Testing conditions and creative performance : meta-analyses of the impact of time limits and instructions

Reference:
Said-Metw aly Sameh, Fernandez-Camilla Bélen, Kyndt Eva, van den Noortgate Wim.- Testing conditions and creative performance : meta-analyses of the impact of time limits and instructions
Psychology of aesthetics, creativity, and the arts - ISSN 1931-390X - (2019), p. 1-65
Full text (Publisher's DOI): https://doi.org/10.1037/ACA0000244

Testing Conditions and Creative Performance: Meta-Analyses of the impact of Time Limits and Instructions

Sameh Said-Metwaly [1,2], Belén Fernández-Castilla [1,2], Eva Kyndt [1,3], and Wim Van den Noortgate [1,2]

[1] KU Leuven (University of Leuven)

[2] imec-itec, KU Leuven (University of Leuven)

[3] University of Antwerp

Author Note

Sameh Said-Metwaly, Faculty of Psychology and Educational Sciences & imec-itec, KU Leuven (University of Leuven); Belén Fernández-Castilla, Faculty of Psychology and Educational Sciences & imec-itec, KU Leuven (University of Leuven); Eva Kyndt, Faculty of Psychology and Educational Sciences, KU Leuven (University of Leuven) & Faculty of Social Sciences, University of Antwerp; Wim Van den Noortgate, Faculty of Psychology and Educational Sciences & imec-itec, KU Leuven (University of Leuven).

This research was supported by a grant from the Ministry of Higher Education and Scientific Research in Egypt.

Correspondence concerning this article should be addressed to Sameh Said-Metwaly, Faculty of Psychology and Educational Sciences, KU Leuven (University of Leuven), 8500 Kortrijk, Belgium. E-mail: sameh.metwaly@kuleuven.be

Abstract

This study attempts to clarify the effects of varying testing conditions on creative performance in divergent thinking and product-based tasks. Two meta-analyses, one for time limits (short vs. long) and one for instructions (standard vs. explicit), were conducted. Moderator analyses were performed to examine whether the effects of time limits and instructions differ by measurement method, domain of creative performance, gender, study's country of origin, educational level, study quality, and scoring method of originality. It was also investigated whether the effects of time limits vary depending on whether long time condition is timed or untimed, and whether the effects of instructions vary under timed versus untimed conditions. A total of 57 effect sizes (12 studies) for time limits and 165 effect sizes (38 studies) for instructions were analyzed using a meta-analytic three-level model. The time limits meta-analysis showed that long time limits (vs. short time limits) significantly enhanced creative performance, with smaller effects in studies from the USA (vs. non-USA). Analyses on divergent thinking indicators showed that long time limits (vs. short time limits) yielded higher originality scores, with no significant differences for fluency or flexibility. Long time limits showed greater effects on fluency and flexibility in timed (vs. untimed) long conditions. The instructions meta-analysis revealed that explicit (vs. standard) instructions significantly enhanced creative performance. An examination of divergent thinking indicators showed that explicit (vs. standard) instructions resulted in higher originality scores, with no significant differences for fluency or flexibility. Explicit instructions demonstrated larger effects on originality in non-college (vs. college) subjects, as well as in untimed (vs. timed) conditions.

Keywords: creative performance, divergent thinking, creative product, testing conditions, time limits, instructions, meta-analysis

Testing Conditions and Creative Performance: Meta-Analyses of the impact of Time Limits and Instructions

The measurement of creativity has received great interest in recent years, mainly due to the increasing value attributed to creative individuals and their novel ideas in the modern technological society (Bart, Hokanson, Sahin, & Abdelsamea, 2015; Plucker & Makel, 2010; Rietzschel, Nijstad, & Stroebe, 2014). Multiple methods have been employed for measuring different aspects of the creativity construct, from personal to contextual factors (Belcher, Rubovits, & Di Meo, 1981; Park, Chun, & Lee, 2016; Said-Metwaly, Kyndt, & Van den Noortgate, 2017a).

Divergent thinking assessment has been commonly used to evaluate an important aspect of creative potential (Said-Metwaly, Kyndt, & Van den Noortgate, 2017a; Zeng, Proctor, & Salvendy, 2011). In his structure of intellect model, Guilford (1967, 1975) proposed that creativity mainly depends on divergent production, in which a broad search process is used to generate logical alternatives. Divergent thinking tests, such as the Wallach-Kogan Creativity Tests (Wallach & Kogan, 1965) and the Torrance Tests of Creative Thinking (Torrance, 1966), require individuals to generate as many responses as they can to open-ended tasks (e.g., alternative uses for common objects, consequences of hypothetical events, or picture construction). The resulting responses are scored with respect to different indicators, typically including: fluency (the total number of responses generated), flexibility (the number of the categories into which the responses fall), and originality (often operationalized by the number of uncommon or infrequent responses) (e.g., Said-Metwaly, Fernández-Castilla, Kyndt, & Van den Noortgate, 2018). Another method to measure creativity is to use product-based assessment. This method mainly focuses on tangible products or outcomes rather than inferred cognitive processes. Product-based assessment, such as Amabile’s (1982) consensual assessment technique (CAT), typically requires subjects to create an actual product (e.g., a story, a poem, a collage, or a musical composition), and the creativity of these products is then judged by experts in relevant areas (Hennessey, Amabile & Mueller, 2011).

