

Methodological Issues in Measuring Creativity: A Systematic Literature Review*

Sameh Said-Metwaly

University of Leuven, Belgium

E-mail address: sameh.metwaly@kuleuven.be

Eva Kyndt

University of Leuven, Belgium

E-mail address: eva.kyndt@kuleuven.be

Wim Van den Noortgate

University of Leuven, Belgium

E-mail address: wim.vandennoortgate@kuleuven.be

ARTICLE INFO

Keywords:

Creativity measurement
Creativity assessment
Creativity ratings
Creativity testing

Article history:

Received 08 September 2017

Received in revised form 05 December 2017

Accepted 07 December 2017

ISSN: 2354-0036

DOI: 10.1515/ctra-2017-0014

ABSTRACT

The growing body of creativity research has raised several challenging issues with regard to the measurement of this construct. This paper aims to provide a review of current challenging methodological issues related to measuring creativity. Five methodological issues are discussed: selecting measurement instruments, sampling, testing conditions, psychometric properties and domain-general/specificity of creativity. This paper reveals that there remain a number of unresolved issues and serious questions surrounding the measurement of creativity. Research gaps and suggestions for future research are discussed.

INTRODUCTION

In recent years, creativity has increasingly attracted attention as one of the twenty-first century skills that students need to develop in order to attain success in the information age. Creative individuals and their unique ideas constitute a powerful element in facing rapid and complex changes from various worldwide sources of competition (Kilgour, 2006). Thus, a great deal of research effort has been directed towards the understanding of creativity and its determinants.

The measurement of creativity is one of the focal topics in creativity research (Kaufman, Baer, Cole, & Sexton, 2008; Park, Chun, & Lee, 2016; Plucker & Makel, 2010). Hocevar (1981) stated that creativity is probably more difficult to measure than any other psychological construct. The search for reliable and valid instruments to measure

* This research was supported by a grant from the Government of Egypt

this construct has been the most challenging issue confronting researchers and educators interested in creativity (Batey, Chamorro-Premuzic, & Furnham, 2010; Davis & Belcher, 1971; Mouchiroud & Lubart, 2001). Over the past decades, many instruments have been developed for measuring creativity, which look at different aspects of this construct, including creative processes, creative products, personality traits of creative individuals or the climate where creativity occurs (Barbot, Besançon, & Lubart, 2011; Batey, 2012; Fishkin & Johnson, 1998; Plucker & Renzulli, 1999; Rhodes, 1961).

Although significant progress has been made over the last decades, there are still many issues and challenges surrounding the measurement of creativity (Lemons, 2011; Park et al., 2016; Plucker & Runco, 1998). Researchers interested in measuring creativity are faced with a variety of methodological decisions that might have serious impact on the resulting outcomes. They need to make decisions regarding which instruments to use to measure creativity, whether these instruments have adequate psychometric properties, how to create ideal testing conditions to maximize creativity scores and whether to measure creativity as domain-general or domain-specific (Cropley, 2000; Lemons, 2011; Plucker & Runco, 1998; Zeng, Proctor, & Salvendy, 2011). A review of these issues is required to reach a clearer understanding of the nature of the current debate over these issues and the significance of this debate to creativity research (Lemons, 2011). In our recent paper (Said-Metwaly, Kyndt, & Van den Noortgate, 2017), we reviewed the existing approaches to measuring creativity, pointing out commonly used instruments and the advantages and weaknesses of each approach. In this paper, we try to provide an extended overview of the measurement of creativity by shedding light on challenging methodological issues facing researchers in this area. Through this literature review, we seek to provide researchers with an up-to-date account of the current state of creativity measurement, highlight some of the methodological issues and challenges surrounding this measurement, discuss possible causes and ways to overcome these challenges and lay the basis for revealing a series of gaps and questions that could be addressed in future research.

METHOD

We searched the creativity literature published up to December 31st, 2016 in accordance with these four steps: first, we searched the following databases: ERIC, Google Scholar, JSTOR, PsycINFO and Web of Science using the following search string: (“creativity” OR “creative thinking” OR “creative performance” OR “creative ability” OR “creative potential”) AND (“measurement” OR “assessment” OR “evaluation” OR “testing”). Second, we checked the reference lists of the papers identified in the first step for further relevant papers (i.e. “backward search”). Third, additional publications were

retrieved through searching databases for papers that referred to the previously identified papers in steps 1 and 2 in their citations (i.e. “forward search”). Fourth, we searched the following key journals in creativity: *Creativity Research Journal*, *Gifted Child Quarterly*, *Psychology of Aesthetics, Creativity, and the Arts*, *The Journal of Creative Behavior* and *Thinking Skills and Creativity*.

The papers identified using the search process were first screened as to their relevance on the basis of their titles and abstracts. The remaining papers were included if they met the following criteria: (1) were written in English, (2) were journal articles, conference papers or dissertations, (3) addressed or discussed the measurement of creativity and methodological issues relevant to this measurement, and for which (4) the full text was available. Studies using qualitative approaches for measuring creativity and intervention studies were excluded.

The analysis of the data was informed by the guiding objectives of the review. A thematic approach was employed to analyze the final list of included papers. The analysis involved categorizing the content of the included papers according to the aim, issues discussed and conclusions of each paper.

RESULTS

An initial review of titles and abstracts of 2,064 papers identified through the search process yielded 198 papers. Of these, 131 papers met the inclusion criteria. The 131 included papers addressed the measurement of creativity and issues related to this measurement. Five challenging methodological issues pertaining to measuring creativity were identified: selecting measurement instruments, sampling, testing conditions, psychometric properties and domain-generalty/specificity of creativity. These issues are discussed below and summarized in Table 1.

Table 1
A Summary of Methodological Issues in Measuring Creativity

Aspect	Limitations
Selecting Measurement Instruments	<p>Divergence of instruments used as a result of researcher's dissimilar conceptions of creativity.</p> <p>Narrow aspects of creativity outlined by each instrument.</p> <p>Using instruments in isolation, neglecting the multidimensional nature of creativity.</p>
Sampling	<p>Inconsistent results of creativity research due to different instruments used.</p> <p>Small number of studies conducted on children.</p> <p>Limited number of studies carried out in African and American countries other than the USA.</p> <p>Small sample size in some studies.</p> <p>Inconsistent results for reliability and validity due to sampling from different sub-populations.</p>
Testing Conditions	<p>Different performance on creativity instruments due to varying testing conditions.</p> <p>Inconsistent results regarding the influence of testing conditions on creative performance.</p> <p>Unexplored reasons behind the influence of testing conditions on creative performance.</p>
Psychometric Properties	<p>Conflicting evidence of validity.</p> <p>Lack of updated and developmentally and culturally appropriate norms.</p> <p>Limited application of modern psychometric analyses.</p>
Domain-Generality/ Specificity of Creativity	<p>Mixed results on whether creativity is domain-general or domain-specific.</p> <p>More research is needed to obtain a definitive answer to this issue.</p>

Selecting measurement instruments

Selecting appropriate instruments to measure creativity is an essential and ongoing challenge for creativity researchers. The variety in definitions of this construct complicates the selection of suitable instruments to measure it (Batey et al., 2010; Mumford & Gustafson, 1988; Piffer, 2012; Treffinger, Renzulli, & Feldhusen, 1971; Zampetakis, 2010). The definition of creativity that a researcher adopts will greatly influence the selection of the aspects of this construct that are to be studied and, in turn, the instrument chosen to measure them (Treffinger, Young, Selby, & Shepardson, 2002). For instance, a researcher who asserts the process-based view will probably measure creativity using divergent thinking tests, while a researcher who asserts the person-based view might measure creativity in terms of personality traits (Batey, 2012). This also applies to those who assert the product- or press-based views of creativity. Accordingly, the instruments used by these researchers often have a particular focus and reflect only a narrow range of the aspects of the creativity construct (Baer & McKool, 2009; Horn & Salvendy, 2006; Lemons,

