Identifying and Assessing Creativity as a Component of Giftedness

James C. Kaufman¹, Jonathan A. Plucker², and Christina M. Russell²

Abstract
Most theories of giftedness include creativity as a central component. Creativity assessment has a key role, therefore, in measuring giftedness. This article reviews the state of the creativity assessment, from divergent thinking tests (including the Torrance Tests of Creative Thinking) to the consensual assessment technique to rating scales and self assessments. Despite the many flaws present in every type of creativity measurement, there are compelling reasons for including creativity as part of a gifted assessment battery.

Keywords
creativity, assessment, giftedness, creative assessment

Although educational assessment generally pays little attention to creativity, creativity moves from being a background player to a key role within the context of giftedness and gifted education. Most theories of giftedness include creativity as a central component. The federal government proposed a multifaceted definition of giftedness in the early 1970s that appears to have been based on the trait-specific view of giftedness (as opposed to state-specific views). This delineation suggested that giftedness and talent are manifest in six areas: general intellectual ability, specific academic aptitude, creative or productive thinking, leadership ability, visual and performing arts, and psychomotor ability (Marland, 1972). The Marland definition has been extremely influential and is still used by many school districts in their identification of talented students. In a large national study, Callahan, Hunsaker, Adams, Moore, and Bland (1995) found that nearly 50% of districts based their gifted education identification procedures on this definition, making it far and away the most popular definition in this setting. Our review outlines major theories of creativity in addition to ways to identify and assess creative people (e.g., for admission to a program or school) and creative products.

Other prominent theories of giftedness also include creativity: Gagné’s (2005) Differentiated Model of Giftedness and Talent (DMGT) conceptualizes “gifts” as the innate abilities (or

¹California State University, San Bernardino, CA, USA
²Indiana University, Bloomington, IN, USA

Corresponding Author:
James C. Kaufman, Learning Research Institute, California State University at San Bernardino, Department of Psychology, 5500 University Parkway, San Bernardino, CA 92407, USA
Email: jkaufman@csusb.edu
aptitudes) in at least one domain area (intellectual, creative, socioaffective, and sensorimotor) and “talents” as demonstrated mastery of these gifts. Renzulli’s (1986, 2005) Three-Ring Conception views giftedness as emerging from the interaction of well above average ability, creativity, and task commitment, with each characteristic playing a critical role in the development of gifted behavior. In Sternberg’s (2005) model, giftedness is conceptualized as a synthesis of wisdom, intelligence, and creativity (WICS).

Prominent theories of creativity and giftedness have been applied to create numerous assessments to identify products, environments, and people. Gifted identification systems are, for all practical intents and purposes, the main application of creativity assessments in the United States (Kaufman, Plucker, & Baer, 2008; Plucker & Renzulli, 1999). Much of our review of the field of creativity assessment will be geared to emphasize measures used in evaluating giftedness.

What Is Creativity?

The first step to discussing the assessment of creativity is to first define the construct. It is notable that so many studies on creativity do not actually specify what is meant by “creativity.” Plucker, Beghetto, and Dow (2004) selected 90 different articles that either appeared in the top creativity journals or were articles in a different peer-reviewed journal with the word “creativity” in the title. Of these articles, only 38% explicitly defined what creativity was. Based on current theoretical and empirical work, Plucker et al. (2004) proposed the following definition, which we will use for the purposes of this article:

Creativity is the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context. (p. 90)

Our conceptions of creativity assessment were shaped by early pioneers in the field. In 1950, Guilford placed creativity into a larger framework of intelligence in his Structure of the Intellect Model. He attempted to organize all of human cognition along three dimensions: operations, content, and product. “Operations,” the first dimension, were seen as the general intellectual processes or ‘mental gymnastics’ needed for any kind of task. The second dimension, “content,” referred to the general subject area, while “product” represented the actual products that might result from thinking abstractly in different kinds of subject matters. With five operations, four contents, and six products, Guilford’s (1967) model had 120 different possible mental abilities. Guilford later introduced the notion of divergent thinking, otherwise known as the ability to answer open-ended questions with both novel and useful responses. Today, divergent thinking remains a very common measure among creativity assessments (Hunsaker & Callahan, 1995).