One major issue in creativity research concerned the effect of testing conditions on the assessment of creativity (Busse, Blum, & Gutride, 1972; Plucker & Makel, 2010; Said-Metwaly, Kyndt, & Van den Noortgate, 2017b). Previous work in this area has primarily explored how various testing conditions influence performance on divergent thinking and product-based tasks (Said-Metwaly, Kyndt, & Van den Noortgate, 2017b). Runco and Albert (1985) stated that creative performance is easily affected by the conditions under which such performance is measured (e.g., test-like vs. game-like atmosphere, individual vs. group setting, varying instructions, and different scoring methods of originality). They also suggested that creative performance in one testing condition is not necessarily informative of creative performance in another condition (Runco & Albert, 1985). There has been a great deal of research devoted to identifying the optimal conditions for measuring creativity (Barron & Harrington, 1981; Ezzat, Agogué, Masson, & Weil, 2016; Hattie, 1980). This research has mostly focused on how to create ideal testing conditions to maximize creativity scores or to yield distinct patterns of correlations between creativity and other cognitive constructs (Barron & Harrington, 1981; Treffinger, Renzulli, & Feldhusen, 1971). Based on a review of the literature, Wallach and Kogan (1965) came to the conclusion that the inconsistency in reported results pertaining to the convergent validity of creativity measures and the relationship between creativity and intelligence might be attributable to the different testing conditions used in measuring creativity. Several subsequent studies showed that the psychometric properties of creativity measures (including reliability, construct, discriminant, and predictive validity) could be dependent on testing conditions (e.g., Benedek, Mühlmann, Jauk, & Neubauer, 2013; Chand & Runco, 1993; Forthmann, Lips, Szardenings, Scharfen, & Holling, 2018; Harrington, 1975; Nusbaum, Silvia, & Beaty, 2014).

One theoretical paradigm that has guided the research on external influences and creativity is Amabile's (1983, 1996) componential model of creativity. Besides expertise and skills inherent in individuals, Amabile went further to propose how situational factors could enhance or hamper creativity. Generally, it has been suggested that situational factors might affect individuals' intrinsic motivation to engage in a task or activity, and subsequently their creative performance (Hennessey & Amabile, 2010). According to Amabile (1983), the higher the individuals' sense of external constraints on a given task, the lower the level of their intrinsic motivation, and hence the less likely they are to think creatively. In this regard, Andrews and Smith (1996) pointed out that generating creative responses often depends on analytic processing that requires a great deal of cognitive effort. Thus, individuals should be motivated to make an effort to engage in such analytic processing (Andrews & Smith, 1996). However, externally imposed constraints might result in a loss of motivation and engagement in mere surface-level thinking, ending up with creativity being killed (Amabile, Hadley, & Kramer, 2002; Rosso, 2014). Additionally, as proposed by the controlled-attention theory of creativity, creative performance is affected by executive functions including working memory capacity, inhibition, and fluid intelligence (Benedek et al., 2014). These functions are suggested to be responsible for supplying top-down control over attention during task performance through actively maintaining a task goal in memory, inhibiting stereotypical but not original ideas, adopting effective cognitive search strategies, and judging and refining generated ideas (Beaty & Silvia, 2012; Benedek et al., 2014). On that account, it is not excluded that varying testing conditions could affect the capacity of these executive functions to direct an individual's attentional and motivational resources to task targets, which might, in turn, be associated with change in performance on creativity tasks.

A number of external testing-related factors that potentially affect creative performance could be identified in the literature. Among these, time limits and test instructions have received considerable attention in creativity research (Johns & Morse, 1997; Morse, Morse, & Johns, 2001; Plucker & Makel, 2010; Sajjadi-Bafhgi, 1986). However, there is an ongoing debate regarding the effects of these two factors on creative performance (Baer, 1994; Hong, O'Neil, & Peng, 2016; Morse et al., 2001).

Setting time limits on creativity tasks has been the subject of many studies examining its impact on measured creative performance (Morse et al., 2001; Sajjadi-Bafhgi, 1986; Treffinger et al., 1971). Many creativity tasks are administered under a specific time limit (Morse et al., 2001). However, experience of time pressure might arouse stress that could force individuals to pull their attention away from the task at hand and instead monitor progress over time, leading to exhaustion of mental resources involved in that task (Roskes, Elliot, Nijstad, & De Dreu, 2013). Moreover, Amabile et al. (2002) stated that creativity essentially relies on a combinatorial process in which multiple associations between concepts are shaped and then evaluated in an individual's mind, thus trying to find the most interesting and useful associations. Yet, a successful combinatorial process depends on having enough time for exploring concepts and playing with ideas (Amabile et al., 2002). Consequently, the presence of time constraints could severely undermine creativity through inhibiting exploration and reinforcing the maintenance of status quo approaches of thinking (Rosso, 2014). In this regard, Wallach and Kogan (1965) recommended allowing participants as much time as they desire on creativity tasks. On the other hand, Hattie (1977) pointed out that standardizing testing conditions would be difficult to attain without imposing time limits. Furthermore, Guilford (1971) indicated that when creativity tasks are given under liberal time, participants might invent strategies that would overly affect their performances. This, in turn, might lead to a change in what tasks claim to be measuring (Christensen & Guilford, 1963; Guilford, 1971). The empirical work that investigated the effect of time limits on creative performance yielded a rather equivocal picture (Baer, 1994; Lemons, 2011; Said-Metwaly, Kyndt, & Van den Noortgate, 2017b). Some researchers reported that creativity tasks produced higher scores when they were given under longer amounts of time (e.g., Johns & Morse, 1997; Khatena, 1971; Preckel, Wermer, & Spinath, 2011; Roskes et al., 2013), other researchers reported contradictory results (e.g., Johns, Morse, & Morse, 2000; Khatena, 1972; Madjar & Oldham, 2002; Morse et al., 2001; Sajjadi-Bafhgi, 1986; Sajjadi-Bafhgi & Khatena, 1985).

