

# INVESTIGATIONS OF THE VALIDITY OF DIVERGENT THINKING TESTS

by

Sue Hyeon Paek

(Under the Direction of Mark A. Runco)

## ABSTRACT

Despite the prevalent usage of divergent thinking (DT) tests in creativity research, DT tests have been plagued by the lack of discriminant and criterion-related validity. In previous literature, the complex nature and scoring procedure of DT tests were not fully dealt with in examining the discriminant and criterion-related validity of DT tests, which may have weakened empirical evidence on validity. Therefore, the current studies aimed to investigate the discriminant and criterion-related validity of DT tests by taking into account the complexity and procedure of DT tests.

The first study examined the discriminant validity of DT indices, ruling out a potential method effect entailed by scoring procedure in DT tests using multitrait-multimethod analysis. Results indicated that the method effect was present, and fluency and originality were discriminable against one another when the method effect was ruled out. This finding provides empirical evidence indicative of the discriminant validity of DT indices. The second study investigated the criterion-related validity of DT tests using latent profile analysis. Results indicated that three groups with their own DT profiles were present, and this group membership significantly explained individual differences in creative achievement and ideational behaviors. This finding implies that the low correlations between DT and creative performance in a

conventional approach do not necessarily indicate a lack in the criterion-related validity of DT tests where heterogeneous groups are present.

INDEX WORDS: divergent thinking, discriminant validity, criterion-related validity, method effect, divergent thinking profiles

INVESTIGATIONS OF THE VALIDITY OF DIVERGENT THINKING TESTS

by

SUE HYOEN PAEK

B.A., Seoul National University of Education, Korea, 2003

M.A., Seoul National University, Korea, 2008

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

© 2017

Sue Hyeon Paek

All Rights Reserved

# INVESTIGATIONS OF THE VALIDITY OF DIVERGENT THINKING TESTS

by

SUE HYOEN PAEK

Major Professor:	Mark A. Runco
Committee:	Bonnie Cramond
	Zhenqiu Lu
	Selcuk Acar

Electronic Version Approved:

Suzanne Barbour  
Dean of the Graduate School  
The University of Georgia  
August 2017

## DEDICATION

To my family and friends.

## ACKNOWLEDGEMENTS

I would like to thank Drs. Mark A. Runco, Bonnie Cramond, Zhenqiu Lu, Selcuk Acar, and Sarah Sumners.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	V
LIST OF TABLES .....	VIII
LIST OF FIGURES .....	IX
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW .....	1
The Definition and Measurement of Divergent Thinking .....	2
The Strengths of Divergent Thinking Test .....	3
The Weakness of Divergent Thinking Test .....	5
Objects of the Present Studies.....	7
References.....	9
2 THE DISCRIMINANT VALIDITY OF A DIVERGENT THINKING TEST:	
MULTITRAIT-MULTIMETHOD ANALYSIS .....	14
Abstract .....	15
Introduction.....	16
The Present Study .....	22
Method .....	23
Results.....	28
Discussion .....	29
Conclusion .....	35



References .....	36
3 LATENT PROFILE ANALYSIS OF THE CRITERION-RELATED VALIDITY OF A DIVERGENT THINKING TEST.....	49
Abstract .....	50
Introduction.....	51
The Present Study .....	57
Method .....	58
Results.....	63
Discussion .....	65
Conclusion .....	71
References .....	72
4 CONCLUSION.....	87
References.....	91

## LIST OF TABLES

	Page
Table 2.1: Descriptive Statistics of Divergent Thinking Test Scores .....	45
Table 2.2: Model Comparison .....	46
Table 2.3: Correlations between Latent Factors and Composite Scores of Divergent Thinking ..	47
Table 3.1: Descriptive Statistics and Correlations of the DT Indices, CAAC, and RIBS Scores .....	81
Table 3.2: Fit Indices for Mixture Models.....	82
Table 3.3: Correlations and Means of Factor Scores of DT Tests in Three Groups.....	83
Table 3.4: Mean Differences in the CAAC Subscales and the RIBS by Group Membership.....	84

## LIST OF FIGURES

	Page
Figure 2.1: Path diagram of model with a method effect .....	48
Figure 3.1: Mixture model examining the groups with different profiles of divergent thinking in children .....	85
Figure 3.2: Profiles of divergent thinking.....	86

## CHAPTER 1

### INTRODUCTION AND LITERATURE REVIEW

Divergent thinking (DT) is one of the most popular constructs for assessing creative thinking (Hocevar, 1981). Valid assessments of DT are important in making valid inferences regarding creative thinking and performance. Also, because DT assessments provide information for interventions to enhance creativity development among individuals, including children, it is important that they should be valid (Crocker & Algina, 1986). However, it is challenging to assess DT in a reliable and valid manner due to the complex nature of DT. For instance, DT is not only multi-dimensional but also affected both by personal and environmental factors (Michael & Wright, 1989). The complex nature of DT is a potential obstacle in investigating the psychometric qualities of DT tests. If this complex nature of DT is not adequately addressed, the psychometric qualities of DT tests such as validity could be over- or underestimated. Despite the popularity of DT tests, the validity of DT tests, in fact, have been questioned based on conflicting findings and the interpretation of those findings (see Hocevar, 1981; Michael & Wright, 1989; Zeng, Proctor, & Salvendy, 2011 for a review of critiques on measurement issues of DT). Yet, the complexity of DT is often neglected in examining the validity of DT tests. The current studies, therefore, attempted to take the complexity of DT into account in examining the validity of DT tests. In the following section, previous literature is reviewed regarding the definition and measurement of DT, and strengths and weakness of DT tests to address gaps to be filled in the current studies.

## **The Definition and Measurement of Divergent Thinking**

The creative thinking process consists of multiple sequential phases such as problem discovery, divergent thinking, and the evaluation of ideas (Runco, 1991). At the problem discovery phase, a problem is discovered to be solved. Then, solutions to the problem are generated at the divergent thinking phase, followed by selecting the best solution which is novel as well as appropriate at the idea evaluation phase. But, it should be noted that these phases are not always linear but can nonlinearly recur (Mumford, Mobley, Uhlman, Reiter-Palmon, & Doares, 1991). Given this creative thinking process, divergent thinking is defined as generating solutions or ideas to given problems in different directions (Guilford, 1956).

Guilford (1956) originally suggested the concept of DT based on the Structure of Intellect (SOI) model as well as tests to quantify DT. In DT tests, Guildford suggested multiple indices such as fluency, flexibility, originality, and elaboration, which represent conceptually different properties of DT. Research has mainly used these indices (Getzels & Jackson, 1962; Torrance, 1974, 1981; Wallach & Kogan, 1965). However, it is still controversial whether creative thinking is unidimensional or multidimensional (see Hocevar, 1979a, 1979b; Kim, 2006 for a review of conflicting findings). On one hand, it is argued that creative thinking is ‘g’ factor since a general creative thinking ability involves multiple facets of DT (e.g., Hocevar, 1979a; 1979b). On the other hand, it is claimed that creative thinking is a multifaceted construct in which facet has a distinct cognitive property (e.g., Kim, 2006).

There are four major DT tests which are commonly used to measure the creative thinking process: SOI Divergent Production Tests (Guilford, 1967), Torrance Tests of Creative Thinking (Torrance, 1974), Wallach-Kogan Creativity Tests (Wallach & Kogan, 1965), and Getzels-Jackson Creativity Tests (Getzels & Jackson, 1962). These DT tests share a common operational

definition of DT and multiple DT indices but include dissimilar contents and tasks in tests (Zeng et al., 2011). Despite the significant contributions of these DT tests to creativity research, a few critical questions about the psychometric qualities of DT tests are still unresolved. For example, validity is one of the critical issues in DT tests that many leading scholars have addressed, because how to define and measure ideation is still controversial (Clapham, 2004; Guilford, 1971; Hargreaves & Bolton, 1972; Hocevar, 1979a, 1979b; Plucker, 1999; Runco, 1986; Silvia et al., 2008; Treffinger, Renzulli, & Feldhusen, 1971).

Messick (1995) defined validity as “an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessments.” Simply put, validity indicates the extent to which a test measures well what we believe it measures. In other words, poor validity of DT tests fundamentally threatens credibility of what we measure in DT tests. Also, the well-established validity of DT tests is necessary because a measurement is used in deciding on interventions and inference (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Given the concerns about the validity and complex nature of DT, the validity of DT tests should, therefore, be examined in rigorous ways to enhance their credibility.

### **The Strengths of Divergent Thinking Tests**

The strengths of DT tests are discussed in three respects. First, DT tests provide a framework to scientifically measure the creative process (Guilford, 1956). DT tests allow us to quantify divergence in ideation as four indices. This framework has spurred research on measuring the creative process to incorporate various creativity related constructs. For instance, DT tests are modified to measure creative thinking (Torrance, 1974; Wallach & Kogan, 1965),

problem solving (Okuda, Runco, & Berger, 1991), problem finding (Runco & Okuda, 1987), problem finding in arts (Csikszentmihalyi & Getzels) and science (Hoover & Feldhusen, 1990), and associative thinking (Acar & Runco, 2014; Mednick, 1962).

Second, that framework of quantifying divergence in ideation suggests multiple facets of DT. Although abilities measured in DT tests are often referred to as DT, DT is a collective label for multiple properties of DT. Because creative thinking is complex in nature, tools measuring creative thinking should embrace this nature. Multiple indices in DT tests representing different properties in DT contribute to better understanding the complex nature of creative thinking and its relationships with other covariates. For instance, multiple indices are used to explain groups with different creative performance (von Stumm, Chung, & Furnham, 2011), the relationships of DT to intelligence (Kim, 2008), and the relationships of DT to creative personality, behaviors, and performance (Plucker, Qian, & Schmalensee, 2014). In these studies, patterns in their relationship to other covariates are not identical across indices.

Lastly, the reliability of DT tests is adequate as supported in many studies using different methods including inter-item consistency (Chronbach's  $\alpha = .81-.89$  in Runco, 1986), proportion of true variance explained by latent DT factors ( $H = .64-.87$  in Silvia, 2011), and test-retest reliability ( $r = .9$  in Torrance, 1990). Also, the validity of the DT tests in some studies is also supported, despite the controversy that still remained in other studies. Behind those popular usages of DT tests, there is the theoretical notion that creative individuals are more likely to have this ability (Guilford, 1968; Torrance, 1974; Wallach & Wing, 1969). Simply put, this divergence in ideation may serve as one of keys in creative thinking as well as creative performance. Several longitudinal studies support that indices measured in DT tests certainly explain individual differences in creative performance both as single composite scores (Plucker,

1999) and individual indices (Runco, Millar, Acar, & Cramond, 2010; Torrance, 1969).

Successfully predicting creative performance by these DT indices serves as evidence, in part, supporting the criterion-related validity of DT.

### **The Weakness of Divergent Thinking Tests**

Due to these advantages of DT tests, DT testing is one of the most prevalent measures assessing creative thinking (Hocevar, 1981). However, the validity of DT tests have been plagued with conflicting findings on this validity (see Hocevar, 1981; Zeng et al., 2011 for a review). This chapter specifically discusses two validity issues, the discriminant and criterion-related validity of DT tests, which are consistently critiqued in the comprehensive reviews (Hocevar, 1981; Zeng et al., 2011).

The discriminant validity of DT tests involves the relationships among the indices of DT tests. Ever since Guilford (1956, 1967) suggested four indices of DT, research has mainly used these indices (Torrance, 1974, 1981; Wallach & Kogan, 1965). Despite the conceptually different properties of these indices, many empirical studies found high interdependency among indices (Hocevar, 1981; Zeng et al., 2011) and even suggested DT indices equal to linearly weighted fluency (Hocevar, 1979a, 1979b). This argument supports unidimensionality and serves as a rationale to use a fluency score as a single indicator even though originality is closer to creativity (Runco & Acar, 2012). In contrast, Kim (2006) found strong empirical evidence that DT consists of more than one dimension. Runco (2011) was also against using fluency as a single indicator and suggested using multiple indices to better capture different properties of DT corresponding to their operational definitions. This issue raises the question of whether these indices are truly discriminable, or if specific scoring procedures in DT tests may cause this



problem. Still, this issue has not yet been investigated with a framework ruling out a possible obstacle due to this specific scoring procedure.

The criterion-related validity of DT tests on creative performance has also been questioned in studies showing low correlations between DT test scores and creative performance. Cropley (2000), for example, criticized that DT did not predict creative performance to a sufficient magnitude. One possible factor for such low criterion-related validity could be the heterogeneity in a population due to the complex nature of DT. DT consists of multiple traits, and a constellation of these traits in DT is most likely different across individuals. For instance, some individuals may be more productive in their ideation, fluency, while others may be more novel, originality (Martinsen, Kaufmann, & Furnham, 2011). Despite the possible heterogeneity in a population, many studies have traditionally examined general principles underlying the relationships between DT indices and creative performance across individuals. This approach assumes that a population is homogeneous regarding these relationships. Where a population is potentially not homogeneous, recent research suggests that a constellation of traits may outperform individual traits in predicting outcomes by indicators (Asendorpf & Denissen, 2006). Given the complexity of DT as well as individuals' idiosyncratic characteristics in DT, a constellation of DT indices may be more meaningful than individual indices of DT in predicting creative performance. Studies have yet to use a constellation of DT indices in predicting creative performance.

In previous literature, findings and interpretations on the validity of DT tests are not constant across studies, which may have empirical evidence weak. A possible reason is that the complex nature of DT and procedure of scoring DT tests have not been fully addressed in examining the validity of DT tests. Another reason is probably that previous literature relies on

rules of thumb in determining whether empirical evidence is sufficient to support the discriminant and criterion-related validity of DT tests, which is common in conventional approaches. These two possible reasons are the gaps to be filled. Therefore, the current studies attempted to provide strong empirical evidence in model-based approaches taking into consideration the complexity and procedure of DT tests in investigating the discriminant and criterion-related validity of DT tests.

Findings in the current studies would contribute to expanding our knowledge on the validity of DT tests in two respects. On one hand, the current studies would uncover obstacles that probably prevent us from accurately examining the discriminant and criterion-related validity of DT tests. On the other hand, the current studies would further suggest ways to examine the discriminant and criterion-related validity of DT tests by ruling out those probable obstacles.

### **Objectives of the Present Studies**

The goal of the present studies is to examine the validity of DT indices in order to provide empirical evidence on two controversial issues of the discriminant and criterion-related validity in DT tests.

In the first study, the discriminant validity of DT indices was examined using multitrait-multimethod (MTMM) analysis in confirmatory factor analysis (CFA). MTMM analysis in CFA allowed us to disentangle a probable method effect from true interdependencies between DT indices which is solely determined by underlying cognitive contingency across DT indices. This analysis reduces bias in examining the discriminant validity between pairs of DT indices. Thus, the first study attempted to answer the following questions:

1. Is a method effect present in DT tests?

2. Are fluency and originality discriminable with and without ruling out a method effect?

Empirical evidence on the discriminant validity would contribute to bolstering the multidimensionality of DT as suggested in creativity theories (Guilford, 1956; Torrance, 1974; Wallach & Kogan, 1965).

In the second study, the criterion-related validity of DT was examined using latent profile analysis (LPA). LPA permitted us to classify groups with their own profiles in DT based on high/low patterns in DT indices. Specifically, the second study examined whether more than one group with its own profile in DT was present, and these group memberships explained differences in creative performance. Thus, the second study attempted to answer the following questions:

1. How many groups with distinct DT profiles are present?
2. Do identified groups with their own DT profiles explain individual differences in creative performance?

Results would provide empirical evidence supporting that the low correlations between DT indices and creative performance do not necessarily indicate the low criterion-related validity of DT indices. Given the probable heterogeneity in a population regarding DT, this evidence would contribute to dealing with idiosyncrasy in assessing DT.

Taken together, the current studies would contribute to establishing reliable and valid assessment of DT, which is fundamental to identify children's creative potential and provide appropriate interventions. Consequently, this study would enhance creativity development among children.

