Creativity Research Journal

ISSN: 1040-0419 (Print) 1532-6934 (Online) Journal homepage: http://www.tandfonline.com/loi/hcrj20

Objectively Scoring Divergent Thinking Tests for Originality: A Re-Analysis and Extension

Denis Dumas & Mark Runco

To cite this article: Denis Dumas & Mark Runco (2018) Objectively Scoring Divergent Thinking Tests for Originality: A Re-Analysis and Extension, Creativity Research Journal, 30:4, 466-468

To link to this article: https://doi.org/10.1080/10400419.2018.1544601

CREATIVITY RESEARCH JOURNAL, *30*(4), 466–468, 2018

ISSN: 1040-0419 print/1532-6934 online
DOI: https://doi.org/10.1080/10400419.2018.1544601

Taylor & Francis
Taylor & Francis Group

COMMENTS AND CORRECTIONS

Objectively Scoring Divergent Thinking Tests for Originality: A Re-Analysis and Extension

Denis Dumas
University of Denver

Mark Runco
Southern Oregon University

Divergent thinking (DT) tests, so often used in the research on creativity, when administered and scored correctly, are reliable estimates of creative potential. They do provide reliable scores about ideation, including the originality, flexibility, and fluency of ideas. Scores from DT tests are far from synonymous with creativity per se, which is why they are not called *creativity tests*, but *estimates of creative potential* instead. The three ideational indices usually coded from DT responses are justified by the more comprehensive theories of DT (Guilford, 1967; Runco, 1991; Torrance, 1995; Wallach & Kogan, 1965; see Runco & Acar, in press, for a recent review).

There is controversy concerning how to score DT. Sometimes the theories just cited are ignored and only fluency (i.e., the number of ideas produced) is used. There are several reasons to avoid this practice, in addition to the fact that it contradicts theory. The *standard definition* of creativity, for example, points specifically to originality as a prerequisite for all creativity (Runco & Jaeger, 2012), which means that fluency is not the most important index of DT. Even if there were reasons to use only one index, that index should probably be originality and not fluency. Fluency is sometimes used alone because it can be highly correlated with originality and flexibility (Hocevar, 1979). Yet there are empirical, as well as theoretical, reasons to avoid relying on fluency. One relevant line of work uses explicit instructions with DT tests (Runco, Illies, & Eisenman, 2005a, 2005b; for a recent meta-analysis see Acar, Runco, & Park, 2018), one finding of which was that when examinees are directed to original ideas, their fluency decreased. This implies that the ideational processes are distinct, for otherwise originality and flexibility would react to instructions in the same direction (increasing or decreasing) as fluency. The fact that originality and flexibility can go the opposite direction from fluency suggests that they are not dependent on fluency.

Another line of work uses partial correlations (Hocevar, 1979; Runco & Albert, 1985). The logic here is that the variance shared by fluency and originality can be statistically controlled, leaving residual variance that represents originality entirely independent of fluency. When this method was first applied, the unique variance of originality showed questionable reliability (Hocevar, 1979), but later more than adequate reliability was found, albeit in a gifted sample (Runco & Albert, 1985).

The partial correlational research was quite some time ago, and in the interim research has demonstrated the usefulness of computer generated scores for DT (Beketayev & Runco, 2016; Hass, 2017; Dumas & Dunbar, 2014; Dumas & Strickland, in press). The high objectivity of these computer-generated DT scores suggested that it would be informative to replicate the partial correlational research and to check again the reliability of originality, after fluency is controlled, but now using the new computer-generated scores. That was the purpose of our research.

METHOD

Participants

This is a reanalysis of data reported by Dumas and Strickland (in press). That earlier study had a completely different objective (i.e., exploring correlations among malevolent ideation and sex) and was completed before we realized that the data provided an opportunity to replicate the work of Runco and Albert (1985) using computer-generated originality scores. The participants in that earlier work and this research were 84 undergraduate students at a large mid-Atlantic American university (48 women;

Address correspondence to Denis Dumas. E-mail: Denis.Dumas@du.edu
University of Denver, 1999 East Evans Avenue, Denver, CO 80208-1700

57.1%). They were recruited via introductory psychology courses for which students are required to participate in research. Participants ranged in age from 18 to 41, with a mean age of 19.98 (*SD* = 2.71). Slightly less than half of participants were European American (*n* = 35; 41.7%) with the remaining indicating ethnicity as African American (*n* = 18; 21.4%), Asian/Pacific Islander (*n* = 12; 14.3%), Hispanic/Latino (*n* = 10; 11.9%), and more than one ethnicity (*n* = 9, 10.7%). The majority of the sample reported their first language as English (*n* = 76; 90.5%).