Task instructions given to participants are also likely to affect their creative performance (Chen et al., 2005; Ezzat et al., 2016; Hong et al., 2016; Hung, Chen, & Chen, 2012; Lemons, 2011). Variations in instructions wording might influence test takers' perception of the task, choice of processing strategies, and consequently performance (Chand & Runco, 1993; Di Mascio, Kalyuga, & Sweller, 2016; Runco, Illies, & Eisenman, 2005). Two types of instructions have dominated creativity research over the last several decades: standard instructions and explicit instructions to "be creative". Researchers have made use of goal setting theories to interpret the effect of these instructions on creative performance (Niu & Liu, 2009). It has been theorized that goals regulate individuals' effort through shifting their attention toward noteworthy aspects of the task and effective strategies that may conduce to the achievement of the assigned goals (Madjar & Shalley, 2008). As a result, when individuals are given creativity oriented instructions, they are more likely to maintain a "be creative" goal in mind, which would direct them to filter ideas generated in favor of creative ones (Beaty & Silvia, 2012; Chen et al., 2005; Chua & Iyengar, 2008; Katz & Poag, 1979; Paulus, Kohn, & Arditti, 2011). Conversely, without explicit creativity instructions, individuals might direct their attention and effort toward idiosyncratic goals and care less about the level of creativity of the ideas generated, as long as they are relevant (Chua & Iyengar, 2008; Nusbaum et al., 2014). Consistent with this view, Ward (1994) hypothesized that when engaging in a creativity task, individuals tend to adopt the path of least resistance and move toward easily obtainable ideas, and then quit when it becomes challenging to generate ideas. Unfortunately, the easily obtainable ideas are most likely to be the least creative ones (Rietzschel et al., 2014; Ward, Patterson, Sifonis, Dodds, & Saunders, 2002). In this manner, motivating individuals to brush the path of least resistance aside and put much effort to attain predetermined creativity goals might increase the possibility of creative ideas (Rietzschel et al., 2014; Ward, 1994). Results of previous studies that looked at the effect of test instructions on rated creativity are inconsistent (Hong et al., 2016; Morse et al., 2001; O'Hara & Sternberg, 2000-2001; Said-Metwaly, Kyndt, & Van den Noortgate, 2017b). Some studies revealed that better performance on creativity tasks was achieved under explicit instructions than under standard instructions (e.g., Christensen, Guilford, & Wilson, 1957; Evans & Forbach, 1983; Gerlach, Schutz, Baker, & Mazer, 1964; Katz & Poag, 1979; Madjar & Shalley, 2008; Niu & Sternberg, 2001). Other studies indicated that the effect of explicit instructions was limited to some aspects of creative performance such as the quality of responses (e.g., Bartis, Szymanski, & Harkins, 1988; Gilchrist & Taft, 1972; Harrington, 1975; Hong et al., 2016; Runco & Okuda, 1991) or was not significant overall (e.g., Chua & Iyengar, 2008; Datta, 1963; Johns & Morse, 1997; Niu & Liu, 2009; Ward, Saunders, & Dodds, 1999).

The Present Study

Accurate measurement of creativity should be based on maximum performance of individuals (Harrington, 1975). However, this performance is not only a function of the tasks used to measure creativity, but also of testing conditions (Trentham, 1979; Van Mondfrans, Feldhusen, Treffinger, & Ferris, 1971). As indicated earlier, there is conflicting evidence regarding the influence of testing conditions on creative performance. In a recent systematic review of methodological issues relative to the measurement of creativity, Said-Metwaly, Kyndt, and Van den Noortgate (2017b) concluded that there is a clear indication that different testing conditions yield varied performance on creativity tasks, yet what remains unsettled is the most probable outcome under each condition and the causes of these variations. They also emphasized the need for further studies that take the characteristics of the individual and the task into account to arrive at a clear picture of the variations in creative performance due to testing conditions (Said-Metwaly, Kyndt, and Van den Noortgate, 2017b). Thus, the main aim of this study is to undertake a meta-analysis of studies that addressed the effects of testing conditions on creative performance. Specifically, this study aims to clarify the effects of varying time limits and instructions on performance on divergent thinking and product-based tasks and examine potential moderator variables of these effects. Following this aim, this study seeks to answer the following research questions: (1) Does the manipulation of time limits (short vs. long) affect creative performance measured using divergent thinking or product-based tasks? (2) Does the manipulation of instructions (standard vs. explicit) affect creative performance measured using divergent thinking or product-based tasks? (3) Are there moderator variables that explain the variability in the effect sizes? By answering these questions, this study helps to resolve inconsistencies found across previous studies concerning the influence of testing conditions on creative performance. Moreover, this meta-analysis takes possible effects into consideration and identifies factors accounting for variations between individual studies. As such, this meta-analysis allows to study the generalizability of the conclusions, in contrast to individual studies investigating particular tests or populations.

Method

Study Variables

Dependent variable. The dependent variable of interest is creative performance, measured in terms of either divergent thinking indicators (e.g., fluency, flexibility, or originality) or product-based assessment (e.g., CAT). Creative performance was defined in this study as a person’s ability to solve problems, or shape ideas, perceptions, procedures, or products in a novel and task-appropriate way (Gardner, 1993; Oldham & Cummings, 1996; Vernon, 1989).

