Creativity Research Journal

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

Divergent Thinking as an Indicator of Creative Potential

Mark A. Runco & Selcuk Acar

To cite this article: Mark A. Runco & Selcuk Acar (2012) Divergent Thinking as an Indicator of Creative Potential, Creativity Research Journal, 24:1, 66-75, DOI: 10.1080/10400419.2012.652929

To link to this article: https://doi.org/10.1080/10400419.2012.652929

CREATIVITY RESEARCH JOURNAL, *24*(1), 66–75, 2012

ISSN: 1040-0419 print/1532-6934 online
DOI: 10.1080/10400419.2012.652929

Routledge
Taylor & Francis Group

Divergent Thinking as an Indicator of Creative Potential

Mark A. Runco and Selcuk Acar
University of Georgia, Athens

Divergent thinking (DT) tests are very often used in creativity studies. Certainly DT does not guarantee actual creative achievement, but tests of DT are reliable and reasonably valid predictors of certain performance criteria. The validity of DT is described as reasonable because validity is not an all-or-nothing attribute, but is, instead, a matter of degree. Also, validity only makes sense relative to particular criteria. The criteria strongly associated with DT are detailed in this article. It also summarizes the uses and limitations of DT, conceptually and psychometrically. After the psychometric evidence is reviewed, alternative tests and scoring procedures are described, including several that have only recently been published. Throughout this article related processes, such as problem finding and evaluative thinking, are linked to DT.

The concept of *divergent thinking* is attractive for a number of reasons. It is, for example, a good metaphor of the kind of cognition that should lead to original ideas. Divergent thinking is easy to contrast with *convergent thinking*, which typically leads to conventional and "correct" ideas and solutions rather than original options. Yet divergent thinking, or what Guilford (1950, 1968) called *divergent production*, is more than a metaphor. In fact, one reason the concept is so attractive is that it leads directly to testable hypotheses and allows reliable assessment of the potential for creative thought.

The key idea is that of potential. Divergent thinking is not the same as creative thinking. Divergent thinking often leads to originality, and originality is the central feature of creativity, but someone can do well on a test of divergent thinking and never actually perform in a creative fashion. This argument parallels a theme of this special issue of the *Creativity Research Journal*, though at first blush that may not be entirely obvious. The theme involves *indicators* of creative talent. What exactly is an indicator? It is a kind of predictor; and any time a prediction is made, there is some uncertainty. Even someone who has always performed at a high level in the past may slip up in the future, especially if circumstances change. On a more macro level, an organization or cultural group which has maintained a high level of innovation in the past may miss the boat in the future. Predictions are estimates because the future is always unknown. Indicators are best guesses. They are predictors. Good indicators are reliable and accurate, but there is always some uncertainty. Similarly, when looking at or measuring individuals, it is possible to determine the degree of potential, but that potential may or may not be fulfilled. There is the same kind of uncertainty with estimates of potential, such as a test of divergent thinking, as there is with an indicator. Psychometric research suggests that tests of divergent thinking provide useful estimates of the potential for creative thinking. They are good indicators of future creative performance. They are far from perfect, but (a) no test is perfect, and (b) the degree of accuracy can be objectively determined.

This article briefly reviews the research of relevance to the accuracy of divergent thinking tests. Much of it focuses on reliability or validity, the two primary psychometric criteria of accuracy and usefulness, but some focuses on the procedures for insuring that divergent thinking test are administered and scored in the best possible fashion. Newer tests and scoring procedures are described and the conclusion points to future research considerations.

THE MEASUREMENT OF CREATIVITY

Hocevar (1981) identified four types of creativity tests: divergent thinking tests, attitude and interest inventories, personality inventories, and biographical

Correspondence should be sent to Mark A. Runco, Torrance Creativity Center, University of Georgia, Aderhold Hall, Athens, GA 30602. E-mail: runco@uga.edu

inventories. Each offers useful information, but tests of divergent thinking have dominated the field of creativity assessment for several decades. This has created one problem, namely that occasionally they are regarded as tests of creativity. As noted, that is not a tenable view.

The general idea of divergence of thought can be found in theories going back to the 19th century (e.g., Binet & Simon, 1905), but it was J. P. Guilford (1950, 1968) who clearly tied divergent production to creative potential. He hypothesized several components for divergent thinking (Guilford & Hoepfner, 1971) but empirical evidences supported some more than the others. Most tests of DT now look only to fluency, originality, flexibility, and elaboration. In fact, too often only fluency is used. That is defined in terms of productivity. A fluent individual gives a large number of ideas. Originality is usually defined in terms of novelty or statistical infrequency. Flexibility leads to diverse ideas that use a variety of conceptual categories. Elaboration, the least common, is suggested when the individual follows an associative pathway for some distance. For most purposes, fluency should not be used alone. There is unique and reliable variance in the other indices (Runco & Albert, 1985) and fluency is not as closely tied to creativity as is originality and flexibility. More on this is presented in the following.

Guilford (1950, 1968) developed a large number of tasks, as did Torrance. In fact, the Torrance Tests of Creative Thinking (TTCT) are still enormously popular. Other batteries were developed by Wallach and Kogan (1965) and Williams (1980). Meeker, Meeker, and Roid (1985) took up where Guilford left off and revised and extended his efforts. Later, Urban and Jellen (1996) developed the Test of Creative Thinking (Divergent Production; TCT-DP), and most recently Auzmendi, Villa, and Abedi (1996) developed a multiple choice test to assess DT.

Which is the best test? One way to address that question is to consider which is the most reliable. Reliability is a good place to start any discussion of tests because it is a prerequisite for validity. An unreliable test cannot be valid.

