

Which Test of Divergent Thinking Is Best?

Mark A. Runco

University of Georgia, Athens, GA USA

E-mail address: runco@uga.edu

Sue Hyeon Paek

University of Georgia, Athens, GA USA

E-mail address: suehpaek@gmail.com

Hanadi N. Alsuwaidi

Ministry of Education, UAE

E-mail address: h.n.alsuwaidi@hotmail.com

Ahmed M. Abdulla

Arabian Gulf University, Kingdom of Bahrain

E-mail address: ahmedy82@uga.edu

Fatima A. Al-Jasim

Arabian Gulf University, Kingdom of Bahrain

E-mail address: gazran5552@gmail.com

ARTICLE INFO

Keywords:

Creativity
Divergent Thinking
Originality
Creativity Testing
Uses test
Problem Finding
Creative Potential
Runco Ideational Behavior Scale
Quick Estimate of Convergent Thinking
Standard Definition of Creativity

Article history:

Received 22 November, 2015

Received in revised form 16 February 2016

Accepted 18 February 2016

ISSN: 2354-0036

DOI: 10.1515/ctra-2016-0001

ABSTRACT

Divergent thinking (DT) tests are probably the most commonly used measures of creative potential. Several extensive batteries are available but most research relies on one or two specific tests rather than a complete battery. This may limit generalizations because tests of DT are not equivalent. They are not always highly inter-correlated. Additionally, some DT tests appear to be better than others at eliciting originality. This is critical because originality is vital for creativity. The primary purpose of the present study was to determine which test of DT elicits the most originality. Seven measures of DT were administered on a sample of 611 participants in eight Arabic countries. The tests were Figural, Titles, Realistic Presented Problems, Realistic Problem Generation, Instances, Uses, and Similarities. The *Quick Test of Convergent Thinking*, *Runco's Ideational Behavior Scale*, and a demographic questionnaire were also administered. A linear mixed model analysis confirmed that the originality scores in the DT tests differed by test. Post-hoc tests indicated that the *Titles* and *Realistic Problem Generation* tests produced the highest mean originality scores, whereas the *Realistic Presented Problems* test produced the lowest mean originality scores. These differences confirm that research using only one DT test will not provide generalizable results.

Creative behaviour has been recognized as among the most important forms of human capital. Its role in innovation, invention, design and advance in a wide range of domains is now broadly recognized. Efforts to enhance creative skills are on the rise, as are efforts to accurately assess creative potential and creative behaviour.

The difference between creative *potential* and actual creative *behaviour* should be recognized in all efforts to enhance or assess creativity (Runco, 2008). The former is latent, sometimes difficult to see and widely distributed. The latter takes the form of overt action, observable performances and socially-recognized accomplishments. Certainly the former is required for the latter, which is where enhancement efforts come in. Most enhancement efforts can be seen as attempts to fulfill potentials such that actual performance is likely.

The most widely-used assessments for creative potential require divergent thinking (DT). This sometimes involves an actual divergence of thought, with associative paths exploring a variety of directions, conceptual spaces and possibilities (Acar & Runco, 2014, 2015), although most of the time DT is operationalized more simply in terms of the number, originality and flexibility of ideas produced (Guilford, 1968; Torrance, 1995).

Various tests of DT are available, including those of Guilford (1968), who coined the phrase *divergent production*, Torrance (1962, 1995), Milgram (1990) and Wallach and Kogan (1965). These are DT test *batteries* in that each is actually a set of tasks rather than one single test. Wallach and Kogan, for example, had Instances, Uses, Similarities, Line Meanings and Pattern Meanings tests. The use of multiple tests follows in part from theories of creativity, and also from psychometric theory, which posit that the most reliable assessments are based on multiple indicators (Anastasi, 1988; Cronbach, 1989).

The primary purpose of the present study was to determine which test of DT elicits the most originality. This objective was deemed to be important because (1) originality is the key component of creativity (see Runco & Jaeger, 2012, for a review); (2) enhancement efforts may be the most useful if differences among tests are recognized; and (3) so many different tests of DT are being used in the research, it would be useful to know which is best specifically for originality. It is possible that different DT tests would elicit approximately the same level of originality, in which case it would be reasonable to use one test rather than a battery. The results from that one test would generalize to other DT tests. More likely, originality will vary from test to test, in which case results from any one test should not be interpreted as indicative of results from other tests of DT. In that light, the present research is relevant to questions of generalizability and to research findings that have used specific tests.

