

SUBJECTIVE AND OBJECTIVE MEASUREMENT IN CREATIVITY:

COMPARISON STUDIES

by

BURAK TURKMAN

(Under the Direction of Mark A. Runco)

ABSTRACT

Developing technologies and twenty-first century opportunities have helped build computerized scoring systems for assessments of creativity. Surely, computerized scoring is the future direction for such new creativity tests, especially since computers can address the problem of more laborious manual scoring procedures for such tests. However, quality and quantity are distinct values for creativity measurement, and many objective scoring systems do not capture this difference, thereby causing validity problems. Therefore, subjective scoring is also vital to stable creativity measurement. Through three interrelated yet distinct studies, my dissertation aimed to bring new perspectives to measurements of creativity by comparing and building upon existing objective and subjective scoring methods in creativity judgment. Hence, comparing subjective and objective scoring methods and outcomes was selected as the overarching theme for the three studies below. These comparisons critically examined existing scoring methods to determine gaps where new measurement techniques may be needed.

INDEX WORDS: Objective Scoring, Subjective Scoring, Divergent Thinking, Associative Process, Explicit Instructions, Creativity Judgments

SUBJECTIVE AND OBJECTIVE MEASUREMENT IN CREATIVITY:
COMPARISON STUDIES

by

BURAK TURKMAN

B.A., Istanbul University, Turkey, 2006

M.S.Ed., Purdue University, USA, 2011

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

2016

© 2016

Burak Turkman

All Rights Reserved

SUBJECTIVE AND OBJECTIVE MEASUREMENT IN CREATIVITY:
COMPARISON STUDIES

by

BURAK TURKMAN

Major Professor:	Mark A. Runco
Committee:	Bonnie Cramond
	Seock-Ho Kim
	Selcuk Acar

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016

DEDICATION

To my parents Ayfer & Faik Turkman for their endless love, prayers, and support.

ACKNOWLEDGEMENTS

I owe a great debt of gratitude to those who have accompanied me on my journey towards completing this dissertation. Firstly, this research would not have been possible without these kind and generous people.

To my parents, Ayfer and Faik Turkman, your unrelenting support and sacrifice in my pursuits were priceless. You have always believed in me and given your time, money, and love to walk on the path that I wanted to be. Your grandchildren, especially Anastasya will never forget her “Babaanne’s” friendship and love on the journey we all walked together. Thank you so much Babaanne for taking care of our daughter and son to give us time and courage to complete my dissertation journey.

To my loving wife Sonya, princess daughter Anastasya, and pasha son Ahmet Evan. Your support and existence were my biggest cheer for me to complete my dissertation. Without my daughter’s loving voice, or my son’s heavenly smile or my wife’s love, this journey wouldn’t be coming to an end. I also would like to thank my wife Sonya for reading and editing my countless drafts and revisions.

Dr. Mark A. Runco, I was very fortunate to have the opportunity to study under your direction. Your sharp analysis and encouragement always brought a source of inspiration to complete my degree. Your scholarly perspective and research discipline were the most important lessons that I learned from grad school. Dr. Bonnie Cramond, you will never let your Dr. Torrance flag down. Your positive energy will ignite many lights to enlighten dark futures. Thank you for your advice through my graduate life at the University of Georgia and your

support to complete my dissertation. Dr. Seock-Ho Kim, you are a role model for many young academicians like me. I will never forget your professionalism, and your ethic and discipline inspired me to be someone like you towards my students. I am very thankful for your valuable advise and insight on this journey. Dr. Selcuk Acar, you are an amazing person, academician, and a friend. Thank you for encouraging me to come to the UGA. I am grateful for your recommendations without which this work could not have been done.

My special thanks goes to my friend Mustafa Veysi Nural. Thank you for helping me to develop the computer algorithm and for making this research possible. Dr. Sarah Sumners, Dr. Desiree Sharpe, Sureyya Yoruk, Sue-hyeon Paek, Mustafa Yesil, and Sarah Marie Catalana thanks for providing great company and helping hands.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	V
LIST OF TABLES	IX
LIST OF FIGURES	X
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW	1
Current Issues in Creativity Measurement	2
Objectivity and Subjectivity Discussions in Creativity Measurement	5
The Most Popular Creativity Tests and Their Pros and Cons	6
Objectives of the Current Studies	19
2 ASSOCIATIVE PROCESSES IN DIVERGENT THINKING A SPECIAL CASE:	
FLEXIBILITY	21
Abstract	22
Introduction	23
Research Questions	28
Method	29
Results	32
Discussion	38
3 KEYWORDS STUDY	43
Abstract	44

Introduction	44
Research Questions	47
Method	47
Results	52
Discussion	55
4 EXPLICIT INSTRUCTIONS & CREATIVITY JUDGMENT	59
Abstract	60
Introduction	60
Research Questions	65
Method	65
Results	68
Discussion	69
5 CONCLUSIONS	72
REFERENCES	77
APPENDICES	89
A THE MANY USES GAME	89
B THE CREATIVITY WORKSHOP SLIDES	90
C THE ATTITUDES AND VALUES SCALE	95

LIST OF TABLES

	Page
Table 2.1: Descriptive statistics for the divergent thinking tasks	32
Table 2.2: Descriptive statistics for all indices defined using seven judges	34
Table 2.3: Descriptive statistics for all indices defined using the IF, WAN, and WN.....	35
Table 2.4: Alpha coefficients for close and remote associations for each judge	35
Table 2.5: Bivariate correlations between ID and composite DT scores for remote associations ..	36
Table 2.6: Bivariate correlations between ID and composite DT scores for close associations	37
Table 2.7: Alpha coefficients for close and remote associations for each source	37
Table 2.8: Bivariate correlations between ID and composite DT scores for close and remote associations	38
Table 3.1: The Experts' initial keywords lists	50
Table 3.2: The finak keywords lists	52
Table 3.3: Descriptive statistics for the divergent thinking tasks	52
Table 3.4: Descriptive statistics for the essays' ID values	53
Table 3.5: Bivariate correlations among keywords, ID and DT scores	54
Table 3.6: Descriptive statistics for TED talks study	54
Table 3.7: Descriptive statistics for the Ludwig (1995) study	55
Table 4.1: Alpha coefficients for the essays	68

LIST OF FIGURES

	Page
Figure 2.1: Theoretical demonstration of close, remote, and extremely remote (irrelevant) associations involved in DT	26

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

For decades, measuring creativity has befuddled researchers and practitioners alike. The lack of a widely accepted definition of creativity and the resulting lack of appropriate criteria to assess creativity make creativity measurement research both significant and complicated. Through three interrelated yet distinct studies, my research brought new perspectives to measurement in creativity by comparing and building upon extant objective and subjective scoring methods. These comparisons critically examined existing scoring methods to determine gaps where new measurement techniques may be needed.

The three studies in this dissertation contributed to methods of assessment in creativity judgments. All three studies compared objective and subjective scoring mechanisms and their effects on quantifying creativity for potential research purposes. The first study measured flexibility in terms of semantic distance, the closeness in meaning between two terms, to discern creative thinking. It also used human judges to obtain subjective scoring, who could then explain semantic associations. The second study identified and used a set of keywords to quantify the number of original ideas in an essay. The third study explored the causal relationship of Explicit Instructions (EIs) on the quality of creativity judgments made by non-expert judges.

Providing a comparison of subjective and objective scoring practices is the integral goal of these three studies. Each scoring system comes with its pros and cons. Throughout these three studies, different scoring methods were utilized and analyzed for their respective potential to assess creativity.

Current Issues in Creativity Measurement

The measurement of creativity has attracted the attention of researchers for decades because few other psychological concepts have proven so difficult to measure (Hocevar, 1981). According to Runco (2010a), usefulness and originality are two main aspects of creativity, and because they are generally unpredictable, it is very difficult to formulate measures for them. This challenge has inspired deep and controversial explanations of what creativity is and what it is not (2010a). For example, creativity has been presented as a complex psychological construct requiring multifaceted assessment techniques (Runco, 2004). First, it is difficult to capture all domains of creativity in a singular test, even a test with multiple sections. Secondly, it is difficult to determine what behaviors are creative. This criterion issue (Runco, 2010a) creates several problems because of the uncertainty regarding what to look for or what to expect. These are only two of the core reasons why researchers find it very difficult to measure creativity.

Definition Challenge

In consequence, there is no one widely accepted definition for creativity. Thus, many creativity tests and measures do not agree with each other (Runco, 2010a). Without knowing what to look for or what to expect, it is very challenging to get valid results or data. According to Williams (1999), the clarity of the definitions of higher-order cognitive constructs is connected to the usefulness of the terms used. Skinner (1970) argued that science should predict possible outcomes and control for variances; otherwise, one cannot have a good understanding of cause and effect. Runco (2010a) pointed out the importance of having an explicit definition for creativity. He opined that this problem with a lack of any explicit, universal definition in creativity research might produce highly conflicting results in two different studies for the same

aspect of creativity when any two varying definitions are applied. Additionally, this lack of uniformity undermines the construct validity of creativity tests themselves (Runco, 2010b).

The Criterion Problem

One of the biggest issues that creativity research faces is the criterion problem. The criterion problem in creativity assessment refers to the lack of criterion-validity in creativity tests; in other words, creativity tests should be able to predict certain analogous criteria (Runco, 2010a). This problem has many roots. The greatest source of this problem is the lack of a more broadly accepted everyday definition of what creativity is (Runco, 2010b). Therefore, the criterion problem discourages many researchers from using some creativity tests (i.e., Divergent Thinking tests) because of validity issues. Not having a sound criterion for creativity discernment causes judges of creativity to have problems discriminating creativity from other attributes. Hocevar (1981) pointed out that this also makes it difficult to discriminate between various dimensions of creativity. This lack of clear norms causes a “halo effect” in creativity measurement in that certain subjective impressions can influence an observer’s judgment. Subjective judgments depend on the observer’s overall opinion about what creativity is. Of course, an individual’s definition of creativity influences all such judgments (Hocevar, 1981). It can be concluded that peer and/or teacher nominations, supervisor ratings, and product judgments are not impressive indicators of creativity because of the rater's potential inability to discriminate creativity from other traits, such as intelligence, personality.

Validity and Reliability Issues

Evidence of reliability for the Torrance Tests of Creative Thinking (TTCT), the Structure of Intelligence (SOI) and the Wallach and Kogan test are fairly convincing, according to Plucker and Makel (2010) because these tests’ psychometric evidence of reliability values are in

acceptable range. However, the aforementioned criterion and perceived lack of predictive validity problems (Baer, 1993) have led some educators and researchers to avoid using creativity tests at all.

The validity issues (e.g., predictive validity, discriminant validity) are connected to the criterion and explicit definition problems in creativity research, and thus, many researchers have attempted to address the criterion problem in creativity research. For example, Runco (2010b) attempted to discriminate between some overlapping aspects of creativity dimensions (e.g., discovery, invention, innovation, and originality). However, he concluded that no single core criterion could explain all of creativity (Runco, 2010b). In other words, as presented, creativity is a complex construct that is interconnected with varying aspects of creative thinking. Hocevar (1981) stated that many researchers have worked on different aspects of an intricate framework for creativity, including contributing factors such as cognition, attitudes, interests, personality, biography, etc. However, these aspects of creativity are not enough to use in the full assessment of creative potential, although these are appropriate concerns to explain different assumptions about creativity. Hocevar did conclude that diversity is the most common feature of creativity tests, and that diversity represents the most complex dimension of creativity (Hocevar, 1981).

On the other hand, some researchers think that Amabile's Consensual Assessment Technique (CAT) is the answer to the "criterion problem" in creativity research (Plucker & Makel, 2010). Sternberg and Kaufman (2010) argued that the criterion problem is a result of the field's challenges in defining its terms uniformly. Amabile's CAT strives to solve this problem. In the CAT measure, expert opinion is the criterion, but this technique is not infallible because of expert variability. The CAT brings to the fore such questions as expertise level and why expert criteria would be considered more valid than self-rating or non-expert evaluations, such as peer,

teacher. These controversial ideas about the CAT will be discussed further below.

Overall, looking at the studies and concerns mentioned above as a whole, it could be concluded that no one measure can capture all the components of creativity. Therefore, it can be posited that creativity is a syndrome or complex construct that must be measured by applying more than one scale (Runco, 2010b).

Objectivity and Subjectivity Discussions in Creativity Measurement

According to Runco (2010a), creative ideas must be both original and useful, which means that many original ideas are not creative, so they can slip through the cracks of the objective scoring system (Silvia, Winterstein, Willse, Barona, Cram, Hess, & Richard, 2008). In response pools, it is difficult to filter out unconventional responses. Additionally, the uniqueness scoring system of traditional divergent thinking (DT) tasks is not adept at identifying mundane responses. For example, “make a soup bowl” is an obvious use for a bowl, but it could be unique in the small sample sizes that are typical of creativity research. This feature also reduces the reliability of DT tasks.

On the other hand, subjective scoring would deem “make a soup bowl” uncreative. In Hocevar and Michael’s (1979) study, the uniqueness scores of the raters were statistically lower than the objective uniqueness scores, indicating that the raters made greater discriminations in judging uniqueness.

The methods that have been designed to overcome the confounding effects of fluency (number of ideas) generally penalize idea generation and/or large samples (Silvia et al., 2008). For example, dividing DT indices by fluency penalizes idea production. Therefore, subjective ratings play an important role in creativity assessment in spite of raters’ disagreements in DT tasks. Indeed, Runco and Mraz (1992) demonstrated the importance of training raters in their

study. Providing clear instructions can help ensure a high agreement among raters. However, Csikszentmihalyi and Wolfe (2005) opined that the cognitive process of creativity is “one that takes place in a context of previous cultural and social achievements, and is inseparable from them... Creativity is not produced by single individuals, but by a social system making judgments about individuals’ products” (p. 81-82). Hence, subjective ratings and raters’ criteria are not separable from existing constructs of creativity (e.g. fluency, originality, usefulness). Csikszentmihalyi and Wolfe (2005) concluded that objectivity is not an issue in creativity measurement. The product must be judged according to “the effect it is able to produce in others who are exposed to it (p. 82).

The Most Popular Creativity Tests and Their Pros and Cons

Divergent Thinking Tests

DT tests are the most widely used measures in creativity research (Plucker & Makel, 2010) and have been described as the backbone of creativity assessment for many decades (Kaufman, Plucker, & Baer, 2008). Runco’s (1999) definition of DT offers some helpful insights for understanding the concept: “Divergent Thinking is cognition that leads in various directions” (p. 577). As can be understood from Runco’s DT definition, the concept of DT is open-ended. This particular feature distinguishes DT tests from traditional IQ tests that focus on one right answer. Fluency (number of ideas), originality (number of new and unusual ideas), flexibility (number of connections among different areas and ideas), and elaboration (number of details provided about ideas) are the four main aspects of DT that are frequently mentioned in creativity literature (Runco, 1999).

Although DT was recognized in early mental ability tests (Runco, 2011), the general idea of “divergent production” comes from the 1950s, specifically from J.P. Guilford’s studies on

creativity. According to Guilford (1950, 1968), divergent production is highly connected to creative potential. Guilford defined it as the "generation of logical alternatives from given information, where the emphasis is upon variety, quantity, and relevance of output from the same source" (Guilford & Hoepfner, 1971, p. 20). By looking at this definition, it can be assumed that DT involves thinking in different directions and seeking out variety. Taylor (1988) used a very similar approach, defining DT as "one's ability to go off in different directions when faced with a problem."

Although Guilford (1988) hypothesized the existence of 36 different DT abilities in his latest SOI model study, many DT tests only focus on fluency, originality, flexibility, and elaboration. Fluency measures the number of ideas; originality is the number of unseen or unique ideas; flexibility measures the diversity of ideas from different categories; and elaboration measures the extent to which details are given about ideas (Runco & Okuda, 1991). The TTCT are the most popular DT tests in creativity research. Torrance, like Guilford, developed a large number of tasks that depended on the notion that "DT is not a uniform ability." The TTCT maintain a popularity because of many revisions based on the results obtained (Plucker & Makel, 2010). Guilford and his colleagues developed several other DT tasks (listed chronologically): Unusual Uses (Wilson, Guilford, Christensen, & Lewis, 1954), Consequences (Christen, Merrifield, & Guilford, 1958), and Alternate Uses (Guilford, Christensen, Merrifield, & Wilson, 1978). Others, including Getzels and Jackson (1962), Wallach and Kogan (1965), Cattell and Butcher (1968), Wallach and Wing, (1969), Williams (1980), and Urban and Jellen (1996), also developed Guilford-like tests.