Divergent Thinking (DT) Assessments

A lot of time, energy, and effort have been invested in developing and researching measures of divergent thinking. Ironically, there is not much divergence in the history of creativity assessments. Divergent thinking is the backbone of creativity assessment and has held this position for many decades. The majority of articles measuring creativity use divergent thinking tests; most books on creativity include a long discussion of divergent thinking; and schools frequently use DT tests to identify the creative potential of students. The theoretical launching points for serious development efforts and large-scale application of divergent thinking into assessments were Guilford’s Structure of the Intellect divergent production tests (1967), Wallach and Kogan’s...
(1965) and Getzels and Jackson’s divergent production tests, and Torrance’s (1974, 2008) Tests of Creative Thinking (TTCT).

As previously noted, Guilford (1967) proposed the Structure of the Intellect Model (SOI) in which he identified 24 distinct components of divergent thinking, one type for each combination of the four types of content (Figural, Symbolic, Semantic, Behavioral) and six types of product (Units, Classes, Relations, Systems, Transformations, Implications). The SOI Divergent Thinking battery consists of several dozen tests that correspond with the 24 distinct components. One example of a Guilford DT task is the Sketches subtest, which assesses the Figural Unit dimension by inviting the student to draw as many objects as possible using a basic figure, such as a circle. Many tests of DT were based off of Guilford’s SOI assessments, including the most widely studied of the DT assessments: The Torrance Tests of Creative Thinking.

During the 1960s, researchers published results of studies that relied on “Guilford/SOI-like” assessments. For example, the Instances Test required students to list as many things that move on wheels (that make noise, etc.) as possible (Wallach & Kogan, 1965; Wallach & Wing, 1969). Another such example is the Uses Test in which students provide answers to prompts such as “Tell me all the different ways you could use a chair” (Wallach & Kogan, 1965, p. 31) or bricks, pencils, and toothpicks (Getzels & Jackson, 1962). The greatest difference between the various batteries developed during this time were the conditions under which students took the tests. Wallach and Kogan (1965) preferred a more game-like, untimed administration of divergent thinking tasks as they believed this allowed creativity to be measured distinctly from intelligence due to the creation of “a frame of reference which is relatively free from the coercion of time limits and relatively free from the stress of knowing that one’s behavior is under close evaluation” (p. 24). This test-taking approach is in contrast to timed, test-like protocols used with most other DT measures and is in part based on concerns that “children are so accustomed to the one correct or best answer that they may be reluctant to think of other possibilities or to build up a pool of ideas to be evaluated later” (Torrance, 1970, p. 86).

The Torrance Tests of Creative Thinking (TTCT)
The Torrance Tests of Creative Thinking (TTCT; Torrance, 1974, 2008) remain the most widely used assessment of creative talent (Sternberg, 2006). Torrance focused on divergent thinking as the basis for creativity and constructed tests that emphasized the assessment of divergent thinking. Torrance’s drive for creating the TTCT was not merely to measure creativity, but rather to be used as tools to better understand what fosters and nurtures creativity (Hebert, Cramond, Spiers-Neumeister, Millar, & Silvian, 2002; Kim, 2006). Although based on the SOI Assessments, Torrance (1968) argued that the TTCT represented a sharp departure from Guilford’s factor-type tests associates (Kaufman, Plucker, et al., 2008). This departure is more in regard to Torrance’s early work as many tasks on the most recent versions on the TTCT are similar to Guilford and Guilford-related tests.

Of all tests of creativity, the TTCT are the longest running, continually published assessments of DT, most carefully studied, and most widely used in educational settings of all tests of creativity (Kaufman, Plucker, et al., 2008). The Torrance Tests are commonly used in efficacy studies and meta-analyses of the impact of creativity training programs.

**Test organization.** The TTCT battery is comprised of a Figural and Verbal section with Forms A and B to be used alternatively. The Verbal section examines one’s ability to think creatively with words, whereas the Figural tests assess an individual’s ability to think creatively with pictures. The Figural section includes three subtests:
Kaufman et al.

- **Picture Construction**: participants use basic shape and expands on it to create a picture;
- **Picture Completion**: the student is asked to finish and title incomplete drawings;
- **Lines/Circles**: participant is asked to modify many different series of lines (Form A) or circles (Form B).

