Assessing creative thinking
Practical applications

Haiying Long
FLORIDA INTERNATIONAL UNIVERSITY, USA

Jonathan A. Plucker
UNIVERSITY OF CONNECTICUT, USA

Introduction

There is a general consensus among psychologists and educators that creative thinking improves students’ motivation, metacognitive capacities, interpersonal and intrapersonal skills, and their abilities to solve problems, write creatively, and interpret scientific process (Beghetto, 2010; Plucker, Beghetto, & Dow, 2004; Smith & Smith, 2010; Torrance, 1959). Yet in recent years, with policymakers and stakeholders paying increased attention to students’ scores on standardized achievement tests, creative thinking seems to be “squeezed out of many educational arenas” and to “belong on the endangered species list” (Beghetto & Kaufman, 2010, p. 191). In this context, teachers spend less time on training students’ creative thinking. Even worse, they show unfavorable attitudes towards creativity; therefore, unintentionally discouraging it in the classroom (Cropley, 2006; Oral & Guncer, 1993; Russ, 1993; Scott, 1999). These concerns are not uniquely present in the United States, but rather are widely shared in other cultures, such as China (Niu & Sternberg, 2003), Korea (Kim, 2009), and the United Kingdom (Craft, 2005).

One initial step to address these concerns and increase teachers’ awareness of nurturing students’ creativity in the classroom is to help teachers identify creative thinking and related attitudes and dispositions. For example, taking risks is generally considered to be indicative of creativity, but teachers often find risk-taking behaviors to be undesirable as they seek to maintain an orderly classroom. A better understanding of the nature and identification of creativity can help teachers understand that risk-taking behaviors can be, in certain circumstances, constructive ways to promote learning and creativity. This chapter presents an overview of creativity assessment under the school context. It begins with a discussion of the importance of creative thinking in teaching and further examines a few most widely used creativity assessment approaches. It concludes with recommendations for integrating creativity assessment into teaching and classroom assessment.
Creative thinking in teaching

The importance of creative thinking cannot be overemphasized in teaching. In his famous 1950 Presidential Address for the American Psychological Association, J. P. Guilford stated, “it is our main objective to teach students how to think, and this means also to think constructively. Certainly, if we succeeded in this objective, there should be much evidence of creativeness in the end product” (p. 448). Later, Torrance (1959) indicated four reasons why psychologists should not neglect studies on creativity. Besides its significant contribution to the advancement of society, creative thinking promotes personality development, enhances information acquisition, academic achievement, and future career success. Vygotsky, the eminent Russian psychologist, also noted that “exercise of the imagination” (1967/2004, p. 88) is crucial for preparing school-age children for their future.

Researchers have also developed several theories to understand the essential role of creative thinking in learning. For instance, Sternberg (2002, 2003), in his Theory of Successful Intelligence, defined intelligence as an ability to “achieve success in life in terms of one’s personal standards, within one’s sociocultural context” (Sternberg, 2003, p. 141) and further argued that success is achieved by analytical, creative, and practical abilities. As a crucial component of success, creative ability plays its role when individuals are engaged in novel tasks or situations. In his Three-Ring Conception of giftedness, Renzulli (1978) contended that gifted people have three “interlocking clusters of traits, . . . above-average though not necessarily superior general ability, task commitment, and creativity” (p. 3).

However, the importance of creativity need not be limited to career success. Recently, Beghetto and Kaufman (2007) developed the concept of mini-c, which specifically focuses on students, their creative development and process of learning. It refers to “the novel and personally meaningful interpretation of experiences, actions, and events” (p. 73, emphasis in original). According to this view, learning is not just acquiring and memorizing facts, but demonstrating an in-depth understanding of concepts and flexible use of knowledge that are achieved in the process of creating (Deci, Vallerand, Pelletier, & Ryan, 1991). Therefore, “student imagination and curiosity drive the learning process. Creativity becomes the vehicle for understanding—even in the context of predetermined learning goals” (Beghetto & Plucker, 2006, p. 324). Encouraging mini-c in classroom has at least three benefits for students (Beghetto & Kaufman, 2010): (a) they can build confidence by recognizing their potential of creative abilities; (b) they can develop their own interpretation of ideas during the process of sharing and learning; (c) they can gain insights from others’ understandings and ideas (pp. 196–197).