All of these tests had their roots in Guilford's SOI model, and through them, researchers tried to build an understanding between psychometric approaches to intelligence and to creativity

(Kaufman et al., 2008). For example, Wallach and Kogan are supporters of a game-like, untimed administration of DT tasks. Their reasoning is that a game-like environment allows creativity to be measured as distinct from intelligence. Wallach and Kogan (1965) stated that “a frame of reference which is relatively free from the coercion of time limits and relatively free from the stress of knowing that one’s behavior is under close evaluation” (p. 24) helps participants demonstrate an authentic, creative performance. In this case, Wallach and Kogan’s findings overlapped with Getzels and Jackson’s findings that DT tests should be administered in non-test-like conditions.

All DT tests share at least one common feature that is they are open-ended. By design, there is not one single, right answer (Kaufman et al., 2008). As was mentioned earlier, DT tasks are most often scored for the indices of fluency, flexibility, originality, and elaboration. However, all these indices and their qualities are operational. In other words, the indices do not capture the actual process of DT, but rather are useful estimates of creative potential (Runco, 2011). DT tasks must involve *divergence* in thinking. However, since there are only four indices, measurement of the actual thinking of the respondent is narrow. Brown (1989) addressed this gap and argued that operational definitions that only address the construct of creativity may appear cohesive but are not relevant to real world creativity. He stressed the importance of addressing creativity as a phenomenon, a sentient event (Brown, 1989).

The criterion problem is one of the biggest issues that DT tests face, like other creativity measures. Runco (2007) stated that some researchers use DT tests to predict inappropriate criteria. Therefore, he developed the Runco Ideational Behavior Scale (RIBS) as a response to DT tests’ criterion problem based on the belief that most appropriate criterion would be one that emphasizes ideation (i.e., the generation of ideas). As in a DT test, the RIBs implies that creative

participants show more creative behaviors and ideate more often than non-creatives in their daily lives. However, the RIBs is a self-report measure; a sample idea would be something like: “I think about ideas more often than most people” (Runco, 2007).

The RIBs brings its own drawbacks. For example, the administration of the RIBs might not be logistically feasible with all age groups, such as children. (Plucker & Makel, 2010). Another potential drawback is that, as a self-report instrument, there is some question as to the degree to which respondents are truthful and objective in their self-reporting. This particular disadvantage causes problems directly related to validity. According to social desirability bias studies, certain groups tend to under-report their achievements and abilities in self-report measures in order to be viewed favorably by others (Hébert, Peterson, Hurley, Stoddard, Cohen, Field, & Sorensen, 2001). Other groups may over-report or have mistaken ideas of what their behaviors are in reference to the norm. However, self-reports can also provide advantages. For example, they are easily administered and are great assessments of how the respondents view the topics.

The reliability of DT tests is fairly convincing (See, for example, Torrance, 1981; Torrance, Khatena, & Cunnington, 1973). However, there are some concerns about the influence of fluency on other scores, especially originality. It has been clearly shown that fluency contaminates originality (Clark & Mirels, 1970; Hocevar & Michael, 1979). In other words, fluency is acting as a factor in DT tests’ scoring, especially in originality scores. Runco and Albert (1985) explained the contaminating effect of fluency and argued that this is not always the case; they asserted that originality is reliable, even after controlling for fluency. Runco and Albert (1985) found evidence of reliability after removing the fluency effects on nonverbal tasks.

Additionally, Hocevar and Michael (1979) found evidence of reliability in originality and flexibility scores after controlling for the fluency effects in other DT scores.

Invariably, validity is a concern in DT tests. Validity basically asks, “Are you measuring what you think you are measuring?” From the validity perspective, DT tests face discriminant and predictive validity issues. For example, there is an obvious relationship between creativity and intelligence on some level. However, intelligence alone is not sufficient to explain creative performance (Torrance, 1966). As mentioned earlier, Wallach and Kogan (1965) emphasized a game-like environment to administer DT tests in order to achieve maximum efficiency and authenticity. This method is entirely different from the standardized test approach used in the administration of traditional psychometric IQ tests. The testing environment is important in supporting the discriminant validity of DT tests. Creativity tests should be distinguished from IQ tests in order to show the main characteristics of creativity as separate from a measure of intelligence. In light of these facts, DT tests do not have major issues with discriminant validity unless Convergent Thinking (CT) tests and DT tests are administered together. CT tests do focus on one right answer and are closely related to IQ tests. However, DT and CT represent two necessary phases of creative problem solving, and they are described as two ends of a continuum (Eysenck, 2003). For example, problem solving requires both divergent and convergent thinking. Also, Runco (2014) opined that, in a natural environment, it is unusual to find a problem that completely relies on one or the other. Therefore, it can be concluded that CT and DT might work together, causing discriminate validity problems between DT and IQ.

Predictive validity is an additional concern for researchers using DT tests (Baer, 1993, 1994; Kogan & Pankove, 1974). Predictive validity in DT research shows the strength of the relationship between DT tests and defined criteria. However, creativity is presented as a complex

construct or syndrome. This signifies that no one single measure can explain creative potential. Therefore, it is important to define criteria by considering the multifaceted nature and complex attributes of creativity. In other words, DT tests are only a predictor for creative potential; it is difficult to use only one method to explain creativity. The RIBs is not the only solution that addressed this criterion problem. For example, Kogan and Pankove (1974) used extracurricular activities and accomplishments as criteria, and Torrance (2002) used activities and accomplishments as well as creative lifestyle as criteria to assess creative potential.

However, Runco, Plucker, and Lim (2001) stated that previous assessments of the predictive validity of DT tests were ill-conceived. In other words, the criteria used previously were not always appropriate criteria for DT. Runco et al. (2001) opined that better predictive validities would be found if the criteria relied solely on ideation and not on opportunities and extra-cognitive factors.

Traditional methods to score DT tests are very straightforward. DT tests are often scored according to the aforementioned four criteria of fluency, originality, flexibility, and elaboration. On the other hand, the TTCT-Figural has additional measures and excludes flexibility because of flexibility's high correlation with fluency (Torrance & Ball, 1984). However, these four criteria are highly correlated and require labor to parse out (Runco & Acar, 2012). Hence, Wallach and Kogan (1965) recommended using only two criteria: uniqueness and fluency. Even this reduction to only two criteria will still require considerable work in scoring and in providing inter-rater reliability (Runco & Acar, 2012).

Like the many definitions of creativity, there are many ways to assess DT and what it indicates for creativity and for creative potential. These researchers have all attempted to isolate the features of creative potential in order to better predict creative behavior.

Torrance Tests of Creative Thinking

Explaining creativity in terms of DT is the most commonly used method for studying creativity (Hocevar, 1981). As previously mentioned, the Torrance Tests of Creative Thinking (TTCT) are the most used DT tests (Plucker & Makel, 2010). The TTCT were developed by E. Paul Torrance with the objective of having a valid and reliable test to be administered to individuals from kindergarten through adulthood (Torrance, 2008). The administration and re-norming of the TCTT have been refined several times according to the current thinking in the field in 1974, 1984, 1990, 1998, and 2008 respectively. This updating of the norms perhaps explains why the TTCT are the most widely used tests of DT (Plucker & Makel, 2010). Creativity can be measured as a general way of thinking, and the TTCT have the highest predictive validity in comparison to other creativity measures (Kim, 2011). Furthermore, Torrance's longitudinal study showed that the TTCT are better predictors of creative achievement than IQ measures (Kim, 2011). In many states, the TTCT are used as selective test to qualify for gifted and talented programs in education.

The TTCT have two different forms: The TTCT-Verbal and the TTCT-Figural. The TTCT-Verbal consists of two different forms, A and B, including six activities each: ask-and-guess (ask questions and guess causes), product improvement, unusual uses, unusual questions, and just suppose. In these activities, test takers respond to prompts in each task by writing (Torrance, 1974). The TTCT's figural portion also has two parallel forms, A and B, including three activities each: picture construction, picture completion, and repeated figures of lines or circles (Torrance, 1974). Torrance revised the TTCT and scoring measures several times. In 1984, Torrance added two norm-referenced measures of creative potential and 13 criterion-referenced measures in the TTCT's scoring (Hébert, Cramond, Neumeister, Millar, & Silvian,

2002). Additionally, Torrance eliminated the measure of flexibility (number of connections among different areas and ideas) in the TTCT's figural 1984 version because it correlated very highly with fluency (number of ideas) (Hébert et al., 2002). These revisions aimed to enhance the TTCT's measure of all the different manifestations of creativity that Torrance observed (Hébert et al., 2002). These new subscales helped to predict creative achievement and improved the TTCT's validity (Ball & Torrance, 1984).

Like Wallach and Kogan (1965) and Getzels and Jackson (1962), Torrance (1966) highlighted the importance of creating a game-like, stress-free environment for participants in order to get the best results from the test. Accordingly, participants should be motivated and encouraged to “have fun”. Also, the environment should encourage them to utilize their real potential without any external or internal distraction.

Torrance (1966) developed the TTCT in order to better understand and support the different ways in which children express their creativity and to map out an individualized teaching method for children who have different instructional needs. Therefore, these tests aimed to be a tool for the enhancement of creativity Hébert et al., 2002). Torrance (1966, 1974) suggested the following uses for the tests:

1. To understand how the human mind works and functions and its development;
2. To discover effective rationale for individualizing instruction;
3. To provide clues for remedial and psychotherapeutic programs;
4. To evaluate the effects of educational programs, materials, curricula, and teaching procedures;
5. To learn the significance of latent potentialities.

The main advantages of the TTCT are their ease of application and use, large norming longitudinal samples (Davis, 1997), and predictive validity (Kim, 2006; 2008). The sensitivity of the structure of the TTCT allows for children from low SES and minority groups to be identified as gifted (Cramond, 1994; Torrance, 1977).

Several critics have found fault with the TTCT. For example, the TTCT used the same participant responses to derive multiple scores, which created scores that come with a high number of inter-correlations (Heausler & Thompson, 1988). These critics contended that the lack of consistent construct validity in the TTCT comes from this inter-correlation issue. Therefore, these response sets need to be reevaluated (Plucker & Makel, 2010). Baer (2009) noted the TTCT's low correlation ($r = .06$) between performance on the TTCT-Verbal and TTCT-Figural. This low correlation disproves the domain generality of creativity from the TTCT (Baer, 2009). According to Baer, this situation also conflicts with Torrance's claim that the TTCT-Verbal and -Figural are a complete battery to measure creative potential and different cognitive abilities (Baer, 2009).

Cramond (1994) and Kim (2011) addressed the critics of the TTCT. Cramond (1994) criticized Baer for excluding the TTCT's longitudinal studies that showed a substantial amount of validity. And Kim (2011) elaborated on Torrance's reasoning in the study. The main point that Baer missed is that Torrance developed these two forms of the TTCT in order to measure different aspects of creative potential (Torrance, 1966). The TTCT also refers to the domain specificity of creativity. Therefore, the TTCT's emphasis is on different cognitive abilities and is not proposed to produce correlated achievement scores because these are different scales. Furthermore, it does not follow that the TTCT are not a complete battery of tests; it only means that these two TTCT-Verbal and -Figural forms measure different aspects of creative potential.

Remote Associates Test

The Remote Associates Test (RAT) was developed by Sarnoff Mednick in 1962. Mednick's (1962) main purpose was to deal with the problem of discerning creative products: differentiating between the actually creative and the merely unusual (Brown, 1989).

The RAT was developed in order to measure the ability to make remote associations and to evaluate their appropriateness (Brown, 1989). The RAT relies on an associative theory framework. Mednick hypothesized that CT is necessary in creativity to ensure that the actual creativeness of a product meets foreseen requirements. Therefore, Mednick took CT into consideration when evaluating creative products (Brown, 1989).

Mednick (1968) pointed out the importance of forming new combinations from mutually distant associative elements for creative thinking and that these new combinations should be useful, not based on mere whimsy.

The RAT consists of 30 items to be responded to within 40 minutes. Every item includes three stimuli words, and the task is to find a fourth word that connects with the three stimuli words (Mednick & Mednick, 1967). The structure of the RAT is completely different from the traditional DT tests with its convergent structure. In contrast, the RAT is very useful in finding the distance between two ideas in terms of associative theory.

The RAT has very promising reliability data (Mednick, 1968); however, the evidence for validity in the RAT is lacking (Fasko, 1999). According to Mednick's (1962) report, the RAT scores are negatively correlated with school grades and some intelligence scores. The RAT scores were high for those people who were more flexible and held atypical views (Kaufman et al., 2008).

Even though the RAT is no longer published, it is still used in the field. Taft and Rossiter (1966) showed that the RAT is largely in line with convergent thinking assessments. In 2003, Bowden and Jung-Beeman revised the norms, and the RAT continues to be used to measure abilities outside of creative talents.

Consensual Assessment Technique

The CAT was developed by Amabile (1982) and presented a clever solution for the criterion problem in creativity assessment (Plucker & Makel, 2010). The CAT is often used with other products and tends to have sufficient reliability ratings (inter-rater reliability ranges from .70 to .90) (Runco, 2010a). The CAT can be described as the “gold standard” because it takes actual creative performances into account, and it is not connected to any specific theory of creativity. It mimics the way creativity is measured in the “real world” (Sternberg & Kaufman, 2010). The CAT is structured on the idea that the best measure of the creativity of a work is the combined assessment of experts in that field (Kaufman et al., 2008). Additionally, the CAT is a domain specific assessment (Runco, 2010a); therefore, it allows comparisons within a group of artifacts.

MacKinnon (1978) is one of the researchers who emphasized the importance of creative products and mentioned that “the starting point, indeed the bedrock of all studies of creativity, is an analysis of creative products, a determination of what it is that makes them different from more mundane products” (p. 187). There are many other researchers who also valued the importance of creative products like MacKinnon (see, for example, Guilford, 1967; Wallach, 1976), and this emphasis on creative products was interpreted as a response to the perceived need for external criteria to which researchers could compare other methods of assessing creativity in order to help establish evidence of validity (Kaufman et al., 2008). However, Runco (1989)

pointed out the potential problem of evaluating creative products. The measurement of creative products may involve the inconsistent psychometric quality of DT. Runco (1989) argued that products might reflect specific talents, rather than a global expression of creativity.

Amabile (1982) stated, “A product or response is creative to the extent that appropriate observers independently agree it is creative” (p. 1001). This means that experts know what is and what is not creative. Hence, using experts to evaluate a product’s creativity should theoretically eliminate the criterion problem. However, the idea of using experts comes with other considerations. One such issue is determining the necessary level of expertise, and that depends on a variety of factors: the target domain, the purpose of the assessments, and the skill of the subjects (Amabile, 1996; Runco, McCarthy, & Svensen 1994; Runco & Smith, 1992). Amabile (1996) recommended determining appropriate judges by relying on the experts’ training about and knowledge of the target domain. Amabile (1996) advised that an expert should have “at least some formal training and experience in the target domain” (p.73).

In considering the difficulty of using expert raters in creativity assessment, some researchers attempted to replace expert raters with non-expert raters. However, using novice or peer raters is problematic because it cannot give the construct validity, like the CAT claims, and expert and novice raters tend not to agree closely enough on their ratings (Kaufman et al., 2008). In general, expert and novice judges produce different ratings of product creativity although the domain in which the product is created impacts the degree to which the groups’ ratings overlap (Plucker & Makel, 2010). However, expert and novice judges have shown a high correlation when evaluating the creativity of short stories (Kaufman, Baer, & Cole, 2009), but not in ratings for poetry (Kaufman, Baer, Cole, & Sexton, 2008).

The CAT shows a little domain generality. Some researchers argued that if creativity varies by domain, a single creativity score does not make sense (Kaufman et al., 2008). However, traditional creativity theories view creativity as a general trait or set of skills that can be applied to any field (Kaufman et al., 2008). Domain generality allows one to make projections in creativity related skills. For example, in DT tests, creative activity checklists assume that creativity is more a domain general trait or set of skills (Plucker & Runco, 1998).

In light of these facts, the CAT rating can only compare one person's creativity on a particular task to the creativity of another on that same task. Therefore, the CAT does not allow us to compare students' creativity more generally. The within-group comparison for the CAT is not appropriate for individual testing and does not provide a standardized test score (Kaufman et al., 2008).

The CAT has an acceptable inter-rater reliability score that ranges from .72 to .93 using the Spearman-Brown prediction formula (Amabile, 1983). Amabile (1996) found a different range in her later work when she used Cronbach's coefficient alpha and the intra-class correlation method. Amabile (1996) reported that the number of judges has a direct effect on the inter-rater reliability correlations. According to this report, more judges bring a higher inter-rater reliability correlation (Amabile, 1996). Also, Amabile (1996) mentioned that five to ten experts represent a sufficiently large group. However, having fewer than five experts as raters risks having a low level of inter-rater reliability (Amabile, 1996). Due to financial and logistical obstacles, this may be a challenge for many local school districts to find five or more experts.