2011; Plucker, Beghetto, & Dow, 2004; Zeng et al., 2011). In addition, these instruments are often used separately, and each instrument has its specific limitations (Horn & Salvendy, 2006; Kaufman et al., 2007; Lemons, 2011). The isolation of the four aspects of creativity may give the impression that these aspects are independent (Batey, 2012; Soroa, Balluerka, Hommel, & Aritzeta, 2015); however, the four aspects are mutually related (Batey, 2012). Given the multidimensional nature of creativity and the interrelationships among its different aspects, no single instrument can comprehensively capture this construct (Cropley, 2000; Fishkin & Johnson, 1998; Hennessey & Amabile, 2010; Lemons, 2011; Treffinger et al., 2002). Therefore, measuring creativity using single or isolated instruments may be considered as an extreme and somewhat misleading oversimplification. This may explain some of the inconsistent results in the creativity literature (Batey, 2012; Batey & Furnham, 2006).

There have been increasing calls for the use of a multidimensional approach for measuring creativity (Ambrose & Machek, 2015; Barbot et al., 2011; Lemons, 2011) by applying different instruments that exemplify various sources of information (Ambrose & Machek, 2015; Feldhusen & Goh, 1995; Fishkin & Johnson, 1998; Nakano, Primi, Ribeiro, & Almeida, 2016; Park et al., 2016; Treffinger et al., 2002). Triangulation of information sources provides an advantage to the measurement of creativity by compensating for the limitations in a single instrument through the parallel strengths of the other instruments (Jick, 1979). However, before making decisions regarding the instruments that will be included in creativity measurement, it is first important to reach a consensus on what “creativity” is and which creativity aspects need to be measured (Lemons, 2011). This mutual conception of creativity should reflect the multidimensionality (i.e. includes process, product, person and press aspects) of the creativity construct and the characteristics of the specific domain (domain-specific creativity skills). In addition, it is important to keep in mind that an individual’s whole creativity cannot be obtained through the simple sum of the scores of its aspects, because of the interrelated forces among these aspects (Hennessey & Amabile, 2010; Sternberg & Lubart, 1995). Moreover, Sternberg and Lubart (1995) have suggested that there may be a threshold for each of these aspects below which creativity is unattainable, regardless of an individual’s score on the other aspects. Resolving this issue necessitates constructing a clear picture of how each of these aspects contributes to an individual’s creativity and how the interrelationships among these aspects work.

Sampling

The process of selecting a representative sample from a targeted population and its impact on measuring creativity has received some attention in creativity research. This issue is crucial because the accuracy and the generalization of results depend on the extent to which the sample is representative of the population of interest. Some points should be highlighted with regard to sampling in creativity research. First, although there has been a recent increase in creativity studies conducted on children (e.g. Cheung & Lau, 2010; Dziedziewicz, Gajda, & Karwowski, 2014; Sayed & Mohamed, 2013), the number is still lower compared to that of studies on older subjects (Darvishi & Pakdaman, 2012; Han & Marvin, 2002; Long, 2014). This is regrettable considering the importance of early recognition and nurturing of creative children for individual and social development purposes (Han & Marvin, 2002). Second, creativity studies have rarely included samples from African or American countries other than the USA, and cross-cultural studies have been restricted to comparisons between the USA and a few Asian countries, in particular China and Korea (Long, 2014). Third, the sample size in some creativity studies (e.g. Amabile, 1982; Baer, 1991, 1994a, 1994c; Johns, Morse, & Morse, 2000; Runco & Okuda, 1991) was small (i.e. less than 30). This may raise considerable concerns about the conclusions based on the results of these studies.

Finally, the characteristics of the subpopulation appear to influence the results of creativity studies with respect to reliability and validity. Milgram and Milgram's (1976a) study, for example, indicated that the validity of the Wallach-Kogan Creativity Tests (WKCT), as assessed by correlations with the scores from a self-report questionnaire measuring nonacademic accomplishments in nine performance domains, was different for male and female high school students. The correlations for males were significant in social leadership, writing, community service and fine arts domains, while for females they were only significant in writing and fine arts domains. In addition, Runco's (1986a) study revealed that the WKCT only had validity in gifted samples. Furthermore, Runco and Albert (1985) compared gifted and non-gifted children in terms of the reliability and validity of originality scores measured by the WKCT. The results indicated that the reliability and validity of originality scores were higher for gifted children than for non-gifted children. Similarly, Runco (1985) found that flexibility scores, measured by the WKCT, were reliable and valid only for high-achieving children when compared to low-achieving children. In line with these findings, Kim, Cramond, and Bandalos (2006) reported different factor structures of the Torrance Tests of Creative Thinking (TTCT) for three samples of kindergarten children, third graders and sixth graders. Also, the unidimensional factor structure of the

TTCT was more prevalent in studies with a smaller sample size (less than 400) (e.g. Clapham, 1998; Heausler & Thompson, 1988; Hocevar, 1979), while the multidimensional factor structure of the TTCT was more prevalent in studies with a larger sample size (more than 400) (e.g. Kim, 2006; Kim et al., 2006; Krumm, Aranguren, Filippetti, & Lemos, 2014; Krumm, Lemos, & Filippetti, 2014). Similar findings regarding the effect of different samples on the reliability and validity of measuring creativity were reported in earlier studies (e.g. Dewing, 1970; González, Campos, & Pérez, 1997; Howieson; 1981).

Testing conditions

Individuals' performance on instruments measuring creativity appears to be influenced by the conditions under which these instruments are administered (Beghetto, 2005; Hattie, 1977, 1980; Plucker & Renzulli, 1999; Runco & Albert, 1985; Treffinger et al., 1971). There has been a great deal of research on the optimal conditions for administering creativity instruments (see Chen et al., 2005; Hattie 1977; Lemons, 2011). This body of research has been concerned with comparing different testing conditions, including atmosphere (test-like vs. game-like), time limit (timed vs. untimed), setting (individual vs. group) and instruction (explicit "be creative" vs. standard) (Plucker & Renzulli, 1999). Wallach and Kogan's (1965) claim that game-like and untimed conditions enhance performance on creativity instruments has guided a large part of this research. However, empirical findings regarding this issue have been mixed. Some studies re-asserted Wallach and Kogan's claim with regard to game-like (e.g. Dansky & Silverman, 1973; Dentler & Mackler, 1964; Drwal, 1973; Sandlund, Linnarud, & Norlander, 2001; Ward, Kogan, & Pankove, 1972), untimed (e.g. Cropley, 1972; Johns & Morse, 1997; Kelly & McGrath, 1985; Torrance, 1969) or both conditions (e.g. Adams, 1968; Vernon, 1971), while other studies reported contrary results (e.g. Busse, Blum, & Gutride, 1972; Channon, 1974; Hattie, 1980; Johns et al., 2000; Kogan & Morgan, 1969; Leith, 1972; Van Mondfrans, Feldhusen, Treffinger, & Ferris, 1971; Williams & Fleming, 1969).