## References

- Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal*, 26, 229-238.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- Asendorpf, J., & Denissen, J. J. (2006). Predictive validity of personality types versus personality dimensions from early childhood to adulthood: Implications for the distinction between core and surface traits. *Merrill-Palmer Quarterly*, 52, 486-513.
- Clapham, M. M. (2004). The convergent validity of the Torrance Tests of Creative Thinking and creativity interest inventories. *Educational and Psychological Measurement*, 64, 828-841.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt, Rinehart & Winston.
- Cropley, A. J. (2000). Defining and measuring creativity: are creativity tests worth using? *Roeper Review*, 23, 72-79.
- Csikszentmihalyi, M., & Getzels, J. W. (1971). Discovery-oriented behavior and the originality of creative products: a study with artists. *Journal of Personality and Social Psychology*, 19, 47-52.
- Getzels, J. W., & Jackson, P. W. (1962). *Creativity and intelligence: Explorations with gifted students*. Oxford, England: Wiley.
- Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53, 267-293.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.

- Guilford, J. P. (1968). *Intelligence, creativity, and their educational implications*. San Diego, CA: Robert R. Knapp.
- Guilford, J. P. (1971). Some misconceptions regarding measurement of creative talents. *Journal of Creative Behavior*, 5, 77-87.
- Hargreaves, D. J., & Bolton, H. (1972). Selecting creativity tests for use in research. *British Journal of Psychology*, 63, 451-462.
- 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. (1981). Measurement of creativity: Review and critique. *Journal of Personality Assessment*, 45, 450-464.
- Hoover, S. M., & Feldhusen, J. F. (1990). The scientific hypothesis formulation ability of gifted ninth-grade students. *Journal of Educational Psychology*, 82, 838-848.
- Kim, K. (2008). Meta-Analyses of the Relationship of Creative Achievement to both IQ and Divergent Thinking Test Scores. *Journal of Creative Behavior*, 42, 106-130.
- Kim, K. H. (2006). Is creativity unidimensional or multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creativity Research Journal*, 18, 251-259.
- Martinsen, Ø., Kaufmann, G., & Furnham, A. (2011). Cognitive style and creativity. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (pp. 214-221). San Diego, CA: Academic.
- Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69, 220-232.

- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Michael, W. B., & Wright, C. R. (1989). Psychometric issues in the assessment of creativity. In J. A. Glover, R. R. Ronning, C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 53-75). New York, NY: Plenum Press.
- Mumford, M. D., Mobley, M. I., Reiter-Palmon, R., Uhlman, C. E., & Doares, L. M. (1991). Process analytic models of creative capacities. *Creativity Research Journal*, 4, 91-122.
- Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real-world problems. *Journal of Psychoeducational Assessment*, 9, 45-53.
- Plucker (1999). Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data. *Creativity Research Journal*, 12, 103-114.
- Plucker, J. A., Qian, M., & Schmalensee, S. L. (2014). Is what you see what you really get? Comparison of scoring techniques in the assessment of real-world divergent thinking. *Creativity Research Journal*, 26, 135-143.
- Runco, M. A. (1986). Divergent thinking and creative performance in gifted and nongifted children. *Educational and Psychological Measurement*, 46, 375-384.
- Runco, M. A. (1991). The Evaluative, Valuable, and Divergent Thinking of Children. *Journal of Creative Behavior*, 25, 311-319.
- Runco, M. A. (2011). Divergent thinking. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (2nd ed., Vol. 1; pp. 400-403). San Diego, CA: Academic Press.
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, 24, 66-75.

- Runco, M. A., & Okuda, S. M. (1988). Problem discovery, divergent thinking, and the creative process. *Journal of Youth and Adolescence*, 17, 211-220.
- Runco, M. A., Millar, G., Acar, S., & Cramond, B. (2010). Torrance tests of creative thinking as predictors of personal and public achievement: A fifty-year follow-up. *Creativity Research Journal*, 22, 361-368.
- Silvia, P. J., et al. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 68-85.
- Torrance, E. P. (1969) Prediction of adult creative achievement among high school seniors. *Gifted Child Quarterly*, 13, 71-81.
- 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. (1981). Predicting the Creativity of Elementary School Children (1958-80) —and the Teacher Who "Made a Difference". *Gifted Child Quarterly*, 25, 55-62.
- Torrance, E. P. (1990). *The Torrance tests of creative thinking norms—technical manual figural (streamlined) forms A & B*. Bensenville, IL: Scholastic Testing Service, Inc.
- Treffinger, D. J., Renzulli, J. S., & Feldhusen, J. F. (1971). Problems in the assessment of creative thinking. *Journal of Creative Behavior*, 5, 104-112.
- von Stumm, S., Chung, A., & Furnham, A. (2011). Creative ability, creative ideation and latent classes of creative achievement: What is the role of personality? *Psychology of Aesthetics, Creativity, and the Arts*, 5, 107-114.

- Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity–intelligence distinction*. New York, NY: Holt, Rinehart & Winston.
- Wallach, M. A., & Wing, C. W. (1969). *The talented student*. New York, NY: Holt, Rinehart & Winston.
- Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal*, 23, 24-37.



## CHAPTER 2

# THE DISCRIMINANT VALIDITY OF A DIVERGENT THINKING TEST: MULTITRAIT- MULTIMETHOD ANALYSIS<sup>1</sup>

---

<sup>1</sup> Paek, S. H. To be submitted to the *Psychology of Aesthetics, Creativity, and the Arts*

## Abstract

Divergent thinking (DT) tests have been sometimes plagued by low discriminant validity, as evidenced by high “hetero-index correlations” between fluency and originality DT indices. Also disconcerting is that the hetero-index correlations within the same test are often higher than the mono-index correlations across different DT tests. These imply a *method effect* in DT test scores. Such a method effect may prevent us from accurately examining discriminant validity. A new model-based approach to multitrait-multimethod analysis was used in the present study to examine responses of 190 children to three DT tasks. This new approach allowed us to disentangle the method effect from the inter-index correlations for a highly accurate examination of discriminant validity. Results confirmed the hypothesized method effect in the DT scores. In particular, fluency and originality indices were discriminable from one another only when the method variance was controlled. These results indicate that high correlations between pairs of DT indices do not necessarily indicate poor discriminant validity of DT indices. Alternative techniques to rule out a method effect in DT test scores are discussed.

INDEX WORDS: discriminant validity, divergent thinking tests, multitrait- multimethod, method effect

## Introduction

The multidimensionality of divergent thinking (DT) has been questioned because of the lack in empirical evidence. According to theories on the creative thinking process (Guilford, 1968; Runco, 1991; Torrance, 1995; Wallach & Kogan, 1965), DT indices—measuring thinking skills such as fluency, originality, and flexibility used in responding to DT tests—represent distinct cognitive processes. Yet empirical research sometimes indicates notable overlap and interdependency among the DT indices (Hargreaves & Bolton, 1972; Hocevar, 1979a, 1979b; Hocevar & Michael, 1979; Plucker & Runco, 1998; Runco, 1986a; Zeng, Proctor, & Salvendy, 2011). This interdependency has been cited in claiming that the discriminant validity of DT indices is low, and that DT tests, correspondingly, are not a solid measure of the creative thinking process.

Discriminant validity is defined as the degree of distinctness of test scores that theoretically represent different constructs (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). In other words, the more interdependent the test scores, the lower the discriminant validity. What is often overlooked in examining the discriminant validity of DT tests is that there are two causes of an interdependency among DT indices. One is *mental communality*, which suggests that the various indices (e.g., originality) are indeed cognitively contingent on the other DT indices (e.g., fluency). The other is *method communality*, which might arise when the DT indices are scored (i.e., operationally defined) in certain ways with one index relying on another within the same DT task, even though cognitively the indices are distinct (Campbell & Fiske, 1959). Thus, a task, which connotes the scoring procedure, is defined as a method in the present study. Since mental communality represents a cognitive

overlap between two measures, the less mental communality, the higher the true discriminant validity. On the other hand, since method communality does not account for the cognitive foundation of DT indices, greater method communality may increase the potential for error in interpreting true discriminant validity.

In the case of DT, method communality probably arises because originality and flexibility are operationalized such that they depend on fluency in the same DT task (Hocevar, 1979a; Runco, 1991; Silvia, 2008). This operationalization of DT indices increases hetero-index correlations in DT tests and correspondingly constrains to find good discriminant validity. It is no wonder that quite a few studies have found high hetero-index correlations (e.g., between originality and fluency or flexibility and fluency) and concluded that the discriminant validity of DT indices is questionable (Hocevar, 1979c; Hocevar & Michael, 1979; Rudowicz, Lok, & Kitto, 1995; Runco, 1986a). That conclusion might be premature, however, since conventional analytic techniques do not account for solely mental communality ruling out method communality.

In the present study, a new model-based approach to multitrait-multimethod analysis was applied in examining the discriminant validity of DT indices. The present study 1) illustrates how an advanced methodology accurately controls for the confounding of mental and method communality in DT indices and 2) provides a new look at how operationalizations of DT change the interpretation of the discriminant validity of DT indices.

### **Discriminant Validity and Method Effect in Divergent Thinking Tests**

This lack of discriminant validity in DT indices may reinforce the belief that creativity is a fuzzy construct that cannot be measured in an objective and valid fashion (Plucker, Beghetto, & Dow, 2004). For instance, the lack of discriminant validity often serves as a rationale for using

fluency as a single indicator of DT, as if DT is unidimensional (Hocevar, 1979a; 1979b). The correlations reported in previous research deem adequate to justify using the fluency score as a single indicator of ideation (Baer, 1991; Hocevar, 1980; Hong, Milgram, & Gorsky, 1995; Kuhn & Hollings, 2009; Milgram & Milgram, 1976). Hocevar (1979a, 1979b; Hocevar & Michael, 1979) went as far as to argue that all DT indices are a linear function of fluency. This view seems to have caught on, in part because fluency is easy to score. The problem is that the theory of DT, which is quite clear about the different facets or indices (Runco, 1991), is often ignored.

In actuality, the argument for a lack of the discriminant validity among DT indices is not conceptually or empirically justified. There are, for example, several studies showing the importance of conceptually distinguishing fluency from other indices. Runco and Acar (2012) suggested that originality was conceptually distinct from fluency and needed to be emphasized as a central characteristic of creativity against the use of fluency as a single indicator. An experimental study also supports the distinctive construct of originality from fluency (Runco, Illies, & Reiter-Palmon, 2005). Runco et al. (2005) found that explicit instruction directing students to produce original ideas entailed significant increases only in originality scores but no changes in fluency scores. More importantly, method communality in DT indices is inevitable when DT is scored traditionally without any adjustment (Hocevar, 1979a; Runco, 1991; Silvia, 2008). For instance, fluency is scored by counting the number of all ideas produced by an individual, and originality is scored by counting the number of only statistically infrequent ideas produced by the individual. Thus, originality counts are a subset of fluency counts.

Method communality is indicative of a *method effect*: Campbell and Fiske (1959) described a score as contaminated by a method effect if hetero-index correlations are higher than mono-index correlations: hetero-index correlations refer to correlations among two or more traits

when those traits were intended to be conceptually different but are measured by a common method. Mono-index correlations refer to correlations among the same trait across different tests. This kind of method effect is empirically found in studies examining interdependency among DT indices. For instance, hetero-index correlations—the correlation coefficient between different DT indices within the same DT test—have been higher than mono-index correlations—the correlation coefficients between the same DT indices across different DT tests (Aliotti, Britt, & Haskins, 1975; Borland, 1986; Hargreaves & Bolton, 1972; Hocevar, 1979a, 1979b; Hocevar & Michael, 1979; Michael & Wright, 1989; Runco, 1986a; Yamamoto & Frengel, 1966). This pattern in correlations between DT indices implies the presence of a method effect (Campbell & Fiske, 1959).

Similarly, this interdependency is examined in exploratory (Runco, 1986a) and confirmatory factor analyses (CFAs; Hassan, 1986; Kim, 2006). Where exploratory factor analyses are used, fluency and originality measured in the same test tend to load on the same latent factor (Hargreaves & Bolton, 1972; Belcher, Rubovits, & Di Meo, 1981; Primi, Nakano, Morais, Almeida, & David, 2013). On the other hand, only a few studies investigate the discriminant validity of DT indices in CFA while the construct validity of DT tests has been well documented in CFA (Kim, 2006; Kim, Cramond, & Bandalos, 2006; Krumm, Lemos, & Filippetti, 2014; Plucker, 1999). Hassan (1986) tested the structure of the DT indices in CFA and found that two method factors (i.e., verbal and figural method factors) better explained data than three DT-indices factors. In other words, method communality, referring to two method factors, explained more data over mental communality, referring to three DT-indices factors. Hassan's finding indicates a probable method effect and supports poor discriminant validity in DT indices when a method effect is not ruled out.

Efforts have been made to explain the interdependency among DT indices. A few studies have examined the unique variance of originality, controlling for fluency, in order to isolate any interdependency between the two. For instance, previous studies examined ratio scores of originality (Plucker, Qian, & Schmalensee, 2014), partial correlations of originality and outcomes (Runco, 1986), and the weighted sum of responses by their statistical infrequency (Runco & Albert, 1985; Seddon, 1983). However, these analytic techniques failed to separate mental communality from method communality. Rather, the whole of common variance between fluency and originality, containing both mental and method communality, was entirely removed from originality. In these analyses, it is not feasible to examine true discriminant validity solely contributed by mental communality. Therefore, an objective methodological approach that can test the discriminant validity while controlling a method effect is necessary.

### **Multitrait-multimethod Analysis**

Many traits measured in psychology research are, in fact, compounds of variances of traits and methods, which are referred to as trait-method units (Campbell & Fiske, 1959). Decomposing a trait and a method is important in examining the extent to which a test score is solely determined by a trait, and multitrait-multimethod (MTMM; Campbell & Fiske, 1959) analysis is typically used for this decomposition. Recall here that this traditional MTMM analysis allows us to simultaneously examine discriminant validity of a test and the presence of a method effect using a matrix of hetero-index and mono-index correlations in a single framework. This MTMM analysis has been applied in several studies on DT tests (e.g., Hocevar, 1979c; Hocevar & Michael, 1979; Rudowicz, Lok, & Kitto, 1995; Runco, 1986a).

However, this traditional MTMM approach has a limitation of subjective interpretation: It is ambiguous to determine what patterns of correlations sufficiently represent adequate

discriminant validity (Brown, 2006). In addition, previous studies were not able to rule out the method effect when examining the discriminant validity and instead merely explored the possibility of a method effect. These drawbacks can be handled using model-based MTMM analysis, which analyzes a covariance matrix of MTMM to investigate underlying dimensions (i.e., latent factors of traits and methods) of observed variables in the context of CFA.

The CFA model partitions variance of the test scores into the latent factors of the method and the trait, as does traditional MTMM analysis. In the CFA model, test scores are used as the individual observed variables, which load on the latent factors. Then, the CFA model should explain variance of a method in test scores. There are two CFA approaches to analyzing the MTMM matrix depending on the technique of partitioning variance of the test into the method: a correlated traits/correlated methods (CTCM) model and a correlated unique-variables model<sup>2</sup> (Marsh & Grayson, 1995). Although a CTCM model—which allows latent factors of a method effect—is theoretically idealistic, this model often fails to converge due to high complexity in the model (Kenny & Kashy, 1992). For this reason, a correlated unique-variables model is often recommended (Marsh & Grayson, 1995) and used in the present study.

In order to test whether a method effect is present in DT indices, two models (i.e., correlated unique-variable models) are fitted to data and compared against one another: One is a model allowing no method effect, and the other is a model allowing a method effect. A significant improvement in a model fit indicates the presence of a method effect.

---

<sup>2</sup>This model is typically referred to as correlated uniqueness model (Marsh & Grayson, 1995). In structural equation modeling, uniqueness refers to the portion of variance in an observed variable which is not explained by a latent factor. However, uniqueness is also widely used to represent the characteristic of statistically infrequent ideas in creativity research. In order to reduce in possible confusions, unique variable is used replacing uniqueness when describing the CFA model.



This model selection is followed by testing discriminant validity in model-based MTMM analysis. Recall here that discriminant validity in DT indices can be examined in two ways. On one hand, substantially high correlations between pairs of the latent DT factors may indicate the lack in discriminant validity in DT indices. However, high correlation does not allow us to statistically test whether two factors are significantly discriminable against one another.

On the other hand, comparing the fit of two models enables us to statistically test discriminant validity. One model—the model selected as the final model above—allows two separate factors of fluency and originality assuming adequate discriminant validity of fluency and originality. The other model allows a method effect and a single factor across fluency and originality indices assuming poor discriminant validity of fluency and originality. When fluency and originality are forced to be a single factor, the significant decline in the model fit would indicate adequate discriminant validity of the two since two separate factors better explain data over a single factor. Due to statistical tests of discriminant validity by model fit, model-based MTMM analysis is even more sophisticated than traditional MTMM analysis (Flamer, 1983).