Reliability of DT Tests

Reliability is indicative of the stability or consistency of a test result or score. There are different kinds of reliability, including interitem, interrater, and alternative forms reliability. The second of these is most appropriate if there is any subjectivity in the scoring system. Often there is very little subjectivity. The number of ideas produced is determined by simply counting ideas and gives a fluency score; the originality of ideas is determined by compiling ideas within a sample and identifying which are given infrequently (those receive points for originality); and so on. But sometimes it is useful to ask judges to examine the ideas. When this is the case, interjudge agreement must be checked. Fortunately, judges can be objective and give reliable ratings. Meeker (1985) reported an interrater reliability coefficient of well above .90 in her work with Structure of the Intellect Learning Abilities Tests, for example. Urban and Jellen (1996) reported .90, and Wallach and Kogan (1965) .92. This is not to say that there are no concerns and questions. There is some question about how much information to provide to judges (Runco, 1989), for instance, and another question about how to select the best judges (Murray, 1959; Runco, McCarthy, & Svensen, 1994).

Another kind of reliability involves interitem agreement or internal consistency. In one review, these ranged from .42 to .97 (Cropley, 2000), but fortunately most were above .70. Still, at one point reliability analyses brought the usefulness of originality scores into question (Hocevar, 1979a, 1979b). The problem was that originality scores were unreliable when fluency scores were controlled and the conclusion was that fluency contaminated originality (Clark & Mirels, 1970; Hocevar & Michael, 1979). Runco and Albert (1985) determined that this is not always the case and that originality is reliable, even after fluency is statistically controlled, for some tests and for some samples, especially those at higher levels of ability.

Validity

There are various kinds of validity. The question implied by each is, ''Are you testing what you think you are testing?'' Validity is in this sense the opposite of bias. A biased test is influenced by things outside of the target. A test may be verbally biased, for example, in which case linguistic skills determine performances more than creative talents. Highly verbal examinees tend to do well, regardless of creative ability, and examinees with low verbal skills tend to do poorly, again regardless of their creative abilities. There are other potential biases, including SES and experience (Runco, 1991; Runco & Acar, in press).

Discriminant validity is of especially importance for creativity because it is in some ways related to general intelligence. This is suggested by threshold theory, which posits that some level of general intelligence is necessary but not sufficient for creative work (MacKinnon, 1965; Torrance, 1966; Yamamato, 1965). Yet at moderate and high levels, evidence confirms a separation. This supports the discriminant validity of divergent thinking tests.

Wallach and Kogan (1965) reported that test setting and instructions had an impact on the level of discriminant validity such that, with test-like directions, it is low,

but with game-like instructions for DT tests, the separation from intelligence was quite clear. Wallach (1970) later attributed the low correlation between creativity and intelligence above certain intelligence level to the restricted distribution of IQ scores (also see Crockenberg, 1972). Without a doubt, the various DT indexes are differentially related to general intelligence. Torrance (1969) was especially concerned about the elaboration index from the TTCT, but some of the problem with it may have been the difficulty in scoring the tests. Elaboration is difficult for untrained raters to scorer (Cramond, Matthews-Morgan, Bandalos, & Zou, 2005).

The discriminant validity of creativity tests becomes more complicated by the recognition that both divergent and convergent thinking may work together in creative thinking and problem solving (Runco, 2007). They have been described as representing two ends of a continuum (Eysenck, 2003) and, alternatively, as two necessary phases in creative problem solving. In fact, several evolutionary models have described the creative process in two phases, with blind variation responsible for the generation alternative and selective retention based on evaluation of ideas based on their appropriateness and usefulness (Campbell, 1960). This would seem to support a divergent-convergent process. Basadur's (1995) ideation-evaluation model described something very much like this, with ideation and evaluation occurring one after another, but also with problem finding, problem solving, and solution implementation stages. For Basadur, the importance of ideation and evaluation depends both on the stage and the task. Ideation is employed more in problem finding and tasks such as basic research and development; evaluation features more in solution implementation and applied tasks, such as distribution and sales in an organizational structure. Brophy (1998) also argued that creative problem solving requires alternating from divergent to convergent thinking at the right time, but he felt that only a minority can do both with a good balance. Moneta (1994) also hypothesized an optimal balance between the two. Actually, these ideas are not far afield from the original work on DT. Guilford (1968), for example, pointed to evaluation allowed by conceptual foresight, penetration redefinition, and problem sensitivity (Mumford, 2000–2001). Hence, although discriminant validity is a good thing, it is likely that a realistic view of creative problem will recognize that DT does not work in isolation (Runco, 1994).

Predictive validity is based on the strength of relationships between DT tests and various criteria. The trick is finding a good criterion. Kogan and Pankove (1974) used extracurricular activities and accomplishments as criteria and the Wallach and Kogan tests of creativity as predictors. They found poor predictions for 5th grade students and only slightly better predictions for 10th grade students. Wallach and Wing (1969) and Runco (1986) reported much better predictive validity for the same tests. Runco took interactions among indexes into account (e.g., originality $\times$ fluency). Incidentally, after reviewing the research on creativity tests, Hocevar (1981) concluded that self-reported creative activities and achievements are the most defensible tests for the identification of creative talent despite the problems such as determining which activities to label as creative. Crockenberg (1972), on the other hand, criticized the same tests because they include items representing characteristics of well-educated high-IQ people.