With regard to the effect of the testing setting on creativity, Milgram and Milgram's study (1976b) was the only one, to our knowledge, that investigated the impact of individual versus group administration of the WKCT on the performance of gifted and non-gifted children. The study results revealed an adverse effect of the group administration on creativity scores of non-gifted children, with no significant effect for gifted children. Although somewhat prematurely, Milgram and Milgram (1976b) concluded that the group administration of creativity instruments might require subjects with average or high intelligence, but a lower level of intelligence might be sufficient for the individual administration of these instruments. Regarding the effect of testing instruction on creativity, previous studies have yielded mixed results. Some studies found that individuals scored higher

on creativity instruments when given an explicit “be creative” instruction rather than a standard (generic) instruction (e.g. Chen, Himsel, Kasof, Greenberger, & Dmitrieva, 2006; Chua & Iyengar, 2008; Niu & Sternberg, 2003; Nusbaum, Silvia, & Beaty, 2014; Paulus, Kohn, & Arditti, 2011; Rietzschel, Nijstad, & Stroebe, 2014), whereas other studies revealed negative or no effects of explicit instruction on creativity scores (e.g. Chand & Runco, 1993; Johns & Morse, 1997; Runco, 1986b; Runco & Okuda, 1991; Ward, Saunders, & Dodds, 1999).

Although this review shows an inconsistency in results across studies, the most common finding is that variations in testing conditions lead to different performances on creativity instruments. At the same time, what remains unclear is the reasons for these differences and the most likely results under each of these conditions (Hattie, 1980; Treffinger et al., 1971). There are two suggestions that might help in this regard.

First, it is possible that particular personality variables (e.g. motivation, experience or anxiety) interact with testing conditions and, as such, affect individual creativity; individuals with specific characteristics are more likely than others to benefit from one condition or another (Chua & Iyengar, 2008; Kogan & Morgan, 1969). Kogan and Morgan (1969), for example, indicated that test-anxious subjects produced lower creativity scores in a test-like atmosphere, but the opposite was not established in a game-like atmosphere. Likewise, O’Hara and Sternberg (2000-2001) showed an interaction effect of instructions and thinking styles on creativity and found that explicit instruction was less beneficial for individuals with a legislative style (i.e. who prefer to solve problems using their own strategies) when compared to those with a judicial style (i.e. who prefer to analyze and evaluate others’ ideas). Accordingly, to investigate the relationship between testing conditions and performance on creativity instruments, a clear understanding of the potential moderating effects of personality-relevant variables on such relationship is needed.

Second, the influence of testing conditions on creativity might depend on the characteristics of the task (e.g. type, difficulty or familiarity) administered to the subjects (Chen et al., 2005; Harrington, 1975; Kogan & Morgan, 1969). Task-specific characteristics might make an individual more susceptible to changes in testing context when responding to this task. Van Mondfrans et al. (1971) revealed that an untimed condition increased creativity scores on figural tasks, but did not have a significant effect on the scores on verbal tasks. Also, Smith, Michael, and Hocevar (1990) concluded that anxiety-inducing instructions, compared to anxiety-reducing instructions, led to lower creativity scores on mathematical tasks, but not on verbal or figural tasks. Moreover, Chen et al. (2005) reported that the proportion of performance variance explained by explicit instruction varied

greatly across creativity tasks (artistic, verbal and mathematical), ranging from 1% on verbal tasks to 30% on artistic tasks. Hence, the characteristics of the task appear to act as an additional moderating variable that needs to be considered when investigating the variations in creativity under different testing conditions. In summary, research findings in this area emphasize the need for further studies to determine exactly how testing conditions affect performance on creativity instruments (Chua & Iyengar, 2008; Johns et al., 2000; Sandlund et al., 2001).

Psychometric properties

Evaluating the psychometric properties of creativity instruments has received a great deal of attention in the creativity research. However, previous investigations have provided mixed support for the psychometric properties of creativity instruments (Cropley, 2000; Plucker & Makel, 2010). Acceptable estimates of reliability have been reported for creativity instruments, exceeding .70 in most cases (Cropley, 2000; Hocevar & Bachelor, 1989). On the contrary, many concerns have been raised about the validity of these instruments in measuring creativity (Barbot et al., 2011; Hocevar & Bachelor, 1989; Torrance & Haensly, 2003; Treffinger et al., 1971).

With regard to the content validity, each creativity instrument looks at a particular aspect of creativity, assuming that this narrow aspect is representative of the entire creativity construct. Furthermore, most of these instruments regard creativity as a general construct, without considering the specific characteristics of the domain or task of interest (Barbot et al., 2011). Accordingly, Zeng et al. (2011) questioned the content validity of creativity instruments, suggesting that these instruments might be more appropriate for measuring children's creativity as their domain-specific expertise has not yet been developed or become evident. Similarly, Cropley (2000) proposed that it is better to think of these instruments as measures of creative potential rather than creative performance; creative performance relies on further aspects not measured by these instruments.

With respect to the construct validity of creativity instruments, as evaluated by convergent and divergent validity, the existing evidence is not encouraging (Baer, 2015; Brougher & Rantanen, 2009; Fishkin & Johnson, 1998; Hocevar & Bachelor, 1989; Sawyer, 2006). In his review of creativity measurement, Simonton (2003) stated that scores on creativity instruments often correlate weakly with each other (indicating low convergent validity) and correlate highly with intelligence scores (indicating low divergent validity). Clapham (2004) examined the relationship between two divergent thinking tests (verbal and figural TTCT) and two creativity interest questionnaires (How Do You Think? and How Creative Are You?). The results showed very weak correlations between these

types of instruments, which ranged from .06 to .25. More interestingly, Clapham (2004) indicated that the average correlation between these two types of instruments was lower than the average correlation between these instruments and academic aptitude/achievement measures. Moreover, a principal components analysis of the subscale scores on the four creativity instruments resulted in three distinct factors which accounted for 50.78% of the scores' variance and corresponded only to the type of the instrument (verbal divergent thinking, figural divergent thinking and interest), with no factor that combined these two types of instruments. Based on these results, Clapham (2004) stated that it could no longer be assumed that different types of creativity instruments measure the same construct and therefore they should not be used interchangeably. Other researchers have reported similar results (e.g. Baer, 1991, 1994a; Belcher et al., 1981; Busse et al., 1972; Dollinger, Burke, & Gump, 2007; Han, 2003; Kaufman et al., 2010). In contrast, other studies have provided support for the convergent and divergent validity of creativity instruments (e.g. Batey & Furnham, 2006; Davis & Belcher, 1971; Domino, 1994; Harrington, Block, & Block, 1983; Hattie, 1980; Hong, Milgram, & Gorsky, 1995; Kaufman & Baer, 2004; Milgram & Milgram, 1976a; Runco, 1984).

With regard to their predictive validity, creativity instruments have been subjected to considerable criticism regarding their ability to predict future creative achievement (Lemons, 2011). While reviewing the measurement of creativity, Cropley (2000) came to the conclusion that creativity instruments had relatively weak predictive validity with correlation coefficients around .50, compared to coefficients of about .70 for intelligence. However, recent studies have shown that creativity instruments have greater predictive validity than intelligence instruments (e.g. Cramond, Matthews-Morgan, Bandalos, & Zuo, 2005; Gajda, Karwowski, & Beghetto, 2016; Plucker, 1999a; Runco, Millar, Acar, & Cramond, 2010). Besides the conflicting evidence concerning validity, most creativity instruments lack updated and developmentally and culturally appropriate norms for interpreting the performance of different groups (Barbot et al., 2011).