### **The Present Study**

The present study was designed to utilize this newer method to determine whether method communality was in fact confounding correlations, and fluency and originality indices were discriminable ruling out the probable method communality. To this end, I applied a latent variable model—named MTMM model—incorporating multiple indices in multiple tests. This MTMM model allows us to control the probable method effect in DT indices and correspondingly examine the discriminant validity of DT indices as a result of mental communality. Models with and without a method effect were contrasted to determine how the discriminant validity changed by the presence of the method effect. Model-based MTMM

analysis in CFA (Brown, 2006) was conducted to sequentially investigate the following questions:

1. Is a method effect present in DT indices?
2. Are fluency and originality discriminable with and without ruling out a method effect?

The models, which contained the latent DT factors of fluency, originality, and flexibility as suggested in creativity theories (Guilford, 1968; Runco, 1991; Torrance, 1995; Wallach & Kogan, 1965), were tested. Yet, fluency and originality are particularly of interest in interpreting results because fluency and originality are theoretically important as well as the most prevalently used in in creativity literature (Guilford, 1968; Hocevar, 1979a, 1979b; Runco & Acar, 2012; Torrance, 1995).

## **Method**

### **Data and Participants**

Data were collected from 205 children ages 11 to 12 in South Korea. Among them, 9 children who were not granted parental permission and 6 children who did not complete at least one DT task out of three were excluded, which resulted in 190 children in the final analyses. The final sample consisted of approximately equal numbers of boys (49%,  $n = 93$ ) and girls (51%,  $n = 96$ ); one child did not indicate his/her gender. While a majority of children (84%,  $n = 159$ ) had not participated in any gifted program, 15% of children ( $n = 29$ ) had participated in at least one gifted program in or outside of school; there were only two children who did not indicate whether they had taken a gifted program.

The present study heavily relied on online data collection, and as such the open-ended questions (see Measures, below) required typing skills. Children were asked whether they were familiar with typing on computers. All the children were from families owning either computers

or tablet PCs, and 91% of children ( $n = 173$ ) reported that they had used those information technology devices more than once a week. In addition, 94% of children ( $n = 179$ ) had taken computer classes in or outside of school. This high rate of children's exposure to technology classes and devices ensured that children had no trouble typing responses using computers.

## Measures

**Divergent Thinking Test.** Divergent thinking was measured using the Instances test developed by Wallach and Kogan (1965), which consisted of three tasks. Instances tests were chosen since the tests have been broadly used across cultures (Hong & Milgram, 1991; Runco, Abdulla, Paek, Al-Jasim, & Alsuwaidi, 2016; Silvia, 2008), and reliability of these tasks is more than adequate and well documented in previous literature: The reliability of Instances tests was acceptable in Wallach and Kogan's (1965) original study as Spearman-Brown split-half coefficients were .75 for fluency and .51 for originality. In other studies, reliability ranged .65 through .84 for fluency and .37 through .68 for originality (Plucker, Qian, & Wang, 2011; Runco & Okuda, 1991; Runco et al., 2016). Participants were asked to (a) "List all square things you can think of," (b) "List all round things you can think of," and (c) "List all the things you can think of that move on wheels." Each task was scored to yield three DT indices of fluency, originality, and flexibility. Three DT indices were scored following standard guidelines of scoring DT tasks (Guilford, 1968; Runco, 1991; Torrance, 1995; Wallach & Kogan, 1965). Once the ideas pool was established, ideas produced by each individual were counted. For instance, fluency was scored by counting all distinct ideas produced by a child. Duplicate or inappropriate ideas were excluded from counting. Originality was scored by counting statistically infrequent ideas that only less than 3% of children in the sample produced. For instance, responses, that only five or fewer children (3% of 190 children) provided, were counted for originality.

Flexibility was scored by counting different categories of ideas that a child produced. The categories determining flexibility were adapted from a comprehensive lexicon used in previous research (Charles & Runco, 2001). Thus, the flexibility score increased by 1 point when multiple ideas were from the same category. In the present study, reliability of three DT indices was calculated using Cronbach alpha of inter-item consistency: fluency (.64), originality (.58), and flexibility (.55). Reliability of these scores is not as high as required in conventional standard and need to be taken into account in subsequent analyses and interpretations. However, reliability of DT tests can vary depending on samples, tests, and indices ranged from .37 to .9, which are often found in creativity research (Plucker et al., 2011; Runco, 1991; Runco & Okuda, 1991). Since reliability of DT tests in the present study are in this range of reliability reported in previous research, the results can be as trustworthy as findings in previous research.

## **Procedure**

All measures in the survey were translated into Korean and reviewed by a bilingual expert. Then, the survey was back-translated to eliminate possible misinterpretation. Children responded to the online survey of *Qualtrics*, which delivered all measures. The online survey was delivered on a regular school day in computer labs during a creative activities class. Children were given oral and written instructions regarding how to respond to the survey. The oral instructions were scripted by authors, and homeroom teachers were asked to read aloud the instruction to their students. There was no time limitation for the tasks.

## **Data Analyses**

**Descriptive Statistics.** Raw scores of fluency ranged from 7.97 to 18.75, whereas raw scores of originality ranged from 2.62 to 6.44. The means, standard deviations, univariate skewness and kurtosis, and correlations of the three tasks are presented in Table 2.1.

**Normality and Outliers.** Multivariate normality is commonly assumed in structural equation modeling, which implies that each individual variable shows a univariate normal distribution (Mulaik, 2009). Thus, univariate skewness and kurtosis in six DT indices were tested to show univariate normal distributions. No skewness and kurtosis in six DT indices were greater than |2| and |7|, respectively, except originality in the Wheel task. Psychometric quality of originality in the Wheel task may not be sound in some samples. Given that the sample in this study was elementary children who may not have substantial experience with cars or wheels, their responses to the Wheel task might be limited. Furthermore, the Wheel task is more constricted than the other two. Thus, the originality scores in the Wheel task might be highly influenced by knowledge about cars and wheels rather than accurately representing their originality in DT. As seen in Table 2.1, originality in the Wheel task is less than half of originality in the other two tasks. This weak psychometric quality of the Wheel task was consistently found in females in Arabic countries, which reported that originality in the Wheel task is barely correlated to originality in other tasks (Runco et al., 2016).

Multivariate normality was further tested by Mardia's test (Mardia, 1970) using the open source MVN package for the R statistical environment (Korkmaz, Goksuluk, & Zarasiz, 2014). The results of multivariate skewness and kurtosis indicate that data were significantly deviated from multivariate normal distribution. There are the arguments that normality is not critical for many multivariate analyses (Krzanowski, 2000), and structural equation modeling is robust to non-normally distributed data (Enders, 2001). Thus, no corrective techniques were employed to transform data. Instead, all parameters and fit measures were estimated using maximum likelihood estimation with robust standard errors in *Mplus*, which is robust to non-normality (Muthén & Muthén, 1998-2007).

**Missing Values.** There were 9 children who did not complete one or more of the DT tasks. The missing values in DT scores were handled using the imputation method using the expectation maximization (EM) algorithm (Graham, 2009). Granted by the EM algorithm, parameter estimates were adjusted with a weight to deal with missing values.

### **Multitrait-multimethod Analysis in Confirmatory Factor Analysis**

In the present study, a CTCM failed to converge. Thus, we allowed unique variables to be correlated to each other to define a model with the presence of a method effect. Figure 2.1 displays a hypothetical correlated unique-variables model used in the present study. A correlated unique-variables model needs at least three traits and three methods. In the present study, three latent DT factors, fluency, originality, and flexibility, were measured as traits although we only focus on fluency and originality in interpreting results. Also, three tasks, the Round, Square, and Wheel tasks, were used to measure DT and were operationally defined as methods. Each method was loaded on three indices, fluency, originality, and flexibility. Overall, nine DT indices were used as observed variables. A correlated unique-variables model has no latent factor of a method. Instead, a correlated unique-variables model allows correlations among unique variables within the same method, and these correlated unique variables are operationally defined as a method effect in the present study.

When comparing models with and without a method effect, six model fit measures were calculated by maximum likelihood estimation with robust standard errors (i.e., MLR) to determine whether the measurement model fitted well: chi-square, the comparative fit index (CFI; Bentler, 1990), Tucker-Lewis Index (TLI; Bentler, 1990), the root mean square error of approximation (RMSEA; MacCallum, Browne, & Sugawara, 1996), Akaike Information Criterion (AIC; Akaike, 1974), and the Bayesian information criterion (BIC; Schwartz, 1978).

Between two models with and without a method effect, the better model was selected and used in the following analyses. Then, the discriminant validity of DT indices was examined by statistically testing whether fit measures of models changed by the number of factors across three DT indices (i.e., three- versus one-factor model).

## **Results**

### **Method Effect**

In order to test the presence of a method effect, fit measures of two models were compared against one another. Model fits are presented in Table 2.2. First, the model without a method effect showed a poor fit to data (also see the second column of Table 2.2): RMSEA = 0.3, CFI = 0.56, TLI = 0.33. Then, the model with a method effect was fitted to data by allowing unique variables to be correlated to each other within the same method. The model with a method effect better explained data over the model without a method effect,  $\Delta\chi^2 = 191.22, \Delta df = 9, p < .001$  (also, see the third column in Table 2.2). Thus, the model with a method effect was selected as a final model. The final model fitted well to data as supported in model fit measures: RMSEA = 0.059 (95% CI = 0 - .098), CFI = 0.99, and TLI = 0.97. This result indicated the apparent presence of a method effect in DT tests.

### **Discriminant Validity**

In the model without controlling a method effect, the correlation between the latent factors of fluency and originality was almost close to 1 ( $r = 1.00$ ), which indicated that fluency and originality shared all variance. Once the method effect was controlled, the correlation between latent DT factors of fluency and originality went down to  $r = .82$ , which indicated that fluency and originality shared 66.7% of variance. This covariance indicates that the 33.3% of

unique variance can be explained by the unique characteristics of fluency and originality, which is substantial. Table 2.3 presents the correlations between latent DT factors.

However, the shared variance of 66.7% was not sufficient to determine whether two factors were significantly discriminable from one another. Thus, discriminant validity of fluency and originality was further tested by comparing the fit of the final model to the model allowing fluency and originality to be a single factor. By forcing fluency and originality to be a single factor, the model fit significantly declined:  $\Delta\chi^2 = 12.1, \Delta df = 1, p < .001$  (also, see the fourth column in Table 2.2). This significant decline in the model fit implied that the model allowing two separate factors of fluency and originality better explained data over the model with a single factor of fluency and originality. In other words, fluency and originality were not totally interdependent when the method effect was controlled, which supported adequate discriminant validity of fluency and originality in DT tests only when ruling out the method effect.

## **Discussion**

The present study contributes statistical evidence about the valid measurement of DT using fluency and originality as separate indices. In fact, there are two statistical contributions: One involves discriminant validity, and the other concerns a method effect. To our best knowledge, the present study is the first model-based test to provide empirical evidence that originality can be extricated from fluency when the method effect is controlled. Controlling the method effect enables us to distinguish the true discriminant validity as a result of the cognitive distinctness across DT indices (Brown, 2006). The present findings of discriminant validity contribute to the scholarly efforts that have been made to determine whether originality depends on fluency (Hocevar, 1979a, 1979b; Hocevar & Michael, 1979; Runco, 1986a; Runco & Albert, 1985). Recall here that the previous efforts tended to rely on subjective judgments to determine



whether originality was distinct enough from fluency. This subjectivity was apparent in that the correlations had to be judged as large enough for discriminant validity. There are not statistical tests for any form of validity. Inter-index correlations were reported in the previous studies and were not clear-cut to confirm the discriminant validity of DT. The present correlated unique-variables model, relying on a model-based test and controlling method variance, therefore makes a unique contribution and clearly supports the adequate discriminant validity of fluency and originality.

Importantly, the separation of fluency and originality in results is consistent with the comprehensive theories of DT presented by Guilford (1968), Torrance (1995), and Wallach and Kogan (1965). The distinctive characteristics of fluency and originality are also consistent with Mednick's (1962) description that creativity of ideas varies by the degree of ideas' remote associates. Probably, the cognitive process underlying originality is distinct from that of fluency, as Runco and Basadur's (1993) work found that evaluative thinking was more involved when generating original ideas than unoriginal ideas. Similarly, Runco and Acar (2010) found that original ideas more likely relied on personal and social experience over unoriginal ideas.

The present results have implications for both research and practice. On one hand, fluency has often been chosen as the single index of creative thinking in education (Hong, Milgram, & Gorsky, 1995), clinical assessment (Tucha et al., 2011; Zdep, 1966), and psychology (Hocevar, 1980), in part due to its convenience in scoring. This is problematic, as is the practice of collapsing various DT indices into a single score (e.g., Kuhn & Hollings, 2009; Vosburg, 1998). These methods may not be appropriate given that fluency and originality are psychometrically and cognitively distinctive, at least when the right statistics are used to test their separation. Using one index (e.g., fluency alone) or combining various indices into one

composite entail a loss of information that would otherwise be available if originality and fluency were recognized separately.

On the other hand, there are questions about methods such as brainstorming, which typically suggest a focus on producing as many ideas as possible, regardless of their quality (i.e., fluency over originality). Brainstorming may be a useful and widely used method (Hartley & Plucker, 2014) and is highly valued by teachers (Cheung, Tse, & Tsang, 2003) for its own benefits to creative thinking and problem solving (Parnes & Meadow, 1959; Putman & Paulus, 2009). Although extended efforts tend to increase creative ideas (Parnes, 1961), a large number of ideas may not necessarily promise novelty in ideas. For instance, people were asked to produce ideas in DT tests, and the chronological series of ideas were evaluated (Basadur & Thompson, 1986). In the experiment, Basadur and Thompson (1986) found that ideas in the last third were not more creative than the first two thirds. One of the reasons may be that producing more rote ideas may fasten us to conventional ideas in some tasks but not all. For instance, Smith and Linsey (2011) described the sudden occurrence of a “eureka” moment, when an individual breaks out of the mental fixation of a given problem. Smith and Linsey suggested to stop producing ideas if those ideas are counterproductive, which resists finding value merely in being prolific. Rather, they suggested redefining a problem and finding clues or hints rising from the redefined perspectives to trigger new insights. The empirical findings and theoretical arguments together suggest that creative thinking is multifaceted, and different strategies probably are needed to promote the multiple facets.

It should be noted that there is not a single solution to improve creative thinking, given variation in tasks and individual thinking styles. In addition to brainstorming, there are many strategies that we can use to enhance originality and flexibility in DT. Incubation, for example,

may be a useful strategy to explore unanswered questions and take risks that go beyond the conventional imagination (Torrance & Safer, 1990). During incubation, rote ideation and premature conclusions may be inhibited until great “eureka” moments. Transformative thinking may also allow us to take different perspectives, which serves highly flexible thinking (Guilford, 1970). Given the multidimensionality of DT, various strategies are, therefore, necessary to enhance multiple facets of DT in practice.

Recall here that the discriminant validity was affirmed only when the method effect was ruled out. To our knowledge, the present study is also the first to statistically control for a method effect in DT indices. Although traditional MTMM analyses in the previous research pointed to a probable method effect in DT tests and raised questions about the patterns of inter-index correlations (Hocevar, 1979c; Hocevar & Michael, 1979; Runco, 1986a), the method effect was not itself statistically tested. Granted, the scoring procedure for DT tests very frequently relies on the count of ideas for all indices, which is why fluency has been called a “gatekeeper.” When counts are used, fluency and originality cannot completely avoid method communality.

Given the presence of a method effect in DT tests, the high correlation between fluency and originality (e.g., Kim, 2006; Kuhn & Hollings, 2009) is no surprise because method communality may inflate correlation (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Results also supported this inflation in the correlation between fluency and originality due to a method effect. This inflation indicates that discriminant validity of fluency and originality is not appropriately examined when variances in DT indices are confounded with the method effect. Therefore, the high correlation between fluency and originality, frequently found in previous literature, does not necessarily uphold the lack of discriminant validity of DT tests. Given the

statistical test of discriminant validity in results, Hocevar's (1979a, 1979b; Hocevar & Michael, 1979; Zarnegar, Hocevar, & Michael, 1988) claim of a single dimension of DT may not be firm.

