Psychology of Aesthetics, Creativity, and the Arts
2008, Vol. 2, No. 2, 97–99

1931-3896/08/$12.00 DOI: 10.1037/1931-3896.2.2.97

Commentary: The Torrance Tests of Creative Thinking Already Overcome Many of the Perceived Weaknesses That Silvia et al.'s (2008) Methods Are Intended to Correct

Kyung Hee Kim
The College of William and Mary

Silvia et al.'s (2008) primary motivations for exploring and proposing their subjective scoring method are their perceived deficiencies of current divergent thinking tests. First, scores on divergent thinking tests frequently correlate highly with general intelligence. Second, the scoring of divergent thinking tests has changed little since the 1960s. Third, the necessity of instructing people to be creative prior to taking divergent thinking tests is integral to obtaining useful responses and needs to be reaffirmed. Fourth, and finally, the problems posed by uniqueness scoring—confounding with fluency, ambiguity of rarity, and the seeming "penalty" imposed on large samples—that need to be addressed.

First, Kim's (2005) meta-analysis indicated that the relationship between divergent thinking test scores and IQ ($r = .17$) is negligible, which supports the underlying belief that creativity and intelligence are separate constructs. According to Kim's (in press) meta-analysis, divergent thinking test scores predict creative achievement ($r = .22$) better than IQ ($r = .17$). Further, 51.8% of the 274 correlation coefficients incorporated in the study used the Torrance Tests of Creative Thinking (TTCT; Torrance, 1996), and the TTCT predicted ($r = .33$, $p < .0001$) creative achievement better than other measures of creative potential (e.g., Wallach & Kogan Divergent Thinking Tasks [Wallach & Kogan, 1965], Guilford Divergent Thinking Tasks [Guilford, 1967], Sounds and Images [Torrance, Khatena, & Cunnington, 1973], Word Association Tests [Gough, 1976], etc.). In this meta-analysis, art, music, writing, science (including mathematics, medicine, and engineering), leadership, and social skills were used to measure creative achievement. Among these different types of creative achievement, musical achievement was predicted better by IQ than by measures of creative potential, whereas art, science, writing, and social skills were predicted by measures of creative potential better than by IQ. This finding suggests that creativity test scores account for more variance in creative achievement than IQ, and therefore may predict overall creative achievement better.

Second, in Silvia et al.'s (2008) objection, "Methods of administering and scoring divergent thinking tasks have changed little since 1960s," the authors cited Torrance (1967) as one example. However, scoring of the TTCT has changed and improved since the 1960s: The TTCT has been renormed in 1974, 1984, 1990, 1998, and 2007. The 1984 TTCT–Figural manual simplified the scoring procedures and provided a detailed Scoring Workbook (Ball & Torrance, 1984) in addition to the Norms-Technical Manual. Two norm-referenced measures of creative potential, Abstractness of Titles and Resistance to Premature Closure, were added to Fluency, Originality, and Elaboration; and Flexibility was eliminated because it correlated highly with Fluency (Hébert, Cramond, Neumeister, Millar, & Silvian, 2002). Further, (third) the TTCT instructs people to be creative. For instance, the TTCT directions state for Fluency: "Make as many different pictures or objects as you can and put as many ideas as you can in each one," and for Originality: "Try to think of things that no one else will think of" (1966, p. 2), which Torrance specifically emphasized. When Torrance developed the TTCT in 1966, he had a disagreement with Guilford in that Guilford did not want to give clues concerning desired responses when the test taker was asked to think of as many as use of each object. Guilford's instructions did not motivate the participants for divergent thinking, whereas the directions for the TTCT are intended to specifically motivate the test taker for fluency and originality (Torrance, 1994). Torrance, like Silvia et al., considered the motivation given in the instructions to be very important. He explained that we would never attempt to measure jumping ability by measuring how high or far a person just happened to be jumping at a particular time; rather, we would try to motivate him or her to jump as high or as far as he or she can (Torrance, 1994).

Fourth, Silvia et al. (2008) correctly point out the fact that, on most measures of creativity, many research findings support the existence of high correlations between originality and fluency (e.g., Chase, 1985; Dixon, 1979; Heausler & Thompson, 1988; Kim, 2006b; Torrance, 1979). Simonton (1990) indicated that a person's originality is a function of the number of ideas formulated. Torrance and Safter (1999) confirmed that a person who generates a large number of alternatives is more likely to produce original ideas. However, measures of originality usually predict creative behavior more accurately than do measures of fluency (Torrance, 1972a, 1972b, 1972c, 1974). Thus, although fluency increases the chance that original ideas will be produced; there is no guarantee that fluency will generate original ideas (Torrance, 1979). However, more importantly, Silvia et al. could not avoid the high positive correlations between fluency scores and uniqueness scores in their study although they intended to solve their perceived problem.