The CAT measures what it appears to measure (Kaufman et al., 2008); in other words, the CAT has strong face validity. However, finding the evidence for face validity is empirically limiting (Kaufman & Sternberg, 2010). Also, the CAT shows strong construct validity in the

particular domain in which creativity is measured (e.g., technical correctness, aesthetic appeal) (Kaufman et al., 2008). However, lack of predictive validity is still a problem for the CAT. Many CAT projects are marked by this lack of predictive validity (Kaufman et al., 2008). On the other hand, many researchers maintain that CAT scores can predict a particular ability; for instance, a CAT score in poetry writing would predict writing creative poetry – but nothing else (Kaufman et al., 2008). CAT ratings have been shown to remain stable (Baer, 1994), and it is possible to investigate whether past creative work can predict later creative work. In an initial study of Mozart's life, Kozbelt (2005) indicated that creativity prediction is possible. However, CAT scores would not be expected to predict other important outcomes.

In conclusion, CAT is a subjective way to evaluate products' creativity in contrast to DT tests' objective methodology. The following section will address these subjective and objective methods in creativity measurement.

Objectives of the Current Studies

The aforementioned studies reviewed the current issues in creativity measurement and tried to bring new perspectives by addressing most of the essential concerns. The studies placed objective and subjective scoring concerns at the core and built new methods to explain the complex structure of creativity measurement by building on existing research. The studies took the most familiar positive and negative aspects of creativity measurement into consideration in order to bring stronger and more applicable theories into the field of creativity studies.

In addition to the objectivity and subjectivity concerns of creativity measurements, the proposed studies offered new approaches to creativity measurement. The first study provided an extensive explanation about the role of flexibility in terms of semantic distance to explain creative thinking. The second study revealed key words and phrases that people use in order to

introduce original ideas. The third study demonstrated how to make non-expert creativity judges better through the provision of EIs.

Briefly, the proposed studies addressed the following questions:

1. How is semantic distance related to traditional flexibility scores?
2. How can a written work's originality be scored by looking for particular words or phrases that signal original ideas in authors' writings?
3. How can EIs make creativity judges better at creativity evaluation?

These three studies are presented in order to bring new insights to the field and new methods to the administration of creativity measurements. First, these studies can lead to developing new and better ways to measure creativity by considering the comparison of results obtained from subjective and objective scoring. Second, the first study revealed the significant role of flexibility in creative thinking. This can help build the associational process of DT on a stronger theoretical and methodological foundation. Third, a keywords method can change the approach to measuring creativity by doing it in a more practical way. Finally, these studies can help researchers overcome the many problems that existing measures do not address.

CHAPTER 2
THE ASSOCIATIVE PROCESS IN DIVERGENT THINKING
A SPECIAL CASE: FLEXIBILITY¹

¹ Turkman, B. & Runco, M.A. To be submitted to the *Journal of Creative Behavior*.

Abstract

The first study investigated the associative process in DT by focusing on how semantic distance relates to traditional flexibility scores. Originality and flexibility are shown to be more closely tied to creativity than are other DT indices (Runco & Albert, 1985). The first study highlighted the importance of flexibility in semantic distance in order to explain creative thinking. Utilizing both computer algorithms and human scoring to measure semantic distance, the first study helped to clarify the relationship between objective and subjective methods. A similar study (Acar & Runco, 2014) investigated originality and fluency in the same manner; this study added flexibility to that discussion.

In order to investigate the associative process in DT, this study focused on the semantic distance between the DT questions or stimuli and the responses given. These tasks revealed some remote and some close associations to given stimuli. These associations were identified by using seven human judges and by applying three different sources of social associations: WordNet (WN), Word Associations Network (WAN), and IdeaFisher (IF). These sources provided lists of associations, and they served as determining criteria for distinguishing an idea's semantic distance from a stimulus.

The results showed that remote and close associations could be reliably measured when human judges and different social association sources were applied. Inter-rater reliability values among human judges were higher than the three sources of social associations. Inter-item reliability values for remote associations were higher than that for close associations. Statistical analyses showed that traditional flexibility and originality scores, as well as the Idea Density (ID) scores of the participants, were significantly related with remote, but not with close, associations across all human judges and all computer methods. Overall, subjective methods

showed higher inter-item and inter-rater reliabilities and correlations with remote associations and ID values than did computer methods.

Introduction

Mednick's (1962) seminal work, "The associative basis of the creative process," played a critical role in creativity literature leading to an understanding of the associative process in creative thinking. Mednick's central assumption was that creative individuals demonstrate a flatter associative hierarchy (less skewed distribution of association strengths with more subtle differences between different associations) that comes with the production of original ideas. In light of this fact, Mednick developed the Remote Associates Test (RAT) in order to measure the ability to make remote associations among ideas and evaluate their appropriateness (Mednick, 1962). Mednick (1968) pointed out the significance of forming mutually distant associative elements into new combinations in creative thinking and that these new combinations should be useful.

Like Mednick's RAT work, Torrance worked on a concept he called "mental leaps" to help explain the process of making associations between different cognitive areas. Torrance (1995) noted the ability to make mental leaps as an important aspect of DT. Torrance, Ball, and Safter (1992) connected the concept of mental leaps to the resistance of premature closure in order to produce ideas without reaching hasty conclusions. Similarly, Perkins' (1994) recognized that a mental leap is a cognitive process that takes place on a conceptual landscape. Therefore, making mental leaps within a certain category can enhance the cognitive processes involved in creativity.

Like RAT, the Word Association Test (WAT) was also developed to measure associational thinking (Gough, 1976). Respondents rate the associations of given concepts or

word pairs with each other in this method. Subsequently, these ratings are compared with normative data on associative strength. This method requires one single answer like the RAT.

Acar and Runco (2014) argued that open-ended questions create a suitable place to observe the associative process in creative thinking. Wallach and Kogan (1965) also highlighted the (interdependent) relationship between DT and the associative process. In other words, DT tests are valid and reliable measures to assess associative abilities and ideational capacity. DT tests positively link the ability to think about the stimuli and to respond in ways that vary in distance from those stimuli.

DT allows associations that may be remote or close. Remoteness of associations is critical because it is related to both originality and appropriateness of ideas. Stein (1953) proposed a standard definition of creativity that took into account originality and effectiveness to explain the complex structure of creativity. This definition highlighted that originality alone is not enough to be considered creative. An original idea must also be effective or useful to be creative. Therefore, identification of the distance between these ideational pathways in the cognitive process is important to understand. In other words, a very far distance of ideational pathway might represent an unrelated and inappropriate solution to the given task. Acar and Runco (2014) illustrated this relationship in their study as well.

DT's open-ended quality helps researchers study the associative distance between the stimuli and the idea. The father of the associative process in creativity, Mednick (1962), saw creativity as an ability to combine associated elements into new forms. Also, the distance between ideas was linked to a creative potential level. Being able to make more remote associations indicated more creative attitudes than making close associations. Combining elements from different disciplines leads to more creative ideas. This remote association ability

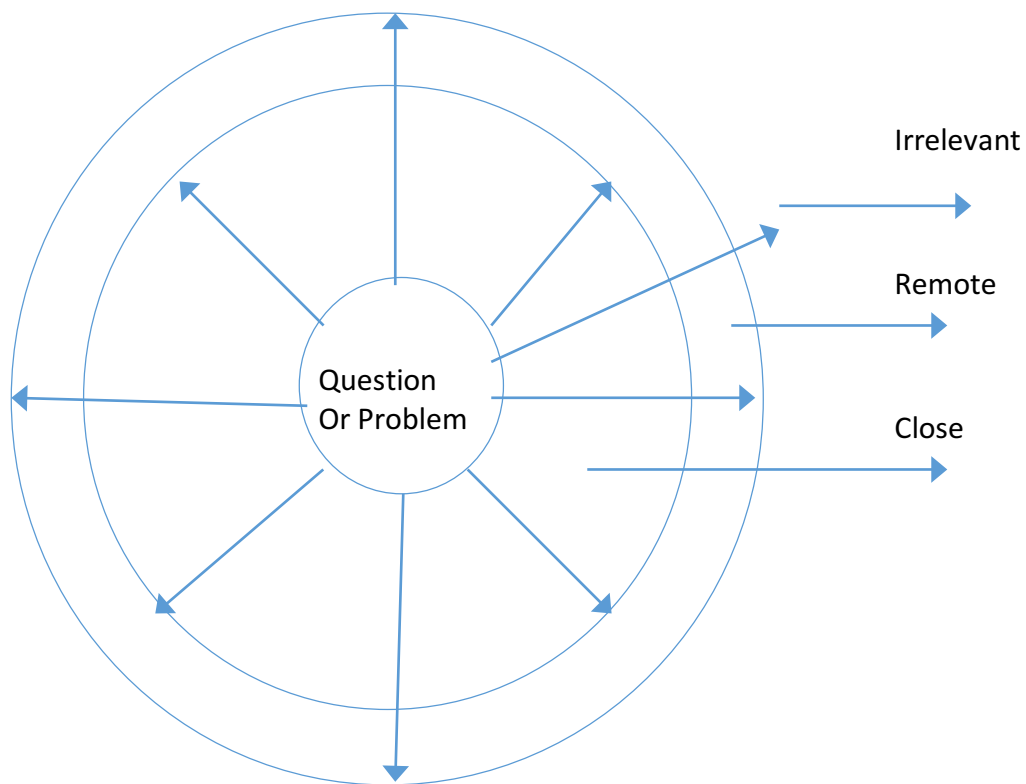
can be exemplified by Shakespeare's 'eye' as a degree of seeing what is not obvious and creating new meaning.

As mentioned earlier, usefulness and originality are two key factors needed to deem a product creative. When we look at this concept from the associative distance perspective, it means that the ideas should be far enough to be original, but not too far as to not be useful or relevant. In figure 1.1, the area outside of the circle represents irrelevant ideas. Therefore, an idea should be within the circle to be useful and relevant. The aforementioned Perkins (1994) and Torrance (1995) mental leap research highlighted the importance of this topic. The distance of the mental leap should be within an optimal range to count as a creative idea or product. For example, the TTCT do not take unrelated responses into consideration. The TTCT first evaluate the relevancy of responses to the given stimuli, thus eliminating more irrelevant associations from considerations of creativity.

Many other studies attempted to investigate the associative process in creative thinking. Benedek, Könen, and Neubauer (2012) attempted to measure associative and DT processes differently. Their results indicated that half of the variability in DT tasks was explained by four associative thinking tasks. This study showed an intertwined relationship between the associative and DT models by applying two different methods. Therefore, Benedek et al. (2012) study comes with "method variance" discussions. In other words, the test format causes the variability among scores instead of individual differences in the targeted processes.

Acar and Runco's (2014) study is one of the very first studies that examined remote and close associations by using traditional instructions. Associational distance is difficult to operationalize. A similar study was conducted by Wilson, Guilford, and Christensen (1953), but the method differed from the Acar and Runco (2014) study. Wilson, et al. asked their participants

to generate six additional responses to elicit remote ideas. On the other hand, Acar and Runco (2014) predetermined the semantic distance to examine close and remote associations. Acar and Runco (2014) showed that the distance between the stimuli and responses can be defined and measured objectively.



*Figure 1: Theoretical demonstration of close, remote, and extremely remote (irrelevant) associations involved in DT. Reprinted from “Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal*, 26, 229-238”.*

Acar and Runco (2014) identified ideational distance through three different online sources of social associations. These online social association sources provided lists of

associations to use in the determination of the remoteness or closeness of an idea to a given stimulus. The results produced a good inter-item reliability for remote associations. Social associations indicate relationships between ideas or concepts that can be defined and explained in an objective or conventional way. Calculating the agreement of “lexical neighborhoods” of any two or more concepts is one of the ways to measure the semantic similarity between ideas (Rapp, 2002). Co-occurrences of concepts or words provide the values to calculate the agreements. For example, some words are associated with other words because they are often used together, such as "soup" and "bowl". These networks are suggestive of close and remote associations. Acar and Runco (2014) used these networks to define associative distance by using computer programs derived from artificial intelligence and linguistic studies. These studies have yielded lexical databases that are used in various fields for different purposes. WordNet, Word Associations Network, and Idea Fisher are some examples of these social associations networks.

The first instrument that was used in this first study is WordNet (WN). This database has served researchers from various fields. For example, Bossomaier, Harre, Knittel, and Snyder (2009) applied the WN lexical database to compute their participants’ “creative quotient” (CQ) from DT tasks. Paletz, Peng, and Li (2011) also used the WN to cluster semantically related terms. Snyder, Mitchell, Bossomaier, and Pallier (2004) applied the WN to compare two response sets from two different individuals and assigned higher scores to given responses from different categories than to those from similar categories.

The second instrument used in this first study is the Word Associations Network (WAN). This associative network clusters words by part of speech: nouns, verbs, adverbs, and adjectives. The WAN’s database contains 3 million associations that are provided for 41,476 word concepts. The WAN takes the typical thinking strategies of the human brain into account and compiles

associations that arise in human memory for any given word.

The third instrument used in this study is the Idea Fisher (IF). The IF tries to unveil original ideas that are already stored in human cognition through successive chains of associations. It aims to retrieve existing schemes and concepts from human cognition and establish new relationships (Fisher, 1996). The IF includes around 700,000 associations and provides a variety of associations for each word.

This study built on Acar and Runco (2014) by focusing on how semantic distance is related to traditional flexibility scores. Runco and Albert (1985) opined that originality and flexibility are more closely tied to creativity than other DT indices, and yet, Acar and Runco's (2014) study did not take flexibility into consideration in order to explain the concept of semantic distance in relation to creativity. Additionally, Acar and Runco (2014) used computerized methods exclusively in their study. Thusly, there was no subjective evaluation of the ideas. This first study used both computer methods and human scoring to analyze how subjective and objective scoring methods identify semantic distance.

Research Questions

The present study aimed to answer the following research questions:

1. How is semantic distance related to traditional flexibility scoring?
2. How do computer and human scoring methods project the process for semantic distance?
3. How are flexibility and fluency scores related to remote and close associations? Is one of them more reliable in explaining the associative process?

Method

Sample

The sample included 49 university students majoring in various fields. Most of the participants ($n = 40$) majored in humanities, social sciences, or art; the remaining ($n = 9$) majored in the biological sciences. Thirty were female, and 19 were male. Their ages ranged from 23-63 years ($M = 27.4$; $SD = 5.53$), and the average GPA was 3.57 on a 4-point scale ($SD = .28$).

Data Collection

Several professors and student organizations were contacted in order to invite their students and members to take part in the study. After securing these participants, individual and group appointments were scheduled to administer the tasks.

Instruments

DT tasks were administered with the demographic information form. The DT tasks used in this study consisted of three items from the Many Uses Game (i.e., “List different uses for a bowl”, then, “tire”, and then “nail.”) from the Runco Creativity Assessment Battery (rCAB, 2011, creativitytestingservices.com). See the appendix a for the many uses game.

Participants were given the following instructions: “*List as many ideas as possible about different uses of for a bowl, a tire, and a nail. Your answers will not be judged or questioned. Have fun!*” They did not have a time limit to complete the DT tasks. Then, they were asked to write a creative essay about any topic they wanted. They were given 15 minutes to finish their essays.

Procedure

Seven human judges and three sources of social associations networks were used to score participants' answers to the DT tasks. The scoring was categorical (remote = 1; close = 0) for all these methods. The objective scoring lists derived from the IF included 444 words or concepts for bowl, 430 for nail, and 253 for tire. The list from WAN included 315 words or concepts for bowl, 380 for nail, and 260 for tire. Finally, the list derived from WN included 709 words or concepts for bowl, 493 for nail, and 123 for tire.

Word associations for WN can be obtained via online or by installing the software. WN provides all sorts of associations, such as hypernymy (i.e., superordinate), meronymy (i.e., part of a whole), and sister term. All words provided by the WN were included in the list, except sample sentences and all prepositions and pronouns. The WAN list was retrieved from the website (<http://wordassociations.net>) and includes all nouns, adjectives, adverbs, and verbs. The IF lists were obtained from the ThoughtOffice version 1.0.125 by using the "Word Associations" function. The IF provides all related word associations when you type in any search term. Newer versions of the IF provide a higher number of word associations.

Subjective scoring values for semantic distance were provided from seven human judges whose expertise is in creativity. These judges included doctoral students, faculty members and researchers in creativity. They read every participant's responses to the DT stimuli and scored them for semantic distance (remote = 1; close = 0). They did not have a criterion by which to score the responses. This is a CAT-level expertise technique that helped to reveal subjective semantic distance values.