Independent variables. Time limits and instructions are the independent variables of this study. Generally, the *time limits* variable has not been clearly defined in past literature; instead it has been operationally specified as a definite period of time (Sajjadi-Bafhgi & Khatena, 1985). Considering that different tasks were used to measure creativity in previous studies and some tasks were administered under untimed conditions, expressing the time limits variable in definite units or intervals of time (e.g., minutes) seemed to make little sense. Hence, we coded this variable into two categories, short and long. In the short time condition, study participants were given a short time limit to complete the assigned task; in the long time condition, participants from the same study were allowed either longer or no time limit to complete the same task. On the contrary, the *instructions* variable has been operationalized in different ways in previous research. Accordingly, we defined explicit instructions as those emphasizing the production of only responses other than the common ones (e.g., creative, original, unusual, ingenious, clever, good, unexpected, or different) through using the two-word "be creative" or the synonyms of the word "creative" (e.g., original, novel, inventive, flexible, or imaginative) (Chen et al., 2005; Di Mascio et al., 2016; Evans & Forbach, 1983; O'Hara & Sternberg, 2000-2001; Runco & Okuda, 1991). Conversely, standard instructions were defined as those that do not address creativity or qualitative criteria explicitly (Chen et al., 2005; Evans & Forbach, 1983; Harrington, 1975).

Moderator variables. Drawing upon previous literature, the potential moderating role of the following variables in the effects of time limits and instructions were studied: measurement method of creative performance, domain of creative performance, gender, study's country of origin, educational level, and study quality, in addition to long time setting (for time limits), time limits (for instructions), and scoring method of originality (for originality indicator). Each of these variables is discussed below.

Measurement method of creative performance. Creativity measures are grounded on different conceptions and might account in part for the inconsistencies in creativity research (Barron & Harrington, 1981; Batey, 2012; Simonton, 2012). So far, it is unclear whether and how the measurement method used moderates the effects of testing conditions on creative performance. We thus sought to shed light on this potential moderating effect. For this study, measurement method was classified into two types: divergent thinking and product-based measures. We also distinguished between different indicators of creative performance (e.g., fluency, flexibility, quality, appropriateness, or total) when reported in the studies.

Domain of creative performance. The conflicting findings on the effects of testing conditions might be due to the different domains of creativity tasks (Chen et al., 2005; Kogan & Morgan, 1969). Previous research indicated that the effects of testing conditions varied across domains of creative performance (Chen et al., 2005; Van Mondfrans et al., 1971). We therefore investigated if the effects of testing conditions could be domain-dependent. Domain of creative performance was included as a categorical variable representing the domain of the tasks used to measure creativity in each study. Two categories for domain were used: verbal creativity (generating solutions to verbal problems or writing something such as essays, stories, or poems) and figural creativity (generating solutions to figural problems or drawing or designing something such as lines or collages). The reason for this is that the vast majority of the tasks employed in the relevant studies fall into those two categories. Studies employing tasks that do not fall into these categories were not included in this moderator analysis.

Gender. Gender might be another moderator variable of the effects of testing conditions. Baer and Kaufman (2008) suggested that there might be differences in the ways males and females respond to various testing conditions due to differences in intrinsic and extrinsic motivation. Although limited, previous studies found gender differences in the effects of testing conditions (e.g., Evans & Forbach, 1983; Katz & Poag, 1979). Hence, we conducted a moderator analysis to examine whether the effect sizes might vary systematically across male and female subjects. The proportion of males in each sample was included as a continuous moderator variable.

Study's country of origin. The effects of testing conditions might also differ as a function of study's country of origin. Research findings on this issue are however limited and inconclusive. While Niu and Sternberg (2001) found that the effects of testing conditions varied by country, Chen and colleagues (2002, 2005) found no evidence of cross-country variations. Given this, we sought to investigate if country of origin moderates the effects of testing conditions on creative performance. Study's country of origin was included as a categorical variable reflecting the country where the study was conducted.

Educational level. Conflicting results from previous studies might also be due to the educational level of the participants in each study. Research regarding this moderator is largely lacking. However, previous research has reported that creativity tends to show an irregular trajectory across age, as there are one or more periods along this trajectory wherein significant drops in creativity ratings take place (Barbot, Lubart, & Besançon, 2016; Gralewski, Lebuda, Gajda, Jankowska, & Wiśniewska, 2016). We thus conducted a moderator analysis of educational level to test whether the effect sizes might vary systematically across educational levels. Educational level was included as a categorical variable indicating the educational level of each sample. When participants' educational level is not explicitly indicated in a study, a decision was made based on participants' age.

Study quality. The information obtained from quality assessment has great significance when evaluating the results and conclusions of studies (Greco, Zangrillo, Biondi-Zoccai, & Landoni, 2013). Moher et al. (1998) indicated that this information needs to be taken into account to avoid or minimize bias. The quality of the studies included in this meta-analysis was assessed using EPHPP, a quality assessment tool for quantitative studies, developed by Thomas, Ciliska, Dobbins, and Micucci (2004). The EPHPP covers the following components of quality: selection bias, study design, confounders, blinding, data collection, and withdrawals. The EPHPP rates the studies from strong quality (if four or more components are rated as strong with no weak ratings), moderate quality (if there are less than four strong ratings and one weak rating), to weak quality (if there are more than one weak rating). This tool was selected for this study because it: (a) encompasses a wide range of quality assessment components, (b) has a detailed dictionary to clarify the criteria for rating each of these components, (c) has been proven to be valid and reliable (Thomas et al., 2004), and (d) has been widely used in previous studies (e.g., Armijo-Olivo, Stiles, Hagen, Biondo, & Cummings, 2010; Farina, Rusted, & Tabet, 2014).