Another pressing issue related to the psychometric properties of creativity instruments is that the field is still living in the past through the almost exclusive dependence on classical psychometric analyses (Plucker & Makel, 2010). Unfortunately, creativity researchers have not reaped the benefits of modern psychometric analyses (e.g. item response theory, generalizability theory and structural equation modelling) (Zampetakis, 2010). Perhaps this is not surprising given that many of the creativity instruments were developed and standardized before the advent of modern psychometric analyses (Fishkin & Johnson, 1998). There have been some efforts to apply modern psychometric analyses

in creativity research, and although limited, their results are promising. These efforts addressed some critical issues in creativity measurement, such as evaluating reliability and validity (Chermahini, Hickendorff, & Hommel, 2012; Karwowski, 2014; Lee, Lee, & Youn, 2005; Primi, 2014; Silvia, 2011; Silvia et al., 2008; Wang, Ho, Cheng, & Cheng, 2014; Zampetakis, 2010), testing domain-specificity of creativity (Barbot, Tan, Randi, Santa-Donato, & Grigorenko, 2012; Chen et al., 2006; Silvia, Kaufman, & Pretz, 2009), evaluating the rater's effect on performance assessment (Hung, Chen, & Chen, 2012) and modelling relationships between creativity and other constructs (Nusbaum & Silvia, 2011; Silvia, 2008). Expanding the use of these modern analyses could provide a better understanding of conflicting results in creativity research.

Domain-generality/specificity of creativity

One of the most contentious issues related to the measurement of creativity is the domain-generality/specificity of creativity (Baer, 2012, 2016; Baer & McKool, 2009; Barbot, Besançon, & Lubart, 2016; Han, 2003; Han & Marvin, 2002; Ivcevic, 2007; Lubart & Guignard, 2004; Plucker, 1998; Plucker & Beghetto, 2004). This issue concerns the extent to which a person's creativity is relevant to the content domain or to the tasks within that domain (Diakidoy & Spanoudis, 2002). The ongoing debate concerns whether creativity is domain-general (i.e. a person who exhibits creativity in one domain is more likely to exhibit creativity in other domains) or domain-specific (i.e. creativity in a certain domain is not related to creativity in other domains) (Mohamed, Maker, & Lubart, 2012; Silvia, Kaufman, & Pretz, 2009). In other words, is there a general factor (or factors) that might be called "c", like the "g" of intelligence, that contributes to creativity regardless of the particular domain or task? Or, alternatively, is there a collection of specific factors with each of them contributing to creativity in only one domain or specific task? (Baer, 1994c, 2012; Kaufman & Baer, 2002). Generally, the domain-generality view has prevailed in creativity research for several decades (Baer, 2015; Plucker, 2004). However, creativity researchers have started to shift their focus towards the domain-specificity view due to the existing evidence that goes against the domain-generality view, such as the notion that the work of eminent creative individuals throughout history is almost exclusively restricted to a single domain (Barbot et al., 2016; Chen et al., 2006; Zeng et al., 2011), the low correlations among creativity ratings in different domains (Chen et al., 2006), the limited power of general creativity instruments in predicting real-life creative achievement (Ayas & Sak, 2014) and the low transferability of the effect of task-specific creativity training programs to other tasks (Baer, 1994a; Barbot et al., 2011; Dow & Mayer, 2004).

Settling the generality/specificity of creativity issue has considerable implications for researchers and practitioners in the field of creativity. Under the domain-generality view,

neither creativity measurement nor creativity training requires to be associated with domains (Baer, 2016). However, following the domain-specificity view would potentially result in questioning the efficiency of current educational practices used for identifying and nurturing creative children (Baer, 1994d). Creativity measurement under the domain-specificity view should be domain-specific and focus on the skills germane to the creative performance in the domain at hand (Baer, 2012, 2015, 2016). By considering creativity as domain-specific, measuring creativity needs to focus not only on the quantitative dimension of creativity (what level?), but also on the qualitative dimension (what domain?). In the same vein, according to the domain-specificity view, creativity training programs should target specific creativity skills relevant to the domain of interest (Baer, 1996, 2012; Baer & Kaufman, 2005; Plucker, 1999b). Baer (2015) outlined how the generality/specificity of creativity issue matters, by stating that “The problem is that much of what we think we have learned from decades of creativity research—using what we now know to be invalid tests—may simply not be true” (p. 172).

Previous studies that addressed the issue of the generality/specificity of creativity have produced inconsistent results. Some of these studies supported the domain-generality view of creativity (e.g. Chen et al., 2006; Hocevar, 1976; Kaufman & Baer, 2004; Plucker, 1999b), while other studies supported the domain-specificity view (e.g. Baer, 1991, 1994a, 1994b, 1996; Han, 2003; Han & Marvin, 2002; Palmiero, Nori, Aloisi, Ferrara, & Piccardi, 2015; Reiter-Palmon, Robinson-Morrall, Kaufman, & Santo, 2012). There are some possible reasons for the dichotomy in the results of these studies. It has been suggested that one of the reasons is that the conclusions of the studies regarding this issue might vary as a result of the method by which creativity is measured (the method effect hypothesis) (Plucker, 1999b, 2004). Generally, product-based studies tend to show evidence of domain-specificity (e.g. Baer, 1991, 1994a, 1994b, 1996; Palmiero et al., 2015), while person-based studies tend to show evidence of domain-generality (e.g. Kaufman & Baer, 2004; Plucker, 1999b). Among the few studies that included these two types of measurements, Runco (1987) compared the generality of self-report (quantitative) and product (qualitative) ratings of creativity across seven domains (writing, music, crafts, art, science, performing arts and public presentation). Self-report ratings exhibited generality of creativity across these domains, while product ratings exhibited domain-specificity. Similarly, Plucker (2004) analyzed Runco’s (1987) data in addition to the data that he collected using confirmatory factor analysis and lent further support to the method effect hypothesis. However, this leaves us with the question of ‘what are the possible causes of that effect?’ (Weisberg, 2006). Baer (1998) suggested that the frequent appearance of the domain-generality finding in person-based studies might be at-

tributed to factors other than creativity. Relying on the response bias in self-reports, Baer (1998) claimed that individuals might adopt certain response styles in responding to self-report questionnaires that might lead to high correlations of creativity ratings across domains. However, this is nothing more than a potential cause that needs to be tested.

Product studies might go against the domain-generalizability of creativity due to the scoring method (i.e. CAT) used in these studies that makes creativity scores only comparable within the sample of interest. Furthermore, the domain-generalizability might be unsupported in product studies due to the insufficient number of tasks used in measuring creativity in these studies (Chen et al., 2006). Product studies (e.g. Baer, 1991, 1994a, 1994b, 1996, Han, 2003; Han & Marvin, 2002) often include one task for each of the performance domains, which might lead to low correlations among creativity ratings in these domains. In their investigation of this issue, Chen et al. (2006) increased the number of tasks by administering a battery of product-based tests that included 21 tasks from three domains (artistic, verbal and mathematical) and found evidence for domain-generalizability.

A further possible reason for the inconsistent results concerning this issue is the methodology used for testing the domain-specificity of creativity. Instead of establishing direct evidence for domain-specificity, domain-specificity is typically concluded through the absence of evidence supporting domain-generalizability; however this absence of evidence for domain-generalizability might be due to the statistical analyses used (Chen et al., 2006). Han and Marvin (2002) have claimed that the use of either bivariate or multilevel analyses could largely guide the findings concerning the generalizability/specificity of creativity question. Empirical results appear to provide support for this claim. Studies that used bivariate analyses often revealed domain-specificity for creativity (e.g. Baer, 1994b; Han, 2003; Han & Marvin, 2002), whereas studies that used multivariate analyses often revealed domain-generalizability (e.g. Chen et al., 2006; Kaufman & Baer, 2004; Plucker, 1999b).

The evidence presented above shows that the debate on the generalizability/specificity of creativity has not been settled yet and the dilemma in choosing a measurement approach and instrument still exists. In attempting to bring about a rapprochement between the two sides of the debate, hybrid models (e.g. Amabile, 1983, 1988, 1996; Baer & Kaufman, 2005; Plucker & Beghetto, 2004) propose that domain-general and domain-specific components act together to produce creative performance. These models are in accord with Sternberg's (1989) hybrid view that the domain-general and domain-specific components are complementary rather than conflicting. This hybrid view of creativity has been supported in recent studies (e.g. Barbot et al., 2016; Diakidoy & Spanoudis, 2002; Hong & Milgram, 2010; Kaufman, Cole, & Baer, 2009; Mohamed et al., 2012). However, this hybrid view is still far from being well established.