METHOD

Participants

The data were collected from 611 participants from three different universities: the Arabian Gulf University and Ahlia University, Kingdom of Bahrain, and the University of Sharjah, the United Arab Emirates. The majority of participants were undergraduate students

(86.4%). 3.3% reported studying or holding a higher diploma degree, 8.5% reported studying or holding a Master's degree, and 1.8% reported studying or holding a higher doctoral degree. Eight departments were represented: Health and Medicine (36.5%), Science, Maths, and Technology (20.7%), Public Relations and Media (10.5%), Education (10%), Business (7.7%), Social Sciences (5.9%), Law (5.3%), Arts & Humanities (3.3%) and others (0.2%). Participants were from more than eight different nationalities including: UAE (34.5%), Kuwait (11.7%), Saudi Arabia (11.7%), Bahrain (9.5%), Syria (7.1%), Jordan (5.7%), Palestine (3.9%), Iraq (3%) and others (13%). The mean age was 21.5 years with a range from 17 to 48, and most were women (71.6%). The survey (see Instruments, below) asked about previous experience, knowledge about creativity and divergent thinking tests. Approximately one third of participants have taken creativity workshop or classes on creativity ($n = 179$, 29.4%) or taken creativity tests before ($n = 213$, 35%).

Instruments

Seven measures of divergent thinking were used: *Figural DT*, *Titles*, *Realistic Presented Problems*, and *Realistic Problem Generation tests from Runco Creativity Assessment Battery (rCAB)*, the *Instances*, *Uses*, and *Similarities tests from Wallach and Kogan (1965)*. These tests are all regularly used in the published research and were chosen such that a wide range of DT tests was covered. The *Quick Estimate of Convergent Thinking (QECT; www.creativitytestingservices.com)* and *Runco's Ideational Behavior Scale (RIBS; Runco, Plucker, & Lim, 2000-2001)* were also given, as was a demographic questionnaire.

The first DT test was adapted from the visual tests of Wallach and Kogan (1965). In the present investigation three visual or "figural DT" items were used. Unlike the *Torrance Test of Creative Thinking-Figural Form*, on the Wallach and Kogan (1965) and rCAB figural tests participants are not required to draw anything. Instead, they are merely asked to look at the figure and list possible representations. The verbatim directions were:

Look at the figure below. What do you see? List as many things as you can that this figure might be or represent. This is NOT a test. Think of this as a game and have fun with it! The more ideas you list, the better.

The figural DT tests (and all tests of DT used in this research) were scored for originality following the methods of Wallach and Kogan (1965) and used many times since (e.g., Runco, Illies, & Reiter-Palmon, 2005). There are alternative methods for scoring originality, as can be seen in the 43 chapters of the two edited volumes on DT (Runco, 1991, 2013). DT tests can be scored by examining "ideational pools," where all ideas from any one examinee are presented in their entirety to judges, for example, or they can be scored statistically and objectively using various cutoffs (e.g., only unique ideas are origi-

nal, or those given by 1%, 2%, 5%, or even 10% of a sample). The decision about which is the best scoring method takes the sample size into account, as well as the relationship of fluency and originality. If too liberal a definition of originality is used, originality scores sometimes approach fluency scores, which indicates that there is poor discrimination. The present study used the 5% cutoff. This determination was made after a lexicon was made for each DT task. These lexicons alphabetically and conceptually organized all ideas given for any one DT task and indicated how many respondents gave each individual idea.

The rCAB *Titles DT Test* was adapted from one of Guilford's (1968) DT batteries. Guilford's Plot Titles tests presented a paragraph story and asked examinees to generate alternative titles. The Titles tests used here consisted of three tasks, each of which asks the participant to list alternative titles for a famous movie, play, and/or a book. Here "Titanic," "Twilight" and "Harry Potter" were used. As is the case for all DT instruments employed in this research, one question is presented at a time. For the *Titles* tasks, the verbatim directions were:

List alternative titles for each movie, play, and book listed below. Spelling does not matter and there are no grades for this. Have fun and list as many alternatives as you can.