Different approaches to assessing creativity

Creative thinking has mainly been assessed by divergent thinking tests, consensual views of judges, and creativity questionnaires (Long, 2014; Plucker & Makel, 2010; Plucker & Renzulli, 1999). These tools can be applied to assess students’ creative thinking in general as well as in specific settings. The introduction of creativity assessment approaches in this chapter is far from exhaustive. Readers who want to find a detailed list of these approaches are referred elsewhere (e.g., Hunsaker & Callahan, 1995; Kaufman, Plucker, & Baer, 2008; Plucker & Makel, 2010; Plucker & Renzulli, 1999).

Divergent thinking tests

Divergent thinking is “cognition that leads in various directions” (Runco, 1999, p. 577). Although not synonymous with creativity, it is an important ability in creative problem solving
Assessing creative thinking

and enables people to think of a variety of ideas (Guilford, 1968). Divergent thinking tests have been extensively used to assess creative thinking among students of varying ages during the past decades (Kaufman et al., 2008; Long, 2014; Plucker & Renzulli, 1999; Plucker & Runco, 1998). For instance, based on the results of divergent thinking tests obtained from a random sample of 100 students from the third through fifth grades, Torrance (1968) concluded that children showed a significant slump in creativity around fourth grade, the famous “fourth-grade slump” of creativity. About four decades later, Kim (2011) examined the developmental trend of creative thinking with six normative samples of 272,599 students from kindergarten through twelfth grade and adults and found that individuals’ creative thinking scores have significantly decreased since 1990, with the most significant decline from kindergarten to third grade.

Unlike most standardized tests that require only one correct answer, divergent thinking tests provide individuals with open-ended figural or verbal tasks and require them to generate several ideas. The degree of creative thinking is then assessed through the quantity and quality of the ideas produced in the process. The most commonly used divergent thinking tests include Torrance Tests of Creative Thinking (TTCT; Torrance, 1962, 1974, 1990), Wallach–Kogan Creativity Tests (WKCT; Wallach & Kogan 1965), and Guilford tests (Christensen, Merrifield, & Guilford, 1953; Guilford, 1959, 1967, 1979; Guilford & Christensen, 1956; Wilson, Christensen, Merrifield, & Guilford, 1960). Over the past several decades, these tests have been adapted for different populations and cultures. For instance, they have been translated into different languages and been used in Korea (Jeon, Moon, & French, 2011), Brazil (Wechsler, 2006), Italy (Scibinetti, Tocci, & Pesce, 2011), and Turkey (Aslan & Puccio, 2006), among many other countries. WKCT also has Chinese versions in paper-and-pencil and electronic forms (Lau & Cheung, 2010).

Although there are several divergent thinking assessments in the literature, the majority of the tasks on those assessments are quite similar. For example, in both TTCT and WKCT individuals are asked to generate creative use of some objects (i.e., unusual uses tasks). A figural task of Lines/Circles is included in TTCT and the Line Meanings task is in WKCT and both tasks present individuals with lines or circles to interpret. Participants are asked to list consequences of some situations (e.g., people no longer need to sleep) in Guilford’s test, whereas they are asked to imagine what would happen in similar situations in Just Suppose tasks of TTCT (Plucker & Makel, 2010). Due to these similarities, divergent thinking tasks from different tests are often grouped together. Of all the divergent thinking tasks, unusual uses tasks (e.g., to generate creative uses for a brick or a knife) and instances tasks (e.g., to generate creative instances of things that are round) are used most frequently (Long, 2014).