CONCLUSIONS

Our systematic review reveals a number of challenging methodological issues surrounding the measurement of creativity. First, this review highlights the lack of an agreed-upon measurement instrument for creativity due to the absence of a consensus among researchers regarding the definition of this construct. As a result, researchers often select instruments with a particular focus or capturing only a narrow range of creativity aspects. Second, the picture remains less clear with regard to the impact of different samples on the results of creativity studies with respect to reliability and validity and further research is required, especially for instruments other than the divergent thinking tests. Third, previous studies have reported inconsistent results with regard to the influence of varying testing conditions on creative performance. Fourth, evidence regarding the psychometric quality of creativity instruments is less convincing. Creativity instruments have often been criticized for lacking acceptable indicators of validity and inadequate norms, in addition to the limited application of modern psychometric analyses in examining their psychometric properties. Finally, the issue of whether the measurement of creativity should be domain-general or domain-specific is still open to question.

Based on these findings, we offer the following suggestions that might help in addressing these issues. First, creativity researchers need to come to a consensus on the aspects of creativity that need to be valued and measured on the basis of a mutual definition of this construct. This could reduce confusion and facilitate the selection of suitable instruments to measure this construct. Second, caution is needed when comparing the results of studies done in different subpopulations, because the results of one study are not necessarily informative in relation to the results of another study (Runco & Albert, 1985). Additionally, generalizing the results of a study on a particular sample to a larger population should be done carefully, taking into account the need to represent or model the different sampling factors that may have an influence on the results obtained. Third, further studies are needed to determine exactly how testing conditions affect performance on creativity instruments. These studies need to consider the characteristics of both the individual and the task to arrive at a clear picture of the effect of varying testing conditions on creativity. Such studies might provide valuable information for the development and administration of creativity instruments. It is also worth asserting that researchers need to pay more attention to variations in testing conditions to accurately compare the results of different studies. Fourth, the field of creativity needs to take further steps forward towards the employment of modern psychometric analyses. Reexamining the psychometric properties of the existing creativity instruments and previously published data could contribute to a deeper understanding of inconsistent results in creativity research. This could also

lead to further evidence that either supports the psychometric quality of the existing creativity instruments or highlights the need for developing more reliable and valid instruments. Finally, more research is necessary to obtain a definitive answer to the question of the domain-generality/specificity of creativity. The limitations of the previous studies should be overcome by analyzing data from different sources (e.g. different instruments, samples, domains and contexts). This could contribute towards a better understanding of how creativity works across different domains.

With these suggestions in mind, continued work on the research questions and contentious issues outlined in this review could result in a more detailed picture of the nature of creativity.

REFERENCES

References marked with an asterisk were included in the analysis

- *Adams, J. C. (1968). The relative effects of various testing atmospheres on spontaneous flexibility: A factor of divergent thinking. *The Journal of Creative Behavior*, 2, 187-194.
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43, 997-1013.
- Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45, 357-376.
- Amabile, T. M. (1988). A model of creativity and innovation in organizations. In B. M. Staw, & L. L. Cummings (Eds.), *Research in organizational behavior* (pp. 123-167). Greenwich, CT: JAI Press.
- Amabile, T. M. (1996). Evidence to support the componential model of creativity: Secondary analyses of three studies. *Creativity Research Journal*, 9, 385-389.
- *Ambrose, L., & Machek, G. R. (2015). Identifying creatively gifted students: Necessity of a multi-method approach. *Contemporary School Psychology*, 19, 121-127. doi: 10.1007/s40688-014-0020-z
- *Ayas, M. B., & Sak, U. (2014). Objective measure of scientific creativity: Psychometric validity of the creative scientific ability test. *Thinking Skills and Creativity*, 13, 195-205. doi: 10.1016/j.tsc.2014.06.001
- *Bachelor, P. (1989). Maximum likelihood confirmatory factor-analytic investigation of factors within Guilford's Structure of Intellect model. *Journal of Applied Psychology*, 74, 797-804.
- *Baer, J. (1991). Generality of creativity across performance domains. *Creativity Research Journal*, 4, 23-39.
- *Baer, J. (1994a). Divergent thinking is not a general trait: A multi-domain training experiment. *Creativity Research Journal*, 7, 35-46.

- *Baer, J. (1994b). Generality of creativity across performance domains: A replication. *Perceptual and Motor Skills*, 79, 1217-1218.
- *Baer, J. (1994c). Performance assessments of creativity: Do they have long-term stability? *Roeper Review*, 7, 7-11.
- *Baer, J. (1994d). Why you shouldn't trust creativity tests. *Educational Leadership*, 51, 80-83.
- *Baer, J. (1996). The effects of task-specific divergent-thinking training. *The Journal of Creative Behavior*, 30, 183-187.
- *Baer, J. (1998). The case for domain specificity in creativity. *Creativity Research Journal*, 11, 173-177.
- *Baer, J. (2012). Domain specificity and the limits of creativity theory. *The Journal of Creative Behavior*, 46, 16-29. doi: 10.1002/jocb.002
- *Baer, J. (2015). The importance of domain-specific expertise in creativity. *Roeper Review*, 37, 165-178. doi: 10.1080/02783193.2015.1047480
- *Baer, J. (2016). *Domain specificity of creativity*. San Diego, CA: Elsevier Academic Press.
- *Baer, J., & Kaufman, J. C. (2005). Bridging generality and specificity: The Amusement Park Theoretical (APT) model of creativity. *Roeper Review*, 27, 158-163.
- *Baer, J., & McKool, S. S. (2009). Assessing creativity using the consensual assessment technique. In C. S. Schreiner (Ed.), *Handbook of research on assessment technologies, methods, and applications in higher education* (pp. 65-77). Hershey, PA: IGI Global.
- *Barbot, B., Besançon, M., & Lubart, T. I. (2011). Assessing creativity in the classroom. *The Open Education Journal*, 4, 124-132. doi: 10.2174/1874920801104010058
- *Barbot, B., Besançon, M., & Lubart, T. I. (2016). The generality-specificity of creativity: Exploring the structure of creative potential with EPoC. *Learning and Individual Differences*, 52, 178-187. doi: 10.1016/j.lindif.2016.06.005
- *Barbot, B., Tan, M., Randi, J., Santa-Donato, G., & Grigorenko, E. L. (2012). Essential skills for creative writing: Integrating multiple domain-specific perspectives. *Thinking Skills and Creativity*, 7, 209-223. doi: 10.1016/j.tsc.2012.04.006
- *Batey, M. (2012). The measurement of creativity: From definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal*, 24, 55-65. doi: 10.1080/10400419.2012.649181
- Batey, M., Chamorro-Premuzic, T., & Furnham, A. (2010). Individual differences in ideational behavior: Can the big five and psychometric intelligence predict creativity scores? *Creativity Research Journal*, 22, 90-97. doi: 10.1080/10400410903579627
- *Batey, M., & Furnham, A. (2006). Creativity, intelligence, and personality: A critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs*, 132, 355-429.