Runco, Plucker, and Lim (2000–2001) suggested that previous assessments of the predictive validity of DT tests were ill-conceived. The problem, they argued, was the criteria used previously were often inappropriate for DT. Certainly the criteria in previous studies were related to creativity and quite informative. Torrance (1981), for example, looked at (a) number of high school creative achievements, (b) the number of post high school achievements, (c) number of creative style-of-life achievements, (d) quality of highest creative achievements, and (e) creativeness of future career image as criteria. Still, Runco et al. proposed that better predictive validities would be found if the criteria relied solely on ideation and not on opportunities and extracognitive factors. They examined the predictive validity of DT using the Runco Ideational Behavior Scale, which is a self-report that only asks about ideation. Correlations between it and various DT tests were reassuring. Runco et al. also reported nonsignificant correlations with GPA, which is yet more evidence for discriminant validity.

Tests, Scores, and Scoring

Just as new criteria have been developed, so too have new tests been developed. Several new tests were cited previously (e.g., Urban & Jellen, 1996). The older batteries (e.g., Guilford's, the TTCT, Wallach & Kogan's) are still employed, and some have the advantage precisely because they have been used so many times. There are norms and extensive research that can be consulted for interpretations of scores and results. Yet sometimes there is a need to adapt a test for a specific purpose or sample. Runco, Dow, and Smith (2006) did exactly this to test the role of memory and experience in DT. They developed open-ended tests that, like DT tests, allow numerous responses, but also required that the examinee rely on experience. Examinees (from Orange County, California) were asked for street names, for example, as well as car part, plant names, and book titles. Analyses indicated that scores from the experience-based open-ended tests were correlated with scores on the more standard DT, and, in fact, Runco and Acar (in press) concluded that there is an experiential bias in many DT tests.

Abedi (2002; Auzmendi et al., 1996) developed another new test, called the Abedi-Schumacher Creativity Test. It was designed to assess DT without administering open-ended questions. It is, in fact, a multiple-choice test, but the response options are indicative of tendencies toward fluency, originality, flexibility, and so on. Abedi justified this in that the new test requires very little time to administer. It contains 60 questions, but several multiple choice items can be completed by an examinee per minute, so it is indeed a quick test. Of course, it is not really a test of DT, or at least not a skill or performance test of DT. Another quick test was developed by Cheung (2005). This is an electronic version of Wallach and Kogan DT tests (e-WKCT). Cheung, Lau, Chan, and Wu (2004) have collected enough data to provide norms for Chinese samples. The electronic format allows virtually instantaneous scoring and comparison of individual scores to norms.

Of more significance than the new tests are the alternative scoring systems for DT tasks. That is because there are so many issues with DT indicators. The scoring used by Torrance (1966, 1972, 1980), with four indexes—fluency, originality, flexibility, and elaboration—is still in use in the current version of TTCT. However, those indexes were highly intercorrelated and using all of those indexes is laborious. Therefore, Wallach and Kogan (1965) suggested using only two indexes, namely, uniqueness and fluency. This is still a time-consuming method, and interrater reliability must be determined because judgments must be made some of the time to know whether or not an idea is in fact unique.

Several researchers have suggested that a quality score should be used. Cropley (1967), for example, suggested scoring by giving 4 points to ideas that given by less than 1% of the population, 3 points for those produced by 1–2%, 2 for 3–6%, and 1 for 7–15%, and 0 for the remainder. Harrington, Block, and Block (1983) added quality scoring to the novelty and quantity, and they found that presence of the quality scoring resulted in an increment in the construct validity of DT scores. Quality is also implied by the *appropriateness scores* used by Runco, Illies, and Eisenman (2005) and Runco and Charles (1993) at least in the sense that both recognize the creative things are more than just original. They are also effective, appropriate, or in some way fitting. Ideas that are merely original can be of low quality and not effective.

Vernon (1971) suggested basing originality on the responses from randomly selected group of a hundred people. Responses that are given by only 5 or fewer people are given 2 points, and 0 points are given for ideas produced by 15 or more. Those in between were given a 1.

Hocevar and Michael (1979), Vernon (1971), Runco, Okuda, and Thurston (1987), and Manske and Davis (1968) all suggested that a proportional score could be used. This could be calculated by dividing originality (or flexiblity, for that matter) by fluency. It would, thereby, eliminate any confounding by fluency (Hocevar, 1979b) but would mask an examinee's productivity. Someone who produced two ideas, one of which was unique, would have the same score as someone who produced 10 ideas, 5 of which were unique. Additionally, ratios and proportions are often unreliable.

Another alternative is to simply add together the various indexes of DT. This is probably the least tenable option, with the possible exception of using fluency alone. A simple summation assumes that each index is on the same scale, for example, and without standardization the index with the highest mean would contribute more than the other indexes to the sum. At least as important is that it is contrary to theories of DT that describe each index as a unique aspect of DT.

A confounding by fluency is a serious consideration. Such confounding is implied by bivariate correlations among the various DT indexes but was more dramatically demonstrated by Hocevar's (1979b) analyses of the reliability of originality scores from various divergent thinking tests (alternate uses, plot titles, and consequences). He found that, although reliable before adjustments, when the variance accounted for by fluency was removed from originality, its reliability was quite low.

Hocevar (1979a) suggested dividing ideas into two categories: original or unoriginal. This would help if there is, in fact, overlap and redundancy among indices. Milgram, Milgram, Rosenbloom, and Rabkin (1978) and Moran, Milgram, Sawyers, and Fu (1983) had very good luck with this kind of scoring, though they labeled their DT (nonoverlapping) indexes popular and unusual. In this scoring, an idea can be either original or unoriginal (and fluent).