Over the years, many other divergent thinking tasks or tests have been developed. For instance, Okuda, Runco, and Berger (1991) created real-world divergent thinking questions about school and home. One example is as follows:

Your friend Teddy sits next to you in class. Teddy likes to talk to you a lot and often bothers you while you are doing your work. Sometimes the teacher scolds you for talking, and many times you don’t finish your work because he is bothering you. What are you going to do? Remember to give as many answers as you can.

(p. 47)

Wyrick (1996) constructed a motor creativity test to examine creativity in physical education and required participants to use as many different ways as possible of performing the parallel line test, Ball Wall test, and hoop test (Bournelli, Makri, & Mylonas, 2009; Pagona & Costas, 2008). Wu et al. (1998) developed the Creative Thinking Test to measure creativity among the Chinese population and one of the tasks was unusual uses of chopsticks, a traditional eating utensil in Asian
countries (Cheng, Wang, Liu, & Chen, 2010; Yeh, 2004). However, the basic tasks in these alternative divergent thinking assessments are similar to those in other divergent thinking tasks.

In administrating divergent thinking tests, some differences appear in the instructions and the conditions of taking the tests. In Guilford’s tests individuals are instructed to be as creative as they could. According to Christenson, Guilford, and Wilson’s (1957) study, the prompt to “be creative” increases inventiveness or creativity of the responses. This conclusion has been corroborated by subsequent research (e.g., Runco, Illies, & Eisenman, 2005; Silvia et al., 2008). As another popular test, WKCT allows individuals to take tests in a game-like and untimed condition because Wallach and Kogan (1965) found creativity can be distinguished from intelligence under that condition.

After the ideas are produced in the tests, they are evaluated based on a set of scoring criteria. The traditional four criteria are: fluency, originality, flexibility, and elaboration. The score of fluency comes from the number of ideas produced. Originality focuses on the unusualness of the ideas but tests show some variations on what should be counted as unusual. For example, WKCT gives a 1 to any responses that are generated by only one person and gives a 0 to all other responses. In TTCT, the score of originality is calculated by statistical infrequency of the ideas. Plucker, Runco, and Lim (2006) define the ideas that are produced by 10 percent of students as original. Flexibility refers to the number of categories of responses. Elaboration measures the extension of ideas (Kaufman et al., 2008; Runco, 1999). In TTCT, for instance, participants are given an extra point when a detail is identified on a task.

Meanwhile, some alternative scoring methods of divergent thinking tests are also proposed. Because it is generally agreed in the field that creativity is composed of novelty and appropriateness/usefulness (Plucker, Beghetto, & Dow, 2004), these two aspects are often considered scoring criteria. Amabile’s (1982) Consensual Assessment Technique (CAT, discussed later) was also used to evaluate idea produced in divergent thinking tests (e.g., Dollinger, 2003; Friedman, 2009; Friedman, Forster, & Denzler, 2007; Kousoulas, 2010). Another instance is the use of composite creativity score (CCS). In this method, ratings of novelty and appropriateness in individuals’ responses are obtained and they are then averaged across judges for all solutions to the problems. With this score, an average novelty and appropriateness score is calculated for every individual for each problem and a product moment correlation value between the novelty and appropriateness scores is computed. Based on all these scores, CCS is calculated by using the formula \( \sqrt{N^2 + A^2 + 1rN \times A} \) where \( N \) is novelty score, \( A \) is appropriateness score, \( r \) is the correlation between two scores (Rastogi & Sharma, 2010, p. 143).

The popularity of divergent thinking tests does not suggest that they are not without controversies. One recurring criticism is their lack of predictive validity. More specifically, the tests cannot make good predictions of individuals’ future achievement or behavior, such as creative accomplishments ten years later. Researchers argue that this issue reflects the weakness in methodology more than weaknesses in the psychometric integrity of the tests (Kaufman et al., 2008; Plucker & Makel, 2010; Plucker & Renzulli, 1999). Another issue with the tests is concerned with the response set (Thorndike, 1972), which means that the scores from different evaluation criteria do not yield discrete scores in factor analysis because they are generated from the same pool of responses. Pertain to this problem is the confounding relationship between fluency and originality (Dixon, 1979; Hocevar, 1979; Michael & Wright, 1989), although researchers have recently proposed several strategies for addressing this issue (e.g., Plucker, Qian, & Wang, 2011; Plucker, Qian, & Schmalensee, 2014; Silvia et al., 2008).