In Silvia et al.'s (2008) objection to the ambiguity of rarity the authors stated,

> The objective 0/1 system is not as objective as it seems: It will tend to give 1s to weird responses and to common responses that raters would judge as uncreative. Some evidence for this claim comes from

Kyung Hee Kim, School of Education, The College of William and Mary.

Correspondence concerning this article should be addressed to Kyung Hee Kim, School of Education, Jones Hall, The College of William and Mary, PO Box 8795, Williamsburg, VA 23187. E-mail: KKim@wm.edu

> research that compared the objective 0/1 coding with subjective 0/1 coding. Hocever had 4 raters score responses using a 0/1 unoriginal/original scale. The raters' uniqueness scores substantially lower than the objective uniqueness scores, indicating that the raters had a higher criterion for judging uniqueness. (p. 70)

However, the four raters may have had lower originality scores if they had scored based on the originality lists for the TTCT found in the TTCT scoring manual. In addition, interrater reliability needs to be taken into account. The Torrance Center at the University of Georgia (http://www.coe.uga.edu/torrance/programs.html) sponsors ongoing TTCT training. This training requires that scorers take into account the uniqueness and appropriateness of the response. Torrance balanced the validity of the scoring with being training for interrater reliability. In fact, to earn a certificate for scoring the TTCT from the Torrance Center, a test scorer must obtain an originality score with at least 90% agreement with the standard scores established by Torrance. Therefore, the TTCT does not have a problem with this issue.

Finally, "One of the biggest problems with uniqueness scoring" (Silvia et al., 2008, p. 70) that of the perceived penalty imposed on large samples, is controlled for by the structure of the TTCT. Originality lists for the TTCT have been developed for each item on the basis of normative data, in which each list is a count of the number of statistically frequent ideas. Thus, TTCT originality scores take sample size into account by weighing each response by frequency and do not penalize large samples.

Subjective Scoring of Creativity as an Alternative

Silvia et al. (2008) state,

> Subjective ratings can overcome the three problems faced by uniqueness scoring. First, ratings should, in principle, be unconfounded with fluency: because the raters judge each response separately, generating a lot of responses won't necessarily increase one's creativity score. Second, bizarre, weird and common responses that slip through the cracks of the uniqueness index ought to be caught by the subjective raters. And third, subjective ratings ought to be independent of sample size. (p. 70)

First, according to their results, Silvia et al.'s subjective ratings of uniqueness are not unconfounded with fluency. Second, the high interrater reliability of the TTCT scoring takes into account both appropriateness and uniqueness of a response at the same time. Third, originality scoring for the TTCT is independent of sample size.

One method of subjective scoring examined by Silvia et al. (2008) is the "Top 2 approach," wherein respondents are asked to select their two most creative responses. The assumption is that people will pick the two responses that represent their best, and therefore, most creative efforts. Although this self-rating of creativity is an interesting idea and could be useful because a self-evaluation of responses may give a good indication of a person's ability to recognize their own creative ability, it has inherent problems. For example: What if a participant chooses two common nonunique responses that they believe are most creative? This is problematic because there is no guide for judging the creativity of responses for participants although there is a guide for raters in the study. In addition, if a participant has generated more than two very unique responses, Silvia et al. seem to be suggesting that these responses be disregarded. The authors' suggested benefits the "Top 2 approach," judging only what people consider as their best effort, holding the number of judged responses constant, and the "real-world" similarity of selecting the best ideas, are already present within the structure of the TTCT.

First, scoring the TTCT takes every response into account, determining the best effort. Second, Activity I of the TTCT–Figural is not scored for fluency, although it is scored for originality. Activity II of the TTCT–Figural holds constant the number of responses on which people are evaluated, provided that people complete all of the 10 incomplete figures. On average, if they cannot finish all of the 10, they tend to score higher on abstractness of titles and elaboration. Two types of people are found based on the TTCT scorers' experiences; one type produced quick and novel responses, thus doing better on fluency and originality, whereas the other type gave detailed responses, which indicated greater depth of thought, and did better on elaboration and abstractness of titles (Kim, 2006b). Third, picking one's best ideas does not mean picking the two best ideas.