Long time setting. Previous studies on the effects of time limits compared either short time with long time or timed with untimed conditions. It is possible that the presence of time as a factor in testing situation arouses stress and negatively affects creative performance. On the other hand, Van Mondfrans et al. (1971) claimed that time constraints might be necessary to maintain focused attention during task performance. Therefore, we examined if the effects of time limits differ depending on whether long time condition is timed or untimed.

Time limits. Time limits and instructions might potentially interact to affect creative performance (Christensen et al., 1957; Johns, Morse, & Morse, 2001). Thus, we examined whether the effects of instructions differ as a function of time limits. For studies on the effects of instructions, time limit was included as a categorical moderator variable and was binary-coded as timed or untimed.

Scoring method of originality. There are many different methods for scoring original responses in divergent thinking tasks (Benedek et al., 2013; Forthmann, Holling, Çelik, Storme, & Lubart, 2017; Plucker, Qian, & Wang, 2011; Silvia, Martin, &, Nusbaum, 2009). This scoring variation might play a moderating role in the effects of testing conditions. Scoring method was included as a categorical variable indicating the method used for scoring originality in divergent thinking tasks: uniqueness scoring (statistical infrequency), subjective scoring (subjective ratings of each response), snapshot scoring (holistic ratings), top scoring (ratings of a predefined number of top-responses), or ratio scoring (originality divided by fluency).

Selection of Studies

The papers included in this study were identified by searching creativity literature published up to May 31st, 2017. The search process consisted of the following steps: first, we searched the following databases: ERIC, JSTOR, PsycARTICLES, and Web of Science. On searching, we used the string (“creativity” OR “creative performance” OR “creative thinking” OR “creative ability” OR “creativity test” OR “divergent thinking” OR “creative product” OR “creative production”) AND (“time limits” OR “time press” OR “time pressure” OR “time interval”) to identify time limits studies, and the string (“creativity” OR “creative performance” OR “creative thinking” OR “creative ability” OR “creativity test” OR “divergent thinking” OR “creative product” OR “creative production”) AND (“instructions”) to identify instructions studies. Second, we reviewed the reference lists of the papers identified in the first step for additional papers (i.e., “backward search”). Third, we retrieved more recent references 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 manually searched the tables of content of the following key journals of creativity: *Creativity Research Journal*; *Psychology of Aesthetics, Creativity, and the Arts*; *The Journal of Creative Behavior*; and *Thinking Skills and Creativity*. Finally, we contacted authors of relevant publications for additional published or unpublished papers.

The papers identified using the search process were first screened for their relevance based on their titles and abstracts. In the case that the title and abstract of a paper were insufficient to make a trustworthy decision, the full text was reviewed to permit further evaluation. The retained papers were read in detail, and were included in our meta-analysis if they met the following criteria: the paper (1) reports on an original, empirical, and quantitative study (reviews and qualitative studies were excluded), (2) addresses the effect of time limits (short vs. long) and/or instructions (standard vs. explicit) on creative performance in divergent thinking or product-based tasks, (3) identifies the method used to measure creativity, and (4) reports the statistics needed to calculate the effect size (e.g., descriptive statistics like means and standard deviations; test statistics like *t* or *F*; or effect size measures like Cohen's *d*). Moreover, we only included (5) journal articles, conference papers, or dissertations that (6) were written in English (due to practical reasons including the need for language skills to allow proper interpretation of the research reported), and for which (7) the full text was available. When a study reported comparisons between multiple time limits (e.g., Khatena, 1971, 1972, 1973), only data for the longest and shortest ones were included as discrepancies might be seen most clearly between these two conditions. Also, in the case of studies that compared multiple types of instructions (e.g., Niu & Sternberg, 2003; Di Mascio et al., 2016), only data for standard and explicit instructions were selected. When a study compared originality and flexibility instructions with standard instructions (e.g., Runco & Okuda, 1991), originality instructions were considered as explicit ones.

Coding of Studies

Along with outcome statistics, method, sample characteristics, and information on the potential moderator variables were extracted from each of the eligible studies. In addition, a number of descriptive characteristics were reported for each study, including publication year, sample size, and sample age. The studies were coded for the moderator variables indicated earlier. Data were extracted by the first author. To assess inter-rater agreement, for both meta-analyses 20% of the studies (3 studies/16 effect sizes for time limits and 10 studies/34 effect sizes for instructions) were randomly selected and recoded independently by the second author. The percentage of agreement and Cohen's kappa were used to calculate the inter-rater agreement. The percentage of agreement ranged between 85-100% for time limits and between 82-100% for instructions. Cohen's kappa varied between .83-1.00 for time limits and between .75-1.00 for instructions, indicating a sufficient level of agreement (Landis & Koch, 1977). Disagreements were resolved through subsequent discussion.