Mouchiroud and Lubart (2001) investigated the use of median weights because median scores are less sensitive to mean scores, and this could diminish the influence of fluency scores. Five different originality indices were defined for the TTCT. Three of them were based on the norms reported the TTCT manual, sample-based frequencies (2 points for ideas given by less than 2% and 1 point for those given between 2–5%), and sample-based uniqueness. Two others indexes were based on Runco et al. (1987)'s weighting system. No differences were found between norm-based and sample-based originality indices.

Hocevar (1979a) compared subjective ratings of originality, statistical infrequency scores for originality, and a random method (rating each responses from 0 to 9 in a random number table). Each produced surprisingly adequate levels of interitem reliability. The subjective ratings were far from perfectly correlated with the

objective method that uses statistical infrequency. The subjective scores seemed to have higher discriminant validity than the objective scores.

Seddon (1983) and Snyder, Mitchell, Bossomaier, and Pallier (2004) felt that it should be possible to treat DT as a single compound. Seddon (1983) proposed a method where originality was used as a weight given to the each response and category given, respectively. Snyder et al. (2004) employed a logarithmic function to create a single score, but they combined fluency and flexibility indices rather than originality.

Runco and Mraz (1992) questioned a basic premise of previous scoring systems. They suggested that a more practical method is to ask judges to rate each examinee's total ideational output, rather than basing scores on individual ideas. That would provide judges with much more information—they would see everything any one examinee produced rather than just one idea at a time—and should save a great deal of time as well. Their results, and those of Charles and Runco (2000–2001), confirmed that this method, using *ideational pools*, works quite well.

Runco et al. (1987) compared (a) the summation scores (flexibility, fluency, and originality added together), (b) common and uncommon (the latter ideas produced by less than 5% of the sample), (c) ratio scores (originality or flexibility divided by fluency), and (d) weighted fluency scores (where weights are determined based on rarity of ideas, with less frequent ideas given more weight). They concluded that the weighted system was preferable.

Runco (1986) offered an alternative specifically for scoring flexibility. The basic idea was to look at the number of changes, from ideational category to category. This differs from flexibility defined as the total number of categories. Someone might change from category to category, but only use a few categories, or the person might only use a single category once but cover quite a few of them. Each would seem to relate to flexibility but they represent different cognitive processes.

Apparently the reliability of DT scores can vary in different populations. Runco and Albert (1985) discovered that originality may depend on fluency in nongifted populations, for example, but among the gifted there is separation and originality supplies information that is not contained in fluency. Much the same can be said about flexibility (Runco, 1985).

Clark and Mirels (1970) proposed that tests should be administered such that participants are limited to a certain number of ideas. Fluency would be irrelevant, then. Michael and Wright (1987) limited the number of responses to the three most creative ones, and Zarnegar, Hocevar, and Michael (1987) limited their sample to one single unique idea. These methods do eliminate the concern over confounding by fluency, and they may be realistic in that they require examinees to not only generate ideas but also to select creative or at least unique ones—and that combination (idea generation and evaluation) is probably consistent with creative thinking in the natural environment. Then again, limiting examinees to three or one ideas ignores theories (e.g., Mednick, 1962) and research (Milgram et al., 1978; Runco, 1986) that suggests that time is necessary for finding creative ideas. If only a few ideas are allowed, the individual might not deplete common ideas and find remote associates.

Mumford, Marks, Connelly, Zaccaro, and Johnson (1998) suggested a domain-specific scoring method. They developed a specific scoring method for the domain of leadership and therefore defined scores for "time frame" (decisions that bring immediate or long-term consequences), quality, realism, originality, complexity, use of principles, positive outcomes, and negative outcomes (also see Vincent, Decker, & Mumford, 2002).

Evaluative Skills

Much of the research on DT has gone outside the test and focused, not just on how to make a better test of DT, but instead on what may interact with or even support DT. This makes a great deal of sense because creative thinking is such a complex process. DT plays a role but judgment and evaluation, and perhaps even convergent thinking, are required as well (Cropley, 2006; Runco, 1994). DT alone would lead to wild ideas, some of which might be creative, but many would likely be unrealistic, useless, and irrelevant. Indeed, Eysenck (2003) found psychotic individuals to be highly original, but not creative.

The role of evaluation in creative thought and the relationship with DT has been examined in various ways. Khandwalla (1993), for instance, asked his participants to think aloud while trying to solve various problem (i.e., "list objects that are green, liquid, and funny"). They were encouraged to say whatever came to mind. Participants' protocols were recorded and then analyzed in an attempt to infer what specific processes lead to an original idea. Khandwalla identified five broad categories of divergent thinking from the protocols. He called these problem structuring, searching, feeling, ideating, and evaluating. Ideating was the most common process, followed by evaluating. Interestingly, an associative path involving both ideating and evaluating paths was the most frequent of the five paths identified by Khandwalla. He concluded that "evaluating and ideating form a synergistic rather than an antagonistic combination in complex tasks" (p. 253).

Inaccurate evaluations make it difficult to think creatively. Inaccurate evaluations may turn the individual

away from a potentially fruitful associative path, for example, or simply curtain ideation. This is premature closure (Runco & Basadur, 1993). One function of evaluation is, then, to insure that new ideas can be developed and will be explored. Other functions were described by Dailey and Mumford (2006). For them, evaluative skills allow people to manage their time and cognition effectively because trivial solutions or ideas are discarded and more promising ideas are focused and improved. Another function appears in the improvement of the ideas. Ideators can take context, conditions, constraints, resources, and standards into account to improve ideas and ease the implementation of them. Lonergan, Scott, and Mumford (2004) investigated the role of standards used when judging ideas. They asked their participants to evaluate or suggest revisions to the ideas developed for a new product. Different standards were applied during either evaluating ideas or when suggesting new revisions to them. Highly original ideas were more accurately judged in well-structured tasks.