The presence of a method effect directs us to revisit operationalizations of DT that have been conventionally used in creativity research. Indeed, Runco (1986b) pointed to the operational definitions of DT indices in traditional scoring as a cause of the method effect. In other words, new operational definitions of DT indices may bring a downturn in the method effect. Scholarly efforts have been made to eliminate the method effect in DT indices by redefining originality in four ways (see Runco, 1991 for a review). Hocevar and Michael (1979) suggested dividing originality by fluency to compute ratio scores for originality (also see Plucker et al., 2014). Seddon (1983) suggested inversely weighting ideas by the occurrence rate (the occurring frequency of an individual idea divided by sample size). Harrington, Block, and Block (1983) suggested eliminating ideas directed by the immediate environment. Most recently, Silvia et al. (2008) suggested a subjective scoring method to score only the top two best ideas for originality. These redefinitions abate a high dependency of originality on fluency, but it may change what exactly the originality index represents. With respect to both a method effect and the meaning of originality, the technique used in the present study is suggestive. Given theories of creativity and of DT, the last option, demonstrated in the present study, may be best: The method effect can be controlled by allowing unique variables to be correlated to each other within the same method in the context of CFA or, more rigorously, allowing a bi-factor model which includes latent factors of methods (Podsakoff, MacKenzie, & Podsakoff, 2012). In other words, this technique allows statistically controlling for the method effect without changing the meaning of originality.

## **Limitations and Future Research Directions**

There were three limitations in the present study. First, the present study only accounted for different tasks of the Instances test in empirically testing a method effect, although various other methods are available for measuring DT: verbal, figural, action-and-movement, problem-presented, problem generation, scientific-hypotheses formulation, and real-world problem tests (Runco et al., 2016). Even verbal DT tests vary and include Uses, Instances, Similarities, Titles tests, and several others. Thus, the probable variation in the method effect across DT tests needs to be accounted for to achieve generalizability of the findings. Furthermore, the interaction of tasks and cultures is probably present as found in the Wheel task. Therefore, it is still necessary to test the method effect in other cultures in addition to Korean children examined in the present study.

Second, the latent factor of flexibility is often highly correlated with the latent factor of fluency, which may be contraindicative of the discriminant validity between fluency and flexibility. Runco et al. (2005), for example, reported a correlation of .71. Given the high correlation between fluency and flexibility, sometimes the flexibility score is eliminated, at least on the Torrance Figural Tests (Hèbert, Cramond, Neumeister, Millar, & Silvian, 2002). Kuhn and Hollings (2009) also discarded flexibility in testing the factor structure of DT tests for the same reason. Because flexibility was not the focus of the present study, the discriminant validity of fluency and flexibility was not examined. Yet, this does not necessarily mean that flexibility is trivial. Flexibility is cognitively different from fluency (Guilford, 1968; Runco, 1985; Torrance, 1974) and should provide useful information, above and beyond what is supplied by fluency and originality. In fact, flexibility better discriminates gifted children from non-gifted children over fluency (Runco, 1985) and contributes to explaining creative performance in a real-world setting

(Runco, 1986c). For these reasons, future research should further investigate whether flexibility is discriminable from fluency when different DT tests are examined.

As a final note, the reliability and normality of the DT scores were not impressive in the present study. Moderate reliability (Runco & Okuda, 1991) and non-normality (Khun & Holling, 2009) are often found in the DT research. This marginal reliability and normality may reflect situational motivation (Runco, 1986b), situational interest (Jeon, Moon, & French, 2011), and prior knowledge and experience coming into play when taking the tests (Jeon et al., 2011; Runco & Acar, 2010). Future research should, therefore, reduce any possible errors in measurement, motivational and situation, to improve the psychometric basis of DT tests.

### **Conclusion**

It is not easy to believe in something invisible and even harder when it is intricate. The nature of psychological constructs is a good example. One way to measure these invisible constructs is quantifying them through operationalization as is done in DT. Scholarly efforts seek to find the adequate operationalization of DT best reflecting the nature of DT but may partially achieve such adequate operationalization of DT or fail to do so entirely. Empirical evidence based on the probable imperfect operationalization of DT has reinforced the argument against a dimensional structure of DT. Thus, the present study suggests a new way of operationalizing DT and examines discriminant validity of DT indices: Correlated unique-variable model affirmed a method effect present in DT indices, and originality was discriminable from fluency after disentangling a method effect from true interdependency between fluency and originality. This work supports the discriminant validity in DT indices against the prevailing argument that DT indices are linear functions of fluency. This work further suggests a possible technique to handle a method effect in DT indices and contributes to valid assessment of DT.

## References

- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19, 716-23.
- Aliotti, N. C., Britt, M. F., & Haskins, G. P. (1975). Relationships among creativity, intelligence, and achievement measures in Upward Bound students. *Psychology in the Schools*, 12, 423-427.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- Baer, J. (1991). Generality of creativity across performance domains. *Creativity Research Journal*, 4, 23-39.
- Basadur, M., & Thompson, R. (1986). Usefulness of the ideation principle of extended effort in real world professional and managerial creative problem solving. *Journal of Creative Behavior*, 20, 23-34.
- Belcher, T. L., Rubovits, J. J., & Di Meo, P. A. (1981). Measurement of creativity: A factor analytic study. *Psychological Reports*, 48, 819-825.
- Bentler, P.M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, 107, 238-46.
- Borland, J. H. (1986). A note on the existence of certain divergent-production abilities. *Journal for the Education of the Gifted*, 9, 239-251.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Charles, R. E., & Runco, M. A. (2001). Developmental trends in the evaluative and divergent thinking of children. *Creativity Research Journal*, 13, 417-43
- Cheung, W. M., Tse, S. K., & Tsang, H. W. (2003). Teaching creative writing skills to primary school children in Hong Kong: Discordance between the views and practices of language teachers. *Journal of Creative Behavior*, 37, 77-98.
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6, 352-370.
- Flamer, S. (1983). Assessment of the multitrait-multimethod matrix validity of Likert Scales via confirmatory factor analysis. *Multivariate Behavioral Research*, 18, 275-308.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549-576.
- Guilford, J. P. (1970). Creativity: Retrospect and prospect. *Journal of Creative Behavior*, 4, 149-168.
- Guilford, J. P. (1968). *Creativity, intelligence, and their educational implications*. San Diego: Knapp/ EDITS.
- Hargreaves, D.J., & Bolton, H. (1972). Selecting creativity tests for use in research. *British Journal of Psychology*, 63, 451-462.
- Harrington, D. M., Block, J., Block, J. H. (1983). Predicting creativity in preadolescence from divergent thinking in early childhood. *Journal of Personality and Social Psychology*, 45, 609-623.



- Hartley, K. A., & Plucker, J. A. (2014). Teacher use of creativity-enhancing activities in Chinese and American elementary classrooms. *Creativity Research Journal*, 26, 389-399.
- Hassan, M. A. (1986). *Construct validity of Torrance Tests of Creative Thinking: A confirmatory factor-analytic study* (Unpublished doctoral dissertation). Claremont Graduate School, Claremont.
- Hébert, T. P., Cramond, B., Neumeister, K. L. S., Millar, G., & Silvian, A. F. (2002). *E. Paul Torrance: His life, accomplishments, and legacy*. Storrs: The University of Connecticut, The National Research Center on the Gifted and Talented (NRC/GT).
- 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. (1979c, April). *Multitrait-multimethod analyses of tests of divergent thinking*. Paper presented at the meeting of the Western Psychological Association in San Diego, California.
- Hocevar, D. (1980). Intelligence, divergent thinking, and creativity. *Intelligence*, 4, 25-40.
- Hocevar, D., & Michael, W. B. (1979). The effects of scoring formulas on the discriminant validity of tests of divergent thinking. *Educational and Psychological Measurement*, 39, 917-921.
- Hong, E., & Milgram, R. M. (1991). Original thinking in preschool children: A validation of ideational fluency measures. *Creativity Research Journal*, 4, 253-260.
- Hong, E., Milgram, R. M., & Gorsky, H. (1995). Original thinking as a predictor of creative performance in young children. *Roeper Review*, 18, 147-49.

- Jeon, K. N., Moon, S. M., & French, B. (2011). Differential effects of divergent thinking, domain knowledge, and interest on creative performance in art and math. *Creativity Research Journal*, 23, 60-71.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.
- Kim, K. H. (2006). Is creativity unidimensional or multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creativity Research Journal*, 18, 251-259.
- 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, 459-477.
- Korkmaz, S., D. Goksuluk, and G. Zarasiz, 2014. An R package for assessing multivariate normality. *The R Journal*, 6, 151-162.
- 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, 72-81.
- Krzanowski, W. J. (2000). *Principles of multivariate analysis* (2nd ed.). New York, NY: Oxford University Press.
- Kuhn, J., & Holling, H. (2009). Measurement invariance of divergent thinking across gender, age, and school forms. *European Journal of Psychological Assessment*, 25, 1-7.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130- 149.

- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530.
- Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177-198). Thousand Oaks, CA: Sage.
- Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69, 220-232.
- Michael, W. B., & Wright, C. R. (1989). Psychometric issues in the assessment of creativity. In J. A. Glover, R. R. Ronning, C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 53-75). New York, NY, US: Plenum Press.
- Milgram, R. M., & Milgram, N. A. (1976). Creative thinking and creative performance in Israeli students. *Journal of Educational Psychology*, 68, 255-259.
- Mulaik, S. A. (2009). *Foundations of factor analysis*. Boca Raton, FL: CRC press.
- Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.
- Parnes, S. J. (1961). Effects of extended effort in creative problem solving. *Journal of Educational Psychology*, 52, 117-122.
- Parnes, S. J., & Meadow, A. (1959). Effects of "brainstorming" instructions on creative problem solving by trained and untrained subjects. *Journal of Educational Psychology*, 50, 171-176.
- 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.

- Plucker, J. A., & Runco, M. A. (1998). The death of creativity measurement has been greatly exaggerated: Current issues, recent advances, and future directions in creativity assessment. *Roeper Review*, 21, 36-39.
- Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39, 83-96.
- Plucker, J. A., Qian, M., & Schmalensee, S. L. (2014). Is what you see what you really get? Comparison of scoring techniques in the assessment of real-world divergent thinking. *Creativity Research Journal*, 26, 135-143.
- Plucker, J. A., Qian, M., & Wang, S. (2011). Is originality in the eye of the beholder? Comparison of scoring techniques in the assessment of divergent thinking. *Journal of Creative Behavior*, 45, 1-22.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539-569.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879-903.
- Primi, R., Nakano, T. C., Morais, M. F., Almeida, L. S., & David, A. M. (2013). Factorial structure analysis of the Torrance Test with Portuguese students. *Estudos De Psicologia*, 30, 19-28.
- Putman, V. L., & Paulus, P. B. (2009). Brainstorming, brainstorming rules and decision making. *Journal of Creative Behavior*, 43, 23-39.

- Rudowicz, E., Lok, D., & Kitto, J. (1995). Use of the Torrance Tests of Creative Thinking in an exploratory study of creativity in Hong Kong primary school children: A Cross-cultural Comparison. *International Journal of Psychology*, 30, 417-430.
- Runco, M. A. (1985). Reliability and convergent validity of ideational flexibility as a function of academic achievement. *Perceptual and Motor Skills*, 61, 1075-1081.
- Runco, M. A. (1986a). The discriminant validity of gifted children's divergent thinking test scores. *Gifted Child Quarterly*, 30, 78-82.
- Runco, M. A. (1986b). Divergent thinking and creative performance in gifted and nongifted children. *Educational and Psychological Measurement*, 46, 375-384.
- Runco, M. A. (1986c). Flexibility and originality in children's divergent thinking. *Journal of Psychology*, 120, 345-352.
- Runco, M. A. (Ed.). (1991). *Divergent thinking*. Norwood, NJ: Ablex.
- Runco, M. A., & Acar, S. (2010). Do tests of divergent thinking have an experiential bias?. *Psychology of Aesthetics, Creativity, and the Arts*, 4, 144.
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, 24, 66-75.
- Runco, M. A., & Albert, R. S. (1985). The reliability and validity of ideational originality in the divergent thinking of academically gifted and nongifted children. *Educational and Psychological Measurement*, 45, 483-501.
- Runco, M. A., & Basadur, M. (1993). Assessing ideational and evaluative skills and creative styles and attitudes. *Creativity and Innovation Management*, 2, 166-173.
- Runco, M. A., & Okuda, S. M. (1991). The instructional enhancement of the flexibility and originality scores of divergent thinking tests. *Applied Cognitive Psychology*, 5, 435-441.

- Runco, M. A., Abdulla, A., Paek, S. H., Al-Jasim, F., & Alsuwaidi, H. N. (2016). Which test of divergent thinking is best? *Creativity. Theories–Research–Applications*, 3, 4-18.
- Runco, M. A., Illies, J. J., & Reiter-Palmon, R. (2005). Explicit instructions to be creative and original: A comparison of strategies and criteria as targets with three types of divergent thinking tests. *Korean Journal of Thinking and Problem Solving*, 15, 5-15.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Seddon, G. M. (1983). The measurement and properties of divergent thinking ability as a single compound entity. *Journal of Educational Measurement*, 20, 393-402.
- Silvia, P. J. (2008). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). *Creativity Research Journal*, 20, 34–39.
- Silvia, P. J., et al. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 68-85.
- Smith, S. M., & Linsey, J. (2011). A three-pronged approach for overcoming design fixation. *Journal of Creative Behavior*, 45, 83-91.
- 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. (1995). *Why fly?* Norwood, NJ: Ablex.
- Torrance, E. P., & Safter, H. T. (1990). *The incubation model of teaching: Getting beyond the aha!* Buffalo, NY: Bearly Limited.
- Torrance, E.P. (1962). *Guiding creative talent*. Englewood Cliffs, NJ: Prentice-Hall.

- Tucha, L., Tucha, O., Sontag, T. A., Stasik, D., Laufkötter, R., & Lange, K. W. (2011). Differential effects of methylphenidate on problem solving in adults with ADHD. *Journal of Attention Disorders, 15*, 161-173.
- Vosburg, S. K. (1998). The effects of positive and negative mood on divergent thinking performance. *Creativity Research Journal, 11*, 165–172.
- Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity–intelligence distinction*. New York, NY: Holt, Rinehart & Winston.
- Yamamoto, K. & Frengel, B. A. (1966). An exploratory component analysis of the Minnesota Tests of Creative Thinking. *California Journal of Educational Research, 17*, 220-229.
- Zarnegar, Z., Hocevar, D., & Michael, W. B. (1988). Components of original thinking in gifted children. *Educational and Psychological Measurement, 48*, 5-16.
- Zdep, S. M. (1966). Intelligence, creativity and anxiety among college students. *Psychological Reports, 19*, 420-420.
- Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal, 23*, 24-37.

Table 2.1

*Descriptive Statistics of Divergent Thinking Test Scores*

	1	2	3	4	5	6	7	8	9
1. SFLU	—								
2. SORG	0.59	—							
3. SFLX	0.84	0.54	—						
4. RFLU	0.48	0.26	0.39	—					
5. RORG	0.43	0.41	0.40	0.74	—				
6. RFLX	0.42	0.22	0.42	0.81	0.71	—			
7. WFLU	0.32	0.27	0.34	0.42	0.39	0.38	—		
8. WORG	0.12	0.20	0.19	0.27	0.31	0.17	0.62	—	
9. WFLX	0.15	0.11	0.18	0.26	0.19	0.33	0.67	0.35	—
<i>M</i>	18.75	2.94	6.42	14.89	4.03	6.44	7.97	1.35	2.62
<i>SD</i>	7.36	2.93	1.82	8.16	3.71	2.5	3.89	2.19	1.14
Skewness	0.07	1.57	-0.02	0.44	0.98	0.27	0.85	4.39	0.28
Kurtosis	-1.24	3.57	-0.45	-0.86	0.20	-0.66	1.33	27.13	-0.25

*Note.* SFLU = The fluency score in the Square task, RFLU = The fluency score in the Round task, WFLU = The fluency score in the Wheel task, SORG = The originality score in the Square task, RORG = The originality score in the Round task, WORG = The originality score in the Wheel task, SFLX = The flexibility score in the Square task, RFLX = The flexibility score in the Round task, WFLX = The flexibility score in the Wheel task.