**The Consensual Assessment Technique**

The Consensual Assessment Technique (CAT) is another widely used approach to assess creative thinking. This technique does not focus on creative thinking *per se* but on the products
Assessing creative thinking

which are indicative of creative process and of individuals. According to Amabile (1982) who initially developed this technique:

A product or response is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated. Thus, creativity can be regarded as the quality of products or responses judged to be creative by appropriate observers, and it can also be regarded as the process by which something so judged is produced.

(p. 1001)

Based on this definition, the procedure of the CAT was evolved. Individuals produce their works, which are later evaluated by a panel of experts. The judges are not required to follow any criteria of creativity during evaluation or justify their criteria after evaluation, but only base the evaluation on their subjective criteria. In order to make the technique reliable, tasks should be open-ended. There are five special requirements for the procedure, mainly with regard to judges. First, the judges may not be highly creative but they should have at least some familiarity with the domain of interest. Second, the judges must assess the products independently and without any training. Third, the judges should assess other dimensions of the work along with the creativity of the works, such as technical aspects and aesthetic appeal, in order to examine how independent these dimensions are. Fourth, the judges should rate products relative to one another instead of using absolute criteria. Fifth, the judges should rate the works in a random order.

Since it was created, the CAT has been applied to assess creativity in different fields, such as writing (e.g., Baer, 2003; Kaufman, Baer, & Gentile, 2004; Kaufman, Gentile, & Baer, 2005), art (e.g., Amabile, 1982; Stanko-Kaczmarek, 2012), music (e.g., Brinkman, 1999; Eisenberg & Thompson, 2011), and engineering (e.g., Fodor & Carver, 2000). It has also been used in evaluating creativity of the solutions to simulated real-life problems. For instance, Shalley (1991) created an in-basket task which asked individuals to work as a human resource director in a steel company and to solve interpersonal and security issues. This assessment technique was also found in studies on other countries, including China (e.g., Chen et al., 2005; Niu, 2007; Niu & Sternberg, 2003), France (e.g., Botella, Zenasni, & Lubart, 2011), among others. Many studies (e.g., Friedrich & Mumford, 2009; Palmiero, Cardi, & Belardinelli, 2011) employed a procedure that is similar to the CAT. Interestingly, researchers (e.g., Dollinger, 2003; Friedman, 2009; Freidman, Forster, & Denzler, 2007; Kousoulas, 2010) began to assess ideas generated from divergent thinking tests with this technique (Long, 2014). Due to its wide popularity, the CAT was praised as the “gold standard” of creativity assessment (Carson, 2006; Kaufman et al., 2008).

Meanwhile, the CAT was adopted to meet the needs of the researchers in specific domains. For example, Barbot, Orriols, and Pouyade (2008–2010) computerized CAT procedure and created Consensual Assessment Technique-Interface (CAT-i) in which products were presented in a random order in the computer and a 7-point rating scale was provided for all the judges. Barnard (1992) developed Consensual Assessment of Interior Design Creativity (CAIDC) with which the interior design products are evaluated on 12 dimensions. Cropley and Kaufman (2012) created the Creative Solution Diagnosis Scale (CSDS) to measure functional creativity. Horng and Lin (2009) constructed an eight-category Creative Culinary Product Criteria Matrix to judge products in culinary contests.