Analyses

Two meta-analyses were conducted, one for the effects of time limits on creative performance and one for the effects of instructions. For studies comparing two groups (short time vs. long time or standard instructions vs. explicit instructions), we calculated Cohen's *d*, the mean difference between two groups divided by the pooled (within-groups) standard deviation (Borenstein, 2009). For studies with matched groups, formulas of Morris and DeShon (2002) were used to calculate a standardized mean difference that is comparable to Cohen's *d*. Calculating the sampling variance for the matched groups effect size requires a correlation value between creativity scores under different conditions (Morris & DeShon, 2002). However, in almost all cases, this correlation value was not reported. Therefore, we performed three meta-analyses for those data sets, using correlation values of .2, .5, and .8 respectively, in order to study to what extent the conclusions depend on the imputed correlation. It should be noted here that the correlation of .2 is the most conservative test in this case. If means and standard deviations were not provided, the effect size was computed from t or F values using Borenstein's (2009) formulas. Since Cohen's *d* tends to overestimate the population effect size, particularly for small samples, effect sizes were corrected for bias by converting *d* to Hedges' *g* (Hedges, 1981). A positive effect size reflects a higher creative performance in either long time or explicit instructions conditions.

The resulting corrected effect sizes and their sampling variances were then combined across studies using a random effects model. Overall effect sizes, weighted by the inverse of their variances, and corresponding confidence intervals were calculated. Given that the majority of studies in both meta-analyses reported more than one effect size, using traditional random effects models might yield flawed statistical inferences as a result of neglecting statistical dependence of these effects (Becker, 2000). Hence, we employed a meta-analytic three-level model to account for dependence within studies (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013, 2014). With this model, we differentiate three sources of variance: between-study variance (σ_V^2) (differences among studies in effect size estimates over outcomes), within-outcome variance (σ_U^2) (differences within studies in effect size estimates across multiple outcomes), and sampling variance (σ_E^2) (differences in effect size estimates due to sampling error) (Cheung, 2014; Van den Noortgate et al., 2013). By using a meta-analytic three level model, effect sizes within studies are not required to be aggregated, and accordingly we would gain insight into variations in effect sizes within studies besides variations among studies and could further test for the effects of within-study moderator variables (Rapp, Van den Noortgate, Broekaert, & Vanderplasschen, 2014). To further explore the degree of variation in effect sizes caused by heterogeneity, a likelihood ratio test comparing models with and without between-study variance and also models with and without within-study variance was used. Significant results of the likelihood ratio test indicate that effect sizes are heterogeneous, and thus moderators of these effect sizes are more likely to exist (Van den Bussche, Van den Noortgate, & Reynvoet, 2009). When conducting moderator analyses, the expected effect size for each category of a moderator variable was tested using a Wald test (versus a null hypothesis of no effect), and differences between categories were tested using Type III *F*-tests.

To investigate the impact of outliers, we ran a sensitivity analysis through excluding extreme effect sizes (2 *SD* above and below the mean) one by one and estimating the corresponding overall effect sizes. Finally, for exploring potential publication bias, we relied on the visual inspection of the symmetry of funnel plots (Light & Pillemer, 1984), followed by a three-level extension of Egger's regression intercept test (Egger, Smith, Schneider, & Minder, 1997) to account for dependency among effect sizes. The analyses were carried out using *metafor* package in *R* (Viechtbauer, 2010).

Results

Time Limits

Papers meeting the inclusion criteria. A total of 4,221 papers were initially resulted from the search process. From these, 20 potentially relevant papers were identified after the titles and abstracts were screened. Next, eight papers were excluded for the following reasons: six did not include sufficient information necessary for effect sizes calculation (Beaty & Silvia, 2012; Christensen et al., 1957; Johns et al., 2001; Sajjadi-Bafhgi, 1986; Sajjadi-Bafhgi & Khatena, 1985; Van Mondfrans et al., 1971), one focused on convergent thinking (Roskes et al., 2013), and one compared timed condition with untimed, take home condition (Torrance, 1969). Thus, the final number of time limits studies selected for this meta-analysis was twelve. These studies were published between 1971 and 2015. Ten (83.34%) of these were carried out in the USA, and one (8.33%) each in Canada and Germany. Sample sizes ranged from 20 to 261, with varying proportions of males (16% to 55.17%). Participants ranged in mean age from 14.15 to 46.32 years and were mainly college students (53.85%). Eleven (91.67%) studies used divergent thinking measures and one (8.33%) used product-based measures. Verbal tasks (73.33%) were used in most of these studies. Five (41.67%) studies were of moderate quality and seven (58.33%) were of weak quality. Nine (75%)

studies used untimed long conditions and three (25%) used timed long conditions. Among divergent thinking studies, five (45.45%) used uniqueness scoring, one (9.10%) used top scoring, and five (45.45%) did not report data for originality.

The twelve studies reported 57 effects; many individual studies provided more than one effect size from multiple measures or from independent subgroups. The values of the extracted effect sizes ranged from -0.91 to 3.74. Among these, 47 were positive favoring long time limit, three were equal to zero, and seven were negative favoring short time limit. Table 1 provides descriptive information and effect sizes for the twelve studies selected for the time limits meta-analysis. A forest plot of these studies is presented in Figure 1.

Overall analyses. Table 2 summarizes the results of the time limits meta-analysis. For the effect of time limits on creative performance, an initial three-level analysis incorporating all 57 effect sizes was conducted. This analysis revealed an overall mean effect size of 0.81 ($p = .01$) (creative performance is 0.81 *SD* higher under long time limits compared to short time limits), which is rather large according to the rules of thumb of Cohen (1988). The between-study variance estimate was rather large: about 69% of the total variance in observed effect sizes was systematic variance between studies, 24% systematic variance within studies, and 7% sampling variance. The likelihood ratio test revealed that both systematic variances were statistically significant ($\chi^2 = 30.2$ and 80.2, respectively, $df = 1$, $p < .0001$), suggesting that the size of the effect might be affected by moderator variables.