The Verbal section of the TTCT has seven subtests. The first three tasks are considered part of Ask-and-Guess:

- **Ask-and-Guess**: the examinee is asked to look at a picture (e.g., Form A has a picture of an elf staring into a pool of water) at the beginning of the test booklet to complete three subtasks:
  - **Asking**: participant asks as many questions as possible about the picture;
  - **Guessing Causes**: participant lists possible causes for the pictured action;
  - **Guessing Consequences**: participant lists possible consequences for the pictured action;
- **Product Improvement**: participant asked to make changes to improve a toy;
- **Unusual Uses**: participant asked to think of as many different uses for an ordinary item (e.g., cardboard box) as possible;
- **Unusual Questions**: participant asks as many questions as possible about an ordinary item (this subtest does not appear in later editions);
- **Just Suppose**: participant is asked to “just suppose” an improbable situation has happened then list the possible ramifications.

**Scoring recommendations.** The TTCT can be administered in a group or individual setting from kindergarten through higher education or beyond (Kim, 2006). The administration, scoring, and score reporting of the tests and forms are standardized and include detailed norms that were revised accordingly (Torrance, 1974, 2008; Torrance & Ball, 1984). While reading and understanding the manual allows novice raters to produce reliable scores, Torrance recommended for scorers to be trained. Untrained raters deviate from the scoring protocol when assessing originality as they tend to unintentionally allow their personal judgments to affect scoring of individual responses.

**Scoring dimensions.** Although the original test produced scores in four traditional DT areas, current versions of the Figural TTCT (Torrance, 2008) produce scores in the following five areas:

- **Fluency**: The number of responses to a given stimuli;
- **Elaboration**: The extension of ideas within a specific category of responses to a given stimuli;
- **Originality**: The uniqueness of responses to a given stimuli. The most common responses receive a score of 0, whereas all other legitimate responses receive a score of 1;
- **Resistance to Premature Closure**: The degree of psychological openness when processing information; to keep an “open mind” (Kim, 2006);
- **Abstractedness of Titles**: A measurement of the degree to which a title is more abstract/moves beyond a labeling of a picture (Kim, 2006).

Earlier revisions to the scoring protocol (Torrance & Ball, 1984) streamlined the scoring process through the removal of Flexibility scores as these were undifferentiated from Fluency scores (Hebert et al., 2002). The Verbal TTCT produces scores for Fluency, Flexibility, and Originality. Another change in the scoring protocol allowed for the Figural tests to be scored for Resistance to Premature Closure. This score is determined by a participant’s tendency not to immediately...
close the incomplete figures on the Figural Picture Completion Test. Torrance believed that this score reflected the examinee’s ability to “keep open and delay closure long enough to make the mental leap that makes possible original ideas” (Torrance & Ball, 1984, p. 20). On the Picture Completion task, the responses are scored zero points for finishing the picture with the easiest, most direct route. One point is scored for indirectly finishing the figure. Two points are awarded for never completing the picture or completing it with irregular lines that form part of the figure rather than with simple straight or curved lines.

Scoring for Abstractedness of Titles for the figures in the Picture Construction and Picture Completion tasks range from zero points for common, obvious titles (“Shoe” or “River”), one point for more descriptive titles (“The Dancing Cat”), two points for more descriptive or imaginative titles that reach beyond concrete labels (“The Giant’s Finger Puppet”) or three points for abstract but appropriate titles that go beyond the picture and tell a story (“The Time of Your Life”).

**Other Divergent Thinking Assessments**

Most DT assessments are borrowed from or are very similar to the TTCT and Guilford’s SOI Assessments. One example is that the DT scores from the Profile of Creative Abilities (Ryser, 2007) are derived from Guilford’s work and have tasks similar to those on the TTCT. One departure is real-world divergent thinking items, which are similar to those of Guilford and Wallach and Kogan verbal tasks, but are placed in a realistic, applied context. This format assumes that a more realistic assessment of DT skills should take place within a realistic context (Plucker, Runco, & Lim, 2006). Runco and his colleagues have developed a number of these realistic DT tasks and gathered evidence of adequate reliability (Chand & Runco, 1993).