Table 2.2

*Model Comparison*

	Three-Factors Model without a ME	Three-Factors Model with a ME	Single-Factor Model with a ME
RMSEA	0.3	0.06	0.11
CFI	0.56	0.99	0.96
TLI	0.33	0.97	0.91
AIC	8434.37	8069.94	8088.32
BIC	8531.78	8196.57	8211.71
ABIC	8436.75	8073.03	8091.34
LL	-4187.18	-3995.97	-4006.16
$df$	30	39	38
$\Delta\chi^2$		191.27	10.19
$\Delta df$		9	1
$p$		< .001	.001

*Note.* ME = Method effect

Table 2.3

*Correlations between Latent Factors and Composite Scores of Divergent Thinking*

	Fluency-Originality	Fluency-Flexibility	Originality-Flexibility
Model without a ME	1	1	1
Model with a ME	.817	.982	.77

*Note.* ME = Method effect.

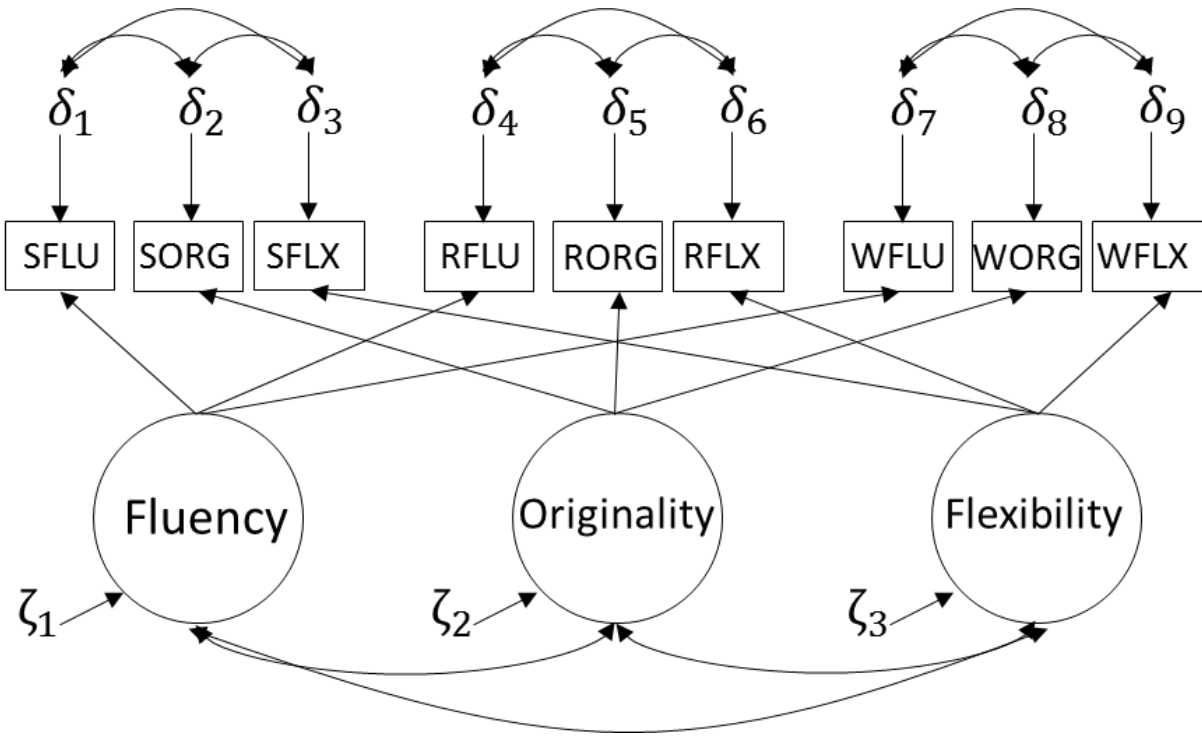


Figure 2.1. Path diagram of model with a method effect

*Note.* SFLU = The fluency score in the Square task, RFLU = The fluency score in the Round task, WFLU = The fluency score in the Wheel task, SORG = The originality score in the Square task, RORG = The originality score in the Round task, WORG = The originality score in the Wheel task, SFLX = The flexibility score in the Square task, RFLX = The flexibility score in the Round task, WFLX = The flexibility score in the Wheel task.  $\delta$  = Unique variable.  $\zeta$  = Disturbance.

CHAPTER 3

LATENT PROFILE ANALYSIS OF THE CRITERION-RELATED VALIDITY OF A  
DIVERGENT THINKING TEST<sup>3</sup>

---

<sup>3</sup> Paek, S. H. To be submitted to the *Creativity Research Journal*

## **Abstract**

The criterion-related validity of divergent thinking (DT) tests has often been questioned because of relatively low correlations between DT test scores and creative performance. A possible inhibiting factor of testing the correlations—the presence of heterogeneity of the sample—is, however, often overlooked. If groups with different DT profiles are present, low correlations may occur as a result of inconsistencies in the relationships across groups. Low correlations correspondingly would not indicate actual relationships (and criterion-related validity). Thus, this study identified groups with their own DT profiles and examined their creative performance. To this end elementary children's ( $N = 325$ ) responses in three DT tasks were analyzed using latent profile analysis. Results suggested three groups with their own DT profiles were present: one group scored moderate on fluency and low on originality; another group scored moderate on fluency and high on originality; the third group scored uniformly low across fluency, originality, and flexibility. While only the low-originality group significantly excelled in creative achievement, only the high-originality group significantly excelled in ideational behaviors. Taken together, results confirm that low correlations between DT indices and creative performance do not necessarily indicate a lack of the criterion-related validity of DT tests. Idiosyncratic properties in DT profiles and their corresponding creative strengths were discussed as educational implications.

**INDEX WORDS:** criterion-related validity, divergent thinking tests, latent profile analysis, divergent thinking profiles, creative performance

## **Introduction**

Ever since Guilford (1968) proposed that divergent thinking (DT) is extricable from convergent thinking and general intelligence, DT has been used in assessing the creative thinking process. DT is operationalized in terms of fluency, originality, and flexibility and has proven to be especially useful in cross-sectional research on the relationships between creative thinking process and performance—often referred to as criterion-related validity (Clapham, Cowdery, King, & Montang, 2005; Plucker, Qian, & Wang, 2011; Runco, 1986a) as well as longitudinal research (Cramond, Matthews-Morgan, Bandalos, & Zuo, 2005; Plucker, 1999; Runco, Millar, Acar, & Cramond, 2010; Torrance, 1980). These studies mainly investigated the general principles of the relationships between DT and creative performance across people. Despite these scholarly efforts, the criterion-related validity of DT tests has been questioned because of the low correlations between DT test scores and creative performance (Baer, 1993; Han 2003; Simonton, 2007).

Although the magnitude of correlations is often used to examine criterion-related validity (Messick, 1995), there are several factors probably contributing to low correlations: the correlations are likely low when the relationships are examined in longitudinal research over many years, or actual performance is correlated to the scores on pencil-and-paper tests (Cramond, 1994; Torrance, 1981). Further, low correlations may not necessarily indicate low criterion-related validity when there is not a single and general relationship between DT and creative performance across people (Laursen & Hoff, 2006). A possible factor of low correlations—the presence of heterogeneity of the sample—is, in fact, often overlooked. A single and general relationship is less likely due to the complexity and multidimensionality of DT as well as variety in creative performance (Csikszentmihalyi & Larson, 1984; Davis, 1989). For

instance, it is likely that not all people high in fluency are high in originality. Just using three traditional indices—fluency, originality, and flexibility—there are a number of possible profiles of DT. The presence of various DT profiles would mean that a population might be heterogeneous, and groups within that population may have idiosyncratic properties of DT. Also, not all people high in each DT index may be high in all types of creative performance. Very likely, some people are good at only generating ideas while others successfully accomplish creative achievement beyond just ideation. Linear correlations between DT and creative performance may not represent the complex relationship between the two due to this probable heterogeneity. Despite this complexity of DT and creative performance, there was no effort to explain various ways of examining criterion-related validity. Thus, we contend that the low correlations do not necessarily indicate a lack of criterion-related validity if more than one group with its own DT profile was present in a population.

The present study therefore provides a new approach to uncover a probable factor—the presence of groups with their own DT profiles—misguiding the interpretation of the criterion-related validity of DT tests. To this end, past work was reviewed to show that a population may be heterogeneous regarding DT profiles, and conventional methodological approaches are at a disadvantage in investigating criterion-related validity given the probable presence of heterogeneity. A new methodological approach—testing the presence of DT profiles—was then introduced, and results illustrate how DT profiles were identified and explained individual differences in creative performance.

### **Heterogeneous Groups and Divergent Thinking**

The average effect of DT on creative performance—referred to as the correlation between DT test scores and creative performance—is as low as .2 in a meta-analysis (Kim, 2008). This

small effect is consistently found in a longitudinal study when students are examined throughout the period from fifth through twelfth grade (Kogan & Pankove, 1972). A typical technique testing this effect is correlation. Correlation tests aim to find a single line best summarizing scattered observations in terms of two variables (Crocker & Algina, 1986).

One of the probable reasons for low correlations is that relationships between DT and creative performance vary across groups. Patterns in creative thinking are very likely to vary from group to group (Kim, 2006; Kirton, 1976). For instance, eminent historical figures exhibited very conflicting characteristics: While some creative individuals were very productive by adapting conventional ideas, others were less productive by pursuing only novel ideas (Simonton, 1997, 2000). Differences in creative thinking patterns are not limited to the eminent. Groups of random adults and college students also exhibit different patterns in thinking or thinking preference in business (Groves, Vance, Choi, & Mendez, 2008) and engineering (Lumsdaine & Lumsdaine, 1995). Such differences no doubt apply to DT.

Those heterogeneous relationships may offset each other, which makes those relationships, in turn, inconspicuous. This assertion is convincing given the considerable heterogeneity among the 274 effects as found in Kim's (2008) meta-analysis (Kim, 2008). In her study, some portion of heterogeneity in the effects was significantly explained by various study characteristics such as DT tests, creative achievement type, and domains. Yet, considerable heterogeneity remained in the effects.

Besides study characteristics, the characteristics of people in groups may also contribute to heterogeneity in the relationship between the two. Interestingly, Kim's meta-analysis indicated that the average effect of DT on creative performance was larger when the effect was estimated within a group over when the average effect was estimated across groups. More specifically,



Runco (1986b) found that the effect varied across gifted and non-gifted populations. Similarly, Torrance and Wu (1981) found this effect varied by students' intelligence in their longitudinal study. These empirical findings imply that the presence of probable groups with their own characteristics may have led to varying relationships between DT and creative performance. Some of these characteristics may not have been explored since nominal groups are not often examined in a conventional research approach.

One of the characteristics may be individual groups' idiosyncrasy in DT indices. The relationships between pairs of DT indices are not always consistent across groups (Runco, 1986a, 1986b; Runco & Albert, 1985). Despite the strong correlations between fluency and originality found in many studies (Hocevar, 1979; Plucker, Qian, & Wang, 2011; Seddon, 1983; Silvia, 2008), those correlations are considerably spread from .3 (Borland, 1986) to .67 (Hargreaves & Bolton, 1972). Interestingly, the participants and measures were very similar in the two studies (Borland, 1986; Hargreaves & Bolton, 1972): The participants in both studies were children from a comparable age range, from 9 to 11 years old, and the measures were both figural DT tests.

Inconstant relationships between pairs of DT indices across groups may imply that groups are heterogeneous regarding the combination of multiple DT indices, which is termed DT profiles in the present study. In other words, some individuals may very likely be highly fluent in DT, but not all that original or flexible, while others are less fluent in DT, but exceptional in either high originality or flexibility, or both. In short, a DT profile is idiosyncratic in each group instead a single DT profile in a population.

## **Person-centered Approaches in Creativity Research**

The conventional statistical techniques in DT research (Clapham, Cowdery, King, & Montang, 2005; Plucker, Qian, & Wang, 2011; Runco, 1986a) tend to be variable-centered. They focus on relationships among individual variables across groups assuming that these relationships apply broadly as found in correlation tests. This, in turn, assumes that groups are homogeneous, and all represent one population (Laursen & Hoff, 2006). Yet given the probable heterogeneity in DT profiles of specific groups, these variable-centered approaches may not be suitable for explaining the idiosyncratic relationship between DT and creative performance across all individuals. Thus, it might be more reasonable to allow qualitatively different groups in a population.

Person-centered approaches have a focus on “the identification of groups of individuals who function in a similar way at the organism level and in a different way relative to other individuals at the same level” (Magnusson, 2003, p. 16). In other words, person-centered approaches focus on demonstrating differences across groups in how variables are related to each other within individuals of a single group (Laursen & Hoff, 2006). In a heterogeneous population, person-centered approaches have advantages for better configuring the complex nature of multi-dimensional constructs, such as DT, and explaining the criterion performance over what individual variables, such as DT indices, can do (Marsh, Ludtke, Trautwein, & Morin, 2009).

Typical techniques of person-centered approaches are cluster, latent class, and latent profile analysis (Laursen & Hoff, 2006). Although a few studies classified individuals into groups by creativity-related traits in cluster analyses (Gilson & Salley, 2004; Helson, 1965; Tan, 2001; Truhon, 1983), only one study used DT in classification (Khandwalla, 1993). For instance, Khandwalla investigated cognitive styles in business students using a DT test of problem

solving. He found three cognitive styles—intuitive ideator, anxious analyst, and random scanner—which manifested differences in six mechanisms of problem structuring. However, Khandwalla's (1993) classification relied on subjective criterion of conventional cluster analysis. With respect to creativity-related traits, studies used cluster analyses in investigating team engagement in technicians (Gilson & Salley, 2004), childhood interest in adults (Helson, 1965), teachers' perception of fostering students' creativity in teachers (Tan, 2000), and playfulness, play, and creativity in children (Truhon, 1983).

Despite the significant contributions of the traditional cluster analyses in creativity research, these traditional cluster analyses lack objectivity in determining the number of groups which attenuate flexible modeling (Pastor, Barron, Miller, & Davis, 2007). In marked contrast, model-based cluster analyses, such as latent profile analysis (LPA), provide more rigorous criteria in determining the number of groups by comparing fit measures of models with different numbers of groups. Thus, the model-based cluster analyses have gradually increased in education research (Pastor et al., 2007).

Despite the good deal of research using traditional cluster analyses in creativity research (e.g., Khandwalla, 1993), only a few studies applied model-based cluster analyses. For example, Sen (2016) found three groups, regular ideators, idea-producers, and idea-aversers, in ideation ability using mixed Rasch model. Other studies sorted individuals into groups using the patterns in domain-specific creative achievements and found three groups (Silvia, Kaufman, & Pretz, 2009; von Stumm, Chung, & Furnham, 2011): "Visual Arts," "Performing Arts," and "No Creativity" groups. In the same vein, von Stumm et al. (2011) expanded Silvia et al.'s (2009) study using DT tests, ideation, and a broad range of personality traits. Using the patterns of domain-specific creative achievement, they found three groups: non-creative people, average

creative achievers with high interpersonal competence, and high creative achievers. Yet, none of these studies classify individuals in regards to profiles of DT and predict creative performance by those profiles.

### **The Present Study**

The present study therefore aimed to uncover a probable factor, the presence of groups with their own DT profiles, which may have previously misguided the interpretation of the criterion-related validity of DT tests. To this end, the following two questions were tested.

3. How many groups with distinct DT profiles are present?
4. Do identified groups with distinct DT profiles explain individual differences in creative performance?

LPA was used to identify the presence of groups with their own DT profiles. LPA is a model-based cluster analysis, which classifies individuals with similar profiles of a multidimensional construct, such as DT, into groups based on continuous indicators in confirmatory factor analysis (CFA; Marsh et al., 2009). The present study used a three-step approach in LPA. First, the models of three DT factors were tested to determine the model best explaining data by allowing different numbers of groups (i.e., groups with their own profiles). Models with different numbers of groups were sequentially fitted. Once the best model was selected, group membership estimated in the best model was assigned to each child corresponding to his/her profile of DT. Then, fluency and originality scores were described across groups to show how these groups are distinct in terms of the DT profiles. Lastly, whether group membership (i.e., DT profiles) was tested as a predictor to explain differences in creative performance.

Creative performance was operationalized with two established criterion measures: creative achievement (Milgram & Milgram, 1976; Paek, Park, Runco, & Choe, 2016; Runco, 1986a) and ideational behaviors (Runco, Plucker, & Lim, 2000-2001). Creative achievement included both quantity and quality scales of creative achievement. Quantity of creative achievement is represented by a frequency of ordinary creative activities in children's life. Quality of creative achievement is represented by a frequency of creative accomplishments that are socially recognized. On the other hand, ideational behaviors were assessed as a criterion of DT because ideational behaviors represented the direct outcome of DT—defined ideas as products (Runco et al., 2000-2001). Runco et al. (2000-2001) contended that ideational behaviors were quantifiable products since any final products inevitably originated from these ideas given the mechanism to create those final products. They also described that ideation better represented everyday creativity that ordinary people can manifest. Thus, ideational behaviors are valid criterion performance in children.