- *Beghetto, R. A. (2005). Does assessment kill student creativity? *The Educational Forum*, 69, 263.
- *Brougher, S. J., & Rantanen, E. M. (2009). Creativity and design: Creativity's new definition and its relationship to design. *Proceedings of the Human Factors and Ergonomics Society*, 1, 605-609. doi: 10.1177/154193120905301005
- *Busse, T. V., Blum, P. & Gutride, M. (1972). Testing conditions and the measurement preschool children of creative abilities in lower-class. *Multivariate Behavioral Research*, 7, 287-298.
- *Chand, I., & Runco, M. A. (1993). Problem finding skills as components in the creative process. *Personality and Individual Differences*, 14, 155-162.
- *Channon, C. E. (1974). The effect of regime on divergent thinking scores. *British Journal of Educational Psychology*, 44, 89-91.
- *Chen, C., Himsel, A., Kasof, J., Greenberger, E., & Dmitrieva, J. (2006). Boundless creativity: Evidence for the domain generality of individual differences in creativity. *The Journal of Creative Behavior*, 40, 179-199.
- *Chen, C., Kasof, J., Himsel, A., Dmitrieva, J., Dong, Q., & Xue, G. (2005). Effects of explicit instruction to "be creative" across domains and cultures. *The Journal of Creative Behavior*, 39, 889-1009.
- *Chermahini, S. A., Hickendorff, M., & Hommel, B. (2012). Development and validity of a Dutch version of the Remote Associates Task: An item-response theory approach. *Thinking Skills and Creativity*, 7, 177-186. doi: 10.1016/j.tsc.2012.02.003
- Cheung, P. C., & Lau, S. (2010). Gender differences in the creativity of Hong Kong school children: Comparison by using the new electronic Wallach-Kogan Creativity Tests. *Creativity Research Journal*, 22, 194- 199. doi: 10.1080/10400419.2010.481522
- *Chua, R. Y., & Iyengar, S. S. (2008). Creativity as a matter of choice: Prior experience and task instruction as boundary conditions for the positive effect of choice on creativity. *The Journal of Creative Behavior*, 42, 164-180. doi: 10.1002/j.2162-6057.2008.tb01293.x
- *Clapham, M. M. (1998). Structure of the figural forms A and B of the Torrance Tests of Creative Thinking. *Educational and Psychological Measurement*, 58, 275-283.
- *Clapham, M. M. (2004). The convergent validity of the Torrance Tests of Creative Thinking and creativity interest inventories. *Educational and Psychological Measurement*, 64, 828-841.

- *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.
- *Cropley, A. J. (1972). Originality scores under timed and untimed conditions. *Australian Journal of Psychology, 24*, 31-36.
- *Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using? *Roeper Review, 23*, 72-79.
- *Dansky, J. L., & Silverman, I. W. (1973). Effects of play on associative fluency in pre-school-aged children. *Developmental Psychology, 9*, 38-43.
- Darvishi, Z., & Pakdaman, S. (2012). Fourth grade slump in creativity: Development of creativity in primary school children. *GSTF International Journal of Law and Social Sciences, 1*, 40-48.
- *Davis, G. A., & Belcher, T. L. (1971). How shall creativity be measured? Torrance Tests, RAT, Alpha Biographical, and IQ. *The Journal of Creative Behavior, 5*, 153-161.
- *Dentler, R. A., & Mackler, B. (1964). Originality: Some social and personal determinants. *Behavioral Science, 9*, 1-7.
- *Dewing, K. (1970). The reliability and validity of selected tests of creative thinking in a sample of seventh-grade West Australian children. *British Journal of Educational Psychology, 40*, 35-42.
- *Diakidoy, I. N., & Spanoudis, G. (2002). Domain specificity in creativity testing: A comparison of performance on a general divergent-thinking test and a parallel, content-specific test. *The Journal of Creative Behavior, 36*, 41-61.
- *Dollinger, S. J., Burke, P. A., & Gump, N. W. (2007). Creativity and values. *Creativity Research Journal, 19*, 91-103.
- *Domino, G. (1994). Assessment of creativity with the ACL: An empirical comparison of four scales. *Creativity Research Journal, 7*, 21-33.
- *Dow, G. T., & Mayer, R. E. (2004). Teaching students to solve insight problems: Evidence for domain specificity in creativity training. *Creativity Research Journal, 16*, 389-398.
- *Drwal, R. L. (1973). The influence of psychological stress upon creative thinking. *Polish Psychological Bulletin, 4*, 125-129.
- Dziedziewicz, D., Gajda, A., & Karwowski, M. (2014). Developing children's intercultural competence and creativity. *Thinking Skills and Creativity, 13*, 32-42. doi:10.1016/j.tsc.2014.02.006
- *Feldhusen, J. F., & Goh, B. E. (1995). Assessing and accessing creativity: An integrative review of theory, research, and development. *Creativity Research Journal, 8*, 231-247.

- *Fishkin, A. S., & Johnson, A. S. (1998). Who is creative? Identifying children's creative abilities. *Roepers Review*, 21(1), 40-46. doi: 10.1080/02783199809553925
- *Gajda, A., Karwowski, M., & Beghetto, R. A. (2016). Creativity and academic achievement: A meta-analysis. *Journal of Educational Psychology*. Advance online publication. doi: 10.1037/edu0000133
- *González, M. A., Campos, A., & Pérez, M. J. (1997). Mental imagery and creative thinking. *The Journal of Psychology*, 131, 357-364.
- *Han, K. (2003). Domain-specificity of creativity in young children: How quantitative and qualitative data support it. *The Journal of Creative Behavior*, 37, 117-142.
- *Han, K., & Marvin, C. (2002). Multiple creativities? Investigating domain-specificity of creativity in young children. *Gifted Child Quarterly*, 46, 98-109.
- *Harrington, D. (1975). Effects of explicit instructions to "be creative" on the psychological meaning of divergent thinking test scores. *Journal of Personality*, 43, 434-454.
- *Harrington, D. M., Block, J., & Block, J. H. (1983). Predicting creativity in preadolescence from divergent thinking in early childhood. *Journal of Personality and Social Psychology*, 45, 609-623.
- *Hattie, J. A. (1977). Conditions for administering creativity tests. *Psychological Bulletin*, 84, 1249-1260.
- *Hattie, J. A. (1980). Should creativity tests be administered under test-like conditions? An empirical study of three alternative conditions. *Journal of Educational Psychology*, 72, 87-98.
- *Heausler, N. L., & Thompson, B. (1988). Structure of the Torrance Tests of Creative Thinking. *Educational and Psychological Measurement*, 48, 463-468.
- Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, 61, 569-598. doi: 10.1146/annurev.psych.093008.100416
- *Hocevar, D. (1976). Dimensionality of creativity. *Psychological Reports*, 39 (3), 869-870.
- *Hocevar, D. (1979). The unidimensional nature of creative thinking in fifth grade children. *Child Study Journal*, 9, 273-278.
- *Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of Personality Assessment*, 45, 450-464.
- *Hocevar, D., & Bachelor, P. (1989). A taxonomy and critique of measurements used in the study of creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 53-75). New York: Plenum Press.
- *Hong, E., & Milgram, R. M. (2010). Creative thinking ability: Domain generality and specificity. *Creativity Research Journal*, 22, 272-287. doi: 10.1080/10400419.2010.503535