DT measure: the alternate uses task (AUT)

The AUT is a psychometric measure in which participants are asked to generate as many unique uses for an object as possible within a certain amount of time (i.e., 2 min per object in this case). The AUT has been used for assessing divergent thinking and creative ability for many years (Guilford, 1967; Hudson, 1968; Torrance, 1972), and remains one of the most-often utilized task within the creativity research literature (Puryear, Kettler, & Rinn, 2017). The following 10 object names were presented to participants in a randomized order: book, brick, fork, hammer, pants, shoes, shovel, table, truck, and trumpet. The AUT was administered online through the Qualtrics online platform, with participants completing the task remotely from any computer (but not smartphone or tablet) that was connected to the Internet.

SCORING AND RESULTS

Most important for our purposes was reliability. To that end, the number of uses generated by each participant for each object was tallied, and then summed across all 10 items on the AUT. Fluency counts across the 10 items on the AUT exhibited a high level of scale reliability (α = .95). semantic analysis (LSA) incorporating a very large corpus (i.e., 37,651 included documents and more than 11 million included words) of English-language texts. This corpus (*general-reading-up-to-the-first-year-in-college*: Kintsch, 2000; Landauer, Foltz, & Laham, 1998; http://lsa.colorado.edu/) was built to approximate the reading experience of the typical Anglophone undergraduate, and is used regularly within the research literature on divergent thinking (e.g., Dumas & Dunbar, 2016; Hass, 2017). To measure originality, LSA models both an AUT prompt (e.g., book) and a given participant response (e.g., throw like a frisbee) as vectors in a high-dimensional semantic space. The semantic distance between the AUT prompt and the response is quantified by taking the cosine of the angle between the vectors, yielding semantic distance scores that range from −1 to 1, with scores close to 1 indicating that that the response is very similar to the prompt, and therefore not very original. Then, semantic distance scores are subtracted from 1 to produce quantities ranging from 0 to 2, with scores close to 2 indicating that a given response was highly original (i.e., the semantic distance between the response and the prompt was high). These originality scores were calculated for each individual idea generated by participants. Originality scores for each participant were then averaged within each AUT object-prompt, and then those averages were summed across object-prompts to produce a total originality score. Originality scores across each of the ten items on the AUT exhibited satisfactory scale reliability (α = .82).

Residualized originality

To partial the variance from the originality scores that could be attributed to fluency, 10 separate regression models were fit in which the fluency scores from each single AUT object-prompt (e.g., book) predicted the originality scores from the same prompt. The bivariate correlation between fluency and the LSA originality score was (r = .28, p = .01). Individual R-square statistics from each of these regression models were generally small and are reported in Table 1. Most important for our purpose were the new scores based on the residuals of each of these 10 regression models. The 10 new scores represent residualized originality for each object-prompt on the AUT (i.e., originality that has no overlap with fluency). This is the same method used by Runco and Albert (1985), except using LSA generated originality scores. These residualized originality scores also exhibited satisfactory reliability (α = .81).

DISCUSSION

Because the reliability coefficient for residualized originality (.81) was only slightly lower than that from the nonresidualized originality coefficient (.82), it can be inferred that the observed reliability of the nonresidualized originality scores were not confounded by fluency. This conclusion is corroborated by the small R-square coefficients presented in Table 1, which illustrate that only a very small amount of the variance in originality scores was attributable to fluency. The adequate reliability of originality, even after overlap with fluency is controlled, is consistent

TABLE 1
R-square coefficients from each of 10 linear regression models used to create residualized originality scores

AUT Object-Prompt	*R-square*
Book	.13
Brick	.01
Fork	.01
Hammer	.08
Pants	.01
Shoes	.04
Shovel	.14
Table	.01
Truck	.05
Trumpet	.04