The word association lists from IF, WAN, and WN were used to determine remote and close associations. A response was defined as a close associate if it was included in the

corresponding list. If there was not any match in the corresponding list, it was defined as a remote associate. A new computer algorithm was coded to screen the responses and to find matching words and concepts in the three word associations lists.

Human judges applied their own sole judgments and expertise in order to determine remote and close associations.

Additionally, in order to find each participant's essay ID value, the Computerized Propositional Idea Density Rater (CPIDR 5, pronounced "spider") was applied. CPIDR 5 is a computer program that determines the propositional idea density (ID) of an English text. Participants' essays were first converted to plain text format in order to analyze them with CPIDR 5.

Runco, Turkman, Acar, and Nural (in press) conducted three studies that focused on the relationship between literal characteristics and creativity. Both oral and written works were taken into consideration, and ID values for these oral and written works were calculated using CPIDR 5. ID is a method used to analyze concept frequencies in written or oral works. An ID value is estimated by the total number of verbs, adjectives, adverbs, prepositions, and conjunctions divided by the total number of words. This theory came from Kintsch (1974) and Turner and Greene's (1977) studies. According to Kintsch, English text tends to be constructed from propositional bases. "Propositions represent ideas and language expresses propositions, and hence ideas" (Kintsch, 1974, p. 5). ID, operationalized in terms of propositions, has proven to be useful in the linguistic research of Turner and Greene (1977) and Covington (2009). Runco et al. (in press) also investigated the usefulness of the ID method in terms of its relationship to DT, citation impact (CI), and eminence. Their study clearly showed that written or oral works having more ideas in them tend to be cited more frequently by other scholars, and the higher frequency

of ID is related to both reputational change obtained from encyclopedic entries and to Ludwig's Creative Achievement Scale (CAS).

Analyses

Inter-item and inter-rater reliability analyses were applied in order to investigate all indices and raters' reliability coefficient alpha. Also, Pearson product-moment correlations with the ID and the bivariate correlations among all indices including flexibility, originality, and fluency scores were analyzed. Paired samples *t*-tests were applied in order to discern any significant difference between remote and close association means. All these analyses were done using IBM SPSS 23.0.

Results

Forty-nine participants gave a total of 1136 responses for the bowl, nail, and tire stimuli: 426 for bowl, 333 for nail, and 377 for tire. Descriptive statistics of the three verbal DT tasks are provided in Table 2.1.

Statistical analyses indicated that the alpha coefficient was .88 for the three fluency scores (bowl, nail, and tire).

Table 2.1

Descriptive Statistics for the Divergent Thinking Tasks

	Min.	Max.	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Bowl	1	21	8.12	4.362	0.930	0.755
Tire	1	17	6.41	3.741	0.786	0.125
Nail	2	20	7.29	3.253	1.497	3.672

The flexibility values on the DT tasks were scored by two experts independently. The experts categorized participants' responses into various areas in order to see the number of connections between ideas that made by the participants. The categories that were created by experts are as follows: entertainment, art, protection, furniture, gun, protection, science fiction, home, sport, game, tool, item, clothing, science, maintenance, container, personal tool, nest, social, support, medical, shelter, and weight. The participants' flexibility values were scored by the experts who looked at the diversity of ideas in the aforementioned categories. The bivariate correlation revealed that these two experts' scorings was highly correlated ($r = .98, p < .001$). Additionally, a paired t -test investigated the difference between these two experts scorings but the difference was not found to be significant ($t(48) = 2.26, p = .01, d = .38$).

Statistical analyses for the human and computer scoring are reported in Table 2.2 and Table 2.3 respectively. Descriptive statistics, reliability (alpha coefficient), correlations with ID, and a comparison of the remote and close associations were conducted for each method. Inter-rater reliability analyses were conducted for both the human and computer scoring methods. Statistical analyses indicated that inter-rater reliability (intra-class reliability alpha) for the human judges was .99 for remote associations and .98 for close associations. Also, the statistical analyses showed that the computer methods had high inter-rater reliability for remote and close associations: .96 and .92 for remote and close associations respectively. Additionally, the composite scores for all human judges and computer methods across three DT items were calculated to see the inter-rater reliability between these two different methods, and the alpha was .92 for remote and .44 for close associations.

Table 2.2
Descriptive Statistics for All indices Defined Using Seven Human Judges

		Bowl		Nail		Tire	
		Remote	Close	Remote	Close	Remote	Close
Judge 1	Min.	1	0	0	0	1	0
	Max.	15	11	14	9	19	3
	<i>M</i>	4.82	3.31	4.39	2.00	6.49	.80
	<i>SD</i>	3.51	2.74	3.39	1.83	3.16	.935
Judge 2	Min.	0	0	0	0	1	0
	Max.	14	11	15	8	18	5
	<i>M</i>	4.92	3.20	4.88	1.51	5.84	1.45
	<i>SD</i>	3.80	2.46	3.61	1.50	2.93	1.15
Judge 3	Min.	0	0	0	0	1	0
	Max.	15	9	15	8	14	7
	<i>M</i>	5.78	2.35	4.51	1.51	4.61	2.67
	<i>SD</i>	3.88	2.10	3.41	1.50	2.65	1.80
Judge 4	Min.	0	0	0	0	2	0
	Max.	14	8	16	8	19	2
	<i>M</i>	5.82	2.33	5.00	1.39	6.76	.53
	<i>SD</i>	3.48	2.08	3.62	1.61	3.15	.68
Judge 5	Min.	0	0	0	0	2	0
	Max.	14	11	16	7	16	4
	<i>M</i>	4.69	3.33	5.08	1.31	6.37	.92
	<i>SD</i>	3.48	2.41	3.59	1.47	2.76	1.07
Judge 6	Min.	0	0	0	0	2	0
	Max.	14	10	16	8	19	3
	<i>M</i>	5.57	2.61	4.94	1.45	6.69	.59
	<i>SD</i>	3.71	2.57	3.72	1.60	3.13	.762
Judge 7	Min.	1	0	0	0	2	0
	Max.	14	9	15	8	20	3
	<i>M</i>	5.43	2.69	4.90	1.49	6.55	.73
	<i>SD</i>	3.51	2.27	3.61	1.62	3.18	.861

A paired *t*-test investigated the difference between human and computer scoring methods, and the difference was determined to be significant for both remote and close associations: ($t(48) = 4.54, p < .001, d = 0.64$), ($t(48) = -11.23, p < .001, d = -1.6$) respectively. The correlations between human and computer scoring methods were ($r = .91, p < .001$) for remote associations and ($r = .53, p < .001$) for close associations.

Table 2.3

Descriptive Statistics for All Indices Defined Using the IF, WAN, and WN

		Bowl		Nail		Tire	
		Remote	Close	Remote	Close	Remote	Close
IF	Min.	1	0	1	0	0	0
	Max.	14	17	9	9	15	5
	<i>M</i>	4.84	3.29	3.96	2.43	5.57	1.71
	<i>SD</i>	3.09	2.72	2.41	2.15	2.85	1.22
WAN	Min.	0	0	0	0	1	0
	Max.	14	8	8	10	14	6
	<i>M</i>	5.10	3.02	3.20	3.18	4.61	2.67
	<i>SD</i>	3.31	2.26	2.38	2.35	2.54	1.53
WN	Min.	0	0	0	0	1	0
	Max.	15	16	7	17	16	4
	<i>M</i>	5.12	3.00	2.22	4.16	6.18	1.10
	<i>SD</i>	2.99	2.80	1.91	3.03	2.75	1.08

Note: IF= IdeaFisher; WAN= Word Associations Network; WN= WordNet

Human Judges

Alpha coefficients were calculated for close and remote associations for each judge.

Alpha reliabilities of the three DT items are provided in Table 2.4 for each judge.

Composite scores were generated across three DT tasks as the average score in order to conduct further analyses. A paired *t*-test investigated the difference between remote and close associations, and the difference was found to be significant ($t(48) = 8.18, p < .001, d = 1.17$).

Table 2.4

Alpha Coefficients for Close and Remote Associations for Each Judge

	Close	Remote
Judge 1	.66	.85
Judge 2	.72	.86
Judge 3	.66	.82
Judge 4	.60	.87
Judge 5	.71	.86
Judge 6	.70	.86
Judge 7	.66	.85

Bivariate correlations were calculated to see the relationship among ID, traditional flexibility and fluency scores, and remote and close associations. The results indicated that ID was highly correlated with flexibility ($r = .31, p < .05$), fluency ($r = .30, p < .05$), and remote associations ($r = .36, p < .05$), but not with close associations ($r = -.03, p = .83$). All bivariate correlations for each judge are provided in Table 2.5 and Table 2.6.

Table 2.5
Bivariate Correlations Between ID and Composite DT Scores for Remote Associations

	ID	1	2	3	4	5	6	7	8
1. Fluency	.30*								
2. Flexibility	.31*	.90**							
3. Judge 1	.36*	.86**	.81**						
4. Judge 2	.38**	.80**	.77**	.97**					
5. Judge 3	.35*	.73**	.67**	.89**	.87**				
6. Judge 4	.35*	.89**	.82**	.97**	.94**	.86**			
7. Judge 5	.34*	.84**	.82**	.97**	.97**	.89**	.97**		
8. Judge 6	.36*	.86**	.82**	.98**	.97**	.87**	.99**	.98**	
9. Judge 7	.36*	.86**	.81**	.98**	.97**	.88**	.98**	.98**	.99**

Note: ID= Idea Density; 1= Fluency; 2= Flexibility; 3-8= Judges from 1 to 8

* $p < .05$ level (2-tailed)

** $p < .01$ level (2-tailed)

The bivariate correlations clearly showed that remote associations are related positively to ID, fluency, and flexibility. Also, the bivariate correlations between fluency, flexibility, ID, and the scorings of the judges were higher than for close associations. Total fluency scores were significantly related with both close ($r = .40$ to $r = .55$) and remote ($r = .73$ to $r = .90$) associations. However, the flexibility did not follow the same pattern. Traditional flexibility scores were related with remote associations only, not with close associations.

Table 2.6
Bivariate Correlations Between ID and Composite DT Scores for Close Associations

	ID	1	2	3	4	5	6	7	8
1. Fluency	.30*								
2. Flexibility	.31*	.90**							
3. Judge 1	.01	.50**	.23						
4. Judge 2	.08	.44**	.18	.93**					
5. Judge 3	.01	.52**	.27	.85**	.87**				
6. Judge 4	.02	.47**	.20	.89**	.85**	.82**			
7. Judge 5	.02	.55**	.27	.93**	.90**	.88**	.86**		
8. Judge 6	.05	.40**	.10	.94**	.90**	.86**	.92**	.91**	
9. Judge 7	.05	.48**	.23	.92**	.87**	.84**	.91**	.90**	.94**

Note: ID= Idea Density; 1= Fluency; 2= Flexibility; 3-8= Judges from 1 to 8

* $p < .05$ level (2-tailed)

** $p < .01$ level (2-tailed)

Computer Scoring (Objective Methods)

Alpha coefficients were calculated for close and remote associations for each method. Alpha reliabilities of the three DT items are provided in Table 2.7 for each source. Composite scores were generated in order to conduct further analyses. A paired t -test investigated the difference between remote and close associations, and the difference was significant ($t(48) = -7.97, p < .001$).

Table 2.7
Alpha Coefficients for Close and Remote Associations for Each Source

	Close	Remote
IF	.61	.79
WAN	.65	.77
WN	.42	.69

Note: IF= IdeaFisher; WAN= Word Associations Network; WN= WordNet

Bivariate correlations were calculated to see the relationship among ID, traditional flexibility and fluency scores, and remote and close associations. The results indicated that ID was highly correlated with flexibility ($r = .31, p < .05$), fluency ($r = .30, p < .05$), remote

associations ($r = .30, p < .05$), but not with close associations ($r = -.03, p = .83$). All bivariate correlations for each method are provided in Table 2.8.

Table 2.8
Bivariate Correlations Between ID and Composite DT Scores for Close & Remote Associations

	ID	1	2	3	4	5	6	7
1. Fluency	.30*							
2. Flexibility	.31*	.90**						
3. IF Close	.22	.81**	.69**					
4. IF Remote	.29*	.85**	.85**	.50**				
5. WAN Close	.14	.82**	.63**	.76**	.69**			
6. WAN Remote	.35*	.91**	.91**	.67**	.89**	.54**		
7. WN Close	.36*	.91**	.80**	.83**	.77**	.79**	.81**	
8. WN Remote	.21	.86**	.86**	.68**	.91**	.74**	.89**	.71**

Note: IF= IdeaFisher; WAN= Word Associations Network; WN= WordNet

* $p < .05$ level (2-tailed)

** $p < .01$ level (2-tailed)

The bivariate correlations clearly showed that remote associations are related to ID, except WN social associations. Also, the bivariate correlations for remote associations have higher fluency, flexibility, and ID values than close associations, except for WN social associations. Total fluency scores were significantly related with both close and remote associations. The flexibility followed a different way in computer methods. Traditional flexibility scores were only related with remote associations in human scoring methods, but it did not follow the same pattern with computer scoring method. In the computer scoring method flexibility was significantly correlated with both close and remote associations.

Discussion

Subjectivity and objectivity in DT has been addressed by many researchers (Christensen, Guilford, & Wilson, 1957; Silvia, 2008; Acar & Runco 2014; Benedik et al., 2012). So, the focus of this first study was to take both subjective and objective methods into consideration to explain

the associative process from different perspectives by particularly focusing on traditional flexibility scores. Acar and Runco's (2014) social association approaches brought an objective perspective to the associative process concept. This study built on that study to investigate subjectivity and the role of flexibility in associative thinking in order to more thoroughly explain creative thinking.

In order to see merits of these two objective and subjective methods, a variety of statistical analyses were conducted. Inter-rater reliability values showed that human scoring had higher values than computer methods for both close and remote associations (.99 vs. .96 for remote associations and .98 vs. .92 for close associations). Also, composite scores were generated to calculate inter-rater reliability between human and computer methods, and the alphas were .92 and .44 for remote and close associations respectively. Paired *t*-tests showed significant difference between these two methods ($t(48) = 4.54, p < .001$) and ($t(48) = 11.23, p < .001$) for remote and close associations respectively. These results clearly showed that the human scoring method had higher alpha values than the computer scoring method. Paired *t*-tests also found a significant difference between these two methods.

Alpha coefficient indices defined in this first study ranged from .66 to .87 for human judges and from .42 to .79 for computer methods. This finding clearly showed that human scoring had higher alpha values than computer methods. These results indicated that remote and close associations can be reliably measured when human scoring and computer methods were applied, except for the WN. The WN had much lower scores than other social associations networks. When we excluded the WN from our results, the alpha values ranged from .61 to .79 instead of from .42 to .79. The WN did not follow a reliable pattern like other computer and human scoring methods did in this study. Acar and Runco's (2014) study found a similar pattern

with the WN. In their study, the WN had the lowest alpha coefficient. Overall, human scoring methods had higher alpha values than computer methods.

These findings indicated that human scoring method is more reliable than computer methods. The computer methods showed the greatest variability. This finding indicates that better close and remote associations are needed. However, when the WN was excluded from consideration, there was no discrepancy. Therefore, the WN needs to be more thoroughly examined in order to refine the definition of close and remote associations in social associations networks.

The correlations with ID and remote and close associations showed that ID was significantly related with remote associations in both human and computer scoring methods, except for the WN. ID did not show any significant correlation with close associations in human and computer scoring methods, except for the WN. Pearson r values for remote associations ranged from .34 to .38 in human scoring methods, and from .21 to .35 in computer scoring methods. These results indicated that human scoring methods had higher correlations with ID.

The correlations with fluency scores and the remote and close associations indicated that fluency was significantly related with both remote and close associations in both human and computer methods. However, fluency showed higher correlation values for remote associations. Traditional flexibility scores did not follow the same pattern as fluency in human scoring methods. The flexibility scores showed significant correlations with only the remote associations, but not with the close associations. This finding can be interpreted to indicate that traditional flexibility scores are more reliable than fluency scores in order to explain creative potential in the associative thinking approach. However, the correlation values for flexibility in the computer scoring methods were different from those in the human scoring methods.

Flexibility showed significant correlations with both remote and close associations in computer scoring methods. At the same time, the ID and traditional flexibility scores showed a significant correlation. Overall, the traditional flexibility scores acted as a determining factor for creative potential with ID in human scoring methods. Also, the correlation of flexibility scores with ID was supported by this result.

Hence, ID is related to creative potential and creative thinking. Runco et al. (in press) showed that ID is significantly related with originality scores. Also, ID has a significant correlation with creative achievements, citation impact, and popularity. In other words, more ideas bring higher creative production. Therefore, the findings in this first study are vital. ID showed significant correlations with remote associations, but not with close associations. These findings support how important making remote associations are for creative thinking.