Mumford, Blair, Dailey, Leritz, and Osborn (2006) identified more than 35 evaluative errors. These included isolation, sufficing, discounting anomalies, premature reservation of ideas, and optimism. Licuanan, Dailey, and Mumford (2007) detected the errors in evaluation of ideas in cases of high versus low originality, high versus low complexity, employment of active processing, and creativity framing. Intriguingly, people tended to make more errors in their evaluations of highly original ideas by underestimating their merit. Also, as the complexity of the setting where the idea is developed increased, errors in appraisal of creative ideas also increased. Further analyses provided evidence that people tend to underscore the originality of the ideas when they focus on performance under the complex conditions. Mumford et al. used these findings to suggest methods for improving the accuracy of the evaluations. They showed that creativity framing (training about the value of creativity in team settings and some variables involving interactional processes which improve creativity in team settings) and active processing (having the participants elaborate on the originality of the ideas deeply) can improve the accuracy of the evaluation in highly original ideas.

Evaluations are also involved when there is a need to predict the necessary resources (e.g., time, money) and outcomes of ideas when implemented. Dailey and Mumford (2006) proposed that implementation-based evaluation can decrease errors in estimating the resources needed for the new ideas and their outcomes. They found that people tended to underestimate the resources needed and overestimate positive outcomes of idea implementation. Also, personal characteristics (e.g., neuroticism) influenced the accuracy of evaluations. Domain familiarity decreases the accuracy of evaluations.

Blair and Mumford (2007) focused their research on the attributes of ideas that are taken into account when evaluating. They found that people favored ideas that were consistent with the current social norms, led expected outcomes, and were complex to execute but easy to understand. They rejected the ideas that were original and risky as well those had detailed descriptions, probably because these were too obvious and simple. Under the time pressure, people preferred ideas that were consistent with social norms and required time, in addition to those that had detailed descriptions and were original. Original and risky ideas were also preferred when nonstringent evaluation criteria were emphasized and when there was limited time available. Participants who expected that their evaluations were to be compared to others tended to prefer the ideas that were more extreme.

Runco and Smith (1992) compared inter- and intrapersonal evaluations of ideas. Not surprisingly, these two kinds of evaluations were related to one another, so someone who is accurate at judging his or her own ideas is also accurate when judging ideas given by others. The more surprising finding was that intrapersonal evaluations were significantly correlated with divergent thinking scores while interpersonal evaluations were not. The interpersonal evaluations, on the other hand, were correlated with preference for ideation. Interestingly, participants were more accurate when judging the uniqueness than the popularity of their own ideas (popularity being inversely related to originality) but were more accurate when judging the popularity than the uniqueness of ideas given by others.

Basadur, Wakabayashi, and Graen (1990) defined evaluation as a kind of thinking style. They then examined style as a moderator of the impact of training. Four styles of creative problem solving were compared: generator (experience and ideation), conceptualizer (thinking and ideation), implementor (experience and evaluation), and optimizer (ideation and evaluation). Analyses indicated that the optimizers gained significantly more from the training than the other three styles.

Runco and Basadur (1993) also examined the benefits of training but they actually focused the instruction o the evaluations of ideas (not just on DT or ideation in general). Managers who received this training improved in the sense of increasing correct choices and decreasing incorrect choices. They also examined the relationship between ideational and evaluative skills and found that the latter were most strongly related originality (58% of common variance) than with fluency (13% of common variance). There was a higher correlation between the posttreatment evaluative skills and ideational skills than the pretreatment counterparts. The strongest association when style was analyzed was with the conceptualizer style.

Runco and Charles (1993) compared subjective evaluations of originality, appropriateness, and creativity with objective ratings of the same ideas. Appropriateness was of special interest because creative things are always more than original; they are in some way appropriate. It may be effectiveness or aesthetic appeal, but in some way creative things are all appropriate, as well as original. Runco and Charles discovered that the subjective evaluations paralleled the objective ratings. More interesting was that originality and appropriateness were not correlated, as would be expected, given what was just said about creative things. In fact, the lowest originality ratings were given to the most appropriate ideas. Additionally, (a) originality was central to the creativity judgments, and (b) appropriateness without originality was judged to be indicative of low levels of creativity.

Later Charles and Runco (2000–2001) examined developmental changes in divergent thinking and evaluative skills among 3th through 5th grade children. The 4th grade was of special interest because Torrrance (1968) and others previously found a slump at that age. Charles and Runco found that evaluative skills for originality and preference for appropriateness scores tend to increase by age. However, the relationship between divergent thinking fluency scores and evaluative skills was quite small, except for one test item in which an increasing preference of appropriateness accounted for a decline in the proportion of high-quality ideas. Interestingly, there was no 4th grade slump in this sample. Divergent thinking actually peaked in the 4th grade.

Runco (1991) also investigated children's evaluative skills. He designed this study as a test of the theory that children's creativity reflects lack of discretion (about what is conventional) rather than intentional creative expression. If this theory held up, and children were only unintentionally creative, they would tend to be poor judges of the quality of ideas. Runco, therefore, asked one group of children to rate the creativity of ideas, and asked a second group to rate the same ideas for popularity. Popularity was examined because it might be more meaningful and workable a concept than *creativity* for children. Children from 4th through 6th grades rated ideas (some of which were entirely unique) that had been generated by other children. Analyses indicated that evaluative accuracy had discriminant validity in that it was unrelated to intelligence test scores. Statistical comparisons of the accuracy ratings indicated that children were, in fact, more accurate when rating ideas for popularity rather than creativity. Evaluative accuracy was related to DT, but as is the case in research with other samples (e.g., Runco & Smith, 1992), the relationship is only moderate. Being good at DT does not, then, insure accurate evaluations of ideas. Runco interpreted the findings of this study as confirming that evaluative accuracy can be improved with appropriate instruction.