To gain a more nuanced understanding of the effects of time limits, further analyses were performed to independently examine the mean effect size for each indicator of creative performance. Analyses on indicators were conducted by pooling data across studies for each indicator. Given that only one of the included studies used product-based measures, subsequent analyses were restricted to divergent thinking indicators. Only three divergent thinking indicators (fluency, originality, and flexibility) were reported in many of the included studies. As can be seen in Table 2, all the three effect sizes were rather large, but only the mean effect size for originality was statistically significant (0.89, p = .01). The mean effect sizes for fluency (1.02, p = .16) and flexibility (0.74, p = .22) were not significant. Although the mean effect size estimates for the three indicators were comparable, only the mean effect size for originality reached statistical significance, because it was based on a larger number and more homogeneous effect sizes. This finding suggests that long time limits enhance originality scores, but there is no convincing evidence of an effect on fluency or flexibility scores. The likelihood ratio test showed significant between-study variance for fluency (χ^2 = 12, $df = 1$, p = .0005) and originality (χ^2 = 16.2, $df = 1$, p < .0001), but not for flexibility (χ^2 = 2.3, $df = 1$, p = .13). Within-study variance was found to be significant for fluency (χ^2 = 24.4, $df = 1$, p < .0001), and not significant for flexibility (χ^2 = 0, $df = 1$, p = 1) and originality (χ^2 = 3.7, df = 1, p = .054).

Sensitivity analyses. We found that the choice of the correlation coefficient between matched groups (.2, .5, or .8) did not affect the statistical significance of the resulting effect sizes, in addition the estimates of the effect sizes and their standard errors were still comparable. Therefore, we present the results when using a correlation of .5 for the corresponding data sets. Regarding the impact of outliers, by removing three potential outliers (all belonging to Preckel et al., 2011) one by one, the overall effect size ranged between 0.78 (95% CI [0.25, 1.32]; p = .008) and 0.80 (95% CI [0.24, 1.36]; p = .009). No outliers were identified for any of the indicators. Indeed, the three outliers identified in the overall analysis were located in fluency, however the large standard deviation of this indicator allowed the range of non-extreme values to spread far from the mean. The findings of the sensitivity analysis indicate that the estimates obtained were fairly robust.

Publication bias. Figure 2 shows the funnel plots for the overall and each indicator analysis in the time limits meta-analysis. A visual examination of the overall analysis plot shows that although the effect sizes are somewhat evenly dispersed around the overall mean effect size, there are three extreme effect sizes (all belonging to Preckel et al., 2011) on the right-hand side that do not have counterparts on the opposite side. The same three extreme effect sizes emerged in the funnel plot of fluency. This might cause some publication bias for the overall analysis and fluency. The funnel plots of flexibility and originality are somewhat symmetrical around the mean effect size. This was confirmed by the Egger's test that showed that publication bias was significant for the overall analysis (t = 3.02, df = 55, p = .004) and fluency (t = 4.78, df = 9.53, p = .0009), but not for flexibility (t = 3.89, df = 1, p = .16) and originality (t = 1.22, df = 20, p = .24). This suggests that the true mean effect size for the overall analysis and fluency might be overestimated. As indicated earlier, the asymmetry found for the overall analysis and fluency might have been caused by the three extreme effect sizes from Preckel et al.'s (2011) study. Therefore, it is possible that there is no publication bias, but rather that the effects in this study are larger than those in other studies.

Moderator analyses. The effect of each of the potential moderator variables was tested in an overall moderator analysis incorporating all effect sizes and in separate moderator analyses for each indicator separately. Measurement method of creative performance moderator was only tested for the overall analysis. For study's country of origin and educational level moderators, the studies were divided into only two categories because of the limited number of effect sizes for some categories. Regarding study's country of origin, most of the studies in our sample were conducted in the USA, with few studies conducted outside the USA. Accordingly, we divided the studies according to country into two categories: USA and non-USA. Similarly, for educational level, most of the studies were conducted on college students, and only a few targeted either elementary or high school students. Therefore, the studies were classified according to educational level into two categories: college (including undergraduate and postgraduate students) and non-college (including elementary and high school students). With regard to study quality, all the studies were rated as either moderate or weak; none was rated as strong. Additionally, there were insufficient data available to carry out the moderator analyses on domain of creative performance for both fluency and flexibility and also on country for flexibility. It should also be noted that there was a high association between measurement method and domain of creative performance moderators (see Supplementary Table S1 for the associations between the moderator variables in the time limits meta-analysis). Table 3 reports the results of the analyses of the moderator variables. The categories of each categorical variable are ordered by coding values. As Table 3 shows, only study's country of origin was significant for the overall analysis (p = .003). The obtained mean effect size was higher for the non-USA studies (2.40, p = .0002) than for the USA studies (0.55, p = .02). This moderator explained 47.33% of the between-study variance for the overall analysis. In addition, long time setting was found to be significant for fluency (p = .04) and originality (p = .03). Regarding fluency, studies using timed long periods had significantly higher mean effect size (2.25, p = .03) relative to those using untimed long periods (0.10, p = .83). Similarly, for flexibility, studies using timed long periods (1.53, p = .004) yielded significantly higher mean effect size than those using untimed long periods (.33, p = .25). Long time setting moderator was able to explain 100% of the between-study variance for fluency and flexibility, and 49.23% of the within-study variance for flexibility.