Also, originality and flexibility have been shown to be more closely tied to creativity than other DT indices (Runco & Albert, 1985). In light of this fact, this first study demonstrates the huge significance of the connection between flexibility and associative thinking for creative ideation. However, computer methods did not catch this particular feature of flexibility, and such methods did not reveal a difference between remote and close associations. On the other hand, computer scoring methods did not show high reliability and correlations as were shown with the human scoring methods.

The findings indicated that subjective scoring is more reliable than objective scoring and had higher correlation values. However, the objective scoring values also had good reliability and correlation values, but not as high as the subjective methods. When the difficulty of finding expert judges is taken into account, objective scoring can be a viable solution for many cases because of its high reliability and correlation values. Also, this first study showed how important

flexibility is to determining creative thinking in comparison to fluency. Fluency scores are always linked to originality; however, this first study showed that fluency scores do not play a real determining role in creative thinking because it was linked significantly to both human and computer scoring methods.

CHAPTER 3
KEYWORDS STUDY²

² Turkman, B. & Runco, M.A. To be submitted to the *Journal of Creative Behavior*.

Abstract

The second study investigated the common words or phrases that are used to introduce original ideas in written works by examining language and attitude studies in creativity literature. Some language studies revealed key relationships with regard to using creative ideas in language. This keywords study took these relationships into account and found specific words and phrases that are used with regularity when people introduce a new idea. This second study provided an understanding of how creativity is expressed in written works. These keywords were tested by using different sources of writings and speeches. The statistical analyses indicated that such keywords successfully identified original ideas in written works and speeches.

Introduction

There is scant literature on predicting the originality of an idea or written work. Few studies explored the various kinds of potential predictors, such as behavior or words or phrases that introduce original ideas. In this section, extant studies and the related literature are outlined.

Kris (1952) focused on primary and secondary process of thinking to define creative behavior. These two modes of thinking counterpoint one another. Secondary thinking accompanies the more dominant primary thinking until the point that they switch roles and the secondary thinking becomes the more important mode. Primary process represents imaginative thinking and secondary process represents realistic thinking (Freud, 1900). Kris argued that creative people are better able to shift between these two different process modes of thinking than are less creative people. According to Kris, creative inspiration involves regression to a primary process state of consciousness. The elaboration stage of the creative process involves a return to a secondary process state. Therefore, it can be concluded that creative people are better able to shift along the primary and secondary process continuum to generate creative ideas.

Rothenberg's (1973) study of creativity and language is one of the earliest studies about the relationship between language and creativity. Rothenberg examined the most extreme form of atypicality in his study, antonymy, because the use of antonymy (e.g., sweet sorrow) can represent a high creative potential. He linked creativity to "Janusian thinking", which he defined as "the ability to use two or more opposite or contradictory concepts simultaneously". This study showed that highly creative groups were more likely than less creative groups to respond with an opposite idea. Therefore, he concluded that creativity is related to conceptual opposites or contradictions.

Martindale (1975) built on these foundations and found that creative people show more variability along a primordial-conceptual axis than do less creative people. Martindale (1975) developed "The English Regressive Imagery Dictionary (RID)" in order to measure primordial versus conceptual thinking. For Martindale, conceptual thought represents logical or reality-oriented thinking. Alternatively, primordial thought is associative and not oriented in reality. Martindale developed this dictionary as he saw psychological processes reflected the content of a text. Martindale's rationale and methods of measurement for primary-secondary process modes of thinking are the theoretical background for this study.

More recently, Estes and Ward (2002) approached creativity from a linguistic perspective. They determined that two linguistic factors affected the emergence of novel attributes in concept modification: typicality and irrelevance. Typicality influenced novel attributes in concept modification. In this case, the researchers pointed out that the typicality spectrum lies in various degrees of antonymous combinations (e.g., friendly enemy). This overlapped with Rothenberg's (1973) findings, which asserted that conceptual contradictions

facilitate creativity. In Estes and Ward (2002), the influence of irrelevance on the creative emergence of novel attributes was equivocal.

As previously mentioned, a product must be original and useful to be creative. Amabile (1982) opined that these indices, originality and usefulness, are to be observed by external observers. Amabile and Gitomer (1984) questioned whether intrinsic motivation is a personality trait or is influenced by external factors such as competition or expected evaluation.

Additionally, Ruscio, Whitney, and Amabile (1998) inquired what task behaviors stem from intrinsic motivation and result in those recognizably original and useful outcomes. They found that both measures of behavioral and verbal processes measures were strong predictors of creativity. The behavioral measures they identified include: *involvement, set breaking, pace, confidence, playfulness, enjoyment, and planning*. Verbal processes measures included: *goal statement, Aha!, transition, irrelevant to task exploration, concentration, difficulty, repetition, and evaluation*.

Ideas play an important role in creative thinking and innovation. Hence, many creative potential tests focus on ideas. As mentioned above, DT is one of the most common creative potential tests used to assess ideas. Runco et al. (in press) investigated this relationship between ideas and creativity. They used the idea frequency approach to explain creative ideation. The software that they used examined the language and analyzed the notion frequencies in the texts. As stated earlier, this method estimated the idea frequency by the total number of verbs, adjectives, adverbs, prepositions, and conjunctions divided by the total number of words. They took both oral and written texts into account in 4 different studies to investigate the relationship between ID and DT, citation impact (CI), and eminence. Their study clearly showed that written or oral works having more ideas in them tend to be cited more frequently by other scholars, and

the higher frequency of ID is related to both reputational change obtained from encyclopedic entries and to Ludwig's Creative Achievement Scale (CAS).

The keywords study contributed to this knowledge base by creating a keywords approach to predict the originality levels of written works. The keywords study revealed general tendencies (and particular keywords) that are used with regularity when people introduce a new idea. The keywords study examined language and attitudes studies in creativity literature in order to develop a unique, objective approach to finding what words or phrases introduce original ideas.

Research Questions

1. Can the originality of a written work be predicted by looking for particular words or phrases that introduce original ideas in authors' writings?
2. Are fluency and flexibility scores related to the frequency of keywords and the originality of written works?
3. Are original ideas as represented in written works more dependent on sentence length/number or on simple word counts?
4. Are there variances in different modes of communication to sustain an original idea?
5. To what extent does this computerized keywords method decrease the labor intensity in scoring DT tasks?

Method

Sample

The sample was ($n=26$) university students majoring in various fields. Most of the participants ($n=18$) majored in humanities, social sciences, or art, and the remaining ($n=8$)

majored in biological sciences. Twenty were female, and 6 were male. Their ages ranged from 23-63 years (mean, 28; *SD*, 7.50) and the average GPA was 3.71 on a 4-point scale (*SD* = .24).

Data Collection

Several professors and student organizations were contacted to invite their students and members to take part in the study. After securing participants for this study, individual and group appointments were scheduled to meet with each participant.

Instruments

DT tasks were administered with the demographic information form. The DT tasks used in this second study consisted of three items from the Many Uses Game (i.e., “List different uses for a bowl”, then, “tire”, and then “nail.”) from the Runco Creativity Assessment Battery (rCAB, 2011, creativitytestingservices.com).

Participants were given the following instructions: “*List as many ideas as possible about different uses for a bowl, a tire, and a nail. Your answers will not be judged or questioned. Have fun!*” They did not have a time limit to complete the DT tasks.

Then, participants were asked to review the ideas they had given in the DT tasks and to select the ideas that they deemed original. They were then instructed to write an essay about these original ideas; they could write the essay about any topic they wanted as long as they incorporated their original ideas from the DT tasks. They were given 15 minutes to finish the essays.

Procedure

The participants’ essays were analyzed by two experts from the field of creativity assessment in order to discern any general tendencies used to introduce an original idea in written works. The experts did not have any predetermined criteria by which to analyze the

essays. The experts reviewed the participants' DT responses and read their essays to conceptualize the participants' ways of thinking.

The essays included only original ideas from the DT tasks. Therefore, the experts did not create a response pool from which to identify original ideas. Also, all participants stated that they had taken at least one creativity-related class in the past. Therefore, they were familiar with creativity and originality concepts.

Two experts identified two different sets of keywords, and this difference came from their subjective definitions of originality and creativity. The experts' original keywords lists included both mutual and differing words, but the final keywords list included only the words they mutually agreed upon. Both stated that some of the keywords they had found were words often used to introduce an original idea, but other keywords functioned to introduce a contradiction or possibility (e.g., "however", "would be"). Even though the contradiction or possibility introduction words were not always used to introduce an original idea, they had a high potential to be used in such a way or to at least be proximate to an original idea in a sentence. See Table 3.1 for the judges' initial lists of keywords.

Again, the final keywords list consisted of the words that the experts mutually agreed upon. All statistical analyses were conducted based on these 17 words or phrases. See Table 3.2 for these keywords.

After reaching a consensus regarding the identified keywords, a computer algorithm was developed in order to search for these words or phrases in written works. This computer algorithm provided all the information needed to count the number of keywords in written works.

The ID method was also used in this second study to test the functionality of the keywords. Runco et al. (2014) used the ID concept in three different eminence and creativity-related studies. The results revealed that ID represents a useful measure of ideation and creative potential. Also, as the first study indicated, ID is significantly related to making remote associations, which is connected to creative ideation. In light of these facts, ID was the criterion used to test the functionality of keywords in identifying original ideas.

Table 3.1
The Experts' Initial Keywords Lists

Expert 1	However	Change	If	Led	Lead
	Expect	Possible	In addition	Should be	Dimension
	Way to	Would be	Easier	Either or	Easier
	Either or	Instead of	Another	Influence	As well
	Most probably	Can be	Cause	A new way	New method
	Would use	Can use	What else	To Create	
Expert 2	However	Led	Expect	Dimension	New Method
	Way to	Either or	Influence	Most Probably	Would use
	To Create	But	Fashioned	Would be	Beside
	Lead	Possible	Solution	Conflict	Could make
	Could also	Can also	But	Can even	

The functionality of these keywords was tested in different sources, which included both written and oral samples. For oral sources, TED talks were retrieved from www.ted.com/talks. More than 100 speeches from a variety of disciplines (e.g., art, innovation, entrepreneurship, literature, technology, personality, design, and education) were sequestered using the search term “creativity”. The keywords numbers in these speeches were calculated by using the computer algorithm that was developed specifically for this study. Also, as in study 1, CPIDR 5 was used to calculate the ID values in the speeches. It was additionally observed that the speeches

exhibited a different way to introduce original ideas than found in written works. For example, “if” is one of the words that was often used to introduce an original idea in the speeches, but this word was not in the keywords list that had been derived from the participants’ essays.

CPIDR 5 includes a special module only for speeches. This speech module rejects most repetitions, not counting them as new propositions; it also rejects hesitations and interjections. These are necessary functions for working on unedited speeches, such as the TED talks transcriptions, which inevitably display such features of oral discourse.

Another method to test the functionality of the keywords was to use eminent authors’ writings. Runco, Kaufman, Halladay, and Cole (2010) used a subsample from Ludwig’s (1995) eminent people study. Runco et al. (2010) investigated the reputational change of eminent people over the time. This sample list included 100 eminent people, who were identified from Ludwig’s (1995) original sample of 1004 individuals. Digital samples of their writings were found in different online sources. Project Gutenberg (www.gutenberg.org), Internet Archive Search (archive.org), and local university sources were utilized to find the writings of these 100 eminent authors. However, some of the names from the list, mainly those of painters and artists, did not produce a digital sample of their writings. Additionally, some of the writings were excluded from this study because they did not produce an idea reflection writing sample (e.g., geographical reports, daily work summaries etc.).

The computer algorithm provided all the keywords from these digital writing samples, and CPIDR 5 provided the ID values for these writing samples. Additionally, the computer algorithm was modified for a second time here in order to also count the number of sentences in these digital writing samples so that a proportional number of keywords could be obtained regardless of the length of any one writing sample; otherwise, the number of keywords would

grow larger, as the length of the writing sample increased. Another method was tested in this second study by dividing the total number of keywords by the total number of words used in each writing sample. However, using the total number of sentences method is more logical in a keywords study because an original idea is developed over the course of a sentence; a sentence naturally includes the story behind the ideas.

Table 3.2
The Final Keywords List

Keywords				
However	Led	Expect	Dimension	New Method
Way to	Either or	Influence	Most Probably	Would use
To Create	But	Fashioned	Would be	Beside
Lead	Possible			

Results

Twenty-six participants gave 503 responses for the bowl, nail, and tire stimuli: 203 for bowl, 149 for nail, and 151 for tire. Descriptive statistics of the three verbal DT tasks are provided in Table 3.3.

Table 3.3
Descriptive Statistics for the Divergent Thinking Tasks

	<i>N</i>	Min.	Max.	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Bowl	26	2	28	7.96	5.63	2.01	5.38
Nail	26	9	19	5.69	4.23	1.50	2.95
Tire	26	1	20	5.69	3.99	1.96	5.94

Also, descriptive statistics of the ID of the essays, the number of words, and the number of sentences used in the essays are provided in Table 3.4.

Table 3.4
Descriptive Statistics for the Essays' ID Values

	<i>N</i>	Min.	Max.	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
ID	26	0.434	0.582	0.489	0.034	.522	0.585
Number of Sentence	26	4	52	22	12.22	.653	0.079
Number of Words	26	97	445	245.54	84.09	.482	-0.103
Number of Keywords	26	0	12	3.12	2.67	1.58	3.54

Note: ID= Idea Density

Statistical analyses indicated that the alpha coefficient was .93 for the three fluency scores (bowl, nail, and tire).

Another statistical analysis was conducted to discern the role of traditional flexibility scores in the keywords method. The alpha coefficient was .82 for the three traditional flexibility scores (bowl, nail, and tire). Composite scores were generated in order to conduct further analyses.

Bivariate correlations were calculated to see the relationship among ID, traditional flexibility and fluency scores, and the keywords ($n=26$). The results indicated that ID was significantly correlated with the keywords ($r = .70, p < .001$) when the total number of keywords in an essay was divided by the total number of sentences. Also, the bivariate correlation was significant ($r = .59, p < .005$) when the total number of keywords in an essay were divided by the total number of words. The results further indicated that traditional flexibility scores were significantly related with ID ($r = .40, p < .05$) and the keywords method ($r = .43, p < .05$). However, the bivariate correlations did not reveal any significant relationship between fluency and other indices. All bivariate correlations for each variable are provided in Table 3.5.

Table 3.5
Bivariate Correlations Among Keywords, ID and Composite DT Scores

	ID	1	2	3
1. Keywords/Sentence	.70**			
2. Keywords/Words	.59**	.66**		
3. Fluency	.25	.04	.24	
4. Flexibility	.40*	.27	.43*	.79**

Note: ID= Idea Density

* $p < .05$ level (2-tailed)

** $p < .01$ level (2-tailed)

Statistical analysis clearly demonstrated that there is a large and significant correlation between ID and the keywords. This also showed that using the sentence method is a more accurate way to capture the ideation process. Therefore, the sentence method was used for the rest of the analyses. These results followed Runco et al.'s (in press) findings, which showed ID is a strong predictor of creative potential. Additionally, the results indicated that the keywords method was significantly related with flexibility, but not related with fluency cores.

The descriptive statistics for the TED talks study are provided in Table 3.6.

The relationship between the ID values and the keywords in the TED talks were investigated ($n=103$). The statistical analyses indicated that the keywords were significantly related with TED talks ID values ($r = .32, p = .001$). These results also followed the same patterns as found in the study mentioned above. An additional method was further applied in the TED talks study, the keyword/speech method.

Table 3.6
Descriptive Statistics for the TED Talks Study

	<i>N</i>	Min.	Max.	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
ID	103	0.485	0.58	0.53	0.02	0.157	-0.181
Number of Sentences	103	29	433	177.53	98.07	0.48	-0.63
Number of Keywords	103	3	55	22.44	13.14	0.525	-0.533

Note: ID= Idea Density

As mentioned above, the speeches followed a different pattern in presenting original ideas. This second study did not particularly focus on speeches; however, some observations were made based on that sample. When the word “if” was added to the keywords list, the bivariate correlation increased ($r = .43, p < .001$). This result substantiated the expert opinions about varying ways to identify original ideas in different modes of communication.

The descriptive statistics for the Ludwig (1995) study are provided in Table 3.7. The relationship between eminent persons’ ID values and the number of keywords was investigated ($n=45$). The bivariate correlations revealed that ID was significantly related with the keywords in the writings of eminent authors ($r = .32, p < .05$). This third method to test the functionality of keywords also follows the same pattern as the two previous participant essay method and the TED talks method.