Runco and Vega (1990) used a similar methodology but examined evaluative accuracy of parents and teachers when judging ideas given by children. The accuracy of evaluations of popularity and creativity was again compared. Interestingly, parents with more of their own children provided more accurate evaluations than other parents. Evaluative accuracy and DT were again correlated, as was the case in other samples, but contrary to expectations, parents and teachers did not differ significantly in the accuracy of their evaluations.

Evaluative accuracy seems to vary with the specific task, or at least when figural, verbal, and realistic tasks are compared. Runco and Dow (2004) found evaluative accuracy was the lowest in realistic problems and highest in verbal DT tests. The figural test was in the middle in terms evaluative accuracy. These findings are not at all surprising given other empirical research on DT which also shows task differences in fluency, originality, and flexibility. Runco et al. (2005), for example, administered both realistic and unrealistic tasks and asked participants to produce as (a) many ideas as possible, (b) only original, (c) only appropriate, or (d) only creative ideas. Results indicated that the participants produced more original but fewer appropriate ideas in unrealistic tasks than realistic tasks in all instructional groups. Flexibility and fluency scores were highest for the unrealistic tests in all instructional groups except when instruction focused on originality. Appropriateness and originality were not strongly related (.26). What is most important is no doubt that the realistic tasks elicit the lowest originality scores. This tendency had been explained by the fact that individuals are likely, or at least able, to draw on experience and memory, and therefore do not need to actually use creative thinking to generate new ideas, when faced with realistic tasks, but evaluations certainly could also play a role.

CONCLUSIONS

Thinking about DT has following the motion of a pendulum. There was great enthusiasm early on, many people apparently thinking that DT tests could replace IQ. Soon critics showed that predictive and discriminant validities were moderate at best, and the pendulum went to the other extreme. At that point, DT and all tests of creativity were questioned and largely dismissed. The pendulum swung again when more careful psychometric studies demonstrated that the validities were, in fact, as good as is found throughout psychological assessment (some coefficients in excess of .6), and as good as would be expected with a reasonable theory. The bedrock of that reasonable theory is that DT tests are not tests of creativity. They are estimates of the potential for creative problem solving. DT is not synonymous with creativity.

To our reading of the literature, creativity studies are ready to progress beyond that point. The word is out, DT is not synonymous with creativity but DT tests provide useful estimates of meaningful potential. Not surprisingly, with this in mind, there is now research that starts with that premise and looks beyond validity and reliability. The most impressive of those focus on the underpinnings of DT. Consider in this regard recent research on the relationship between divergent thinking and speed of relatedness judgment. Vartanian, Martindale, and Matthews (2009) proposed that creative people must be faster at judging the relatedness of concepts, as this would enable judgments about promising ideas and ideational pathways. Vartanian et al. were confirmed that people with higher divergent thinking ability were faster in their relatedness judgments. Furthermore, relatedness judgments were not related to IQ. Genetic factors underlying divergent thinking have been uncovered by Reuter et al. (2002) and Runco et al. (2010). Runco et al., for example, found that verbal and figural fluency were related to several genes, even after controlling intelligence, but originality was not. The key genes involve dopamine reception. One intriguing thing about this is that the genetic correlate seems to be limited to fluency and productivity, and not originality. Additionally, dopamine reception is associated with tendencies towards obsessions and addictions, which may be relevant to the health issues in creativity studies (Richards & Runco, 1997).

The point is the same one we used as our opening sentence: There is great value in the concept of divergent thinking. Much of the research focuses on DT tests, and their reliability and validity, but additional research tells us more broadly how DT is various social and psychological factors (e.g., IQ, personality, family background) and how it is associated with problem solving, ideation, and creative potential. Ideas are useful in many aspects of our lives, and the research on divergent thinking remains useful for understanding the quality of ideas and the processes involved.

REFERENCES

Abedi, J. (2002). A latent-variable modeling approach to assessing reliability and validity of a creativity instrument. *Creativity Research Journal*, *14*, 267–276.

Auzmendi, E., Villa, A., & Abedi, J. (1996). Reliability and validity of a newly constructed multiple-choice creativity instrument. *Creativity Research Journal*, *9*, 89–96.

Basadur, M., Wakabayashi, M., & Graen, G. B. (1990). Individual problem solving styles and attitudes toward divergent thinking before and after training. *Creativity Research Journal*, *3*, 22–32.

Basadur, M. (1995). Optimal ideation-evaluation ratios. *Creativity Research Journal*, *8*, 63–75.

Binet, A., & Simon, T. (1905). The development of intelligence in children. *L'Annee Psychologique*, *11*, 163–191.

Blair, C. S., & Mumford, M. D. (2007). Errors in idea evaluation: Preference for the unoriginal? *Journal of Creative Behavior*, *41*, 197–222.

Brophy, D. R. (1998). Understanding, measuring and enhancing individual creative problem-solving efforts. *Creativity Research Journal*, *11*, 123–150.

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, *67*, 380–400.

Charles, R. E., & Runco, M. A. (2000–2001). Developmental trends in the evaluation and divergent thinking of children. *Creativity Research Journal*, *13*, 417–437.

Cheung, P. C., Lau, S., Chan, D. W., & Wu, W. Y. H. (2004). Creative potential of school children in Hong Kong: Norms of the Wallach-Kogan Creativity Tests and their implications. *Creativity Research Journal*, *16*, 69–78.