Although Amabile (1982) specified the procedure and requirements of the CAT, researchers modified some of them. One modification was concerned with judges. Unlike using experts as recommended by Amabile, judges with different levels of expertise were selected. For example, Kaufman et al. (2005) used high school gifted creative writers as quasi-experts to evaluate
writing tasks, and found solid agreement with the experts. However, two follow-up studies that explored novice–expert agreement found different results. One hundred and six college students were employed as novice judges to assess SciFaiku poems and short stories (Kaufman, Baer, & Cole, 2009; Kaufman, Baer, Cole, & Sexton, 2008); however, their assessments did not correspond with expert ratings. Kaufman and Baer (2012) proposed that “quasi-experts” – people with some level of training in the field being assessed – could be used in lieu of actual experts. Kaufman, Baer, Cropley, Reiter-Palmon, and Sinnett (2013) found solid agreement between quasi-experts and experts in rating both creative writing and inventions; this finding is consistent with Plucker, Kaufman, Temple, and Qian’s (2009) investigation of movie reviews. Users on a movie website showed strong agreement with professional critics, whereas novice raters (college students) did not.

Amabile (1982) suggested that no criteria were offered to raters and the judgment was only based on their subjective conception of creativity. But several researchers provided criteria for judges in their studies. For instance, Paletz and Peng (2009) asked coders to rate the products on originality and plausibility. Studies that used consensus of the judges also indicated the scoring criteria, which mostly included two essential components of creativity: originality (or novelty) and appropriateness (or usefulness, fit) (Long, 2014). Additionally, training sessions were added to the procedure although Amabile (1982) required no training for judges before evaluation. In Dollinger and Shafran’s (2005) study, non-experts received four-minute pretraining and were shown 16 prototype drawings from earlier studies prior to evaluation. Likewise, Reiter-Palmon, Mumford, O’Connor Boes and Runco (1997) trained raters to use the criteria of quality and originality as well as the anchors of the scales to evaluate sample problems. They further brought the raters together for a panel meeting and discussed discrepancies.

Compared with divergent thinking tests, the CAT has many advantages, such as, it focuses on products and represents a “clever solution for the ‘criterion problem’ in creativity research” (Plucker & Makel, 2010, p. 59). However, this technique has a few caveats. First, it involves more subjective judgment, which cannot transform into any standardized scores, thus, cannot be compared with scores obtained from other methods of assessing creativity. Second, the procedure is time-consuming and hiring experts is costly.

Questionnaires/scales

Besides using divergent thinking tests and the CAT to measure creative thinking, researchers also utilized a wide range of questionnaires to assess other aspects of creativity, including creative personality, behavior/activity, affect and motivation, and creativity in teaching. Table 26.1 presents several questionnaires or surveys that are not only widely used but also most relevant to teaching. This section only introduces the first questionnaire listed for each category in Table 26.1, which was used the most frequently in the field (Plucker & Makel, 2010). Information on other, related instruments can be found elsewhere (e.g., Kaufman et al., 2008; Plucker & Makel, 2010; Plucker & Renzulli, 1999).

Creative personality. Creative Personality Scale (CPS; Gough, 1979) was developed on the assumption that creative people have characteristics in common, such as being independent, confident, and humorous. It is an adjective checklist and consists of 18 positive and 12 negative items. In the scoring process, a score of 1 is given to a positive item, -1 to a negative item, and 0 to not checked items. CPS was mostly used in the form of self-report and with different populations, including both teachers (e.g., Forrester & Hui, 2007) and students (e.g., Charyton & Snelbecker, 2007). Often, researchers asked parents or teachers to report children’s creative personality. For instance, based on the Adjective Check List, an earlier version of CPS, which contained 300 adjectives, Runco
Table 26.1 Creativity questionnaires