Table 3.7
Descriptive Statistics for the Ludwig (1995) Study

	<i>N</i>	Min.	Max.	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
ID	45	0.49	0.578	0.529	0.019	0.443	-0.147
Number of Sentences	45	97	21332	4832.98	4405.642	.1.579	3.281
Number of keywords	45	12	2144	578.13	446.7	1.28	2.527

Note: ID= Idea Density

Discussion

The intricate relationship between language and creativity has been investigated by creativity researchers (Kris, 1952; Rothenberg, 1973; Martindale, 1975; Estes & Wars, 2002), and these studies focused on concept modification (typicality and irrelevance). However, the focus of this second study was to find the general tendencies used to introduce an original idea in writing.

Another motivation for conducting this second study was to find a new approach to alleviate the laborious scoring system of creativity assessments. This newly developed computer algorithm can now provide the originality level for thousands of pages of written works in seconds rather than days.

The roles of traditional flexibility and fluency scores were investigated in this second study to discern their correlation to the keywords method. The first study indicated that flexibility scores followed a more reliable pattern with remote associations and in relation to ID. The functionality of traditional flexibility scores was also tested in this second study. Statistical analyses showed that the alpha coefficients were .82 for the three traditional flexibility scores (bowl, nail, and tire) and .92 for the three fluency scores (bowl, nail, and tire). These results demonstrated that fluency scores had a higher reliability than traditional flexibility scores.

The bivariate correlations revealed that ID was significantly correlated with the keywords ($r = .70, p < .001$) when the total number of keywords in an essay was divided by the total number of sentences. Also, the bivariate correlation was significant ($r = .59, p < .005$) when the total number of keywords in an essay was divided by the total number of words. However, fluency scores did not have a significant correlation with the keywords method and ID. The bivariate correlations were $r = .04$ for the keywords/sentence method and $r = .24$ for the keywords/word method, and $r = .25$ for ID. On the other hand, traditional flexibility scores displayed significant correlations with the keywords/word method ($r = .43, p < .05$) and with ID ($r = .40, p < .05$). Also, traditional flexibility scores had a high correlation with the keywords/sentence method ($r = .27, p = .18$), but this was not significant.

Statistical analyses clearly proved that there is a large and significant correlation between ID and keywords. It was also demonstrated that using the sentence method is a more accurate

way to capture the real ideation process. However, traditional flexibility scores did not have a significant correlation with the keywords/sentence method, but did with the keywords/word method. The relationship between the ID and keywords methods is helpful to understand this difference. The ID showed a higher correlation with the keywords/sentence method ($r = .70$) than with the keywords/word method ($r = .59$). The reason behind this is ID examines the idea integrity of a sentence. Therefore, the sentence method is more accurate to capture creative ideation in written works.

Two more cases were investigated to test the functionality of the keywords method. The first case study was done by using 103 TED talks categorized under the topic of creativity. The speeches were from different disciplines (e.g., art, innovation, entrepreneurship, literature, technology, personality, design, and education). The TED talks study revealed a significant correlation between ID and the keywords in the speeches ($r = .32, p = .001$). Another significant finding of the TED talks study was to capture the explicit differences between different modes of communication to present original ideas. The original keywords list was modified here by adding one more word to it (“if”). Expert observations about the word “if” increased the correlation between the ID and the keywords method ($r = .43, p < .001$). This finding clearly showed that different modes of communication can have their own unique style in the presentation of original ideas.

In the second case, the digital writing samples of certain eminent authors were investigated from Ludwig’s (1995) eminent people study ($n = 45$). The results revealed a significant correlation between the ID values of the writing samples and the keywords method ($r = .32, p < .05$). However, this correlation is much lower than in the relationship between the ID values found in the writings of the 26 participants and the keywords method ($r = .70, p <$

.001) even though these samples included only written discourse, not spoken as in the TED talks study.

These results brought another insight to the keywords method. As mentioned above, the experts observed differences in ways to present original ideas in different modes of communication. However, the eminent people study provided much lower correlations with the keywords methods.

The keywords list was originally derived from the essays of the participants. These essays were written in a classroom environment with no planning or preparation on the part of the participants. On the other hand, many novels and well thought out essays were investigated in the eminent people case (Ludwig, 1995). Therefore, it can be concluded that recognizing this type of professional, planned writing as essentially different from quotidian language usage with regard to creativity display is important to using the keywords method properly. In other words, the list of keywords in this second study was productive in capturing the originality level of the less formal writing samples from the participants. However, a more comprehensive study is needed to investigate professional writing samples to reveal their own unique characteristics with regard to proposing original ideas.

CHAPTER 4

EXPLICIT INSTRUCTIONS & CREATIVITY JUDGMENT³

³ Turkman, B. To be submitted to *Psychology of Aesthetics, Creativity, and the Arts*.

Abstract

Assessment methods that rely on expert judges (i.e., the CAT) are resource intensive, and the availability of expert judges, time, and costs are major obstacles to consider. Additionally, individual tendencies and subjective preferences play an important role in assessments submitted by creativity judges. The CAT is one example of a system that used such judges; however, the CAT does not provide a solution to this discrepancy and does not include training or norming for the judges. Furthermore, the transferability of training to another subject or domain is another consideration.

This third study investigated the training of creativity judges by providing explicit instructions (EIs) in order to make them better informed about originality and more aware of the processes involved in creative production. Idea Density and Keywords quantitative methods were used to objectively observe the EIs and the effects of such training in the study.

Introduction

Using judges in creativity assessment is one of the most common methods that researchers utilize in order to evaluate the creativity level of products (Kaufman et al., 2008). The CAT uses expert judges and claims that only experts can know what works are creative. However, finding quality expert judges is not easy under every circumstance. This third study examined the possibility that direct, specific instructions and a workshop on creativity would facilitate better judgments of creativity. EIs manipulate performance and control the task perception and strategies used by individuals (Runco, Illies, & Eisenman, 2005). In this study, creativity judges improved their ability to judge creative and original ideas by participating in a creativity workshop and receiving EIs about how to judge a product's creativity. Additionally,

two objective computerized methods (i.e., the CPIDR 5 & the keyword computer algorithm) were used in this study in order to objectively examine the effects of EIs on the ratings of judges.

Amabile (1982) developed the CAT structured around the idea that the best measure of the creativity of a work is the combined assessment of experts in that field (Kaufman et al., 2008). In order to designate appropriate judges, Amabile (1996) recommended relying on the experts' training and knowledge in the target domain. Amabile (1996) stated that an expert should have "at least some formal training and experience in the target domain" (p. 73). However, the CAT does not include training or norming for the judges. For example, judges cannot be given rubrics to follow in making such judgments or in any other way be interfered with in their unconstrained assessments of an artifact's creativity when using the CAT.

Alternatively, Runco and his colleagues (Runco & Chand, 1994; Runco et al., 1994; Runco & Smith, 1992; Runco & Vega, 1990) questioned the idea of using "experts." They explained why experts would be more valid and useful than self-ratings or other evaluation methods, such as peer or teacher evaluation.

Using experts presents challenges including finding appropriate expert raters and that expense, determining the appropriate expertise level for the area, and planning and organizing appropriate studies in which to apply expert opinions (Plucker & Makel, 2010). Expert methods such as the CAT are very resource-intensive. The CAT requires much more time and financial resources than many other creativity assessment methods. These considerations exemplify Runco's questioning of the feasibility of the CAT. Much of the research done using the CAT includes very few judges. Amabile (1982) suggested 5 experts, but even this number is very difficult to obtain. Therefore, many CAT studies have only 1-2 experts, and having a low number of experts is a concern for having good inter-rater reliability with the CAT.

Runco and Smith's (1992) study discussed the interpersonal ratings of expert judges. They stated that producing art and judging art requires different skills. Therefore, an artist might be able to evaluate his own work, but not others because of his own biases. Another issue that Runco and Smith (1992) identified was that experts, by definition, come from an area in which they were highly trained; therefore, their judgment may rely on high-level, esoteric, idiosyncratic criteria. These judgment criteria reflect their specialized training, but not necessarily creativity per se and might not be appropriate to judge the art of non-professionals. The study by Runco et al. (1994) showed that expert judges were not able to reflect differences among various groups as well as non-expert judges because the criteria used were not directly applicable to the artwork of students and, therefore, were overly critical.

A study by Dollinger and Shaffran (2005) worked with training non-expert judges to evaluate the creativity of ideas and products. In their study, non-expert judges were provided with a set of drawings along with respective ratings as already submitted by experts in order to give them an idea about how to judge a product's level of creativity. The results showed that this training/norming improved inter-judge reliability when compared to results that were obtained from a previous study (Dollinger, Urban, & James, 2004). This study also proved that non-expert judges could indeed be trained. However, this method still requires experts to train non-experts, and it is not known if this method can be transferred to other domains.

Caroff and Besançon (2008) focused on two general traits of creativity judgments: originality and usefulness. They investigated the ways originality and usefulness concepts could affect the ratings of judges and how, too, the personal characteristics of the judges could lead to individual differences in creativity ratings. Caroff and Besançon (2008) investigated how EIs could make creativity judges more inclined to take both originality and usefulness into account

when they rated the creativity of products. In order to achieve this goal, the researchers trained their participants in how to evaluate creativity immediately before the experimental task. They found that the originality level of advertisements had an important impact on creativity ratings. Therefore, creativity judges estimated the creativity of advertisements based on this dimension, which could not be modified by any instruction or training.

EIs are one of the most common ways to enhance a subject's creative performance on any given task. Harrington (1975) showed that when participants were instructed to "be creative", they tended to produce more creative ideas as measured by the Alternate Uses Test (Christen, Guilford, Merrifield, & Wilson, 1960; Wilson, Guilford, & Christensen, 1953) in comparison to those who did not receive such instructions.

Another EI study along these same lines was done by Runco (1986). He compared the DT scores of gifted and non-gifted students by applying EIs. The findings showed that EIs increased the originality scores for all children. However, gifted children showed greater increases in their originality scores than non-gifted children. Additionally, EIs inhibited the fluency and flexibility scores of gifted children more than it did those of non-gifted children.

Chand and Runco (1993) compared the impact of explicit and standard instructions in their study. Real-world DT, real-world problem generation, and a combination of problem generation and DT tests were used in the study. The results showed significant differences among the tests and differences between the explicit and standard instructional groups. Additionally, only the scores elicited by EIs were significantly correlated with, and predictive of, creative activities and accomplishments. These findings aligned with the rationale of using EIs that encourage participants to perform at their best by clearly explaining what is expected from

them because the most reliable estimate of an individual's potential is the person's maximal performance (Chand & Runco, 1993).

Runco and Okuda (1991) used EIs to maximize the different components of DT (e.g. ideational originality and ideational flexibility) in their study in addition to giving inexplicit directions. The directions to maximize the participants' potential led to an increase in originality and flexibility scores. However, contrary to expectations, flexibility scores were low when originality instructions were given. The researchers concluded that EIs do not influence participants' abilities to generate ideas, but rather manipulate the choice of specific ideational strategies (Runco & Okuda, 1991).

This third study investigated how training creativity judges by applying EIs can make them more aware of originality and usefulness concepts and of the processes involved in creative thinking and production. The EIs and the creativity workshop enhanced the judges' attitudes and values regarding creativity. These particular methods and approaches led them to transfer what they learned from this training to other domains.

Objective ID and keywords methods were applied to the same essays in order to observe the impact of EIs and training on the judgments of the participants. Akin to the previous two studies, this comparison showed how objective and subjective scoring differed.

In this study, EIs represented the independent variable, and the qualities of judgments represented the dependent variable. Instructions for EIs included what creativity is, what is important for creative thinking, and other important components of creativity (e.g. appropriateness, originality).

Research Questions

1. Can non-expert judges be trained to be better judges of creativity?
2. Can EIs make judges reliably better and more reliable judges at evaluating creativity?
3. Can EIs and training enhance the beliefs of non-expert judges regarding creativity?
4. Does providing opportunities to apply acquired knowledge make for better internalization of such knowledge with EIs?

Method

Sample

The sample included ($n = 70$) undergraduate level students from a large state university in the Southeastern United States majoring in various fields. Most of the participants ($n = 50$) majored in humanities, social sciences, or art; the remaining ($n = 20$) majored in biological sciences; and three of them did not provide information about their majors. Fifty-nine students were female, and 11 were male. Their ages ranged from 18-20 and from 24-25 years. Only one student was over 25 years old.

Data Collection

Several professors and student organizations were contacted to invite their students and members to take part in the study. A freshman level art-related class volunteered to take part in this study. A whole class appointment was scheduled to meet with these students during their class time. They were briefed on the study during the class and were given instructions about how to complete the study. The participants were exclusively college students in order to demonstrate consistency in the semantic relationship data from participants at a similar level of education.

Instruments

Those students who consented to be part of the study began what will henceforth be referred to as the creativity workshop. First, the participants were given the Attitudes and Values scale (A & V) in order to determine their beliefs with regard to creativity before the training commenced. Subsequently, six different previously written and randomly selected creative essays about different topics were given to the students to rate according to six criteria: creativity, technical skill, aesthetic, originality, flexibility, and fluency. Originality, aesthetic, and technical skill indices were taken from Csikszentmihalyi and Getzels (2014), and a creativity index was taken from Runco (1989).

Each of the participants used a five-level, Likert-type scale with the following response options: (1) Very low, (2) Low, (3) Average, (4) High, and (5) Very High. The essays were presented one at a time, and the students were allowed as much time as needed for the ratings. The first three essays were rated by the participants before EIs, and the remaining three were rated after EIs were given.

Procedure

Upon their completion of the ratings of the first three essays, the participants were given a set of EIs including relevant creativity definitions and concepts. This workshop focused on the following topics: “What is creativity?”, “What is important in creativity?”, “What constitutes Divergent Thinking?”, “What are the four DT indices?”, “What is the role of usefulness and originality in creative ideation?”, “What is subjective and objective scoring in creativity?”, and “What are the comparisons to convergent thinking?”. The creativity workshop slides are shown in appendix b.

All of the topics and terms listed above were defined as EIs, and the tasks used in this study were part of the training. This creativity workshop provided a knowledge base for the participants in order to transfer that knowledge base to other domains by defining terms clearly and providing explicit examples. Hence, general creativity traits were explained and discussed during this training, instead of focusing on a particular domain such as art, literature, or poetry. Researchers like Dollinger and Shaffran (2005) or Caroff and Besançon (2008) trained their novice judges in particular areas in order to designate them expert-like judges; however, they did not get good reliability results or agreement. Additionally, they observed that these judges did not transfer their training to other domains. Therefore, this third study applied a different, more general approach to the training of the novice judges.

After the creativity workshop, the A & V scale was given to the participants again in order to see the effect of the EIs on their beliefs about creativity. Subsequently, a new set of three essays was distributed to the participants to rate them for the same six criteria. These essays were also given out one at a time. A full copy of the A & V scale is shown in appendix c.

Each essay was analyzed using objective and subjective methods to assess their level of creativity. In order to obtain subjective judgments for the creativity level of these six essays, two experts from the field of creativity rated these essays according to the same six criteria provided to the participants, but these experts did not receive EIs, nor did they take part in any creativity workshop.

For the objective judgments, as in studies 1 and 2, CPIDR 5 was used to calculate the ID values of the essays. Also, the keywords method from the second was applied here with these six essays to provide another source of objective ratings.

Results

Seventy participants judged the creativity level of six different essays and were given the A & V scale both before and after the creativity workshop. Statistical analyses indicated that the inter-item alpha coefficient was .75 for the A & V scale. This alpha coefficient was tested again after the creativity workshop, and the results showed that the inter-item alpha coefficient increased to .86.

The A & V results indicated that the total score after the EIs increased by 44 points for all the participants. A paired *t*-test investigated this increase and did not find it to be a significant difference ($t(24) = -0.884, p = .38$).

Inter-rater reliability analyses were conducted for the six essays. Statistical analyses indicated that the inter-rater alpha coefficient for essay 1 was .78, .77 for essay 2, and .87 for essay 3. These three essays were rated by the participants before the EIs. A composite score was calculated across dimensions for the first three essays to conduct further statistical analyses, and the overall inter-rater reliability alpha was .79 for these three essays. The EIs helped to increase the participants' alpha values. Statistical analyses revealed that the alpha values were .88 for essay 4, .92 for essay 5, and .93 for essay 6. Composites scores were also calculated for these essays, and the overall reliability alpha was .89. The EIs increased the overall alpha values from .79 to .89. All alpha reliability values are provided in Table 4.1.