The *Realistic Presented Problem (RPP)* test consisted of three tasks that asked about real-life situations (e.g., school and home). Students were presented with a problem and were asked to provide as many solutions as possible. Before presenting each problem, participants received directions and an example regarding what is exactly was required. The verbatim directions were:

On the next few pages, we will describe a few problems, which may occur at school and work. Your task is to first read about the problem and then try to write down as many solutions as you can for each problem.

The three RPP tests were originally used by Runco and Okuda (1988). Here is one example:

Your friend Pat sits next to you in class. Pat really likes to talk to you and often bothers you while you are doing your work. Sometimes he distracts you and you miss an important part of the lecture, and many times you don't finish your work because he is bothering you. What should you do? How would you solve this problem? Remember to list as many ideas and solutions as you can.

The *Problem Generation (PG) test* was also adapted from Runco and Okuda (1988). They went into detail about the importance of problem finding for the creative process (also see Csikszentmihalyi & Getzels, 1988; Mumford, Reiter-Palmon, & Redmond, 1994) and then described how DT tests could be prepared such that they focus on prob-

lem finding. These tests get at problem finding by asking examinees to generate as many problems as they can. The PG version used here contained three open-ended problems about home and school, living situations, and health and well-being. An example is as follows:

List problems with your friends, peers, schoolmates, or spouses (any individual of the same approximate age). These problems might be real, or they might be hypothetical and imaginary. Do not limit yourself; the more problems you can list, the better. (Do not worry about spelling, and take your time.)

Three tasks from Wallach and Kogan's (1965) Uses Test were administered: Uses for a toothbrush, Uses for a Tire, and Uses for a Spoon. The verbatim directions were:

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 items, list as many alternative uses as you can. The more uses you think of, the better. Do not worry about spelling.

Three tasks from Wallach and Kogan's (1965) Instances Test were administered. These directed participants to (a) "List all square things you can think of," (b) "List all round things you can think of" and (c) "List all the things you can think of that move in wheels."

Three tasks from Wallach and Kogan's (1965) Similarities Test were administered. These asked participants to list similarities for (a) an apple and a pear, (b) a car and a bike and (c) a computer and a phone. The three tasks directions were as follow: "For the following item, list as many similarities as you can. The more similarities you think of, the better. Do not worry about spelling."

The *Quick Estimate of Convergent Thinking* (QECT) was developed by Turkman and Runco (2013) as a way to ensure that tests of DT are in fact assessing DT and not biased by convergent thinking. In short it allows a check of discriminant validity, which is confirmed if correlations between it and DT are low or negative. The QECT consists of 18 items, each of which asks participants to identify a missing part of an image. Each image represents a very common item (e.g., a pencil without an erasure, a keyboard without a space bar, a car without a rear view mirror on the driver's side). The verbatim directions were: "Look at the pictures below. Try to find the missing part in each of them, then, write what you think is missing in the space below each picture." The QECT is scored by counting how many times the examinee correctly identifies the missing part within each image. The maximum score is therefore 18.0.

The RIBS is a self-report designed to assess ideation with a Likert format (Chand O'Neal, Paek, & Runco, 2015; Runco et al., 2000-2001). It is based on the belief that ide-

as can be treated as the products of creative thinking (Runco, 2013). A number of studies have demonstrated that the RIBS is statistically unrelated to GPA and is useful as a criterion of original and divergent thinking (Ames & Runco, 2005; Plucker, Runco, & Lim, 2006). For the present study, the RIBS contained 50 items. A Likert scale ranged from 0 to 4 with the following verbal anchors: 0 (never), 1 (approximately once a year), 2 (once or twice each month, approximately), 3 (once or twice each week, approximately) and 4 (just about every day, sometimes more than once each day). The verbatim directions for the RIBS were:

Use the 0-4 scale (given below) to indicate how often each of the phrases describes your thinking and behavior. You may need to approximate. Please indicate how you really think and behave, not how you would like to. Remember--no names are used. Your responses 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: 0 = Never; 1 = approximately once a year; 2 = once or twice each month (approximately); 3 = once or twice each week (approximately); 4 = Just about every day, and sometimes more than once each day.

Two example items from the RIBS are, "*Approximately how often do you...have ideas for making your work easier?*" and "*Approximately how often do you have ideas about what you will be doing in the future?*"

A demographic questionnaire asked about participants' age, gender, education, college major, nationality and previous exposure to creativity. The questions about previous exposure to creativity were (a) Have you ever taken a class that was focused on creativity? and (b) Have you ever taken a creativity or divergent thinking test?