## **Method**

### **Participants**

The present study collected data from 325 children 11 to 12 years old, who were in the 5<sup>th</sup> or 6<sup>th</sup> grade in South Korea. The sample consisted of 5<sup>th</sup> graders ( $n = 159$ ), 6<sup>th</sup> graders ( $n = 149$ ), and 17 children who did not indicate grade level. Slightly more girls participated in this survey: 145 boys (45%), 161 girls (50%), and 19 children (6%) who did not indicate gender. While approximately 70% of children ( $n = 228$ ) had not participated in any gifted program, 20% of children ( $n = 65$ ) responded that they had participated in at least one gifted program in and outside of school. With regards to computer skills, 87% of children ( $n = 284$ ) had taken computer classes in or outside of school. Every child had either a computer or tablet PC at home,

and approximately 93% of children ( $n = 302$ ) responded that they used those information technology devices more than once a week. Given the children's frequent exposure to computer classes and technology devices, it was reasonable that the children had no difficulties typing responses (e.g., DT tests) using computers.

## **Procedure**

In the survey, all measures were translated into Korean and reviewed and back-translated by a bilingual expert to ensure that the translated version of the survey was equivalent to the original survey. The survey was delivered online to children in computer labs through *Qualtrics* during a class regarding creative activities on a regular school day. The survey instruction was given in both oral and written formats in order to direct children how to respond to the survey. Homeroom teachers read aloud the oral instruction. Response time was not restricted.

## **Instruments**

**Divergent Thinking Test.** Instances tests, which contained three tasks, were used to assess DT (Wallach & Kogan, 1965). Since Instances tests were developed by Wallach and Kogan (1965), the tests have been widely used across cultures (Hong & Milgram, 1991; Runco, Abdulla, Paek, Al-Jasim, & Alsuwaidi, 2016; Silvia, 2008). In previous studies, a range of acceptable reliability was reported in the tests (.65 through .84 for fluency and .37 through .68 for originality; Plucker, Qian, & Wang, 2011; Runco & Okuda, 1991; Runco et al., 2016). Participants were directed to (a) "List all square things you can think of," (b) "List all round things you can think of," and (c) "List all the things you can think of that move on wheels." Each task was scored to yield three indices of DT skills: fluency, originality, and flexibility. Fluency was scored by counting the number of ideas produced. Originality was defined as uniqueness, and uniqueness was scored by counting the number of unique ideas that were produced by only

one child. Flexibility was scored by counting the number of different categories of ideas. Then, the raw scores of three DT tasks were transformed to Z-scores with the mean of zero and the standard deviation of 1, which entailed DT scores indicating a DT score's relationship to the mean of the DT scores in each DT index. Z-scores were chosen because Z-scores represent each child's relative placement on the scales of fluency and originality in a sample while raw scores of uniqueness are always equal to or less than raw scores of fluency. Thus, Z-scores allowed comparing a relative placement of each child across the scales of fluency, originality, and flexibility. Reliability was calculated using Cronbach's  $\alpha$ . Reliability of the DT scores are acceptable: fluency ( $\alpha = .72$ ), originality ( $\alpha = .59$ ), flexibility ( $\alpha = .59$ ).

**Creative Activity and Accomplishment Checklist.** Creative achievement was assessed using a 50-item self-report questionnaire that contained extracurricular activities appropriate for intermediate school children (Milgram & Milgram, 1976; Runco, 1986a; Wallach & Wing, 1969). The Creative Activity and Accomplishment Checklist (CAAC) covers six performance domains, with at least eight items for each domain: Writing (e.g., wrote a short story), Music (e.g., composed music), Visual Arts (e.g., painted a picture), Math/Science (e.g., conducted an original experiment), Technology (e.g., created a Facebook page), and Everyday Creativity (e.g., cooked an original dish). An example of the wording is, "How many times have you .... painted a picture?" The possible responses will be (a) never, (b) once or twice, (c) between three and five times, and (d) six or more times.

The CAAC has 12 subscales, which separately assess quantity and quality of creative achievement in six domains. Quantity of creative achievement was calculated by taking the average rating for the items representing ordinary activities in each domain. Quality of creative achievement was calculated by taking the average rating for the items that required social

recognition (e.g., winning awards) in each domain. This study used the CAAC total score across quantity and quality of creative performance in six domains. The CAAC was highly reliable ( $\alpha = .96$ ): Quantity of the CAAC ( $\alpha = .86$ ) and Quality of the CAAC ( $\alpha = .87$ ).

**Runco Ideational Behaviors Scale.** Ideational behaviors were assessed using the Runco Ideational Behaviors Scale for children (RIBS-Children). The RIBS was originally developed by Runco et al. (2000-2001) and revised to measure children's ideational behaviors (Chand O'Neal, Paek, & Runco, 2015). The RIBS-Children consists of twenty-five items asking about behaviors purely in regards to having ideas but not attitudes, interest, or other correlates of ideation (Runco et al., 2000-2001). The conventional five Likert categories (1-5) were used in the response options for each item: 1 (never), 2 (approximately once a year), 3 (once or twice each month, approximately), 4 (once or twice each week, approximately), and 5 (just about every day, sometimes more than once each day). The RIBS was highly reliable ( $\alpha = .93$ ), and the RIBS score was calculated by taking the average rating of 25 items.

The verbatim directions for the RIBS were:

“Use the 1–5 scale (given below) to indicate how often each of the phrases describes your thinking. Note the focus on your thinking, which might be different from your actual behavior. Also, you may need to approximate. Please indicate how you really think, not how you believe you should act. Remember—no names are used. Your responses are confidential.” They were also given a reminder midway through the RIBS: “Again, you may need to approximate. For each item, circle the response option that is THE CLOSEST to being accurate (Runco et al., 2013, pp. 191).”



## Analyses

In this study, LPA was employed to explore whether a) more than one group with its own DT profile was present, and b) DT profiles, classified by LPA, explained individual differences in creative performance. LPA is a method which uses the pattern of DT factors to classify individuals who have similar DT profiles into the same groups—typically named classes in LPA—and to distinguish those with their own DT profiles into different groups (Lubke & Muthén, 2005; Marsh et al., 2009; Pastor et al., 2007). Although the basic principle behind a LPA is similar to a cluster analysis, LPA is more rigorous because LPA employs a model-based analytic technique whereas traditional cluster analyses rely on subjectivity for sorting (Pastor et al., 2007). LPA is therefore not only more flexible than the traditional cluster analysis in terms of model specification, but also the fit indices allow us to compare models with different numbers of underlying groups (Pastor et al., 2007; Vermunt & Magidson, 2002).

LPA was sequentially executed in three steps (Marsh et al., 2009; Pastor et al., 2007). The first step was selecting the model best explaining data among possibly competing models, which allowed us to determine how many groups with their own profiles were identified. The present study proposed a second-order factor model of DT, which contained three continuous latent DT factors—fluency, originality, and flexibility—measured by three DT tasks (see Figure 3.1). Those three DT factors loaded on its second-order factor, which was a categorical latent factor. This categorical factor contains information about which group a child belongs to, and each group is characterized with its own DT profile. Profiles high and/or low in three DT factors were used to classify children.

An exploratory approach was incorporated to identify the number of groups. Thus, models with a different number of groups were sequentially tested until the best fitting model turned out (see Table 3.2). The criteria of selecting the best model were Vuong–Lo–Mendell–

Rubin likelihood ratio (VLMR-LRT) test and Lo–Mendell–Rubin adjusted likelihood ratio (LMR-adjusted LRT) test. The significant VLMR-LRT and LMR-adjusted LRT tests indicate that having one more group better explains data over the model without the additional group. Besides these two indices, Akaike Information Criterion (AIC; Akaike, 1974), Bayesian Information Criterion (BIC; Schwarz, 1978), adjusted BIC (ABIC), and entropy were comprehensively considered to determine the best fitting model. The lower AIC, BIC, and ABIC and higher entropy, the more a model explains indicating a better model. Once the best fitting model was identified, each child was assigned with the group membership estimated in the best fitting model.

The second step was determining in what way groups classified by LPA were different regarding DT indices. The groups were characterized on the grounds of fluency and originality. Flexibility was not considered because it was not the focus of this study and the patterns of flexibility and fluency are alike. The mean factor scores of fluency and originality were described in each group (see Table 3.3).

The third step is validating the classification of children. We examined differences in creative performance by group membership (i.e., DT profiles) in analysis of variance (ANOVA). Post-hoc tests were followed using pairwise *t*-tests (see Table 3.4). False discovery rate (FDR) was controlled using Benjamini and Hochberg's method to correct *p*-value given multiple comparisons (Benjamin & Hochberg, 1995).

## **Results**

First, correlations between DT indices and creative performance were examined. The correlations were not high, ranged from -.09 to .22 with a median of .15, which is probably interpreted as low criterion-related validity in conventional approaches. Means, standard

deviations, skewness, and correlations of the DT indices, CAAC, and RIBS scores are presented in Table 3.1.

Then, LPA was executed in a three-step approach. A three-DT-factor model was used as the baseline model (see Figure 3.1) because three DT indices are the most commonly used to assess DT (Runco, 1991). Then, models with one, two, three, and four latent groups were sequentially fitted to determine which model best explained the structure of DT. The four-group model initially turned out to best explain DT because this model showed the least ABIC (see Table 3.2). Thus, the five-group model was no longer fitted. However, the increment in model fit of the four-group model from the three-group model was not statistically significant, based on both nonsignificant VLMR-LRT and LMR-adjusted LRT tests ( $\chi^2 = 58.1, p = .09, \chi^2 = 55.7, p = .09$ , respectively), which indicated that the three-group model best explained data with parsimony. AIC, BIC, and entropy further support that the three-group model performed better over the four-group model. The result of the three-group model also corresponded to theory describing various patterns in creative thinking (Kim, 2006; Kirton, 1976) Thus, the three-group model was eventually selected. This result indicates that three groups with their own DT profiles are present.

Three means of DT factor scores are used to display three profiles of DT in Figure 3.2 (also see Table 3.3 for mean factor scores of each group). Group 1 consisted of 17 (5.2%) children moderate in fluency and high in originality, which was referred to as the high-originality group. Group 2 consisted of 121 (37.2%) children moderate in fluency but low in originality, which was referred to as the low-originality group. Group 3 consisted of 187 (57.5%) children all low across fluency, originality, and flexibility, which was referred to as the uniformly-low group. Group membership was assigned to each child corresponding to his or her DT profile.

Last, differences in two measures of creative performance were examined by group membership estimated in LPA using ANOVA. The present study aimed to provide evidence against the argument of low criterion-related validity—based on low correlations in conventional approaches—by showing that the presence of groups with multiple profiles may interrupt showing a single relationship between DT and creative performance. Thus, the uniformly-low group was chosen as a reference group for the other two groups to be compared to. The equal variance assumption as tested in Bartlett’s test of sphericity was not violated in both creative achievement,  $\chi^2(2) = 0.08, p = .96$ , and ideational behaviors,  $\chi^2(2) = 0.17, p = .92$ . Average ratings in two scales significantly differed by group membership: creative achievement ( $F = 4.24, p = .02$ ) and ideational behaviors ( $F = 3.38, p = .04$ ). The comparisons of creative performance by group membership are presented in Table 3.4. Only the low-originality group ( $M = 2.34$ ) significantly outperformed the uniformly-low group ( $M = 2.09$ ) in creative achievement across domains ( $p = .01$ ). In contrast, only the high-originality group ( $M = 3.71$ ) significantly outperformed the uniformly-low group ( $M = 3.19$ ) in ideational behaviors ( $p = .02$ ). These two comparisons remained significant after controlling FDR given the multiple comparisons.

## **Discussion**

The present study aimed to explore a probable factor inhibiting the investigation of the criterion-validity of DT tests. To this end, two issues were examined: whether more than one group with its own DT profile was present, and whether the identified DT profiles explained individual differences in creative performance. This exploration raises a critical question about the typical methodological strategy—zero-order correlation—testing the linear relationship between DT test scores and creative performance. When a population is assumed homogeneous as in a correlation analysis, low correlation coefficients necessarily indicate the relationship

between two variables as weak since a correlation analysis measures the strength of the bivariate association of the two. This weak relationship is correspondingly interpreted as poor criterion-related validity when the two are test scores and criterion performance. A good deal of previous research, typically employing this methodological strategy in examining the criterion-related validity of DT tests, provided evidence about the lack of the criterion-related validity of DT tests, reporting the low correlations of .2 to .3 between DT tests and creative performance (Barron & Harrington, 1981; Sternberg & Lubart, 1996; Zeng, Proctor, & Salvendy, 2011). When groups with their own DT profiles are present as found in the present study, a bivariate correlation is, however, not the ideal since the relationship between DT and creative performance cannot be summarized into a single and general principle across groups.

Results indicate that three groups with their own DT profiles were present when multiple DT indices—fluency, originality, and flexibility—functioned. The finding of bivariate correlation analyses documented in previous literature (see Kim, 2008 for a review) may be true of some but not all subjects within a population given the presence of heterogeneous groups. In other words, the results of correlation analyses may not provide affirmative evidence against the criterion-related validity of DT tests. Because of this lumpy population, as also confirmed in recent works (Silvia et al., 2009; von Strumm et al., 2011), bivariate associations of DT and creative performance may not be as salient as we would hope. The results also confirmed this weak association as correlations ranged from 0 through .22.

Yet, this weak association does not indicate a trivial relationship between DT and creative performance. Rather, to our knowledge, the present study is the first to provide empirical evidence against the claim of the low criterion-related validity of DT tests in the person-centered approach beyond testing correlations in the conventional approach. Results

present that DT profiles—identified as a function of three DT indices—significantly explained individual differences in creative achievement and ideational behaviors. This contribution of DT profiles to creative performance implies that heterogeneity in a population interrupts testing correlations and correspondingly leads us to misinterpret the criterion-related validity of DT tests. In fact, more than one reason has been explored to explain the low correlation between DT and creative performance. Mumford and Gustafson (1988) suggested that creative performance required other cognitive abilities, skills, and opportunities along with DT. On another hand, Baer (1994; Han & Marvin, 2002) asserted that a domain general ability as measured in DT tests is deficient to predict creative performance in different domains. On the other hand, it is hard to avoid low correlations when DT and creative performance are measured across many years and even with different assessments, including pencil-and-paper tests as well as performance measures (Runco et al., 2010). In addition to these probable reasons, the present study informs and expands our notion of the criterion-related validity of DT tests to the idiosyncratic property of DT in each group.

This idiosyncrasy in DT may represent the complex nature of DT. More interestingly, this idiosyncrasy in DT is linked to different creative strengths as supported in the result that different outcomes of creative performance favor different DT profiles. In data, only the low-originality group significantly excelled in creative achievement. This finding implies moderate fluent and low unique DT may prove a greater benefit for creative achievement over highly unique DT. Given that creative achievement is defined as countable outcomes, it can be reasoned that creative achievement requires the ability to select achievable ideas in addition to just generating new ideas (Campbell, 1960; Simonton, 1988). Perhaps, metacognitive ability is central in this self-regulatory process (Pesut, 1990; Runco & Okuda, 1991). When people are

high in metacognition, they are able to recognize what they can and cannot do and eventually plan, monitor, and evaluate their performance (Zimmerman, 2008). Self-monitoring and self-evaluating therefore allow people to weed out wild—mostly unworkable—ideas while generating ideas. In short, creative achievement may not benefit from exclusively unique thinking as some unique ideas are too distant from being achievable.

Creative achievement might be even harder to accomplish without metacognitive ability when quality of creative achievement matters, as in award-winning creative accomplishments. As the roles of metacognition are not testable in the present study, testing differences in quantity and quality of creative achievement was considered separately. These follow-up analyses permitted us to see whether the differences more likely occurred in quality of creative achievement, referring to socially recognizable achievement. These follow-up analyses agreed that the low-originality ( $t = 2.45, p = .02$ ) and high-originality groups ( $t = 2.28, p = .02$ ) both significantly excelled in quantity of creative achievement. In contrast, only the low-originality group significantly excelled in quality of creative achievement ( $t = 2.46, p = .01$ ): In both follow-up analyses, Bartlett's tests of sphericity were satisfied, but no correction was made in the error rate given multiple comparisons. These follow-up analyses corroborated that socially recognizable accomplishment—probably requiring more metacognition—particularly may not favor overly unique DT. Still, it should be noted that the relationship between metacognition and creative performance needs to be directly tested in the future.