Table 4.1
Alpha Coefficients for the Essays

Before the Workshop		After the Workshop	
Essay 1	.78	Essay 4	.88
Essay 2	.77	Essay 5	.92
Essay 3	.87	Essay 6	.93
Composites for 1,2 & 3	.79	Composites for 4,5 & 6	.89

A paired *t*-test investigated the difference between experts' and the participants' scorings. The results indicated that there was a significant difference between the experts' and participants' scorings before the training ($t(2) = 5.20, p < .05$). The results for after the EIs were given revealed that this significant difference had disappeared ($t(2) = 1.77, p = .22$). Statistical analyses indicated that the fluency index had the lowest alpha value ($\alpha = .03$) among the six criteria. However, this alpha value was increased to ($\alpha = .18$) after the creativity workshop. All the other alpha coefficients also increased after the EIs: the creativity index was increased from ($\alpha = .27$) to ($\alpha = .38$); the technical skill index was increased from ($\alpha = .21$) to ($\alpha = .38$); the aesthetic index was increased from ($\alpha = .37$) to ($\alpha = .42$); the originality index was increased from ($\alpha = .20$) to ($\alpha = .30$); and the flexibility index increased from ($\alpha = .33$) to ($\alpha = .41$). The overall reliability values for these six indices were ($\alpha = .44$) for creativity, ($\alpha = .46$) for technical skill, ($\alpha = .44$) for aesthetic, ($\alpha = .38$) for originality, ($\alpha = .50$) for flexibility, and ($\alpha = .47$) for fluency.

Discussion

One of the most popular methods for assessing creativity relies on human judgments. Often professionals (e.g., artists) or other experts are asked to judge creative works (e.g., Csikzentmihalyi & Getzels, 1971; Runco, 1989), but self-ratings (Hocevar, 1981; Rothenberg, 1986) and peer ratings (Speecher, 1964) have also been used with some success.

The training of non-expert judges to evaluate a products' creativity has been investigated by many researchers (Dollinger & Shaffran, 2005; Caroff & Besançon, 2008), but the particular focus in this study was to provide a knowledge base that is transferable to other domains and opportunities for the participants to understand and internalize this process better.

Therefore, general creativity traits were explained and discussed during this workshop, instead of focusing on a particular domain.

Two of the six criteria used by the participants to rate the creativity of the essays were not directly related to traditional dimensions of creativity (technical skill and aesthetic). These two criteria were originally used by Csikszentmihalyi and Getzels (1971) to rate art products.

However, EIs also increased agreement among the participants for these two criteria. The reliability values (alpha coefficients) were increased from ($\alpha = .21$) to ($\alpha = .38$) for the technical skill index and from ($\alpha = .37$) to ($\alpha = .42$) for the aesthetic index. This finding is very important to understanding the effectiveness of EIs. These results indicated that the EIs helped participants think and analyze products from a perspective that takes creativity into account. The workshop did not focus on any particular talent area; it focused only on how creativity is seen by researchers and experts in the modern world. Therefore, the participants added this new perspective to their general view of creativity schema, instead of as applied to any particular area. The new knowledge acquired caused a change in approaches to creativity evaluation.

As was mentioned earlier, the CAT method has some negative and positive aspects, such as the difficulty of finding relevant expert judges and that expense. Therefore, many people work only with 2 judges to evaluate the creativity of products. Using such a low number of judges comes with low inter-rater reliability values. The third study also proved this fact. The alpha value for inter-rater reliability was .75 for expert judges. On the other hand, seventy participants' inter-rater alpha coefficient values were .79 before the creativity workshop and .89 after. The results clearly showed that having more judges comes with higher inter-rater reliability values. It can be concluded that EIs can increase the agreement level among non-expert judges. However, the quality of these judgments is still questionable.

The participants were given the A & V scale in order to discern their beliefs about creativity. The A & V scale was administered before and after the EIs in order to observe the EIs' effect on participants' beliefs about creativity. The A & V results showed that the total score of all participants after the EIs increased by 44 points, and the inter-item alpha coefficient increased from .75 to .86. A paired *t*-test was used to investigate this increase, but it did not find this increase as a significant difference ($t(24) = -0.884, p = .38$). Even though this increase was not significant, it still led to an increase in the inter-rater alpha values, at about .10, for judging the essays.

A paired *t*-test also investigated the difference between the experts' and the participants' scoring. The statistical analysis showed that there was a significant difference between the experts' and participants' scoring before the EIs ($t(2) = 5.20, p < .05$). However, this significant difference disappeared after the creativity workshop ($t(2) = 1.77, p = .22$). The EIs made the participants' ratings closer in alignment with the experts' ratings. This is another piece of evidence to show that non-expert judges can be trained.

During this study, what is expected for someone to judge a products' creativity was explicitly discussed. Also, what is important in creativity and what is expected from a creative product was pointed out several times. This helped to improve participants' general creativity knowledge base from the viewpoint of experts. The EIs proved that a subject's individual awareness and judgments can be enhanced to make them better judges of creativity. The participants additionally had the opportunity to apply what they had learned during this process. This process helped them put this knowledge to task for better internalization of the concepts involved.

CHAPTER 5

CONCLUSIONS

Subjectivity and objectivity in DT have been addressed by many researchers (Christensen et al., 1957; Silvia, 2008; Acar & Runco 2014; Benedik et al., 2012). However, the focus of this first study was to take both subjective and objective methods into consideration in order to explain creative thinking from different perspectives and to contribute to the field of creativity assessment.

All three studies compared objective and subjective scoring mechanisms and their respective effects on quantifying creative potential for research purposes. The first study measured flexibility in terms of semantic distance to explain creative thinking; it also used human judges to obtain subjective scoring to explain semantic associations. The second study identified and used a set of keywords to quantify the number of original ideas in an essay. The third study explored the relationship between EIs and the quality of creativity judgments made by non-expert judges.

The major findings of these three studies were as follows:

1. Remote and close associations can be reliably measured when human raters and different social associations (i.e., the keywords lists from the IF, WAN, WN) were applied. The inter-rater reliability alpha (alpha coefficients) of remote associations was higher than for close associations.
2. Human judge's inter-rater reliability alpha values were higher than with the computer methods. This indicated that human scoring was more reliable than computer scoring.

3. The bivariate correlations indicated that ID was significantly related to remote associations in both human and computer scoring methods, except for the WN. The ID did not show any significant correlation with close associations in human and computer scoring methods, except for the WN.
4. The correlations with fluency scores and the remote and close associations indicated that fluency was significantly related with both remote and close associations in both human and computer methods. However, higher correlation values were shown for remote associations.
5. The traditional flexibility scores did not follow the same pattern as the fluency scores in the human scoring method. The flexibility scores only showed significant correlations with the remote associations, but not with close associations.
6. The originality of a written work can be reliably tested by using the keywords methods.
7. The bivariate correlations revealed that ID was significantly correlated with the keywords when the total number of keywords in an essay was divided by the total number of sentences. Also, the bivariate correlation was significant when the total number of keywords in an essay was divided by the total number of words. However, the sentence method brought higher correlations than the word method.
8. The fluency scores did not have a significant correlation with the keywords method and ID. On the other hand, traditional flexibility scores had significant correlations with keywords/word method and with ID.
9. The keywords/sentence method more accurately captured creative ideation in written works.

10. The TED talks study revealed a significant correlation between the ID of and keywords in speeches. Another significant finding of the TED talks study was the capturing of relevant differences in presenting original ideas in different modes of communication: writing and speech.
11. In the second case, the digital writing samples of certain eminent authors were investigated from Ludwig's (1995) eminence study. The results revealed a significant correlation between the ID values of the writing samples and the keywords method. However, this correlation is much lower than the relationship between the ID values of the 26 participants in the first study, and the keywords method in the second study.
12. From the third study, non-expert judges can be trained, and this acquired knowledge can be transferred to other domains. Additionally, EIs can help raters evaluate a product from a similar perspective even though the criteria are not creativity related.
13. EIs and the creativity workshop caused an increase in inter-rater and inter-item reliability coefficients. The scorings of non-expert judges were more in line with the scorings of experts after EIs.
14. EIs and the creativity workshop also led to an increase in the participants' total score for the A & V scale. However, the *t*-test indicated that this increase was not significant.
15. The creativity workshop was also influential on the relationship between participants' judgments, ID, and keywords methods. The bivariate correlations indicated that the workshop increased the bivariate correlation values for both ID and keywords methods to explain their relation to participants' judgments. Bivariate correlation values between experts and the participants also increased.

Definitional irregularity and the resulting criteria problems create reliability and validity problems in creativity assessment. Scoring is central to any assessment and also for creativity. Better scoring instruments are necessary to improve predictive ability in the assessment of creativity. Broadly speaking, the three studies above investigated and compared objective and subjective scoring methods. Individually, each study closely examined a particular aspect in creativity assessment scoring. Cumulatively, these studies created opportunities for researchers to utilize existing literature and frameworks with the help of technology to better assess creative potential.

The present study has several limitations. First, the sample sizes of the first and second studies were not very large. Hence, a larger sample size might bring some additional insights and keywords to the current list. Additionally, the inter-rater reliability values of experts judges were very high in the first study. All these expert judges come from the same educational training and background. Future studies might consider using expert judges from diverse backgrounds. Second, the writing samples of only 26 participants were investigated to find the keywords. Moreover, different styles of writing and communication have their own unique ways to present original ideas. Therefore, professional writings and speeches need to be investigated to discern their respective originality patterns. Furthermore, this second study applied a more explanatory method to detect the keywords. Future studies might consider finding different theoretical methods to detect keywords. Third, the effectiveness of EIs and the creativity workshop were tested within a single class period. Also, the A & V scale did not significantly increase after the EIs. Hence, a longer training period might produce better results. Additionally, the judges only evaluated six essays (three before the EIs and three after the EIs). Evaluating a limited number of essays prevented a sound correlation analyses to see the correlation between

non-expert and expert judges' evaluation. Therefore, future studies might consider using a higher number essays to evaluate.

REFERENCES

- Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal, 26*, 229-238
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*, 997-1013.
- Amabile, T. (1983). *The social psychology of creativity*. NY: Springer-Verlag.
- Amabile, Teresa M. (1996). *Creativity in context: Update to 'The Social Psychology of Creativity.'* Boulder, CO, US: Westview Press.
- Amabile, T. M., & Gitomer, J. (1984). Children's Artistic Creativity. *Personality & Social Psychology Bulletin, 10*, 209.
- Baer, J. (1993). Why you shouldn't trust creativity tests. *Educational Leadership, 51*, 72-73.
- Baer, J. (1994). Performance assessments of creativity: Do they have long-term stability? *Roeper Review, 17*, 7-11.
- Baer, J. (2009, August). *The Division 10 debate—Are the Torrance Tests still relevant in the 21st century?: Torrance Tests are not still relevant in the 21st century*. Paper presented at the 117th Convention of the American Psychological Association in Toronto, Canada, August 16– 10, 2009.
- Ball, O. E., & Torrance, E. P. (1984). *Streamlines scoring workbook: Figural A*. Bensenville, IL: Scholastic Testing Service, Inc.
- Benedek, M., Könen, T., & Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts, 6*, 273-281.

- Bossomaier, T., Harré, M., Knittel, A., & Snyder, A. (2009). A semantic network approach to the creativity quotient (CQ). *Creativity Research Journal*, 21, 64-71.
- Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, and Computers*, 35, 634-639.
- Brown, R. T. (1989). Creativity: What are we to measure? In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity*. (pp. 3-32). New York, NY, US: Plenum Press.
- Caroff, X., & Besançon, M. (2008). Variability of creativity judgments. *Learning and Individual Differences*, 18, 367-371.
- Cattell, R. B. & Butcher, H. J. (1968). *The prediction of achievement and creativity*. IN: Bobbs-Merrill.
- Chand, I., & Runco, M. A. (1993). Problem finding skills as components in the creative process. *Personality & Individual Differences*, 14, 155-162.
- Christensen, P. R., Merrifield, P. R., & Guilford, J. P. (1958). *Consequences*. Beverly Hills, CA: Sheridan Supply Company.
- Christensen, P. R., Guilford, J. P., Merrifield, P. R., & Wilson, R. C. (1960). *Alternate uses*. Beverly Hills, CA: Sheridan Psychological Services.
- Christensen, P. R., Guilford, J. P., & Wilson, R. C. (1957). Relations of creative responses to working time and instructions. *Journal of Experimental Psychology*, 53, 82-88.
- Clark, P. M., & Mirels, H. L. (1970). Fluency as a Pervasive Element in the Measurement of Creativity. *Journal of Educational Measurement*, 7, 83-86.
- Covington, M. (2009). Idea density — A potentially informative characteristic of retrieved documents. *IEEE Southeastcon 2009*, 201.

- Cramond, B. (1994). We can trust creativity tests. *Educational Leadership*, 52, 70-71.
- Cropley, A. J. (2000). Defining and Measuring Creativity: Are Creativity Tests Worth Using? *Roeper Review*, 23, 72.
- Csikszentmihalyi, M., & Getzels, J. W. (2014). Discovery-Oriented Behavior and the Originality of Creative Products: A Study with Artists. *Systems Model of Creativity*, 1.
- Csikszentmihalyi, M., & Wolfe, R. (2005). Conceptions and research approaches to creativity: Implications of a system perspective approach to creativity in education. In K. Heller, F. Mönks, R. Sternberg, & R. Subotnik (Eds.), *International handbook of giftedness and talent*, (pp. 81-93). Oxford, UK: Elsevier Science Ltd.
- Davis, G. A. (1997). Identifying creative students and measuring creativity. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 269–281). Needham Heights, MA: Viacom.
- Dollinger, S. J., & Shafran, M. (2005). Note on the Consensual assessment technique in creativity research. *Perceptual and Motor Skills*, 100, 592–598.
- Dollinger, S. J., Urban, K. K., & James, T. A. (2004). Creativity and openness: Further validation of two creative product measures. *Creativity Research Journal*, 16, 35-47.
- Estes, Z., & Ward, T. B. (2002). The emergence of novel attributes in concept modification. *Creativity Research Journal*, 14, 149.
- Eysenck, H. J. (2003). Creativity, personality and the convergent-divergent continuum. In M. A. Runco (Ed.), *Critical creative processes*. (pp. 95-114). Cresskill, NJ, US: Hampton Press.
- Fasko, D. (1999). Associative theory. In M. A. Runco & S. Pritzker (Eds.), *Encyclopedia of creativity* (Vol. I; pp. 135-139). San Diego, CA: Academic Press.
- Fisher, M. (1996). *The IdeaFisher: How to land that big idea--and other secrets of creativity in*

- business*. Princeton, NJ: Peterson's/Pacesetter Books.
- Freud, S. (1900). *Interpretation of dreams*. London, England: Hogarth Press.
- Getzels, J.W., & Jackson, P.W. (1962). *Creativity and intelligence*. NY: John Wiley & Sons, Inc.
- Gough, H. G. (1976). Studying creativity by means of word association tests. *Journal of Applied Psychology*, 61, 348-353.
- Guilford, J.P. (1950). Creativity. *American Psychologist*, 5, 444-454.
- Guilford, J. P. (1967). *The nature of human intelligence*. NY: McGraw-Hill.
- Guilford, J.P. (1968). *Creativity, intelligence, and their educational implications*. San Diego, CA: EDITS/Knapp.
- Guilford, J. P. (1988). Some changes in the structure-of-intellect model. *Educational and Psychological Measurement*, 48, 1-4.
- Guilford, J. P. & Hoepfner, R. (1971). *The analysis of intelligence*. New York, NY: McGraw-Hill.
- Guilford, J. P., Christensen, P. R., Merrifield, P. R., & Wilson, R. C. (1978). *Alternate uses: Manual of instructions and interpretation*. Orange, CA: Sheridan Psychological Services.
- Harrington, D. M. (1975). Effects of explicit instructions to 'be creative' on the psychological meaning of divergent thinking test scores. *Journal of Personality*, 43, 434-454.
- Heausler, N. L., & Thompson, B. (1988). Structure of the Torrance Tests of creative thinking. *Educational and Psychological Measurement*, 48, 463-468.
- Hébert, T. P., Cramond, B., Neumeister, K. L. S., Miller, 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).
- Hébert, J. R., Peterson, K. E., Hurley, T. G., Stoddard, A. M., Cohen, N., Field, A. E., &

- Sorensen, G. (2001). Original report: The effect of social desirability trait on self-reported dietary measures among multi-ethnic female health center employees. *Annals of Epidemiology*, 11, 417-427.
- Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of Personality Assessment*, 45, 450.
- Hocevar, D., & Michael, W. B. (1979). The effects of scoring formulas on the discriminant validity of Tests of Divergent Thinking. *Educational & Psychological Measurement*, 39, 917-921.
- Kaufman, J. C., Baer, J., & Cole, J. C. (2009). Expertise, Domains, and the Consensual Assessment Technique. *Journal of Creative Behavior*, 43, 223-233.
- Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and non-expert raters using the consensual assessment technique. *Creativity Research Journal*, 20(2), 171-178.
- Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. New York, NY: Wiley.
- Kaufman, J. C., & Sternberg, R. J. (Eds.) (2010). *The Cambridge handbook of creativity*. New York, NY: Cambridge University Press.
- 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.