**Controversy Regarding Divergent Thinking Assessments**

There is mixed evidence and conflicting opinion as to the psychometric quality and practical importance of DT tests. The DT tests are associated with evidence of reliability and concurrent validity (e.g., Torrance, 1968, 2008). Plucker (1999), for example, reanalyzed data from Torrance’s longitudinal study and found that divergent thinking was much more predictive for how people differed in creative achievement than was traditional IQ. Nonetheless, many scholars have questioned DT tests’ predictive validity (Baer, 1993, 1994; Gardner, 1993; Kogan & Pankove, 1974; Weisberg, 1993). This has led researchers and educators to avoid using DT tests and, broadly, leads to criticisms of the psychometric study of creativity (Plucker & Renzulli, 1999). Another controversial aspect of DT assessments is that people tend to overgeneralize DT test performance to all other aspects of creativity, such as problem identification (e.g., Runco & Okuda, 1988; Wakefield, 1985) and evaluative thinking (e.g., Okuda, Runco, & Berger, 1991; Runco, 1991; Runco & Chand, 1994).

Recent developments in the scoring of DT tasks address many of these concerns. For example, analyses of various scoring systems provide evidence that alternatives to the traditional scores should be considered, such as the calculation of summative scores (i.e., totaling fluency, flexibility, and originality scores), highly uncommon scores (answers given by less than 5% of participants), weighted fluency scores, percentage scores, and scores based on the entire body of each subjects’ answers, as opposed to scoring individual responses in a list of ideas (Hocevar & Michael, 1979; Runco & Chand, 1993; Runco & Mraz, 1992; Runco, Okuda, & Thurston, 1987). Silvia et al. (2008) have had good success with subjective evaluation of DT test results, and Plucker, Qian, and Wang (2011) found evidence that the percentage scoring method (i.e., dividing originality scores by fluency scores) may be the most appropriate scoring strategy for determining originality scores on DT assessments.
Remote Associates Test (RAT)

The Remote Associates Test (RAT), although different from traditional DT tests, is still generally assumed to tap into DT skills (Mednick, 1968). It is based on the associative theory that “creative thinking . . . consists of forming mutually distant associative elements into new combinations which are useful and meet specified as well as unforeseen requirements” (Mednick, 1968, p. 213). Basically, more creative individuals tend to make meaningful, useful associations between disparate concepts and ideas to a greater extent than do less creative individuals.

The initial test consisted of 30 items, each with three stimulus words. Examinees must identify a fourth word that links the groupings of words. An example of a problem is Sleeping, Bean, and Trash. The correct answer is “Bag.” Sleeping bag, bean bag, and trash bag are all common phrases. The RAT is most traditionally used to study problem solving and creative thinking (Ansburg, 2000; Beeman & Bowden, 2000). The test is often used because it is fairly easy to administer and score. Although typically used as a creativity measure, it is more heavily influenced by intelligence and verbal abilities than most other measures (Katz, 1983). The RAT has been updated to modify out-of-date language; Bowden and Jung-Beeman (2003) presented many new items (calling them the Compound Remote Associates Task).

Consensual Assessment Technique (CAT)

Although divergent thinking is the most common way to assess creativity, it is far from the only way. The best assessment of creativity in a particular field is usually derived from the collective judgment of recognized experts in the field (see J. C. Kaufman, 2009, for a fuller discussion). This type of assessment emphasizes the creative product.

The Consensual Assessment Technique (CAT) is one such approach to creative product evaluation. The CAT is based on the idea that the best measure of creativity of a work of art, theory, or any other artifact is the combined assessment of experts in that field to judge creative products (Amabile, 1996). Subjects are asked to create something (a product) and experts independently evaluate the creativity of those products. Poems, collages, and stories are often evaluated in CAT studies. One example of how the CAT was used was in the field of children’s writing (Baer, 1994). Students were given a simple drawing of a boy and a girl and were asked to write an original story involving both the girl and the boy. Experts in this field of children’s writings were asked to evaluate the creativity of the stories written by students on a 1.0 to 5.0 scale. Judges were able to use fractions (4.5) and were not asked to defend their ratings (Baer, 1994).