Yet, the lack in award-winning creative accomplishment does not necessarily indicate the absence of creative potential in the high-originality group. In data, only the high-originality group significantly excelled in ideational behaviors. Despite the frequent use of creative achievement for a criterion of DT, it is often argued that creative achievement may not be the

direct criterion of DT because this achievement certainly requires much more than DT, such as skills, opportunities, and resources (Mumford & Gustafson, 1988; Runco et al., 2000-2001). Amabile (1983) suggested domain-relevant skills, creativity-relevant skills, and intrinsic task motivation for major components for creativity. Given this componential theory of creativity (Amabile, 1983), ideational behaviors are probably the direct and appropriate criterion of DT. Behaviors listed on the RIBS have connections with overt ideation itself that do not necessarily require any tangible outcomes, implying they are not dependent on relevant skills or external resources. When children of the high-originality group continue to invest their efforts in acquiring skills, knowledge, and motivation over time, they would, in turn, attain creative accomplishment in the future. Thus, their current ideational behaviors can be defined as creative potential. The high-originality group may have strong creative potential that has yet to translate into the same types of creative accomplishments for various reasons. Explaining the reasons is still in question and beyond the scope of the present study.

### **Limitations and Future Research Directions**

The reasons may be explained in studies exploring developmental and socio-environmental characteristics of children with many unique ideas. Our sample is elementary children: some of those may yet to have had sufficient training in skills and knowledge to manifest their creative potential. Besides this developmental characteristic of the sample, some of those may have less access to external resources, such as opportunities for special education beyond regular classes (e.g., gifted education) and teacher encouragement and support, as much as they would need. In fact, the children with many unique ideas may not always be favored by classroom teachers as overly unique ideas are likely to deviate from convention. When children voice too many deviant thoughts, teaching can be interrupted. Not surprisingly, a good deal of



previous literature has documented that teachers tend to associate creative children with negative behavioral characteristics such as being easily distracted, nonconforming, disruptive, and rebellious (Aljughaiman & Mowrer-Reynolds, 2005; Chan & Chan, 1999; Fryer & Collings, 1991; Runco, Jonson, & Bear, 1993; Scott, 1999).

Considering that children producing highly unique ideas are not the majority, teachers might easily overlook these few children to continue to teach classes. In results, only 5.2% of children were classified into the high-originality group. In other words, 1-2 students may be identified as highly unique in a class of 20-25 students, which ecologically makes sense in standard classrooms. Taking these findings together, these few children with highly unique thinking may neither be fully recognized as creative nor likely to have support in traditional classrooms despite their strong ideational potential. Although this notion may shed light on teachers' identifying unconventional and hidden potential in children, this notion is still speculative and requires further investigations in relation to teachers' perceptions of creative children. Future research needs to be directed to show how teachers' identification of creative potential varies by children's DT profiles.

Another limitation of the present study needs further investigations in order to fill the gaps for generalizability of the findings. On one hand, it should be noted that the present study measured uniqueness to represent original thinking in DT. Because uniqueness more likely represents novel ideas, the possibility remains that those novel ideas are only deviant but not useful. The standard definition of creativity includes both novelty and usefulness (Plucker et al., 2004; Runco & Jaeger, 2012). Thus, the finding in the present study may lose, in part, generalizability in creative ideas. For instance, this result might not be persistent if usefulness of ideas was taken into account in rating creativity of ideas in addition to uniqueness. Producing

useful ideas is connected to evaluating thinking that is crucial in attaining socially recognizable creative achievement. Therefore, further research should follow to investigate the findings in the present study persistent where novelty and usefulness are both accounted for in rating ideas.

### **Conclusion**

The present study examined whether more than one group with distinct DT profiles was present, and those DT profiles explained variability in creative performance using LPA. Results confirmed the presence of heterogeneous groups regarding DT profiles and differences in creative achievement and ideational behaviors by these groups. The present study contributes to providing empirical evidence against the claim that low correlations between DT indices and creative performance indicates the lack in the criterion-related validity of DT tests. Furthermore, the present study contributes to suggesting idiosyncrasy in creative strengths by showing that favored DT profiles differ by creative performance. These idiosyncratic properties in DT profiles are theoretically important given the complex nature of creative thinking in order to expand our understanding of creative thinking. Also, practices should take this idiosyncrasy into account in providing individualized intervention to enhance creative potential.

## References

- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19, 716-23.
- Aljughaiman, A., & Reynolds, E. (2005). Teachers' conceptions of creativity and creative students. *Journal of Creative Behavior*, 39, 17-34.
- Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45, 357-376.
- Baer, J. (1993). *Creativity and divergent thinking: A task-specific approach*. Hillsdale, NJ: Erlbaum.
- Baer, J. (1994). Divergent thinking is not a general trait: A multidomain training experiment. *Creativity Research Journal*, 7, 35-46.
- Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology*, 32, 439-476.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.
- Borland, J. H. (1986). A note on the existence of certain divergent-production abilities. *Journal for the Education of the Gifted*, 9, 239-251.
- Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, 67, 380 - 400.
- Chan, D. W., & Chan, L. (1999). Implicit theories of creativity: Teachers' perception of student characteristics in Hong Kong. *Creativity Research Journal*, 12, 185-195.

- Chand O'Neal, I., Paek, S. H., & Runco, M. A. (2015). Comparison of Competing Theories about Ideation and Creativity. *Creativity. Theories–Research–Applications*, 2, 145-164.
- Clapham, M. M., Cowdery, E. M., King, K. E., & Montang, M. A. (2005). Predicting Work Activities with Divergent Thinking Tests\_A Longitudinal Study. *Journal of Creative Behavior*, 39, 149-167.
- Cramond, B. (1994). The Torrance Tests of Creative Thinking: From creation through establishment of predictive validity. In R. F. Subotnik, & K. D. Arnold (Eds.), *Beyond Terman: Longitudinal studies in contemporary education* (pp. 229-254). Norwood, NJ: 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, 283-291.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Csikszentmihalyi, M., & Larson, R. (1984). *Being adolescent: Conflict and growth in the teenage years*. New York: Basic Books.
- Davis, G. A. (1989). Testing for creative potential. *Contemporary Educational Psychology*, 14, 257-274.
- Fryer, M., & Collings, J. A. (1991). Teachers' views about creativity. *British Journal of Educational Psychology*, 61, 207–219.
- Gilson, L. L., & Shalley, C. E. (2004). A little creativity goes a long way: An examination of teams' engagement in creative processes. *Journal of Management*, 30, 453-470.

- Groves, K. S., Vance, C. M., Choi, D. Y., & Mendez, J. L. (2008). An examination of the nonlinear thinking style profile stereotype of successful entrepreneurs. *Journal of Enterprising Culture, 16*, 133-159.
- Guilford, J. P. (1968). *Creativity, intelligence, and their educational implications*. San Diego: Knapp/ EDITS.
- Han, K.-S., & Marvin, C. (2002). Multiple creativities? Investigating domain-specificity of creativity in young children. *Gifted Child Quarterly, 46*, 98-109.
- Han, K. -S. (2003). Domain-specificity of creativity in young children: How quantitative and qualitative data support it. *Journal of Creative Behavior, 37*, 117-142.
- Hargreaves, D.J., & Bolton, H. (1972). Selecting creativity tests for use in research. *British Journal of Psychology, 63*, 451-462.
- Helson, R. (1965). Childhood interest clusters related to creativity in women. *Journal of Consulting Psychology, 29*, 352-361.
- Hocevar, D. (1979). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology, 71*, 191-196.
- Hong, E., & Milgram, R. M. (1991). Original thinking in preschool children: A validation of ideational fluency measures. *Creativity Research Journal, 4*, 253-260.
- Khandwalla, P. N. (1993). An exploratory investigation of divergent thinking through protocol analysis. *Creativity Research Journal, 6*, 241-259.
- Kim, K. H. (2006). Is creativity unidimensional or multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creativity Research Journal, 18*, 251-259.
- Kim, K. H. (2008). Meta-analyses of the relationship of creative achievement to both IQ and divergent thinking test scores. *Journal of Creative Behavior, 42*, 106-130.

- Kirton, M. J. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology, 61*, 622-629.
- Kogan, N., & Pankove, E. (1972). Creative ability over a five year span. *Child Development, 43*, 427-442.
- Laursen, B. P., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly, 52*, 377-389.
- Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*, 21–39.
- Lumsdaine, M., & Lumsdaine, E. (1995). Thinking preferences of engineering students: Implications for curriculum restructuring. *Journal of Engineering Education, 84*, 193-204.
- Magnusson, D. (2003). The person approach: Concepts, measurement models, and research strategy. In S. C. Peck & R. W. Roeser (Eds.), *New directions for Child and Adolescent development. Person-centered approaches to studying development in context* (No. 101, pp. 3–23). San Francisco: Jossey-Bass.
- Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. S. (2009). Classical Latent Profile Analysis of Academic Self-Concept Dimensions: Synergy of Person- and Variable-Centered Approaches to Theoretical Models of Self-Concept. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 191-225.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

- Milgram, R. M., & Milgram, N. A. (1976). Creative thinking and creative performance in Israeli students. *Journal of Educational Psychology, 68*, 255-259.
- Mumford, M. D., & Gustafson, S. B. (1988). Creativity syndrome: Integration, application, and innovation. *Psychological Bulletin, 103*, 27-43.
- Paek, S. H., Park, H., Runco, M. A., & Choe, H. S. (2016). The contribution of ideational behavior to creative extracurricular activities. *Creativity Research Journal, 28*, 144-148.
- Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology, 32*, 8-47.
- Pesut, D. J. (1990). Creative thinking as a self-regulatory metacognitive process - A model for education, training and further research. *Journal of Creative Behavior, 24*, 105-110.
- 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.
- Plucker, J. A., Qian, M., & Schmalensee, S. L. (2014). Is what you see what you really get? Comparison of scoring techniques in the assessment of real-world divergent thinking. *Creativity Research Journal, 26*, 135-143.
- Plucker, J. A., Qian, M., & Wang, S. (2011). Is originality in the eye of the beholder? Comparison of scoring techniques in the assessment of divergent thinking. *Journal of Creative Behavior, 45*, 1-22.
- Runco, M. A. (1986a). Divergent thinking and creative performance in gifted and nongifted children. *Educational and Psychological Measurement, 46*, 375-384.
- Runco, M. A. (1986b). The discriminant validity of gifted children's divergent thinking test scores. *Gifted Child Quarterly, 30*, 78-82.

- Runco, M. A., & Albert, R. S. (1985). The reliability and validity of ideational originality in the divergent thinking of academically gifted and nongifted children. *Educational and Psychological Measurement*, 45, 483-501.
- Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, 24, 92-96.
- Runco, M. A., & Okuda, S. M. (1991). The instructional enhancement of the flexibility and originality scores of divergent thinking tests. *Applied Cognitive Psychology*, 5, 435-441.
- Runco, M. A., & Okuda, S. M. (1991). The instructional enhancement of the flexibility and originality scores of divergent thinking tests. *Applied Cognitive Psychology*, 5, 435-441.
- Runco, M. A., Abdulla, A. M., Paek, S. H., Al-Jasim, F. A., & Alsuwaidi, H. N. (2016). Which Test of Divergent Thinking Is Best? *Creativity. Theories–Research–Applications*, 3, 4-18.
- Runco, M. A., Johnson, D. J., & Bear, P. K. (1993). Parents' and teachers' implicit theories of children's creativity. *Child Study Journal*, 23, 91-113.
- Runco, M. A., Millar, G., Acar, S., & Cramond, B. (2010). Torrance tests of creative thinking as predictors of personal and public achievement: A fifty-year follow-up. *Creativity Research Journal*, 22, 361-368.
- Runco, M. A., Plucker, J. A., & Lim, W. (2000-2001). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal*, 13, 393-400.
- Runco, M. A., Walczyk, J. J., Acar, S., Cowger, E. L., Simundson, M., & Tripp, S. (2013). The Incremental Validity of a Short Form of the Ideational Behavior Scale and Usefulness of Distractor, Contraindicative, and Lie Scales. *Journal of Creative Behavior*, 48, 185-197.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.



- Scott, L. C. (1999). Teachers' biases toward creative children. *Creativity Research Journal*, 12, 321–328.
- Seddon, G. M. (1983). The measurement and properties of divergent thinking ability as a single compound entity. *Journal of Educational Measurement*, 20, 393-402.
- Sen, S. (2016). Applying the Mixed Rasch Model to the Runco Ideational Behavior Scale. *Creativity Research Journal*, 28, 426-434.
- Silvia, P. J. (2008). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). *Creativity Research Journal*, 20, 34–39.
- Silvia, P. J., Kaufman, J. C., & Pretz, J. E. (2009). Is creativity domain-specific? Latent class models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics, Creativity, and the Arts*, 3, 139.
- Simonton, D. K. (1988). *Scientific genius: A psychology of science*. Cambridge, MA: Harvard University Press.
- Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, 104, 66-89.
- Simonton, D. K. (2000). Creativity: Cognitive, personal, developmental, and social aspects. *American Psychologist*, 55, 151-158.
- Simonton, D. K. (2007). Creativity: Specialized expertise or general cognitive processes? In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 351–367). Hove, England: Psychology Press.
- Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American Psychologist*, 51, 677-688.

- Tan, A. (2000). A review on the study of creativity in Singapore. *Journal of Creative Behavior*, 34, 259-284.
- Torrance, E. P. (1980). Growing up creatively gifted: A 22-year longitudinal study. *Creative Child and Adult Quarterly*, 5, 148–158, 170.
- Torrance, E. P. (1981). Empirical validation of criterion-referenced indicators of creative ability through a longitudinal study. *Creative Child and Adult Quarterly*, 6, 136–140.
- Torrance, E. P., & Wu, T. (1981). A comparative longitudinal study of the adult creative performances of elementary school children identified as high intelligent and as highly creative. *Creative Child and Adult Quarterly*, 4, 71-76.
- Truhon, S. A. (1983). Playfulness, play, and creativity: A path analytic model. *Journal of Genetic Psychology*, 143, 19-28.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge, UK: Cambridge University Press.
- von Stumm, S., Chung, A., & Furnham, A. (2011). Creative ability, creative ideation and latent classes of creative achievement: What is the role of personality? *Psychology of Aesthetics, Creativity, and the Arts*, 5, 107-114.
- Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity–intelligence distinction*. New York, NY: Holt, Rinehart & Winston.
- Wallach, M. A., & Wing, C. W. (1969). *The talented student*. New York, NY: Holt, Rinehart & Winston.

Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal*, 23, 24-37.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45, 166–183.

Table 3.1

*Descriptive Statistics and Correlations of the DT Indices, CAAC, and RIBS Scores*

	<i>r</i>			<i>M</i>	<i>SD</i>	Skewness
	Fluency	Originality	Flexibility			
CAAC Composite <sup>a</sup>	.14**	.09**	.12**	2.17	0.60	0.61
Quantity Composite <sup>b</sup>	.13***	.22***	.17***	2.49	0.66	0.14
Quality Composite <sup>c</sup>	.2*	.19	.19*	1.86	0.63	1.09
RIBS	.14*	.11***	.16**	3.28	0.87	-0.18

*Note.* <sup>a</sup>Composite score of both the quantity and quality scales across domains. <sup>b</sup>Composite score of the quantity scales across domains. <sup>c</sup>Composite score of the quality scales across domains. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

Table 3.2

*Fit Indices for Mixture Models*

Fit indices	Model			
	1-Group	2-Group	3-Group	4-Group
Number of parameters	28	32	36	40
AIC	7323.09	7169.06	6754.86	7033.23
BIC	7429.04	7290.14	7083.34	7184.59
ABIC	7340.22	7188.64	7219.55	7057.71
Entropy		0.98	0.9	0.88
VLMR-LRT ( <i>p</i> -value)		162.03 (.03)	93.72 (<.001)	58.1 (.09)
LMR-adjusted LRT ( <i>p</i> -value)		155.32 (.03)	89.84 (.001)	55.7 (.09)
Sample size				
$N_{g=1}$	325	16	17	10
$N_{g=2}$		309	121	49
$N_{g=3}$			187	77
$N_{g=4}$				189

*Note.* AIC = Akaike information criterion. BIC = Bayesian information criterion. ABIC = sample-size adjusted Bayesian information criterion. VLMR-LRT = Vuong–Lo–Mendell–Rubin likelihood ratio test. LMR-adjusted LRT = Lo–Mendell–Rubin adjusted likelihood ratio test.