Note: Each of these regression models was formulated with the fluency score as the predictor and the originality score as the outcome

with the findings of Runco and Albert (1986). Our analyses relied on the AUT, and given that the AUT tends to give DT scores that differ from all other tests of DT (Runco, Abdulla, & Paek, 2016), future research should assess the reliability of residualized originality with other DT tests, as well as with other samples of participants. Another possible next step may be to partial the variance attributed to elaboration from originality scores and investigate the reliability of those residualized scores. This may be warranted, because recent evidence (i.e., Forthmann, Oyebade, Ojo, Günther, & Holling, 2018) suggests that elaboration, rather than fluency, may be most related to semantic-network based originality scores. Still, these results are telling and have clear implications. Several times, the ostensible confounding of fluency has been used in arguments that originality is not a useful index of DT, and alternatives have been proposed. Alternatives are not inherently bad, but some of them have used subjective judgments. This ignores the fact that DT is so useful largely because the methods are highly objective and judgment plays a negligible role. Our results confirm that LSA scoring seems to allow DT research to avoid subjective scores and maintain objectivity, with acceptable levels of reliability, even when originality is independent of fluency.

REFERENCES

Acar, S., Runco, M. A., & Park, H. (2018). *What should people be told when they take a divergent thinking test? A meta-analytic review of explicit instructions for divergent thinking*. Under review.

Beketayev, K., & Runco, M. A. (2016). Scoring divergent thinking tests with a semantics-based algorithm. *Europe's Journal of Psychology*, *12*(2), 210–220. doi:10.5964/ejop.v12i2.1127

Dumas, D., & Dunbar, K. N. (2014). Understanding fluency and originality: A latent variable perspective. *Thinking Skills and Creativity*, *14*, 56–67. doi:10.1016/j.tsc.2014.09.003

Dumas, D., & Dunbar, K. N. (2016). The creative stereotype effect. *PLOS ONE*, *11*(2). doi:10.1371/journal.pone.0142567

Dumas, D., & Strickland, A. (in press). From book to bludgeon: A closer look at malevolent responses on the alternate uses task. *Creativity Research Journal*.

Forthmann, B., Oyebade, O., Ojo, A., Günther, F., & Holling, H. (2018). Application of latent semantic analysis to divergent thinking is biased by elaboration. *The Journal of Creative Behavior*. doi:10.1002/jocb.240

Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY, USA: McGraw-Hil1.

Hass, R. W. (2017). Semantic search during divergent thinking. *Cognition*, *166*, 344–357. doi:10.1016/j.cognition.2017.05.039

Hocevar, D. (1979). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology*, *71*(2), 191–196. doi:10.1037/0022-0663.71.2.191

Hudson, L. (1968). *Frames of mind: Ability, perception and self-perception in the arts and sciences*. Oxford, England: W. W. Norton.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, *7*, 257–266. doi:10.3758/BF03212981

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284. doi:10.1080/01638539809545028

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relationships of personality to differential conceptions of creativity: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(1), 59–68. doi:10.1037/aca0000079

Runco, M. A. (1991). *Divergent thinking*. Westport, CT, US: Ablex Publishing.

Runco, M. A., Abdulla, A. M., & Paek, S.-H. (2016). Which test of divergent thinking is best? *Creativity: Theories-Research-Applications*, *3*, 4–18. doi:10.1515/ctra-2016-0001

Runco, M. A., & Acar, S. (in press). Divergent thinking. In J. C. Kaufman & R.J. Sternberg (Eds.), *The Cambridge handbook of creativity*. NewYork, NY: Cambridge University Press.

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. doi:10.1177/001316448504500306

Runco, M. A., Illies, J. J., & Eisenman, R. (2005a). Creativity, originality, and appropriateness: What do explicit instructions tell us about their relationships?. *Journal of Creative Behavior*, *39*, 137–148. doi:10.1002/j.2162-6057.2005.tb01255.x

Runco, M. A., Illies, J. J., & Reiter-Palmon, R. (2005b). 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. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96. doi:10.1080/10400419.2012.650092

Torrance, E. P. (1972). Predictive validity of the Torrance Tests of Creative Thinking. *The Journal of Creative Behavior*, *6*(4), 236–252. doi:10.1002/j.2162-6057.1972.tb00936.x

Torrance, E. P. (1995). *Why fly?* New York, NY: Greenwood Publishing Group.

Wallach, M. A., & Kogan, N. (1965). A new look at the creativity intelligence distinction. *Journal of Personality*, *33*(3), 348–369. doi:10.1111/jopy.1965.33.issue-3