fluency and Originality might also be explained by this reasoning. Simonton (1990) also found that a person's originality is a function of the number of ideas formulated. In addition, the high correlations between all the variables here might partly result from the fact that the five dif-

ferent subscores were based on the same stimuli. Because of the high correlations between subscales, especially the correlation between Fluency and Originality (from .79 and .86), many researchers (Chase, 1985; Clapham, 1998; Dixon, 1979; Heausler & Thompson, 1988; Hocevar, 1979a, 1979b; Hocevar & Michael, 1979; Runco & Mraz, 1992; Treffinger, 1985) have insisted that the TTCT measures only one factor, or the subscales are not independent. However, the analyses we conducted support a two-factor structure in which the factors are correlated at .52.

The second part of this study was designed to examine if the values of model parameters were invariant across gender and grade levels. For this purpose, multiple group analyses were conducted across gender and grades. Even though the χ^2 differences across gender were statistically significant for the factor loadings, measurement error variances, and factor covariance, most of the actual differences in parameter values were small. This suggests that the results of these tests were heavily influenced by the large sample sizes. In addition, none of the differences of the CFI values were greater than .01. Thus, it can be concluded that at a practical level, the two-factor model is sufficient to describe the data from both the boy and girl groups, indicating invariance of the model parameters across gender.

In contrast, the results of the invariance tests across kindergarten, third-grade, and sixth-grade groups indicated that a lack of measurement equivalence existed across the groups. The measurement error variances and factor variances and covariance were statistically significantly different across groups. In addition, many parameter values exhibited fairly substantial differences across the three groups. We examined the MIs to determine which parameter values differed across the three grade levels. These examinations revealed measurement error variances and factor variances and covariance that were not invariant.

In conclusion, the results of the confirmatory analyses indicated that the two-factor model proposed in this study had a much better fit than the one-factor model. This indicates that the TTCT can give more information not only about individuals' strengths and weaknesses from the scores of each of the five subscales but also about their types of creativity: Innovative or Adaptive. The results of the multiple group analyses indicated that the latent structure of TTCT scores showed more differences across grade level groups than gender groups. These findings are consistent with Torrance's conclusion (1977b; see also Cramond, 1993) that the TTCT-Figural was fair in terms of gender. It is interesting that different grade level groups not only have different mean scores but also somewhat different factor structures. This indicates that when TTCT scores are compared among different grade levels, more caution may be needed for interpretation.

Limitations

There are several important limitations of this study. The TTCT is a complex measure and is complicated by the dependence of all scales on the same stimuli as the high correlation coefficients between the subscales have shown before. Second, the subscale of Creative Strengths was not included in this study because of the different

procedures in scoring, which may explain why there have been no published studies using Creative Strengths.

Third, this data was obtained through the Scholastic Testing Service Company, and no information regarding relevant demographics was provided. Demographics should also be considered in confirmatory factor analyses or multiple group analyses using the TTCT. This could give more information about understanding the TTCT, other creativity tests, the nature of creativity, and, ultimately, how to encourage creativity in individuals. Furthermore, motivation (Bamber, 1973; Halpin & Halpin, 1973; Torrance, 1966, 1974) and testing conditions (Bamber, 1973; Callahan, 1991; Halpin & Halpin, 1973) as well as exposure to diverse information (Clapham, 2000-2001) may influence TTCT-Figural scores.

References

Abernathy Tannehill, R. L. (1997). *An analysis of selected creativity tests administered to students affiliated with the Cherokee tribe* (Doctoral dissertation, Mississippi State University). UMI Dissertation Services. Ann Arbor, MI: A Bell & Howell.

Bamber, R. T. (1973). Play, interest, domestication and creativity (Doctoral dissertation, University of Missouri). *Dissertation Abstracts International, 35*, 1013B-1014B. (University Microfilms No. 74-18, 463)

Barron, F. (1961). Creative vision and expression in writing and painting. In D. W. MacKinnon (Ed.), *The creative person* (pp. 237-251). Berkeley: University of California, Institute of Personality Assessment Research.

Bentler, P. M. (1993). *EQS: Structural equations program manual* (Version 4.0). Los Angeles: BMDP Statistical Software.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*, 303-316.

Callahan, C. M. (1991). The assessment of creativity. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 219-235). Boston: Allyn & Bacon.