The CAT does not have standardized scores, rather only comparative scoring among participants. This method is widely used in creativity research and less widely in schools. Creativity may vary a great deal, however, from domain to domain or even on tasks within the same general area (e.g. Baer, 1993). Therefore, CAT ratings may not be the best approach to measure creativity from a domain-general approach. CAT ratings can be used in classrooms, however, to assess creativity for admission to special programs that look for people who excel in an area of creativity (e.g., art). To use the CAT technique, have participants produce some type of creative product. Creative writing and visual art are the most commonly used products although studies have focused on such diverse areas as music composition (Priest, 2006) and math (Baer, 1994). Although most CAT research uses responses to the same prompt (e.g., “write a Haiku about happiness”), it is possible to use products created to different prompts (Baer, Kaufman, & Gentile, 2004).

The next (and most time-consuming) step is to have appropriate experts assign creativity ratings to all of the products. Raters use their own implicit definition of what is creative and are not given specific instructions or allowed to discuss their ratings with each other. Creative products
are compared to each other, as opposed to a specific ideal. Research is currently underway to
determine appropriate levels of expertise for different creativity-judging tasks (Kaufman & Baer,
in press), but there are a few guidelines that are helpful in choosing a judge. Judges should have
a level of expertise clearly higher than the subjects (i.e., a published poet or a musician who has
performed professionally) and have some familiarity with the populations from which the par-
ticipants are drawn (e.g., judges of high school work should have some familiarity with high
school student production in the domain in question). Quasiexperts, such as students with special
experience or aptitude in the domain, have been found to show solid agreement (Kaufman,
Gentile, & Baer, 2005) although novices (such as undergraduates) generally neither agree with
each other or with experts (Kaufman, Baer, & Cole, 2009; Kaufman, Baer, Cole, & Sexton,
2008). Typically, 5 to 10 experts should be adequate; however, more quasiexperts may be needed
to reach strong interrater reliability.

**Parent, Peer, and Teacher Measures: Assessments by Others**

Whereas the CAT focuses on products, assessments by others focus on the creative person (per-
sonality traits, creativity-relevant abilities, motivation, intelligence, thinking styles, emotional
intelligence, or knowledge). This method of assessment can be as simple as having a teacher
globally rank his or her students based on the teachers’ knowledge of the students and implicit
beliefs about the nature of creativity. This method is similar to the CAT in that it is an evaluation
by others; however, the CAT evaluates individuals’ creative products while the assessments by
others focus on the creative person as a whole. This method emphasizes traits and abilities that
are believed to be relevant to creativity and is domain-general (By its nature, CAT is always
domain-specific). The assessors using this method are experts on the child (teachers, parents of
the child), not experts in creativity.

**Creativity Checklists**

One common form of assessment by others is via checklists. When using creativity check-
lists, there is reason to be wary of teacher and parent ratings of creativity based on global
impressions of a student due to unintended bias. For example, when the second author was
a gifted program coordinator, he asked teachers to use a specific rating scale to assess the
creativity behavior characteristics of select students. One teacher completed a checklist for
each student but then discussed the scores with other teachers and revised the instruments
until the scores fit the teachers’ desired ranking of the students. To limit this bias, raters
should be encouraged to rate each child individually and not make post hoc score compari-
sions (and revisions!).

There are many different creativity checklists, most of which were designed for use in
schools. Many creativity checklists, such as the Gifted Rating Scales (see below), are sold
commercially and copyright protected. One freely available assessment is the Creativity
Checklist (Proctor & Burnett, 2004). The Creativity Checklist is composed of characteristics
thought to be indicative of a creative person, both cognitive and dispositional traits, which
load onto 9 scales: fluent thinker, flexible thinker, original thinker, elaborative thinker, intrin-
sically motivated student, curious/immersed in topic, risk taker, imaginative/intuitive, and
engages in complex tasks/enjoys a challenge. The items are defined in terms of classroom
behaviors for teachers to use on elementary students (Proctor & Burnett, 2004). There were
no norms established for this checklist and therefore it is only appropriate for making com-
parisons within groups of students. There also has been no criterion-related or predictive
validity established.
**Gifted Rating Scales**