Clark, P. M., & Mirels, H. L. (1970). Fluency as a pervasive element in the measurement of creativity. *Journal of Educational Measurement*, *7*, 83–86.

Cramond, B., Matthews-Morgan, J., Bandalos, D., & Zuo, L. (2005). A report on the 40-year follow-up of the Torrance Tests of Creative Thinking: Alive and well in the new millennium. *Gifted Child Quarterly*, *49*, 283–291.

Crockenberg, S. (1972). Creativity tests: A boon or boondoggle in education. *Review of Educational Research*, *42*, 27–45.

Cropley, A. J. (1967). Creativity, intelligence, and achievement. *Alberta Journal of Educational Research*, *13*, 51–58.

Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using? *Roeper Review*, *23*, 72–79.

Cropley, A. J. (2006). In praise of convergent thinking. *Creativity Research Journal*, *18*(3), 391–404.

Dailey, L., & Mumford, M. D. (2006). Evaluative aspects of creative thought: Errors in appraising the implications of new ideas. *Creativity Research Journal*, *18*, 367–384.

Eysenck, H. (2003). Creativity, personality, and the convergent-divergent continuum. In M. A. Runco (Ed.), *Critical creative processes* (pp. 95–114). Cresskill, NJ: Hampton.

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*, 444–454.

Guilford, J. P. (1968). *Creativity, intelligence and their educational implications*. San Diego, CA: EDITS/Knapp.

Guilford, J. P., & Hoepfner, R. (1971). *The analysis of intelligence*. New York, NY: McGraw-Hill.

Harrington, D. M., Block, J., & Block, J. H. (1983). Predicting creativity in preadolescence from divergent thinking in early childhood. *Journal of Personality and Social*, *45*, 609–623.

Hocevar, D. (1979a). A comparison of statistical infrequency and subjective judgment as criteria in the measurement of originality. *Journal of Personality Assessment*, *43*, 297–299.

Hocevar, D. (1979b). Ideational fluency as a confounding factor in the measurement of originality. *Journal of Educational Psychology*, *71*, 191–196.

Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of Personality Assessment*, *45*, 450–464.

Hocevar, D., & Michael, W. (1979). The effects of scoring formulas on the discriminant validity of tests of divergent thinking. *Educational and Psychological Measurement*, *39*, 917–921.

Khandwalla, P. N. (1993). An exploratory investigation of divergent thinking through protocol analysis. *Creativity Research Journal*, *6*, 241–259.

Kogan, N., & Pankove, E. (1974). Long-term predictive validity of divergent-thinking tests. *Journal of Educational Psychology*, *66*, 802–810.

Licuanan, B. F., Dailey, L. R., & Mumford, M. D. (2007). Idea evaluation: Error in evaluating highly original ideas. *Journal of Creative Behavior*, *41*, 1–27.

Lonergan, D. C., Scott, G. M., & Mumford, M. D. (2004). Evaluative aspects of creative thought: Effects on appraisal and revision standards. *Creativity Research Journal*, *16*, 231–246.

MacKinnon, D. (1965). Personality and the realization of creative potential. *American Psychologist*, *20*, 273–281.

Manske, M. R., & Davis, G. A. (1968). Effects of simple instructional biases upon performance in the Unusual Uses Test. *Journal of General Psychology*, *79*, 25–33.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220–232.

Meeker, M. (1985). *Structure of Intellect Learning Abilities Test*. Los Angeles: Western Psychological Services.

Meeker, M., Meeker, R., & Roid, G. H. (1985). *Structure of Intellect Learning Abilities Test (SOI-LAT)manual*. Los Angeles, CA: Western Psychological Services.

Michael, W. B., & Wright, C. R. (1989). Psychometric issues in the assessment of creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 33–52). New York, NY: Plenum Press.

Milgram, R. M., Milgram, N. A., Rosenbloom, G., & Rabkin, L. (1978). Quantity and quality of creative thinking in children and adolescents. *Child Development*, *49*, 385–388.

Moneta, G. B. (1994). A model of scientists' creative potential: The matching of cognitive structures and domain structure. *Philosophical Psychology*, *6*, 23–37.

Moran, J. D., Milgram, R. M., Sawyers, J. K., & Fu, V. R. (1983). Stimulus specificity in the measurement of original thinking in preschool children. *Journal of Psychology*, *4*, 99–105.

Mouchiroud, C., & Lubart, T. (2001). Children's original thinking: An empirical examination of alternative measures derived from divergent thinking tasks. *Journal of Genetic Psychology*, *162*, 382–401.

Mumford, M. D. (2000–2001). Something old, something new: Revisiting Guilford's conception of creative problem solving. *Creativity Research Journal*, *13*, 267–276.

Mumford, M. D., Blair, C., Dailey, L., Leritz, L. E., & Osburn, H. K. (2006). Errors in creative thought? Cognitive biases in a complex processing activity. *Journal of Creative Behavior*, *40*, 75–110.

Mumford, M. D., Marks, M. A., Connelly, M. S., Zaccaro, S. J., & Johnson, J. F. (1998). Domain-based scoring of divergent thinking tests: Validation evidence in an occupational sample. *Creativity Research Journal*, *11*, 151–163.

Murray, H. A. (1959). Vicissitudes of creativity. In H. H. Anderson (Ed.), *Creativity and its cultivation* (pp. 203–221). New York, NY: Harper.

Reuter, M., Roth, S., Holve, K., & Henning, J. (2006). Identification of first candidate genes for creativity: A pilot study. *Brain Research*, *1069*, 190–197.