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personality</td>
<td>Creative Personality Scale (CPS, Gough, 1979)</td>
</tr>
<tr>
<td></td>
<td>NEO Personality Inventory-Revised (Costa &amp; McCrae, 1992)</td>
</tr>
<tr>
<td></td>
<td>International Personality Item Pool (IPIP) (Goldberg, 1999)</td>
</tr>
<tr>
<td></td>
<td>What Kind of Person Are You? (Khatena &amp; Torrance, 1976)</td>
</tr>
<tr>
<td>Behavior and Activity</td>
<td>Creative Behavior Inventory (CBI, Hocevar, 1979)</td>
</tr>
<tr>
<td></td>
<td>Runco Ideational Behavior Scale (RIBS, Runco, Plucker, &amp; Lim, 2001)</td>
</tr>
<tr>
<td></td>
<td>Creative Achievement Questionnaire (Carson et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>Creativity Domain Questionnaire (Kaufman &amp; Baer, 2004)</td>
</tr>
<tr>
<td></td>
<td>Creativity Styles Questionnaire-Revised (Kumar et al., 1997)</td>
</tr>
<tr>
<td></td>
<td>Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS; Renzulli, Hartman, &amp; Callahan, 1981)</td>
</tr>
<tr>
<td></td>
<td>Kaufman Domains of Creativity Scale (KDOCS; Kaufman, 2012)</td>
</tr>
<tr>
<td>Affect and Motivation</td>
<td>Affect in Play Scale (Russ, 1993)</td>
</tr>
<tr>
<td></td>
<td>Creative Self-efficacy (Tierney &amp; Farmer, 2002)</td>
</tr>
<tr>
<td>Teaching and Development</td>
<td>The Creativity Checklist (Proctor &amp; Burnette, 2004);</td>
</tr>
<tr>
<td></td>
<td>Fostering Creativity Questionnaire (Teo &amp; Waugh, 2010)</td>
</tr>
<tr>
<td></td>
<td>Teacher Observation Form (Peters &amp; Gates, 2010)</td>
</tr>
<tr>
<td></td>
<td>Creative Learning Characteristics (Cheng, 2011)</td>
</tr>
<tr>
<td></td>
<td>Creativity Fostering Teacher Index (Soh, 2000)</td>
</tr>
<tr>
<td></td>
<td>Inventory of Personal Factors in Technological Creativity Development (Yeh, 1999)</td>
</tr>
</tbody>
</table>

(1989) created a 25-item Parental Evaluation of Children’s Creativity (PECC). Parents were asked to report their children’s creativity on a 1–7 scale of frequency where 1 is rarely and 7 is extremely. In the same manner, Lim and Smith (2008) selected 13 items from the checklist of adjectives and required teachers to assess their students’ creative characteristics on a five-point scale (i.e., 1 “always true” and 5 “not true at all”). Like some divergent thinking tests, CPS was translated into other languages, such as Chinese (Cheung, Lau, Chan, & Wu, 2004) and Korean (Lim & Smith, 2008). Studies showed that this scale has acceptable reliability (Plucker & Makel, 2010).

Creative behavior/activity. Creativity researchers believe that individuals’ past creative experience, activity, or behavior predicts their future creativity (Colangelo, Kerr, Huesman, Hallowell, & Gaeth, 1992; Plucker, 1998). Based on this idea, Hocevar (1976) revised the Holland behavior checklists and created Creative Behavior Inventory (CBI; 1979). By taking this inventory, individuals report their engagement in activities of six domains: science and mathematics, literature and writing, crafts, music, performing arts, and fine arts. There are 90 items in this inventory and some examples of the activities described in the items are: wrote a short story, made a leather craft, applied math in an original way to solve a practical problem. For each item, individuals indicate the frequency of their involvement on a four-point scale ranging from 1 “never” to 4 “more than five times”. An individual’s final score is the ratio of the total score of the responses to the items and the number of items. Hocevar (1979) reported that the internal consistency reliability of the inventory ranged from .63 to .89 and subsequent research consistently supported its relatively high levels of reliability and validity (Hocevar & Bachelor, 1989; Plucker, 1999, 2004; Silvia, Wigert, Reiter-Palmon, & Kaufman, 2012). Because CBI contains many items, Dollinger (2003) removed accomplishments that were more...
relevant to eminent people and shortened it as a 28-item scale that represented activities in visual, crafts, literary, and performing arts. This shortened form gained popularity in the field and was applied to measure creative accomplishments of college students in several studies (e.g., Dollinger, 2003, 2007; Dollinger, Burke, & Gump, 2007; Dollinger, Palaskonis, & Pearson, 2004; Silvia & Kimbrel, 2010). Due to its long-time use in the field and its robust reliability and validity, creativity as measured by CBI was employed as criterion variable in validating other creativity inventories, such as the Minnesota Multiphasic Personality Indicator second edition (MMPI-2; Nassif & Quevillon, 2008). It was also used to provide evidence for the content generality of creativity (Plucker, 1999, 2004).