Procedure

Before collecting the data, all instruments were translated from English into Arabic by the author who is a native speaker of Arabic. The translated versions of all instruments were then checked by two bilingual experts and minor corrections were made. A letter was sent to the three universities in order to obtain permission to collect data. Once permission was received, seven groups were randomly defined within the sample. Each group received the same number of tasks but in different orders, with the exception of the QECT, the RIBS and the demographic questionnaire, which were in the same order across groups. These were interspersed within the DT tests and used as distractor tasks to minimize fatigue effects on the DT tests. Every group received the study instruments in the following order: two DT tasks, the QECT, two DT tasks, a demographic questionnaire, two DT tasks, the first half of the RIBS, two DT tasks, the second half of the RIBS,

and one DT task. Thus, each group received nine DT tasks, the QECT, a demographic questionnaire and the RIBS.

The materials were given to participants as a package. The individual tests could not be given individually, even though that would have allowed more control over the time allocated to each test. Timing can undermine DT, and in fact it is likely to minimize the production of original ideas. If examinees know they are timed, they assume that they are taking a “test” and they follow academic habits and tendencies instead of thinking divergently (Wallach & Kogan, 1965). And if time is at all constrained, they are unable to follow associative paths very far and thus they do not discover the remote associates that are likely to be highly original (Mednick, 1962). For these reasons, and given our interest in originality, the choice was to de-emphasize time and allow the relaxed testing setting--where individual tasks were not timed. Table 1 shows the distribution of the DT tasks across the groups.

Table 1
The Distribution of the DT Tasks Across the Seven Groups

Group	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9
A	<i>Uses</i> Toothbrush	<i>Uses</i> Tires	<i>Uses</i> Spoon	<i>Titles</i> Titanic	<i>Figures</i> Conical	<i>R.P.P</i> Pat	<i>P.G</i> Friends, Peers, etc.	<i>Similarities</i> Apple & pear	<i>Instances</i> Square things
B	<i>Figures</i> Conical	<i>Figures</i> Squares	<i>Figures</i> Lines	<i>Uses</i> Toothbrush	<i>R.P.P</i> Kelly	<i>P.G</i> Living situ- ation	<i>Similarities</i> Car & bike	<i>Instances</i> Round things	<i>Titles</i> Twilight
C	<i>R.P.P</i> Pat	<i>R.P.P</i> Kelly	<i>R.P.P</i> The book	<i>Uses</i> Tires	<i>Figures</i> Squares	<i>Titles</i> Harry P.	<i>Similarities</i> Computer & phone	<i>Instances</i> Move in wheels	<i>P.G</i> Health Issues
D	<i>Titles</i> Titanic	<i>Titles</i> Twilight	<i>Titles</i> Harry P.	<i>Uses</i> Spoon	<i>Figures</i> Lines	<i>R.P.P</i> The book	<i>Similarities</i> Apple & pear	<i>Instances</i> Square things	<i>P.G</i> Friends, Peers, etc.
E	<i>Similarities</i> Apple & pear	<i>Similarities</i> Car & bike	<i>Similarities</i> Computer & phone	<i>Titles</i> Titanic	<i>Figures</i> Conical	<i>R.P.P</i> Pat	<i>Instances</i> Round things	<i>Uses</i> Toothbrush	<i>P.G</i> Living situ- ation
F	<i>Instances</i> Square things	<i>Instances</i> Round things	<i>Instances</i> Move in wheels	<i>Uses</i> Tires	<i>Figures</i> Squares	<i>R.P.P</i> Kelly	<i>Similarities</i> Car & bike	<i>Titles</i> Twilight	<i>P.G</i> Health Issues
G	<i>P.G</i> Friends, Peers, etc.	<i>P.G</i> Living situ- ation	<i>P.G</i> Health Issues	<i>Uses</i> Spoon	<i>Figures</i> Lines	<i>R.P.P</i> The book	<i>Similarities</i> Computer & phone	<i>Instances</i> Move in wheels	<i>Titles</i> Harry P.