Richards, R., & Runco, M. A. (Eds.). (1997). *Eminent creativity, everyday creativity, and health*. Greenwich, CT: Ablex.

Runco, M. A. (1985). Reliability and convergent validity of ideational flexibility as a function of academic achievement. *Perceptual and Motor Skills*, *61*, 1075–1081.

Runco, M. A. (1986). The discriminant validity of gifted children's divergent thinking test scores. *Gifted Child Quarterly*, *30*, 78–82.

Runco, M. A. (1989). Parents' and teachers' ratings of the creativity of children. *Journal of Social Behavior and Personality*, *4*, 73–83.

Runco, M. A. (1991). The evaluative, valuative, and divergent thinking of children. *Journal of Creative Behavior*, *25*, 311–319.

Runco, M. A. (Ed.). (1994). *Problem finding, problem solving, and creativity*. Norwood, NJ: Ablex.

Runco, M. A. (2007). *Creativity: Theories and themes: Research, development, and practice*. New York, NY: Academic Press.

Runco, M. A., & Acar, S. (2010). Do tests of divergent thinking have an experiential bias? *Psychology of Aesthetics, Creativity, and the Arts*, *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., & Basadur, M. (1993). Assessing ideational and evaluative skills and creative styles and attitudes. *Creativity and Innovation Management*, *2*, 166–173.

Runco, M. A., & Charles, R. (1993). Judgments of originality and appropriateness as predictors of creativity. *Personality and Individual Differences*, *15*, 537–546.

Runco, M. A., & Dow, G. T. (2004). Assessing the accuracy of judgment of originality on three divergent thinking tests. *Korean Journal of Thinking and Problem Solving*, *14*, 5–14.

Runco, M. A., Dow, G. T., & Smith, W. R. (2006). Information, experience, and divergent thinking: An empirical test. *Creativity Research*, *18*, 269–277.

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., 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*, 213–211.

Runco, M. A., Noble, E. P., Reiter-Palmon, R., Acar, S., Ritchie, T., & Yurkovich, J. (2011). The genetic basis of creativity and ideational fluency. *Creativity Research Journal*, *23*, 376–380.

Runco, M. A., Okuda, S. M., & Thurston, B. J. (1987). The psychometric properties of four systems for scoring divergent thinking tests. *Journal of Psychoeducational Assessment*, *5*, 149–156.

Runco, M. A., Plucker, J. A., & Lim, W. (2000). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal*, *13*, 391–398.

Runco, M. A., & Smith, W. R. (1992). Interpersonal and intrapersonal evaluations of creative ideas. *Personality and Individual Differences*, *13*, 295–302.

Runco, M. A., & Vega, L. (1990). Evaluating the creativity of children's ideas. *Journal of Social Behavior and Personality*, *5*, 439–452.

Seddon, G. M. (1983). The measurement and properties of divergent thinking ability as a single compound entity. *Journal of Educational Measurement*, *20*, 393–402.

Snyder, A., Mitchell, D., Bossomaier, T., & Pallier, G. (2004). The creativity quotient, an objective scoring of ideational fluency. *Creativity Research Journal*, *16*, 415–420.

Torrance, E. P. (1966). *Torrance tests of creative thinking: Norms-technical manual* (Research Edition). Princeton, NJ: Personnel Press.

Torrance, E. P. (1968). A longitudinal examination of the fourth grade slump in creativity. *Gifted Child Quarterly*, *12*, 195–199.

Torrance, E. P. (1969). Prediction of adult creative achievement among high school seniors. *Gifted Child Quarterly*, *13*, 223–229.

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

Torrance, E. P. (1980). *Norms–technical manual: Demonstrator form for the Torrance Tests of Creative Thinking*. Unpublished Manuscript, Georgia Studies of Creative Behavior, Athens, GA.

Torrance, E. P. (1981). Predicting the creativity of elementary school children (1958–1980)—And the teacher who "made a difference." *Gifted Child Quarterly*, *25*, 55–62.

Torrance, E. P. (1995). *Why fly?* Norwood, NJ: Ablex.

Urban, K. K., & Jellen, H. G. (1996). *Test for Creative Thinking-Drawing Production (TCT-DP)*. Lisse, Netherlands: Swets and Zeitlinger.

Vartanian, O., Martindale, C., & Matthews, J. (2009). Divergent thinking ability is related to faster relatedness judgments. *Psychology of Aesthetics, Creativity and the Arts*, *3*, 99–103.

Vernon, P. E. (1971). Effects of administration and scoring on divergent thinking tests. *British Journal of Educational Psychology*, *41*, 245–257.

Vincent, A. S., Decker, B. P., & Mumford, M. D. (2002). Divergent thinking, intelligence & expertise: A test of alternative models. *Creativity Research Journal*, *14*, 163–178.

Wallach, M. A. (1970). Creativity. In P. H. Mussen (Ed.), *Manual of child psychology* (Vol. 1, pp. 1211–1272). New York, NY: Wiley and Sons.

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

Wallach, M. A., & Wing, C. W. (1969). *The talented student: A validation of the creativity- intelligence distinction*. New York, NY: Holt, Rinehart & Winston.

Williams, F. (1980). *Creativity assessment packet: Manual*. East Aurora, NY: DOK.

Yamamoto, K. (1965). "Creativity"—A blind man's report on the elephant. *Journal of Counseling Psychology*, *21*, 428–434.

Zarnegar, Z., Hocevar, D., & Michael, W. B. (1988). Components of original thinking in gifted children. *Educational and Psychological Measurement*, *48*, 5–16.