**Affect in creativity.** Creativity researchers that developed and used this scale discussed creative thinking from the perspective of child development (Russ & Fiorelli, 2010). The *Affect in Play Scale* (APS; Russ, 1993) is based on a model of creativity that emphasized the importance of cognitive abilities, personality traits as well as affective processes in creative activities. In this model, the term affect, rather than emotion, was used because the latter was considered a more inclusive concept. Children are engaged in play every day, therefore, play becomes a primary arena to study their affective processes. Among various types of play, pretend play that includes pretending behaviors, engagement of fantasy and make-believe is most relevant to creativity (Russ, 1998). This type of play is regarded as “a natural form of creativity” (Russ, p. 469).

The APS targets children of 6–10 years of age and aims to evaluate three aspects of children’s affect expressions displayed in a pretend play task: affect states, affect-laden thoughts, and cognitive integration of affect (Russ, 1993, 1998, 2004, 2013). The task uses two neutral-looking puppets and three small blocks of different colors that are placed on a table. Children are asked to play with the puppets in any way they like for five minutes. In order to provide children with more opportunities to play, the task is designed with less structure. The play session is videotaped and coded later by trained psychologists or graduate students. Five individual scores could be obtained from the coding. Organization score measures the quality and complexity of the plot in the play; imagination score assesses the fantasy elements, block transformations, and number of novel ideas, characters, or events; score of comfort in play rates the level of comfort, involvement, and enjoyment of the play; frequency of affect indicates the number of units of verbal and nonverbal emotions; variety of affect score records the number of affect categories (Russ & Dillon, 2011). But researchers can also choose to utilize some of the indicators. For instance, four scores were generated in Butcher and Niec’s (2005) study, including scores of comfort, imagination (or quality of fantasy), frequency of affect, and variety of affect categories. Only two scores, frequency of affect and variety of affect categories, were analyzed in Russ and Schaffer’s (2010) study. In some studies (Butcher & Niec, 2005; Russ & Kaugars, 2000), the frequency of affect score was further divided into negative and positive affect scores. Previous studies have demonstrated that the interrater reliability of APS scale is mostly between .80 and .90 (Russ, 2004; Seja & Russ, 1999) and its construct validity has also been well supported (Niec & Russ, 2002; Russ, Niec, & Kaugars, 2000).

**The Creativity Checklist.** The Creativity Checklist (Proctor & Burnette, 2004) was developed as a tool particularly for teachers to observe creativity characteristics of students. The researchers that created this checklist believed that creative thinking involves not only cognitive skills but also some motivational and dispositional traits. They also contended that one effective way for teachers to measure students’ creativity is to observe students’ characteristics when they are attending classes and completing tasks in classrooms. Further, the form of checklist is convenient and time efficient. According to the researchers, this measurement can be used as “one component of a range of assessment instruments or procedures encompassing the four Ps

The checklist includes nine items, and each item serves as a descriptor for a list of performance indicators that help teachers identify students’ behaviors in classrooms. For example, item 1 of this checklist is “a fluent thinker,” which corresponds to the description of performance indicators “The student . . . is full of ideas; finds different ways of doing things; answers questions fluently and readily; hypothesizes easily; generally possesses high verbal fluency; can list, tell/retell, label and compile easily; answers (fluently) questions such as How many? Why? What are the possible reasons for? Just suppose . . . ?” (Proctor & Burnette, 2004, p. 426). A three-point Likert scale is used to rate each item. The scale of 1 (rarely) indicates that students have the characteristics described for less than 30 percent of the time; 2 (sometimes) suggests approximately 30–70 percent of the time; and 3 (often) is linked to more than 70 percent of the time. According to the developers, this scale showed moderate test-retest reliability and the factor analysis demonstrated a single factor solution.