- Kim, K. H. (2011). The APA 2009 Division 10 debate: Are the Torrance Tests of Creative Thinking still relevant in the 21st century? *Psychology of Aesthetics, Creativity, and the Arts*, 5, 302-308.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- Kogan, N., & Pankove, E. (1974). Long-term predictive validity of divergent-thinking tests: Some negative evidence. *Journal of Educational Psychology*, 66, 802-810.
- Kozbelt, A. (2005). Factors affecting aesthetic success and improvement in creativity: A case study of the musical genres of Mozart. *Psychology of music*, 33, 235-255.
- Kris, E. (1952). *Psychoanalytic explorations in art*. New York, NY: International Universities Press.
- Ludwig, A. M. (1995). *The price of greatness: Resolving the creativity and madness controversy*. New York, NY: Guilford Press.
- MacKinnon, D.W. (1978). *In search of human effectiveness: Identifying and developing creativity*. Buffalo, NY: Creative Education Foundation.
- Martindale, C. (1975). *Romantic progression: The psychology of literary history*. Washington, DC: Hemisphere.
- Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69, 220-232.
- Mednick, S. A. (1968). The Remote Associates Test. *The Journal of Creative Behavior*, 2, 213-214.
- Mednick, S. A., & Mednick, M. T. (1967). *Examiner's manual: Remote Associates Test*. Boston, MA: Houghton Mifflin.

- Paletz, S. B. F., Peng, K., & Li, S. (2011). In the world or in the head: Internal and external implicit theories of creativity. *Creativity Research Journal*, 23, 83-93.
- Perkins, D. N. (1994). Creativity: Beyond the Darwinian paradigm. In M. A. Boden (Ed.), *Dimensions of creativity* (pp. 119-142). Cambridge, MA: MIT Press/Bradford Books.
- 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. *Roepers Review*, 21, 36-39.
- Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 48-73). New York, NY: Cambridge University Press.
- Rapp, R. (2002). The computation of word associations: Comparing syntagmatic and paradigmatic approaches. In *Proceedings of the 19th International Conference on Computational Linguistics*, Howard International House and Academia Sinica, Taipei, 24 August -1 September (Vol. 1; pp. 1-7). Stroudsburg, PA: Association for Computational Linguistics
- Rothenberg, A. (1973). Word association and creativity. *Psychological Reports*, 33, 3-12.
- Rothenberg, A. (1986). Artistic creation as stimulated by superimposed versus combined-composite visual images. *Journal of Personality and Social Psychology*, 50, 370-381.
- Runco Creativity Assessment Battery (rCAB). (2011). Creativity Testing Services: CTS. <http://creativitytestingservices.com/>
- Runco, M. A. (1986). Maximal performance on divergent thinking tests by gifted, talented, and non-gifted children. *Psychology in the Schools*, 23, 308-315.
- Runco, M. A. (1989). The creativity of children's art. *Child Study Journal*, 19, 177-189.

- Runco, M.A. (1999). Divergent thinking. In M.A. Runco & S. Pritzker (Eds.), *Encyclopedia of creativity* (Vol. I; pp. 557-582). San Diego, CA: Academic Press
- Runco, M. A. (2007) *Creativity: theories and themes: research, development, and practice*. San Diego, CA: Academic Press.
- Runco, M. A. (2010a). Testing creativity. *International Encyclopedia of Education*, 170-174.
- Runco, M. A. (2010b). Parsimonious creativity and its measurement. In E. Villalba (Ed.), *Proceedings of European Council meeting on creativity and innovation*, 393-405. Brussels, Belgium.
- 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. (2014). *Creativity: Theories and themes: Research, development, and practice* (2nd ed.). San Diego, CA, US: Elsevier 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., & 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., & Chand, I. (1994). Problem finding, evaluative thinking, and creativity. In M. A. Runco (Ed.), *Problem finding, problem solving, and creativity* (pp. 40-76). Westport, CT, US: Ablex Publishing.
- Runco, M. A., Illies, J. J., & Eisenman, R. (2005). Creativity, originality, and appropriateness: What do explicit instructions tell us about their relationships? *Journal of Creative Behavior*, 39, 137-148.

- Runco, M. A., Kaufman, J. C., Halladay, L. R., & Cole, J. C. (2010). Change in reputation as index of genius and eminence. *Historical Methods*, 43, 91-96.
- Runco, M. A., McCarthy, K. A., & Svensen, E., (1994). Judgments of the creativity of artwork from students and professional artists. *Journal of Psychology*, 128, 23-31.
- Runco, M. A., & Mraz, W. (1992). Scoring divergent thinking tests using total ideational output and a creativity index. *Educational and Psychological Measurement*, 52, 213–221.
- 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., Plucker, J. A., & Lim, W. (2001). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal*, 13, 393-400.
- Runco, M. A., & Smith, W. R. (1992). Interpersonal and intrapersonal evaluations of creative ideas. *Personality and Individual Differences*, 13, 295-302.
- Runco A. M., & Turkman, B., & Acar, S., & Nural V. M. (in press). Divergent thinking in written works. *The Journal of Genius and Eminence*.
- Runco, M. A., & Vega, L. (1990). Evaluating the creativity of children's ideas. *Journal of Social Behavior & Personality*, 5, 439-452.
- Ruscio, J., Whitney, D. M., & Amabile, T. M. (1998). Looking inside the fishbowl of creativity: Verbal and behavioral predictors of creative performance. *Creativity Research Journal*, 11, 243-263
- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., & Richard, C. A. (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.

- Skinner, B. F. (1970). Creating the creative artist. In A. J. Toynbee (Ed.), *On the future of art* (pp. 61-75). New York, NY: Viking Press.
- Snyder, A., Mitchell, J., Bossomaier, T., & Pallier, G. (2004). The creativity quotient: An objective scoring of ideational fluency. *Creativity Research Journal*, *16*, 415-419.
- Speecher, T. B. (1964). Creativity and individual difference in criteria. In C. W. Taylor (Ed.), *Widening horizons in creativity* (pp. 336-350). New York: Wiley.
- Stein, M. I. (1953). Creativity and culture. *Journal of Psychology*, *36*, 31–322.
- Sternberg, R. J., & Kaufman, J. C. (2010). *The Cambridge handbook of creativity*. New York, NY: Cambridge University Press.
- Taft, R., & Rossiter, J. R. (1966). The Remote Associates Test: Divergent or convergent thinking? *Psychological Reports*, *19*, 1313–1314.
- Taylor, C. W. (1988). Various approaches to and definitions of creativity. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 99-121). New York, NY: Cambridge University Press.
- Torrance, E.P. (1966). *Torrance tests of creative thinking: Directions manual and scoring guide*. Princeton, NJ: Personnel Press.
- Torrance, E. P. (1974). *Torrance tests of creative thinking*. Lexington, MA: Personnel Press.
- Torrance, E. P. (1977). *Discovery and nurturance of giftedness in the culturally different*. Reston, VA: Council on Exceptional Children.
- Torrance, E. P. (1981). *Thinking creatively in action and movement*. Bensenville, IL: Scholastic Testing Service
- Torrance, E. P. (1995). *Why fly?* Norwood, NJ: Ablex.
- Torrance, E. P. (2002). *The manifesto: A guide to developing a creative career*. Westport, CT: Ablex.

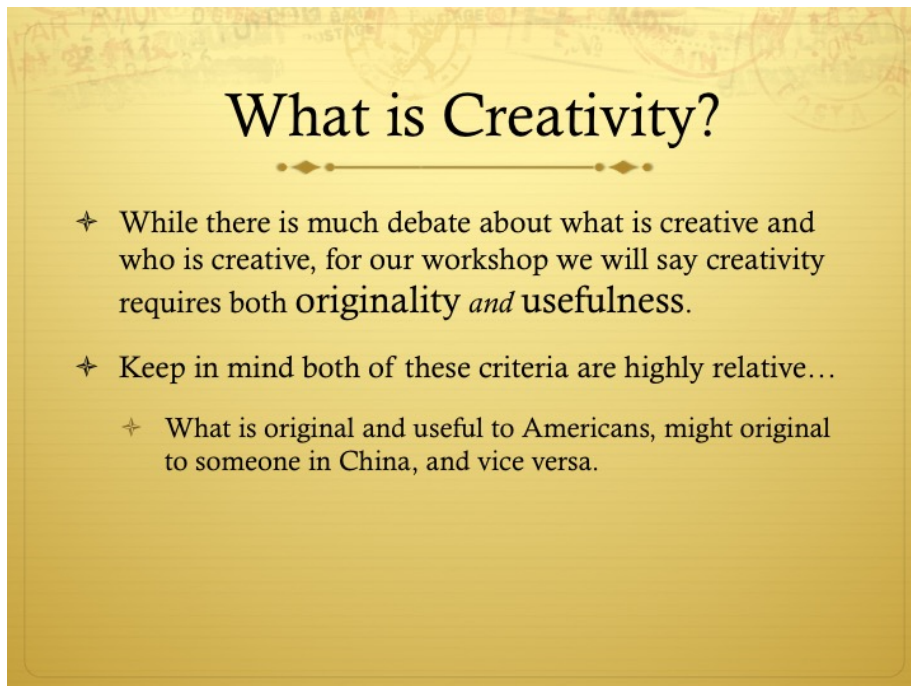
- Torrance, E. P. (2008). *The Torrance Tests of Creative Thinking Norms-Technical Manual Figural (Streamlined) Forms A and B*. Bensenville, IL: Scholastic Testing Service.
- Torrance, P., & Ball, O. (1984). *The Torrance Test of Creative Thinking. Streamlined (revised) scoring guide figural A and B*. Bensenville, IL: Scholastic Testing Service, Inc.
- Torrance, P., Ball, O. & Safter, H. T. (1992). *Torrance Test of Creative Thinking. Streamlined scoring guide figural A and B*. Bensenville, IL: Scholastic Testing Service, Inc.
- Torrance, E. P., Khatena, J., & Cunningham, B. F. (1973). *Thinking creatively with sounds and words*. Lexington, MA.
- Turner, A., & Greene, E. (1977). *The construction and use of a propositional text base* (pp. 77-63). Boulder, Colorado: Institute for the Study of Intellectual Behavior, University of Colorado.
- Urban, K.K., & Jellen, H.G. (1996). *Test for Creative Thinking - Drawing Production (TCT-DP)*. Lisse, Netherlands: Swets and Zeitlinger.
- Wallach, M. A. (1976). Tests tell us little about talent. *American Scientist*, 64, 57-63.
- Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction*. New York, NY: Holt, Rinehart and Winston.
- Wallach, M. A., & Wing, C. (1969). *The talented student*. New York, NY: Holt, Rinehart & Winston.
- Williams, F. (1980). *Creativity assessment packet*. Buffalo, NY.
- Williams, R. L. (1999). Operational definitions and assessment of higher-order cognitive constructs. *Educational Psychology Review*, 11, 411-427.
- Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin*, 50, 362-370.

Wilson, R. C, Guilford, J. P., Christensen, P. R., & Lewis, D. J. (1954). A factor-analytic study of creative-thinking abilities. *Psychometrika*, 19, 297-311.

APPENDICES

APPENDIX A- THE MANY USES GAME

Divergent Thinking Test – rCAB 2011 Copyright Creativity Testing Service
<p style="text-align: center;">“Many Uses” Game</p> <p>People typically use everyday items for specific purposes. Often there are alternative uses for the same objects. For example, a newspaper could be used as a hat or a blanket, and many other things. For the following everyday items, list as many alternative uses as you can. The more uses you think of, the better. Do not worry about spelling.</p> <p><i>Uses for a Bowl -</i></p> <p><i>Uses for a Tire –</i></p> <p><i>Uses for a Nail -</i></p>



Creativity Testing

Researchers use the following “tests” to try to understand someone or something’s creativity:

- ✦ Divergent Thinking Tests
- ✦ Torrance Tests of Creative Thinking
- ✦ Consensual Assessment Technique
- ✦ Remote Associates Test

These tests are a little different than the tests you are used to taking...let’s take a look

Divergent Thinking Tests

Researchers use Divergent Thinking Tests to try to understand and measure if someone or something is creative. The questions are open-ended and really try to understand the originality of an idea.

For example,

Think of all the possible uses for a bowl....

Did you say things like for cereal or for yogurt? Those are good ideas, but not original ideas. If you said something like a hat that would be original.



Divergent Thinking Tests Scoring

Divergent Thinking tasks are scored often for the ideas of fluency, flexibility, originality, and elaboration.

- ✦ **Fluency** measures the number of ideas.
- ✦ **Originality** is number of unseen or unique ideas.
- ✦ **Flexibility** measures the diversity of ideas from different categories.
- ✦ **Elaboration** measures the extent to which details are given about ideas.

Divergent Thinking Tests Relationships

- ✦ Originality usually comes with fluency (# of ideas).
 - ✦ The higher fluency → the higher the originality
 - ✦ This makes sense. People who generate a greater number of ideas tend to have more original ideas
- ✦ Flexibility also brings originality.
 - ✦ People who have diverse ideas from different categories tend to have more original ideas.

Divergent and Convergent Thinking

- ✦ **Divergent Thinking:** Thinking that concentrates on producing a large number of original or unexpected ideas.
 - ✦ Remember we were talking about uses for a bowl...
- ✦ **Convergent Thinking:** Thinking that involves finding the single best answer to a problem or question.
 - ✦ This would be a question like on Jeopardy. For example, what year did the Titanic sink? (1912) There is only one correct answer for a convergent thinking question.

Subjective and Objective Scoring



Subjective means that something is scored based on a judgment. When you were grading those essays earlier you were subjectively scoring them.

Objective means something is scored based on quantitative measures. For example, we could count the number of words in each essay. That would be an objective score.

How tests of creativity (and other subjects too) are scored is important. Be mindful of these differences as we move along.

Creativity Online Workshop

Key Points

- ✦ A product must be *original and useful* to be creative
- ✦ Originality comes with fluency and flexibility
- ✦ Subjective scoring is vital in creativity assessment

APPENDIX C- THE ATTITUDES AND VALUES SCALE

Attitudes and Values

Part of the *Runco Creativity Assessment Battery* (rCAB)©2012

Directions: Use the A-E scale (given below) to indicate how much you agree or disagree with a certain statement. You may need to approximate. Please indicate how you really think and behave, not how you would like to. Remember--no names are used. All of your individually identifiable information are confidential. Again, you may need to approximate. For each item, circle the response option that is THE CLOSEST to being accurate. Here are the options:

1= totally DISAGREE

2 = mostly disagree

3= neutral

4 = mostly agree

5 = totally AGREE

To what degree to you agree with each of the following?

1. Even if some method has worked well in the past, it is a good idea to question and perhaps change it on a regular basis.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

2. One of the advantages of developing expertise is that you can make useful assumptions and work more quickly.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

3. Time is often wasted when everyone involved in a project shares each of his or her own ideas.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

4. Diversity is a good thing to have in an organization that wants to be innovative.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

5. When solving problems it is often beneficial to postpone judgment about possible solutions.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

6. Maybe it is good for mad scientist to be strange, but for the rest of us its best to go along with the crowd.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

7. Solutions and ideas are often improved by considering a variety of perspectives.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

8. It isn't enough to just find an original idea. That idea is only worth something if you test it, verify it, implement it.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

9. If you produce a large number of ideas, you are likely to find some high quality ideas and solutions.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

10. Problem solving and innovation benefit from shifts in perspective.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

11. It can be useful to collect data and obtain new information before solving a problem.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

12. Any group working on all projects should have a person of authority who constantly insures that no time is wasted exploring every option.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

13. It is best to stick with a “tried and true” approach to innovation, once you find something that works.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

14. Good insights often result from concentrating on a problem. It is best not to take time off when immersed in a project.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

15. Useful ideas can often be found if you change the problem; don't just look for solutions to the problem as it is presented.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

16. I look for ways to isolate myself so I can concentrate and think deeply about my work.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

17. There is clear benefit to thinking about ideas that other people will not consider.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

18. I avoid working outside my area of expertise. I do not want to be a beginner again and again.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

19. The important thing at work is to find out what will gain the approval of other people (supervisors, co-workers, clients).

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

20. It is useful to tolerate people who have different views, even if we are trying to solve a particular problem.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

21. It is difficult for me to work with people who have very different backgrounds.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

22. Work can be fun if you approach projects playfully, like they are games and have fun.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

23. Originality can be very useful at work or in school.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

24. Sometimes it is best to be unconventional.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree

25. I am tolerant of people who are different, bohemian, contrarian, odd.

(1) Totally Disagree (2) Mostly Disagree (3) Neutral (4) Mostly Agree (5) Totally Agree