Note. *Uses*: 9 tasks completed by the study participants (*Toothbrush* = Groups A, B, and E; *Tires* = Groups A, C, and F; *Spoon* = Groups A, D, and G); *Figures*: 9 tasks completed by the study participants (*Conical* = Groups B, A, and E; *Squares* = Groups B, C, and F; *Lines* = Groups B, D, and G) *R.P.P*: 8 tasks completed by the study participants (*Pat* = Groups C, A, and E; *Kelly* = Groups B and F; *The Book* = Groups C, D, and G); *Titles*: 9 tasks completed by the study participants (*Titanic* = Groups D, A, and E; *Twilight* = Groups D, B, and F; *Harry Potter* = Groups D, C, and G); *Similarities*: 9 tasks completed by the study participants (*Apple & Pear*= Groups E, A, and D; *Car & Bike* = Groups E, B, and F; *Computer & Phone* = Groups E, C, and G); *Instances*: 9 tasks completed by the study participants (*Square things* = Groups F, A, and D; *Round things* = Groups F, B, and E; *Move in Wheels* = Groups F, C, and G) *P.G*: 9 tasks completed by the study participants (Friend/Peers = Groups G, A, and D; Living Situation = Groups G, B, and E; Health = Groups G, C, and F).

Each participant was asked for his or her consent for the participation in this study. The administration time of completing all instruments was approximately 40 minutes. To encourage subjects to participate in this study, the instructors did not lecture on the day of data collection.

Results

The first analysis checked the inter-item reliability of the tests using Cronbach's alpha. The groups, which completed three items in the tests were selectively used in testing Cronbach's alpha. Pairwise deletion of missing data was used in this analysis, and when possible, for all analyses, to maximize samples sizes. Recall here that the data collection maximized the number of tests that could be compared by staggering the tests across the different groups. Table 2 presents the sample size for each test ($77 \leq ns \leq 90$). Table 2 indicates that the fluency and originality scores for the Instances, Similarities, Titles, Realistic Problem Generation and Realistic Presented Problems were adequately reliable. The Figural DT and Uses tests had very low alphas and were omitted from further consideration. Alpha for the RIBS was .90.

Table 2
Reliability of Fluency and Originality Indices within Groups

Test	<i>n</i>	Valid <i>n</i>	Fluency	Originality
Figures	90	87	.56	.52
Instances	84	81	.84	.68
Uses	90	90	.49	.52
Similarities	90	70	.89	.81
Titles	90	45	.72	.74
RPG ^a	77	55	.79	.81
RPP ^{bc}	90	87	.64	.64

Note. *n* = sample sizes of groups. ^a Realistic Problem Generation. ^b Realistic Presented Problem. ^c RPP test has only two items for group C. All other groups have only one item of RPP. Thus, reliability of RPP is tested using only two items in group C.

The linear mixed models reported below have slightly larger cell sizes than the reliability analyses because the latter included only scores when a participant received three items, while the linear mixed models included scores even if a participant had received one item for any one particular test.

Differences Among DT Tests

Linear mixed model analysis was used to test differences among the DT tests. This is similar to a repeated-measure ANOVA in terms of ruling out dependency among

scores from any one individual. Linear mixed model analysis was chosen because it can use pairwise deletion in handling missing data, whereas repeated-measure ANOVA requires listwise deletion. The mixed linear model analysis confirmed that the originality scores in the DT tests differed, even after score dependency was controlled, $F(6, 1278.9) = 130.166, p < .001$. Omega squared showed that 10.99% of the variance in DT scores was attributed to the difference among the DT tests (Xu, 2003), which was a medium effect size (Kirk, 1996). Table 3 gives the mean scores for each test.

Table 3
Means and Standard Deviations of Fluency and Originality Indices

DT tests	Valid <i>n</i>	Fluency	Originality
Figural	600	4.16 (2.45)	1.98 (2.04)
Instances	586	6.81 (4.07)	2.20 (2.55)
Uses	568	3.66 (2.26)	1.84 (1.74)
Similarities	516	4.72 (2.69)	1.88 (1.90)
Titles	484	3.60 (2.19)	2.96 (2.17)
RPG ^a	545	4.08 (2.91)	3.25 (2.58)
RPP ^b	589	3.26 (1.56)	0.89 (1.00)

Note. Standard deviations are presented in parentheses. ^a Realistic Problem Generation ^b Realistic Presented Problem

For the post-hoc evaluation, paired sample *t*-tests were used to identify which pair of DT tests showed significant mean differences. The significance level was adjusted by Bonferroni correction ($\alpha = .002$) to take into account the familywise error rate. Results indicated that the Titles and Realistic Problem Generation tests produced the highest mean originality scores, whereas the Realistic Presented Problems test produced the lowest originality. Although the mean originality score in the Titles test was slightly higher than in the Realistic Problem Generation test, the difference was not significant. Table 4 presents the results of paired sample *t*-tests and effect sizes (Cohen, 1988).