The Gifted Rating Scales (GRS) were designed with the goals of being user friendly screening instrument, require minimal training to score and interpret, be psychometrically reliable and valid, and include a standardized sample which reflected the current U.S. census demographics (Pfeiffer & Jarosewich, 2007). The forms are linked to the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV) and Wechsler Preschool and Primary Scale of Intelligence–Third Edition (WPPSI-III; Pfeiffer & Jarosewich, 2007). These rating forms were designed for preschool (GRS-P: ages 4.0-6.11) and school-aged children (GRS-S: ages 6.0-13.11). The GRS-P is made up of 60 items that load onto five scales while the GRS-S consists of six scales with 12 items each, for a total of 72 items (Pfeiffer & Jarosewich, 2007). Although these forms were designed for teachers to rate students, research results suggest that a Chinese-translated version of the GRS-S may be acceptable to use parents as raters (Li, Lee, Pfeiffer, & Petscher, 2008). The GRS-S validity was demonstrated as it was compared to IQ scores from the WISC-IV and successful in identifying individuals who are intellectually gifted and not intellectually gifted, particularly when intellectually gifted is operationally defined IQ score within the top 5% (Pfeiffer & Jarosewich, 2007).

**Scales for Rating Behavioral Characteristics of Superior Students**

The Scales for Rating Behavioral Characteristics of Superior Students (SRBCSS) are widely used in the selection of students for gifted and talented programs in Grades K to 12 (Callahan et al., 1995; Hunsaker & Callahan, 1995). The SRBCSS was initially designed to introduce teacher perspectives into the gifted and talented identification process (Bracken & Brown, 2006). These rating scales are based on a multiple talent approach to identifying gifted students (Renzulli, 1986). Included in the SRBCSS are the following 14 scales that help identify student abilities: learning, motivation, creativity, leadership, art, music, drama, planning, communication (precision), communication (expression), math, reading, science, and technology. The development of this scale was based on a literature review as well as feedback from educators. Although the publisher reports no criterion-validity, the reliability has been found to be significant if those completing the assessment have been trained (Center for Creative Learning, 2002). Test-retest and interrater reliability based on teacher ratings have been found to be excellent for the learning, motivation, and creativity scales (Jarosewich, Pfeiffer, & Morris, 2002).

**Guidelines for Using Creativity Checklists**

When using creativity checklists, it is crucial for the assessors to be familiar with the students for whom creativity is being assessed. The raters should have had the opportunity to observe and work with students in a variety of different contests and domains. The validity of creativity checklists depends on how well the assessors know the students, how well they understand the questions and theory of creativity that underlies them, the objectivity of the assessors, and the appropriateness of the questions/theory of creativity that underlies them.

There currently is not a creativity checklist that exists that has the criterion-related concurrent and predictive validity one would like of tests being used for decision-making purposes. It is helpful for several people to independently rate students. While checklists may not be the most psychometrically sound assessments, they can serve as a small piece of an assessment when combined with other measures to help paint a picture of the student’s creative abilities.
Self-Assessment

Self-assessment, in which one asks people to judge their own creativity, is one of the simplest ways to assess creativity. This method seems too easy and good to be true—and dependant on the purpose of the assessment, it may be exactly that.

Creative Personality Assessment

Personality inventories are some of the most prevalent forms of self-assessment. The Five-Factor Theory is the leading theory in personality (Costa & McCrae, 1992). It organizes personality into five components: neuroticism (emotional stability), extraversion, openness to experience, conscientiousness, and agreeableness. Openness to experience is the personality component most associated with creativity. Examples of items on this subscale include “I have a big imagination,” and “I spend time reflecting on things” (Costa & McCrae, 1992). There is a near-universal finding that openness to experience is associated with creativity, whether measured by divergent thinking tests (King, McKee-Walker, & Broyles, 1996; McCrae, 1987), the Consensual Assessment Technique (Wolfradt & Pretz, 2001), or self-report measures (Griffin & McDermott, 1998; Soldz & Vaillant, 1999). The NEO Personality Inventory and briefer NEO Five-Factor Inventory are the most popular measures of the five-factor personality theory (Costa & McCrae, 1992). The International Personality Item Pool (Goldberg et al., 2006) is a free alternative.

Another type of creativity personality assessment is to assess one’s creativity style, which refers to the ways in which people choose to use their creativity (Selby, Treffinger, Isaksen, & Powers, 1993). An example of a style measure is the Kirton Adaption-Innovation Inventory (KAI; Kirton, 1999) which was developed to measure a personality dimension ranging from adaption (ability to do things better) to innovation (ability to do things differently). This 32-item inventory is often used in organizations in the public and private sector as it is relevant to organizational change (Bobic, Davis, & Cunningham, 1999).