Chase, C. I. (1985). Review of the Torrance Tests of Creative Thinking. In J. V. Mitchell Jr. (Ed.), *The ninth mental measurements yearbook* (pp. 1631-1632). Lincoln: University of Nebraska, Buros Institute of Mental Measurements.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

Clapham, M. M. (1998). Structure of Figural Forms A and B of the Torrance Tests of Creative Thinking. *Educational and Psychological Measurement, 58*, 275-283.

Clapham, M. M. (2000-2001). The effects of affect manipulation and information exposure on divergent thinking. *Creativity Research Journal, 13*, 335-350.

Colangelo, N., & Davis, G. A. (1997). *Handbook of gifted education*. Needham Heights, MA: Viacom.

Cramond, B. (1993). 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). Norwood, NJ: Ablex.

Cramond, B. (1994). We can trust creativity tests. *Educational Leadership, 52*(2), 70-71.

Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using? *Roeper Review, 23*(2), 72-79.

Cropley, D. H., & Cropley, A. J. (2000). Fostering creativity in engineering undergraduates. *High Ability Studies, 11*(2), 207-219.

Davis, G. A. (1992). *Creativity is forever* (3rd ed.). Dubuque, IA: Kendall/Hunt.

Davis, G. A., & Rimm, S. B. (1994). *Education of the gifted and talented*. Needham Heights, MA: Simon & Schuster.

Dixon, J. (1979). Quality versus quantity: The need to control for the fluency factor in originality scores from the Torrance Tests. *Journal for the Education of the Gifted, 2*, 70-79.

Gelade, G. A. (2002). Creative style, personality, and artistic endeavor, *Genetic, Social, and General Psychology Monographs, 128*, 213-234.

Gonzales, M. A., & Campos, A. (1997). Mental imagery and creative thinking. *Journal of Psychology, 131*, 357-364.

Guilford, J. P. (1959). *Personality.* New York: McGraw-Hill.

Guilford, J. P. (1962). Factors that aid and hinder creativity. *Teachers College Record, 63*, 380-392.

Halpin, G., & Halpin, G. (1973). The effect of motivation on creative thinking abilities. *Journal of Creative Behavior, 7*, 51-53.

Hassan, M. A. (1986). Construct validity of Torrance Tests of Creative Thinking: A confirmatory factor-analytic study (Doctoral dissertation, Claremont Graduate School). *Dissertation Abstracts International, 46*(8-A), 2233.

Heausler, N. L., & Thompson, B. (1988). Structure of the Torrance Tests of Creative Thinking. *Educational and Psychological Measurement, 48*, 463-468.

Hocevar, D. (1979a). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology, 71*, 191-196.

Hocevar, D. (1979b). The unidimensional nature of creative thinking in fifth grade children. *Child Study Journal, 9*, 273-278.

Hocevar, D., & Michael, W. (1979). The effects of scoring formulas on the discriminant validity of tests of divergent thinking. *Educational and Psychological Measurement, 39*, 917-921.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1998). Fit indexes in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.

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

Isaksen, S. G., & Puccio, G. J. (1988). Adaptation—Innovation and the Torrance Tests of Creative Thinking: The level style-issue revisited. *Psychological Reports, 63*, 659-670.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL 6: Analysis of linear structural relationships by maximum likelihood and least square methods*. Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7 user's reference guide*. Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (2002a). *LISREL 8.53*. Lincolnwood, IL: Scientific Software International, Inc.

Jöreskog, K. G., & Sörbom, D. (2002b). *PRELIS 2.51*. Lincolnwood, IL: Scientific Software International, Inc.

Kang, C. (1989). *Gender differences in Korean Children's responses to the Torrance Tests of Creative Thinking from first to sixth grade*. Unpublished master's thesis, University of Wisconsin–Madison.

Kirton, M. J. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology, 61*, 622-629.

Kirton, M. J. (1978). Have adaptors and innovators equal levels of creativity? *Psychological Reports, 42*, 695-698.

Kirton, M. J. (1987). *Kirton Adaptation-Innovation Inventory manual* (2nd ed.). Hatfield, UK: Occupational Research Centre.

Kirton, M. J. (1989). Adaptors and innovators at work. In M. J. Kirton (Ed.), *Adaptors and innovators: Styles of creativity and problem-solving* (pp. 56-78). London: Routledge.