Table 4
Mean Differences in Originality Scores by DT tests

Pairs of DT tests	ΔM (SE)	t	df	p	d	Effect Size
Figural vs. Instances	-0.22 (0.11)	-2.00	576	.046	-0.17	Small
Figural vs. Uses	0.15 (0.09)	1.65	559	.099	0.14	Small
Figural vs. Similarities	0.10 (0.11)	0.97	506	.332	0.09	Small
Figural vs. Titles	-0.89 (0.12)	-7.72*	478	.000	-0.7	Large
Figural vs. RPG	1.11 (0.09)	12.86*	579	.000	1.07	Large
Figural vs. RPP	-1.29 (0.12)	-10.42*	534	.000	-0.9	Large
Instances vs. Uses	0.40 (0.11)	3.67*	546	.000	0.31	Medium
Instances vs. Similarities	0.32 (0.12)	2.71	494	.007	0.24	Medium
Instances vs. Titles	-0.57 (0.13)	-4.50*	466	.000	-0.47	Medium
Instances vs. RPG	1.32 (0.11)	12.10*	567	.000	1.02	Large
Instances vs. RPP	-1.13 (0.13)	-8.47*	526	.000	-0.74	Large
Uses vs. Similarities	-0.08 (0.09)	-0.92	483	.359	-0.08	Small
Uses vs. Titles	-1.07 (0.11)	-9.57*	452	.000	-0.9	Large
Uses vs. RPG	0.94 (0.08)	12.02*	549	.000	1.03	Large
Uses vs. RPP	-1.45 (0.12)	-11.81*	509	.000	-1.05	Large
Similarities vs. Titles	-0.98 (0.13)	-7.65*	406	.000	-0.76	Large
Similarities vs. RPG	0.99 (0.09)	10.95*	499	.000	0.98	Large
Similarities vs. RPP	-1.38 (0.13)	-10.42*	465	.000	-0.97	Large
Titles vs. RPG	2.02 (0.10)	19.84*	468	.000	1.83	Large
Titles vs. RPP	-0.34 (0.14)	-2.39	438	.017	-0.23	Medium
RPG vs. RPP	-2.39 (0.12)	-20.58*	527	.000	-1.79	Large

Note. d = Cohen's d , RPG = Realistic Problem Generation, RPP = Realistic Presented Problem * $p \leq .002$

Criterion-Related Validity

The various indices of DT were correlated with two criterion variables, namely the QECT and the RIBS. The expectation was that DT would be unrelated to the QECT, which would support the discriminant validity of DT, and related to the RIBS, which would support criterion related validity. Results indicated that several of the correlations were statistically significant, but low ($< .11$). Very importantly, psychometric tests of validity do not depend on statistical significance. The coefficients themselves are interpreted, much like an effect size, in psychometric theory. Thus, even with a few correlations above $.11$, and thus statistically significant, none of the correlations was indicative of validity in psychometric theory. More specifically, the magnitude of the coefficients implies weak predictive (or criterion-related) validity. As a matter of fact, the highest coefficients were between Uses fluency ($r = .28$) and originality ($r = .23$) with the QECT. These indicate poor discriminant validity, at least for the Uses test.

Discussion

The premise of this investigation was that originality is vital for creativity. That is an easy premise to accept, given current definitions used in the literature for creativity. This investigation was designed to compare measures of DT in terms of originality. There are a variety of other tests of DT - only seven were compared here - and future research might examine DT tests that were not examined herein. The tests used here were Figural DT, Titles, Realistic Presented Problems, Realistic Problem Generation, Instances, Uses and Similarities. Even though not all possible tests were examined here, the seven chosen do represent a good start towards a better understanding of the best tests for originality. After all, several of the tests used here (e.g., Uses and Instances) are also included in most well-known batteries (Guilford, 1968; Torrance, 1972; Wallach & Kogan, 1965). Also note the range of DT tests, with verbal, figural, presented problems, discovered problems, similarities and so on. The seven tests represent a fairly wide range of DT testing options.