- *Hong, E., Milgram, R. M., & Gorsky, H. (1995). Original thinking as a predictor of creative performance in young children. *Roeper Review*, 18, 147-149.
- *Horn, D., & Salvendy, G. (2006). Consumer-based assessment of product creativity: A review and reappraisal. *Human Factors and Ergonomics in Manufacturing*, 16, 155-175.
- *Howieson, N. (1981). A longitudinal study of creativity: 1965-1975. *The Journal of Creative Behavior*, 15, 117-134.
- *Hung, S.-P., Chen, P.-H., & Chen, H.-C. (2012). Improving creativity performance assessment: A rater effect examination with many facet Rasch model. *Creativity Research Journal*, 24, 345-357. doi: 10.1080/10400419.2012.730331
- Ivcevic, Z. (2007). Artistic and everyday creativity: An act-frequency approach. *The Journal of Creative Behavior*, 41, 271-290.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24, 602-611. doi: 10.2307/2392366
- *Johns, G. A., & Morse, L. W. (1997). Research note: Divergent thinking as a function of time and prompting to "be creative" in undergraduates. *The Journal of Creative Behavior*, 31, 156-165.
- *Johns, G. A., Morse, L. W., & Morse, D. T. (2000). Divergent production in gifted adolescents using timed vs. untimed stimuli with creative prompting. *Roeper Review*, 22, 165-166.
- *Karwowski, M. (2014). Creative mindsets: Measurement, correlates, consequences. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 62-70. doi: 10.1037/a0034898
- *Kaufman, J. C., & Baer, J. (2002). Could Steven Spielberg manage the Yankees? Creative thinking in different domains. *The Korean Journal of Thinking & Problem Solving*, 12, 5-14.
- *Kaufman, J. C., & Baer, J. (2004). Sure, I'm creative-but not in mathematics! Self-reported creativity in diverse domains. *Empirical Studies of the Arts*, 22, 143-155.
- Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, 20, 171-178. doi: 10.1080/10400410802059929
- *Kaufman, J. C., Cole, J. C., & Baer, J. (2009). The construct of creativity: A structural model for self-reported creativity ratings. *The Journal of Creative Behavior*, 43, 119-134. doi: 10.1002/j.2162-6057.2009.tb01310.x
- *Kelly, J. R., & McGrath, J. E. (1985). Effects of time limits and task types on task performance and interaction of four-person groups. *Journal of Personality and Social Psychology*, 49, 395-407.

- Kilgour, M. (2006). Improving the creative process: Analysis of the effects of divergent thinking techniques and domain specific knowledge on creativity. *International Journal of Business and Society*, 7, 79-107.
- *Kim, K. H. (2006). Is creativity unidimensional or multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creativity Research Journal*, 18, 251-259.
- *Kim, K. H., Cramond, B., & Bandalos, D. L. (2006). The latent structure and measurement invariance of scores on the Torrance Tests of Creative Thinking-figural. *Educational and Psychological Measurement*, 66, 459-477.
- *Kogan, N., & Morgan, F. T. (1969). Task and motivational influences on the assessment of creative and intellectual ability in children. *Genetic Psychology Monographs*, 80, 91-127.
- *Kogan, N., & Pankove, E. (1974). Long-term predictive validity of divergent-thinking tests: Some negative evidence. *Journal of Educational Psychology*, 66, 802-810.
- *Krumm, G., Aranguren, M., Filippetti, V. A., & Lemos, V. (2014). Factor structure of the Torrance Tests of Creative Thinking verbal form B in a Spanish-speaking population. *The Journal of Creative Behavior*, 50, 150-164. doi: 10.1002/jocb.76
- *Krumm, G., Lemos, V., & Filippetti, V. A. (2014). Factor structure of the Torrance Tests of Creative Thinking figural form B in Spanish-speaking children: Measurement invariance across gender. *Creativity Research Journal*, 26, 72-81. doi: 10.1080/10400419.2013.843908
- *Lee, S., Lee, J., & Youn, C.-Y. (2005). A variation of CAT for measuring creativity in business products. *Korean Journal of Thinking and Problem Solving*, 15, 143-153.
- *Leith, G. (1972). The relationship between intelligence, personality, and creativity under two conditions of stress. *British Journal of Educational Psychology*, 42, 240-247.
- *Lemons, G. (2011). Diverse perspectives of creativity testing: Controversial issues when used for inclusion into gifted programs. *Journal for the Education of the Gifted*, 34, 742-772. doi: 10.1177/0162353211417221
- *Long, H. (2014). An empirical review of research methodologies and methods in creativity studies (2003-2012). *Creativity Research Journal*, 26, 427-438. doi: 10.1080/10400419.2014.961781
- Lubart, T. I. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal*, 13, 295-308.
- *Lubart, T. I., & Guignard, J. (2004). The generality-specificity of creativity: A multivariate approach. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: From potential to realization* (pp. 43-56). Washington, DC: American Psychological Association.
- *Milgram, R. M., & Milgram, N. A. (1976a). Creative thinking and creative performance in Israeli students. *Journal of Educational Psychology*, 68, 255-259.

- *Milgram, R. M., & Milgram, N. A. (1976b). Group versus individual administration in the measurement of creative thinking in gifted and nongifted children. *Child Development, 47*, 563-565.
- *Mohamed, A., Maker, J., & Lubart, T. (2012). Exploring the domain specificity of creativity in children: The relationship between a non-verbal creative production test and creative problem-solving activities. *Turkish Journal of Giftedness and Education, 2*, 84-101.
- *Mouchiroud, C., & Lubart, T. (2001). Children's original thinking: An empirical examination of alternative measures derived from divergent thinking tasks. *The Journal of Genetic Psychology, 162*, 382-401.
- Mumford, M. D., & Gustafson, S. B. (1988). Creativity syndrome: Integration, application and innovation. *Psychological Bulletin, 103*, 27-43.
- *Nakano, T. C., Primi, R., Ribeiro, W. J., & Almeida, L. S. (2016). Multidimensional assessment of giftedness: Criterion validity of battery of intelligence and creativity measures in predicting arts and academic talents. *Anales de Psicología, 32*, 628-637. doi: 10.6018/analesps.32.3.259391
- Niu, W., & Sternberg, R. J. (2003). Societal and school influences on student creativity: The case of China. *Psychology in the Schools, 40*, 103-114.
- Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence, 39*, 36-45. doi: 10.1016/j.intell.2010.11.002
- *Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2014). Ready, set, create: What instructing people to "be creative" reveals about the meaning and mechanisms of divergent thinking. *Psychology of Aesthetics, Creativity, and the Arts, 8*, 423-432. doi: 10.1037/a0036549
- *O'Hara, L. A., & Sternberg, R. J. (2000-2001). It doesn't hurt to ask: Effects of instructions to be creative, practical, or analytical on essay-writing performance and their interaction with students' thinking styles. *Creativity Research Journal, 13*, 197-210.
- *Palmiero, M., Nori, R., Aloisi, V., Ferrara, M., & Piccardi, L. (2015). Domain-specificity of creativity: A study on the relationship between visual creativity and visual mental imagery. *Frontiers in Psychology, 6*, 1-8. doi: 10.3389/fpsyg.2015.01870
- *Park, N. K., Chun, M. Y., & Lee, J. (2016). Revisiting individual creativity assessment: Triangulation in subjective and objective assessment methods. *Creativity Research Journal, 28*, 1-10. doi: 10.1080/10400419.2016.1125259
- Paulus, P. B., Kohn, N. W., & Arditti, L. E. (2011). Effects of quantity and quality instructions on brainstorming. *First Quarter, 54*, 38-46. doi: 10.1002/j.2162-6057.2011.tb01083.x

- *Piffer, D. (2012). Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research. *Thinking Skills and Creativity*, 7, 258-264. doi: 10.1016/j.tsc.2012.04.009
- *Plucker, J. A. (1998). Beware of simple conclusions: The case for content generality of creativity. *Creativity Research Journal*, 11, 179-182.
- *Plucker, J. A. (1999a). Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data. *Creativity Research Journal*, 12, 103-114.
- *Plucker, J. A. (1999b). Reanalyses of student responses to creativity checklists: Evidence of content generality. *The Journal of Creative Behavior*, 33, 126-137.
- *Plucker, J. A. (2004). Generalization of creativity across domains: Examination of the method effect hypothesis. *The Journal of Creative Behavior*, 38, 1-12.
- *Plucker, J. A., & Beghetto, R. A. (2004). Why creativity is domain general, why it looks domain specific, and why the distinction does not matter. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: From potential to realization* (pp. 153-167). Washington, DC: American Psychological Association.
- *Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39, 83-96.
- *Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. Kaufman, & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity*. New York: Cambridge University Press.
- *Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of creativity. In R. J. Sternberg (Ed.). *Handbook of human creativity* (pp. 35-60). New York: Cambridge University Press.
- *Primi, R. (2014). Divergent productions of metaphors: Combining many-facet Rasch measurement and cognitive psychology in the assessment of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 461-474. doi: 10.1037/a0038055
- *Plucker, J. A., & Runco, M. A. (1998). The death of creativity measurement directions in creativity assessment has been greatly exaggerated: Current issues, recent advances, and future. *Roeper Review*, 21, 36-39. doi: 10.1080/02783199809553924
- *Reiter-Palmon, R., Robinson-Morrall, E. J., Kaufman, J. C., & Santo, J. B. (2012). Evaluation of self-perceptions of creativity: Is it a useful criterion? *Creativity Research Journal*, 24, 107-114. doi: 10.1080/10400419.2012.676980
- *Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappan*, 42, 109-135.