Table 3.3

*Correlations and Means of Factor Scores of DT Tests in Three Groups*

	<i>r</i>		<i>M</i>		
	Fluency	Originality	Group 1	Group 2	Group 3
Fluency			1.69	1.72	0
Originality	.46		3.55	0.43	0
Flexibility	.95	.52	1.51	1.14	0

*Note.* Group 1 = High-originality group, Group 2 = Low-originality group, and Group 3 = Uniformly-low group.

Table 3.4

*Mean Differences in the CAAC and the RIBS by Group Membership*

	Group 1	Group 2	Group 3	Overall test		Pairwise tests	
	( <i>n</i> = 17)	( <i>n</i> = 121)	( <i>n</i> = 187)			Group1- Group3	Group2- Group3
	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>F</i>	<i>p</i>	<i>p</i> <sup>b</sup>	<i>p</i> <sup>b</sup>
CAAC <sup>a</sup>	2.34 (0.62)	2.28 (0.59)	2.09 (0.6)	4.24*	.015	.09	.01*
RIBS	3.71 (0.93)	3.35 (0.86)	3.19 (0.86)	3.38	.04	.02*	.13

*Note.* Group 1 = High-originality group, Group 2 = Low-originality group, and Group 3 = Uniformly-low group. <sup>a</sup>Composite score of both the quantity and quality scales across domains.

<sup>b</sup>Values indicate non-adjusted *p*-values in pairwise tests. Error rate was adjusted by controlling false discovery rate (FDR) using Benjamini and Hochberg's method to correct *p* -value given four multiple comparisons (Benjamin & Hochberg, 1995). Asterisks indicate the mean differences remained significant after controlling the FDR at the .05 level.

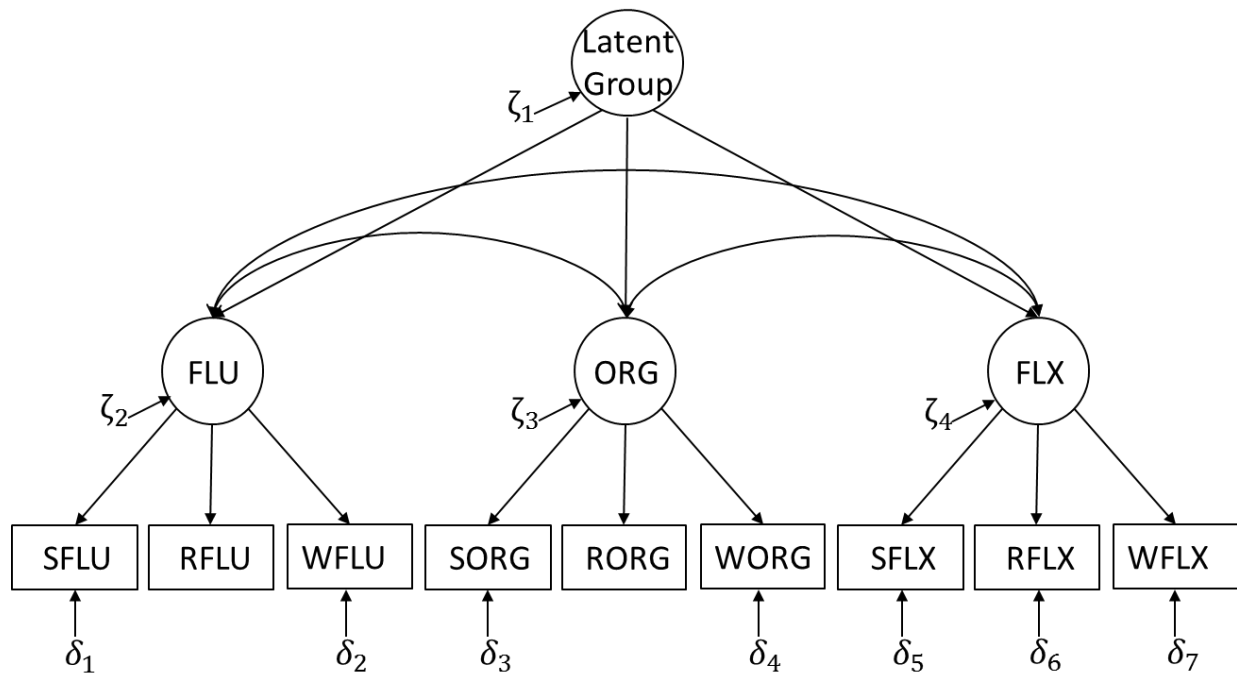
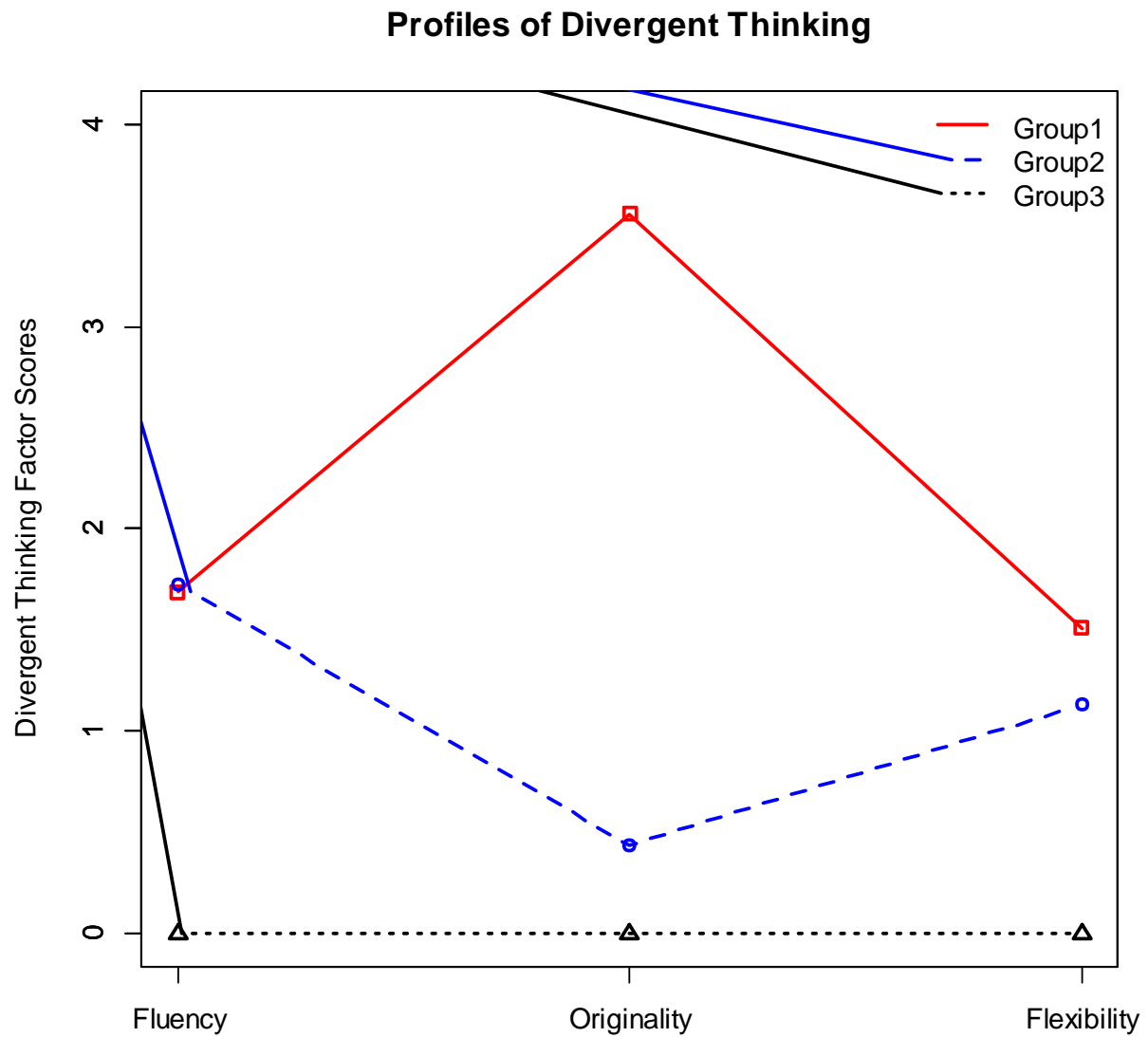


Figure 3.1. Mixture model examining the groups with different profiles of divergent thinking in children

*Note.* FLU = fluency. ORG = originality. FLX = flexibility. SFLU = fluency score in Square test. SORG = originality score in Square test. SFLX = flexibility score in Square test. RFLU = fluency score in Round test. RORG = originality score in Round test. RFLX = flexibility score in Round test. WFLU = fluency score in Wheel test. WORG = originality score in Wheel test. WFLX = flexibility score in Wheel test.  $\delta$  = Unique variable.  $\zeta$  = Disturbance. Unique variables of RFLU and RORG were set to be zero due to the issue of negative residuals in estimation.





*Figure 3.2.* Profiles of divergent thinking

*Note.* Group 1 = High-originality group, Group 2 = Low-originality group, and Group 3 = Uniformly-low group

## CHAPTER 4

### CONCLUSIONS

A divergent thinking (DT) test is one of the most popular measures assessing creative thinking (Hocevar, 1981). The validity of DT tests is necessary to ensure that we adequately measure what is intended. Thus, validity of DT tests has been a long lasting topic in creativity research. (Clapham, 2004; Guilford, 1971; Hargreaves & Bolton, 1972; Hocevar, 1979a, 1979b; Plucker, 1999; Runco, 1986; Silvia et al., 2008; Treffinger, Renzulli, & Feldhusen, 1971). However, the complex nature of DT often challenges a reliable and valid measurement of it. In fact, the validity of DT tests has been plagued with conflicting findings and interpretations of those findings (Hocevar, 1981; Michael & Wright, 1989; Zeng, Proctor, & Salvendy, 2011). Conventional approaches may not enable us to fully deal with this complex nature of DT. Therefore, the current studies used advanced statistical models to take the complex nature into consideration in investigating the validity of DT tests.

The first study (Chapter 2) investigated whether a method effect was present in DT indices, and fluency and originality were discriminable against one another when the method effect was ruled out using MTMM analysis in CFA. Three major findings of this study were as follows:

- 1) A method effect, which was probably entailed by DT tasks, was present in DT indices.
- 2) A correlation between fluency and originality was overestimated when the method effect was not ruled out.

- 3) Fluency and originality were discriminable against one another when the method effect was ruled out in test scores. In marked contrast, fluency and originality were not discriminable when the method effect was not ruled out in test scores.

These findings supported that the method effect in DT indices was an obstacle in testing the discriminant validity of DT indices. Also, the discriminant validity of DT indices was supported when the method effect was ruled out. This empirical evidence on the discriminant validity contributes to upholding the multidimensionality of DT and the usages of multiple indices over a single index in creativity research.

The second study (Chapter 3) examined whether more than one group with distinct profiles in DT was present, and this group membership explained individual differences in creative performance using LPA. Two major findings of this study were as follows:

- 1) Three groups with their own DT profiles were classified based on profiles in DT. One group was moderate in fluency and low in originality of DT, which was labeled the low-originality group. Another group was moderate in fluency and high in originality of DT, which was labeled the high-originality group. The other group was uniformly low across fluency and originality of DT, which was labeled the uniformly-low group.
- 2) Only the low-originality group significantly outperformed the uniformly-low group on creative achievement. On the contrary, only the high-originality group significantly outperformed the uniformly-low group on ideational behaviors.

Results supported that groups with their own profiles in DT were present, and these groups also differed in terms of creative performance. Given the presence of heterogeneous groups, these findings provide empirical evidence against the argument that the low correlations between

DT and creative performance indicate the lack of the criterion-related validity of DT tests. This study contributes to uncovering a probable obstacle of heterogeneous groups and suggesting an approach to handle this heterogeneity in investigating the criterion-related validity of DT tests through person-centered approach.

Two limitations of the current studies need further investigation to achieve generalizability of these findings. First, these studies selectively used only a few DT tests and operational definitions of DT indices among many possible options. Characteristics and measurement qualities are not identical across DT tests (Runco, Abdulla, Paek, Al-Jasim, & Alsuwaidi, 2016) and operational definitions of DT indices (Plucker, Qian, & Schmalensee, 2014; Silvia, 2011; Silvia et al., 2008). Additionally, only responses in children were analyzed to reach these findings. Therefore, further investigation should follow using different DT tests and operational definitions of DT indices in other age groups to generalize these findings. Second, DT test scores may not be free from potential influences by internal factors such as task motivation (Amabile, 1983) and situational interest (Jeon, Moon, & French, 2011), as well as external factors such as environmental cues (Dodds, Smith, & Ward, 2002) and instructions (Runco, Illies, & Reiter-Palmon, 2005). The internal and external factors may probably entail variation in reliability and normality, which often undermines the reliability and normality in some DT tests scores. Most statistical analyses require adequate levels of reliability and normality in variables to ensure accurate results. Hence, the internal and external factors should be managed in measuring DT to improve the reliability and normality in DT test scores.

Despite these limitations, the present studies contributed to statistically testing the presence of probable obstacles in examining the discriminant and criterion-related validity of DT tests such as the method effect in DT tests and heterogeneous groups. These obstacles are connected

to the complex nature of DT and DT tests. In order to deal with this complex nature of DT tests, the present studies also suggested how to empirically examine the discriminant and criterion-related validity of DT tests using advanced methodology. Therefore, findings in these studies contribute to expanding our understanding on the discriminant and criterion-related validity of DT tests beyond conventional approaches.

## References

- Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45, 357-376.
- Clapham, M. M. (2004). The convergent validity of the Torrance Tests of Creative Thinking and creativity interest inventories. *Educational and Psychological Measurement*, 64, 828-841.
- Dodds, R. A., Smith, S. M., & Ward, T. B. (2002). The use of environmental cues during incubation. *Creativity Research Journal*, 14, 278-304.
- Guilford, J. P. (1971). Some misconceptions regarding measurement of creative talents. *Journal of Creative Behavior*, 5, 77-87.
- Hargreaves, D. J., & Bolton, H. (1972). Selecting creativity tests for use in research. *British Journal of Psychology*, 63, 451-462.
- 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. (1981). Measurement of creativity: Review and critique. *Journal of Personality Assessment*, 45, 450-464.
- Jeon, K. N., Moon, S. M., & French, B. (2011). Differential effects of divergent thinking, domain knowledge, and interest on creative performance in art and math. *Creativity Research Journal*, 23, 60-71.
- Michael, W. B., & Wright, C. R. (1989). Psychometric issues in the assessment of creativity. In J. A. Glover, R. R. Ronning, C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 53-75). New York, NY, US: Plenum Press.

- 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.
- Plucker, J. A., Qian, M., & Schmalensee, S. L. (2014). Is what you see what you really get? Comparison of scoring techniques in the assessment of real-world divergent thinking. *Creativity Research Journal*, 26, 135-143.
- Runco, M. A. (1986). Divergent thinking and creative performance in gifted and nongifted children. *Educational and Psychological Measurement*, 46, 375-384.
- Runco, M. A., Abdulla, A., Paek, S. H., Al-Jasim, F., & Alsuwaidi, H. N. (2016). Which test of divergent thinking is best? *Creativity. Theories – Research – Applications*. 3, 4-18.
- Runco, M. A., Illies, J. J., & Reiter-Palmon, R. (2005). Explicit instructions to be creative and original: A comparison of strategies and criteria as targets with three types of divergent thinking tests. *Korean Journal of Thinking and Problem Solving*, 15, 5-15.
- Silvia, P. J. (2011). Subjective scoring of divergent thinking: Examining the reliability of unusual uses, instances, and consequences tasks. *Thinking Skills and Creativity*, 6, 24-30.
- Silvia, P. J., et al. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 68-85.
- Treffinger, D. J., Renzulli, J. S., & Feldhusen, J. F. (1971). Problems in the assessment of creative thinking. *Journal of Creative Behavior*, 5, 104-112.
- Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal*, 23, 24-37.