Results indicated that the Figural DT and Uses tests were less reliable than the other five DT tests. The poor showing of the Uses test may not be all that surprising. Previous research has also found it to be a questionable test of DT. Uses scores were, for example, also the lowest of various tests in the research of Harrington (1975), Hocevar (1979), and Okuda, Runco, and Berger (1991). It is possible that the poor reliability of Uses was due to cultural background. The question about “uses for a tire” showed particularly lower correlations with the other two Uses items: Uses for a tire with Uses for a toothbrush $.21$, Uses for a Tire and Uses for a spoon $.12$, Uses for a toothbrush and Uses for a spoon $.38$. This might reflect the fact that the majority of this sample was fe-

male (71.6%) and for cultural reasons, were not familiar with driving a car. One possible explanation for the low reliability of the Figural DT test is that Arabs rely heavily on verbal communication in expressing their thoughts. A comparative study between Arabic and English languages showed that the number of Arabic word-types is 76 % more than in English (Alotaiby, Foda, & Alkharashi, 2014). In addition, several Arabic studies of creativity have suggested a strong preference for verbal divergent thinking over figural (e.g. Khaleefa, Erdos, & Ashaia, 1996). This might also explain the findings of Kharkhurin and Motalleebi (2008). They studied creative potential in three different countries--the United States, Russia and Iran—and found that participants from Western culture, represented by the American and Russian participants, showed higher divergent thinking scores than participants from Eastern culture, represented by Iranian participants. Clearly more research is needed on the role of culture and on differences among DT tests.

Results indicated that the Titles and Realistic Problem Generation tests had the highest originality scores. There are various possible explanations for this. Runco and Albert (1985) have hypothesized that ideational originality is elicited by the degree of openness of a test, as if the openness or even ambiguity of the stimuli allow more divergence and facilitate the required associative processes. Okuda et al. (1991) raised another possibility, namely that certain tests are more interesting and engaging than others. If a test is engaging, examinees are more likely to actually think about the problem at hand in a mindful way, and this may very well translate into original ideation. Okuda et al. (1991) pointed out that realistic tests are more interesting, precisely because they have a connection to reality and the natural environment. Other research has pointed to experiential influences, the idea here being that some test questions are easy to relate to actual experience, and if such a connection is found, the examinee might find ideas in long-term memory rather than actually constructing an idea spontaneously while taking the test (Runco & Acar, 2010; Runco, Okuda, & Thurston, 1991; Runco, Dow, & Smith, 2006). In this light, tests that remind examinees about their experience are less likely to lead to original ideas than are tests that require spontaneous ideation.

The findings suggest that Titles and the Realistic Problem Generation DT tests might be the tests of choice in future research. But as noted above, future research should be conducted using even more DT tests. Perhaps the different explanations for why some tests elicit higher originality, outlined in the paragraph above (e.g., more open, more engaging, less experiential) could be empirically examined. New computer scoring systems for DT (Beketayev & Runco, 2015) might also help explain differences among tests, given that they provide semantic distances and other useful information. Certainly

different cultures should be sampled in future research on DT, and it would be interesting to compare tests in terms of their flexibility scores, as well as originality. Future research might employ different criteria of creativity given that the DT tests in the present investigation did not show good criterion-related validity, nor good discriminant validity. The present comparison of seven DT tests is a good start, but clearly there are various questions for future research along these lines.

REFERENCES

- Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal*, 26, 229-238.
- Acar, S., & Runco, M. A. (2015). Thinking in multiple directions: Hyperspace categories in divergent thinking. *Psychology of Aesthetics, Creativity, And The Arts*, 9, 41-53.
- Alotaiby, F., Foda, S., & Alkharashi, I. (2014). Arabic vs. English: Comparative statistical study. *Arabian Journal for Science & Engineering (Springer Science & Business Media B.V.)*, 39, 809-820.
- Ames, M., & Runco, M. A. (2005). Predicting entrepreneurship from ideation and divergent thinking. *Creativity and Innovation Management*, 14, 311-315.
- Anastasi, A. (1988). Explorations in human intelligence: Some uncharted routes. *Applied Measurement In Education*, 1, 207-213.
- Beketayev, K., & Runco, M. A. (in press). Semantics-based algorithmic method for assessing divergent thinking. *Europe's Journal of Psychology*.
- Chand O'Neal, I., Paek, S. H., & Runco, M. A. (2015). Comparison of competing theories about ideation and creativity. *Creativity. Theories – Research – Applications*. 2(2), 145-165.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn, R. L. Linn (Eds.), *Intelligence: Measurement, theory, and public policy: Proceedings of a symposium in honor of Lloyd G. Humphreys* (pp. 147-171). Champaign, IL, US: University of Illinois Press.
- Csikszentmihalyi, M., & Getzels, J. W. (1988). Creativity and problem finding. In F. G. Farley & R. W. Neperole (Eds.), *The foundations of aesthetics, art, and art education* (pp. 91–106). New York: Praeger.
- Guilford, J. P. (1968). *Creativity, intelligence, and their educational implications*. San Diego, CA: EDITS/Knapp.
- 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.