- *Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2014). Effects of problem scope and creativity instructions on idea generation and selection. *Creativity Research Journal*, *26*, 185-191. doi: 10.1080/10400419.2014.901084
- *Runco, M. A. (1984). Teachers' judgments of creativity and social validation of divergent thinking tests. *Perceptual and Motor Skills*, *59*, 711-717.
- *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. (1986a). Divergent thinking and creative performance in gifted and nongifted children. *Educational and Psychological Measurement*, *46*, 375-384.
- *Runco, M. A. (1986b). Maximal performance on divergent thinking tests by gifted, talented, and nongifted children. *Psychology in the Schools*, *23*, 308-315.
- *Runco, M. A. (1987). The generality of creative performance in gifted and nongifted children. *Gifted Child Quarterly*, *31*, 121-125.
- *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., Millar, G., Acar, S., & Cramond, B. (2010). Torrance Tests of Creative Thinking as predictors of personal and public achievement: A fifty year follow up. *Creativity Research Journal*, *22*, 361-368. doi: 10.1080/10400419.2010.523393
- *Runco, M. A., & Okuda, S. M. (1991). The instructional enhancement of the flexibility and originality scores of divergent thinking tests. *Applied Cognitive Psychology*, *5*, 435-441.
- Said-Metwaly, S., Kyndt, E., & Van den Noortgate, W. (2017). Approaches of measuring creativity: A systematic literature review. *Creativity. Theories - Research - Applications*, *4*(2), 238-275. doi: 10.1515/ctra-2017-0013
- *Sandlund, E. S., Linnarud, M., & Norlander, T. (2001). Effects of stress versus flotation-REST relaxation on creativity and literacy skills in advanced English as a Second Language (ESL) composition. *International Journal of Language & Communication*, *15*, 95-113.
- Sawyer, R. K. (2006). *Explaining creativity: The science of human innovation*. New York: Oxford University Press.
- Sayed, E. M., & Mohamed, A. H. H. (2013). Gender differences in divergent thinking: Use of the Test of Creative Thinking-Drawing Production on an Egyptian sample. *Creativity Research Journal*, *25*, 222-227. doi: 10.1080/10400419.2013.783760

- Silvia, P. J. (2008). Another look at creativity and intelligence: Exploring higher-order models and probable confounds. *Personality and Individual Differences*, 44, 1012-1021. doi: 10.1016/j.paid.2007.10.027
- *Silvia, P. J. (2011). Subjective scoring of divergent thinking: Examining the reliability of unusual uses, instances and consequences tasks. *Thinking Skills and Creativity*, 6, 24-30. doi: 10.1016/j.tsc.2010.06.001
- *Silvia, P. J., Kaufman, J. C., & Pretz, J. E. (2009). Is creativity domain-specific? Latent class models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics, Creativity, and the Arts*, 3, 139-148. doi: 10.1037/a0014940
- *Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 68-85. doi: 10.1037/1931-3896.2.2.68
- Simonton, D. K. (2003). Expertise, competence, and creative ability: The perplexing complexities. In R. J. Sternberg, & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 213-239). New York: Cambridge University Press.
- *Smith, K. L. R., Michael, W. B., & Hocevar, D. (1990). Performance on creativity measures with examination-taking instructions intended to induce high or low levels of test anxiety. *Creativity Research Journal*, 3, 265-280.
- Soroa, G., Balluerka, N., Hommel, B., & Aritzeta, A. (2015). Assessing interactions between cognition, emotion, and motivation in creativity: The construction and validation of EDICOS. *Thinking Skills and Creativity*, 17, 45-58. doi: 10.1016/j.tsc.2015.05.002
- *Sternberg, R. J. (1989). Domain-generality versus domain-specificity: The life and impending death of a false dichotomy. *Merrill-Palmer Quarterly*, 35, 115-130.
- Sternberg, R. J., & Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity*. New York: Free Press.
- *Torrance, E. P. (1969). Curiosity of gifted children and performances on timed and untimed tests of creativity. *Gifted Child Quarterly*, 13, 155-158.
- *Torrance, E. P., & Haensly, P. A. (2003). Assessment of creativity in children and adolescents. In C. R. Reynolds, & R. W. Kamphaus (Eds.) *Handbook of psychological and educational assessment of children: Intelligence, aptitude and achievement* (2nd ed.) (pp. 584-607). New York: The Guilford Press.
- *Treffinger, D. J., Renzulli, J., & Feldhusen, J. (1971). Problems in the assessment of creative thinking. *The Journal of Creative Behavior*, 5, 104-112.

- *Treffinger, D. J., Young, G. C., Selby, E. C., & Shepardson, C. (2002). *Assessing creativity: A guide for educators*. Sarasota: Center for Creative Learning.
- *Van Mondfrans, A. P., Feldhusen, J. F., Treffinger, D. J., & Ferris, D. R. (1971). The effects of instructions and response time on divergent thinking test scores. *Psychology in the Schools, 8*, 65-71.
- *Vernon, P. E. (1971). Effects of administration and scoring on divergent thinking tests. *British Journal of Educational Psychology, 41*, 245-257.
- *Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity intelligence distinction*. New York: Holt, Rinehart & Winston.
- *Wang, C., Ho, H., Cheng, C., & Cheng, Y. (2014). Application of the Rasch model to the measurement of creativity: The Creative Achievement Questionnaire. *Creativity Research Journal, 26*, 62-71. doi: 10.1080/10400419.2013.843347
- *Ward, T. B., Saunders, K. N., & Dodds, R. A. (1999). Creative cognition in gifted adolescents. *Roeper Review, 21*, 260-266.
- *Ward, W. C., Kogan, N., & Pankove, E. (1972). Incentive effects in children's creativity. *Child Development, 43*, 669-676.
- Weisberg, R. W. (2006). *Creativity: understanding innovation in problem solving, science, invention, and the arts*. Hoboken, N.J.: John Wiley & Sons.
- Williams, T. M., & Fleming, J. W. (1969). A methodological study of the relationship between associative fluency and intelligence. *Developmental Psychology, 1*, 155-162.
- *Zampetakis, L. A. (2010). Unfolding the measurement of the creative personality. *The Journal of Creative Behavior, 44*, 105-123. doi: 10.1002/j.2162-6057.2010.tb01328.x
- *Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? *Creativity Research Journal, 23*, 24-37. doi: 10.1080/10400419.2011.545713

Corresponding author at: Sameh Said-Metwaly, Faculty of Psychology and Educational Sciences, KU Leuven (University of Leuven), 8500 Kortrijk, Belgium.
E-mail: sameh.metwaly@kuleuven.be