Lissitz, R. W., & Willhoft, J. L. (1985). A methodological study of the Torrance Tests of Creativity. *Journal of Educational Measurement, 22*, 1-11.

MacKinnon, D. W. (1961). Creativity in architects. In D. W. MacKinnon (Ed.), *The creative person* (pp. 291-320). Berkeley: University of California, Institute of Personality Assessment Research.

Marcon, R. A. (1995). Fourth-grade slump: The cause and cure. *Principal, 74*(5), 16-17.

Millar, G. W. (2002). *The Torrance kids at mid-life*. Westport, CT: Ablex.

Mouchiroud, C., & Lubart, T. (2001). Children's original thinking: An empirical examination of alternative measures derived from divergent thinking tasks. *Journal of Genetic Psychology, 162*, 382-401.

Nash, W. R. (1974). The effects of a school for the gifted in averting the fourth grade slump in creativity. *Gifted Child Quarterly, 18*, 168-170.

Plucker, J. A. (1999). Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data. *Creativity Research Journal, 12*, 103-114.

Puccio, G. J., Treffinger, D. J., & Talbot, R. J. (1995). Exploratory examination of relationships between creativity styles and creative products, *Creativity Research Journal, 8*, 157-172.

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

Simonton, D. K. (1990). Creativity and wisdom in aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 320-329). San Diego, CA: Academic Press.

Simonton, D. K. (2000). Creativity: Cognitive, personal, developmental, and social aspects. *American Psychologist, 55*(1), 151-158.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Timmel, J. L. (2001). Creativity and acculturation: Psychological and cultural effects on the divergent thinking of Cuban preadolescent immigrants entering the United States. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 61*(9-A), 3468.

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking—Norms, Technical Manual Research Edition—Verbal Tests, Forms A and B—Figural Tests, Forms A and B*. Princeton, NJ: Personnel Press.

Torrance, E. P. (1968). A longitudinal examination of the fourth-grade slump in creativity. *Gifted Child Quarterly, 12*, 195-199.

Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking—Norms, Technical Manual Research Edition—Verbal Tests, Forms A and B—Figural Tests, Forms A and B*. Princeton, NJ: Personnel Press.

Torrance, E. P. (1977a). *Creativity in the classroom*. Washington, DC: National Education Association.

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

Torrance, E. P. (1981). Empirical validation of criterion-referenced indicators of creative ability through a longitudinal study. *Creative Child & Adult Quarterly, 6*, 136-140.

Torrance, E. P. (1984). *The Torrance Tests of Creative Thinking Streamlined (Revised) Manual Figural A and B*. Bensenville, IL: Scholastic Testing Service.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 43-75). New York: Cambridge University Press.

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., & Gupta, R. K. (1964). *Programmed experiences in creative thinking*. Minneapolis: University of Minnesota, Bureau of Educational Research.

Torrance, E. P., & Horng, R. Y. (1980). Creativity and style of learning and thinking characteristics of adaptors and innovators. *Creative Child and Adult Quarterly, 5*, 80-85.

Torrance, E. P., & Safter, H. T. (1999). *Making the creative leap beyond . . .* Buffalo, NY: Creative Education Foundation Press.

Torrance, E. P., & Wu, T. (1981). A comparative longitudinal study of the adult creative achievements of elementary school children identified as highly intelligent and as highly creative. *Creative Child and Adult Quarterly, 6*, 71-76.

Treffinger, D. J. (1985). Review of the Torrance Tests of Creative Thinking. In J. V. Mitchell Jr. (Ed.), *The ninth mental measurements yearbook* (pp. 1632-1634). Lincoln: University of Nebraska, Buros Institute of Mental Measurements.

Walberg, H. J. (1988). Creativity and talent as learning. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 340-361). New York: Cambridge University Press.

Walker, M. L. (1995). Help for the "fourth-grade slump": SPQ2R plus instruction in text structure or main idea. *Reading Horizons, 36*(1), 38-58.

Williams, F. E. (1976). Rediscovering the fourth-grade slump in a study of children's self-concept. *Journal of Creative Behavior, 10*, 15-28.

Yamada, H., & Tam, A. Y. (1996). Prediction study of adult creative achievement: Torrance's longitudinal study of creativity revisited. *Journal of Creative Behavior, 30*, 144-149.