- Hocevar, D. (1979). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology, 71*, 191-196.
- Khaleefa, O. H., Erdos, G., & Ashria, I. H. (1996). Gender and creativity in an Afro-Arab Islamic culture: The case of Sudan. *The Journal of Creative Behavior, 30*, 52-60.
- Kharkhurin, A. V., & Samadpour Motalleebi, S. N. (2008). The Impact of culture on the creative potential of American, Russian, and Iranian college students. *Creativity Research Journal, 20*, 404-411.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746-759.
- Milgram, R. M. (1990). Creativity: An idea whose time has come and gone? In M. A. Runco & R. S. Albert (Eds.), *Theories of creativity* (pp. 215-233). Newbury Park, CA: Sage.
- Mumford M. D, Reiter-Palmon R, & Redmond M. R. (1994). Problem construction and cognition: Applying problem representations in ill-defined domains. In: Runco M (ed.) *Problem finding, problem solving, and creativity*, pp. 3-39. Norwood, NJ: Ablex.
- Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real-world problems. *Journal of Psychoeducational Assessment, 9*, 45-53.
- Plucker, J., Runco, M., & Lim, W. (2006). Predicting ideational behavior from divergent thinking and discretionary time on task. *Creativity Research Journal, 18*, 55-63.
- Runco, M. A. (Ed.). (1991). *Divergent thinking*. Norwood, NJ: Ablex Publishing Corporation.
- Runco, M. A. (2008). Creativity and education. *New Horizons in Education, 56*, 107-115.
- Runco, M. A. (Ed.). (2013). *Divergent thinking and creative potential*. Cresskill, NJ: Hampton Press.
- Runco, M. A., & Acar, S. (2010). Do tests of divergent thinking have an experiential bias? *Psychology of Art, Creativity, and Aesthetics, 4*, 144-148.
- 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., Dow, G., & Smith, W. R. (2006). Information, experience, divergent thinking: An empirical test. *Creativity Research Journal, 18*, 269-277.
- Runco, M. A., Illies, J. J., & Reiter-Palmon, R. (2005). Explicit instructions to be creative and original: A comparison of strategies and criteria as targets with three types of divergent thinking tests. *Korean Journal of Thinking and Problem Solving, 15*, 5-15.
- Runco, M. A., & Jaeger, G. (2012). The standard definition of creativity. *Creativity Research Journal, 24*, 92-96.
- Runco, M. A., & Okuda, S. M. (1988). Problem discovery, divergent thinking, and the creative process. *Journal of Youth and Adolescence, 17*, 211-20.

- Runco, M. A., Plucker, J. A., & Lim, W. (2000-2001). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal*, 13, 393-400.
- Runco, M. A., Okuda, S. M., & Thurston, B. J. (1991). Environmental cues and divergent thinking. In M. A. Runco (Ed.), *Divergent thinking* (pp. 79-85). Norwood, NJ: Ablex Publishing Corporation.
- Torrance E. P. (1972) Can we teach children to think creatively? *Journal of Creative Behavior*, 6, 114–143.
- Torrance, E. P. (1962). *Guiding creative talent*. Englewood Cliffs, NJ: Prentice Hall.
- Torrance, E. P. (1995). *Why fly?* Norwood, NJ: Ablex.
- Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children*. New York: Holt, Rinehart, & Winston.
- Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in medicine*, 22, 3527-3541.

Corresponding author at: Mark A. Runco, The Torrance Center, College of Education, 126A Aderhold Hall, The University of Georgia, 110 Carlton Street, Athens, Georgia 30602 706-255-7992.
E-mail: runco@uga.edu