Creative Behavior Checklists

Rather than asking people questions regarding personality, another method for self-assessment of creativity is through creative behavior checklists, which asks people to rate past or current creative accomplishments. The author of The Creative Behavior Inventory argued that self-reports of activities and attainments are among the best techniques for measuring creativity (Hocevar, 1981). While deciding what activities and attainments are decidedly creative, lists that have been used in research are considered to have reasonable face validity and are recognized by society as important (Hocevar, 1981). This inventory is constructed of 90 items that assess creative behavior in literature, music, crafts, art, performing arts, and math/science. Another creative behavior checklist is the Creative Achievement Questionnaire (CAQ; Carson, Peterson, & Higgins, 2005). This instrument assesses creativity through 96 items across nine domains that load on two factors: the arts (drama, writing, humor, dance, visual arts) and science (invention, science, and culinary). The 10th domain (architecture) did not load on a factor.

Another self-report instrument, which has been used with elementary school students, is the Abedi-Schumacher Creativity Test (CT). The CT is unique in having a Spanish-language version that has been studied extensively (Abedi, 2002; Auzmendi, Villa, & Abedi, 1996). Items include questions such as, “How do you approach a complex task?” with response options of “I come up with a single approach,” “I may be able to come up with a few approaches,” and “I will be able to come up with a variety of approaches.” The CT provides scores in the traditional DT
components of fluency, originality, elaboration, and flexibility, making it a nice complement to DT assessments. Users, however, should be aware that CT results do not correlate significantly with those from the TTCT (Auzmendi et al., 1996).

The Runco Ideation Behavior Scale (RIBS; Runco, Plucker, & Lim, 2000-2001) emphasizes ideation: the use of, appreciation of, and skill of generating ideas. The internal consistency estimates for the 23-item scale tend to be in excess of .90. Runco et al. (2000-2001) concluded that although the RIBS was a sufficiently reliable instrument for use with groups and individuals, the construct validity evidence was somewhat ambiguous. Sample items include “I come up with a lot of ideas or solutions to problems,” “I enjoy having leeway in the things I do and room to make up my own mind,” and “My ideas are often considered ‘impractical’ or even ‘wild.’” Although the psychometric analyses were conducted using scores from college students, the instrument should be appropriate for use with younger children.

**Conclusion**

When gifted identification systems use IQ or other ability tests, they are often using measures that have been extensively tested and validated (A. S. Kaufman, 2009). Yet all the creativity assessments discussed in this review have limitations that prevent them from being the obvious choice to measure creativity as part of a larger assessment. Some have poor reliability or validity evidence; others are impractically long or the psychometrics with young students is unexamined. Given these restrictions, why is it worthwhile to include a creativity assessment as part of a general battery? We argue that there are several key components that a creativity assessment (however flawed) brings to any identification system.

First, there is reason to think that traditional IQ or achievement tests may not tap all of a person’s potential. Divergent-thinking or creativity tests may help give a more comprehensive understanding of a person's overall abilities. Creativity is typically both theoretically and empirically related to intelligence, but at a small (but significant) level; one recent meta-analysis found the average correlation to be $r = .17$ (Kim, 2005). The same can be said of creativity checklists completed by teachers. Second, some minority or female test-takers may be at risk for a stereotype threat reaction to traditional tests (Nguyen & Ryan, 2008). Almost all measures of creativity show less ethnicity and gender biases than standard IQ and achievement tests (J. C. Kaufman, 2006, 2010).

Third, if a test-taker has a learning disability that may affect his or her scores on a traditional ability or achievement measure, creativity tests can add valuable information. Finally, creativity is part of the national guidelines to use multiple selection criteria (rather than rely solely on IQ and achievement data). Most creativity measures, although not necessarily the only solution for such selections, help serve as part of a broader evaluation that can add to the overall picture of each candidate.

It is important to reiterate that we do not support administering a creativity test instead of a traditional IQ, achievement, or behavior test (whether as part of a gifted assessment or any testing scenario). We believe that all of these measures can and should work together to create the fullest possible picture of an individual. Creativity assessment is a work in progress—we know far less about creativity and its measurement than we would like to know—yet there are still a multitude of tools available for a dedicated test administrator.

**Declaration of Conflicting Interests**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References