**Recommendations**

The research on creativity assessment has many lessons and implications for teachers, with the primary message being that creativity can be precisely defined and assessed. This flies in the face of many people’s implicit conceptions of creativity (Plucker et al., 2004), to the point that Beghetto and Plucker (2006) cite these misconceptions as the primary barriers to promoting creativity in the classroom.

Additionally, the conventional wisdom holds that fostering creativity and teaching content knowledge are mutually exclusive actions, but this is not the case (Baer & Garrett, 2010; Beghetto & Plucker, 2006). Learning is the process of constructing knowledge and understanding; therefore, creativity is always present in the classroom, regardless of whether it is explicitly addressed (Beghetto & Plucker, 2006), and the amount of time children spend in schools and classrooms make them optimal contexts for fostering and nurturing creativity.

Although high-stakes standardized tests are unavoidable in the political contexts of most Western countries – tests about which teachers often have little say in creation and administration – teachers have more autonomy in their teaching and self-made classroom assessments, where creativity assessment can be added. Problem-based learning activities are particularly well-suited to CAT-like assessments, as both the activities and assessments model real-world creativity activity: Creativity is often manifested in the solving of problems, which often results in a solution or product that is evaluated by both the creators and other stakeholders (e.g., a teammate, supervisor, critic, or other person or group invested in the effectiveness of the solution). For example, in one problem-based summer experience, high school students working in teams were required to invent working telephone prototypes to facilitate the embedded learning of physics, chemistry, and design content and principles. At the conclusion of the unit, each team presented their products to a panel of experts that included an inventor, a high school physics teacher, and a college professor. This experience modeled real-world creative processes, applied creativity assessment research to the design of constructive, formative and summative assessments, and did so within an engaging, creativity-enhancing context (see Gorman, Plucker, & Callahan, 1998; Plucker & Gorman, 1999). Invention units are another form of problem-based learning that promotes the fostering of creativity and the use of creativity assessments to provide helpful feedback to students (Plucker, 2002).

Second, when teachers seek to assess student creativity – in general, in specific content areas, and on specific tasks – they should realize that many high quality assessment strategies are...
available to them. Regarding general creativity, divergent thinking tests and several questionnaires can be easily adapted to most classroom situations; several questionnaires or inventories are designed for specific content area creativity. And the consensual assessment technique, with its flexible administration and scoring strategies, is well-suited to task-specific creativity assessment.

Finally, researchers should communicate more effectively with teachers about the latest developments in creativity assessment. During the past decade, the study of creativity has received increasing attention from researchers. New psychometric evidence has been accumulated for existing measures, and new questionnaires have been constructed to measure different aspects of creativity. In the five major creativity journals, about 30 new questionnaires were used between 2003 and 2012. Half of the 16 questionnaires employed to assess creativity in teaching were newly produced (Long, 2014). On the other hand, teachers can also engage themselves in creating new methods and scales of assessing creativity. As educators who communicate with students every day, teachers can play an active role in assessing students’ traits.

This proactive stance could be reflected in many ways. Academics could ensure that the fostering of creativity is included in preservice and inservice teaching regarding curriculum and instruction design, and creativity could easily be included as a brief mini-topic when tests and assessments are addressed in teacher preparation programs. Given the importance of creativity and problem-solving to student learning and adult success, it is surprising that creativity has so little of a footprint in preservice and inservice teacher education. And creativity researchers could be more proactive in sharing the latest assessment developments at teacher-focused conferences and in teacher-directed journals and magazines, where this material rarely if ever appears.

References


Haiying Long and Jonathan A. Plucker


Assessing creative thinking