In sum Silvia et al. (2008) argue,

> Subjective scoring should be considered seriously. First, the idiosyncrasies of raters have been overstated: many studies show excellent agreement in the subjective judgments of independent raters. Second, agreement between raters can be enhanced by giving them clear instructions, by providing accepted definitions of creativity, and by training them in the scoring system. Finding low agreement isn't surprising when the raters aren't trained or instructed. Third, variance associated with raters needn't be mere error—rater variance can be modeled, thus reducing overall error. And fourth, the merit of a subjective scoring system is an empirical question. What's important about scores is their reliability and validity, not their ostensible level of objectivity or directness. (p. 71)

They make a reasonable argument; however, their concerns are already addressed by the TTCT.

First, scoring of the TTCT can be considered to be both objective and subjective and has shown excellent agreement (over 90%) between independent raters. Second, as stated earlier, there is ongoing scoring training for the TTCT in which the instruction is systematic and easy to follow. Third, at scoring training for the TTCT, rater variance is modeled and thus overall error is reduced. And fourth, the reliability and validity of the TTCT have been proven (e.g., Hébert et al., 2002; Kim, 2006a; Torrance, 2002). Further, all of the Silvia et al.'s suggested methods of subjective scoring require three to five raters to achieve a dependability rating above .80, whereas scoring for the TTCT does not require more than one rater, provided the rater is trained and certified.

In conclusion, Silvia et al. (2008) present an interesting argument for subjective scoring. However, their motivating concerns are already answered in the structure of the TTCT. The authors do not indicate why they decided to compare their subjective scoring methods with the Wallach and Kogan (1965) uniqueness index, which did poorly against their methods. Their argument would have been far more convincing had they compared their methods with the TTCT, the most widely used and studied, divergent thinking test (Kim, 2006a).

References

Ball, O. E., & Torrance, E. P. (1984). Streamlines scoring workbook: Figural A. Bensenville, IL: Scholastic Testing Service.

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

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

Gough, H. G. (1976). Studying creativity by means of word association tests. *Journal of Applied Psychology, 61,* 348–353.

Guilford, J. P. (1967). The nature of human intelligence. New York: McGraw-Hill.

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., Millar, G., & Silvian, A. F. (2002). *E. Paul Torrance: His life, accomplishments, and legacy.* Storrs: University of Connecticut, National Research Center on the Gifted and Talented.

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education, 16,* 57–66.

Kim, K. H. (2006a). Can we trust creativity tests?: A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal, 18,* 3–14.

Kim, K. H. (2006b). Is creativity unidimensional or multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creativity Research Journal, 18,* 251–260.

Kim, K. H. (in press). Meta-analyses of the relationship of creative achievement to both IQ and divergent thinking test scores. *Journal of Creative Behavior.*

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Crom, J. T., Hess, K. I., et al. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts, 2,* 68–85.

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

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking: Thinking creatively with pictures, Forms A & B.* Bensenville, IL: Scholastic Testing Service.

Torrance, E. P. (1967). The Minnesota studies on creative behavior. *Journal of Creative Behavior, 1,* 137–154.

Torrance, E. P. (1972a). Career patterns and peak creative achievements of creative high school students twelve years later. *Gifted Child Quarterly, 16,* 75–88.

Torrance, E. P. (1972b). Predictive validity of bonus scoring for combinations on repeated figures tests of creative thinking. *Journal of Psychology, 81,* 167–171.

Torrance, E. P. (1972c). Predictive validity of the Torrance Tests of Creative Thinking. *Journal of Creative Behavior, 6,* 236–252.

Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking: Norms-Technical manual.* Lexington, MA: Personnel/Ginn.

Torrance, E. P. (1979). *The search for Satori and creativity.* Buffalo, NY: Bearly.

Torrance, E. P. (1994). *Creativity: Just wanting to know.* Pretoria, South Africa: Benedic Books.

Torrance, E. P. (2002). *The Manifesto: A guide to developing a creative career.* West Westport, CT: Ablex.

Torrance, E. P., Khatena, J., & Cunnington, B. F. (1973). *Thinking creatively with sounds and words.* Bensenville, IL: Scholastic Testing Service.

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

Wallach, M., & Kogan, N. (1965). Modes of thinking in young children. New York: Holt, Rinehart and Winston.

Received November 26, 2007
Revision received November 26, 2007
Accepted November 27, 2007 ■