Instructions

Papers meeting the inclusion criteria. The search process resulted in 1,331 papers. Screening titles and abstracts of these papers yielded 55 papers as potentially relevant to our study. Out of these 55 papers, 17 papers were excluded. The reasons for the exclusions of these papers were that (a) one was not written in English (Lizarraga & Baquedano, 2008); (b) twelve did not report sufficient statistical information to calculate effect sizes (Amabile, 1979; Buyer, 1988; Chen, Himsel, Kasof, Greenberger, & Dmitrieva, 2006; Donnelly, 2013;

Johns et al., 2001; Madjar & Shalley, 2008; Maltzman, Bogartz, & Breger, 1958; O'Hara & Sternberg, 2000-2001; Rosen, Kim, Mirman, & Kounios, 2017; Shalley, 1991; Trentham, 1979; Van Mondfrans et al., 1971); (c) one compared explicit instructions with instructions emphasizing the production of practical solutions to problems (Goncalo & Staw, 2006) and one compared standard instructions with strategy plus explicit instructions (Forthmann, Wilken, Doebler, & Holling, 2016); and (d) one addressed preferences associated with creativity (Oziel, Oziel, & Cohen, 1972) and one addressed creative idea selection (Rietzschel, Nijstad, & Stroebe, 2010), instead of creative performance or production. At last, the overall number of instructions studies retained for analysis was 38. These studies were published between 1957 and 2016. Twenty seven (71.05%) studies were conducted in the USA, three (7.90%) in China, three (7.90%) in both the USA and China, two (5.26%) in Australia, and one (2.63%) each in Canada, Germany, and Netherlands. The number of participants ranged from 28 to 303, with varying proportions of males (16% to 100%). Participants ranged in mean age from 13.83 to 25.4 years and were mainly college students (80%). Thirty (78.95%) studies used divergent thinking measures, seven (18.42%) used product-based measures, and one (2.63%) used both. Verbal tasks (75.56%) were employed in most of these studies. Regarding quality, 25 (65.79%) were weak, 11 (28.95%) were moderate, and two (5.26%) were strong. Twenty nine (76.32%) studies used timed conditions and nine (23.68%) used untimed conditions. Among divergent thinking studies, 14 (46.67%) used subjective scoring, seven (23.33%) used uniqueness scoring, two (6.67%) used ratio scoring, one (3.33%) used snapshot scoring, and six (20%) did not report data for originality.

The 38 studies reported 165 effect sizes. The values of the extracted effect sizes ranged from -8.68 to 9.69. Among these, 114 were positive favoring explicit instructions, five were equal to zero, and 46 were negative favoring standard instructions. Table 4 provides descriptive information and effect sizes for the studies included in the instructions meta-analysis. A forest plot of these studies is shown in Figure 3.

Overall analyses. A summary of the results of the instructions meta-analysis is provided in Table 2. For the effect of instructions on creative performance, an initial analysis incorporating all 165 effect sizes was conducted. This analysis yielded a rather small effect size of 0.26 (p = .03) in favor of explicit instructions. About 98% of the total variance in observed effect sizes was systematic variance within studies, 2% sampling variance, with no systematic variance between studies. The likelihood ratio test showed that systematic variance within studies was significant (χ^2 = 1352.5, df = 1, p < .0001), whereas systematic variance between studies was not significant (χ^2 = 0, df = 1, p = 1).

Analyses on the indicators of creative performance were also conducted. For divergent thinking measures, fluency, originality, and flexibility indicators were reported in many of the included studies, while an aggregate creativity rating was mostly reported in the case of product-based measures. Thus, subsequent analyses were restricted to divergent thinking indicators. The resulting mean effect sizes for divergent thinking indicators followed the same trend that was observed for those of time limits. The mean effect size was significant for originality (0.83, p = .007), but not for fluency (0.02, p = .94) or flexibility (-0.85, p = .37). The likelihood ratio test revealed that both between-study variance and within-study variance were significant for fluency (χ^2 = 16.8 and 61.7, respectively, df = 1, p < .0001), flexibility (χ^2 = 16.7, df = 1, p < .0001; χ^2 = 8.4, df = 1, p = .004, respectively), and originality (χ^2 = 23.2 and 55.9, respectively, df = 1, p < .0001).

Sensitivity analyses. Neither the resulting effect sizes nor their standard errors significantly changed under the different correlation values, thus, a correlation of .5 was used for the presentation and discussion of the results. For the impact of outliers, after removing seven potential outliers (six from Runco (1986) and one from Runco and Okuda (1991)) one by one, the overall effect size ranged between 0.25 (95% [-0.02, 0.52]; p = .07) and 0.29 (95% [0.08, 0.50]; p = .008). For fluency, after removing two potential outliers (Runco, 1986), the mean effect size varied from 0.02 (95% [-0.48, 0.52]; p = .94) to 0.10 (95% [-0.30, 0.49]; p = .61). For flexibility, removing one potential outlier (Runco, 1986) yielded a mean effect size of -0.85 (95% [-2.95, 1.26]; p = .37). For originality, after removing three potential outliers (two from Runco (1986) and one from Runco and Okuda (1991)), the mean effect size varied from 0.70 (95% [0.18, 1.22]; p = .01) to 0.74 (95% [0.27, 1.20]; p = .004). On that account, the adjusted mean effect sizes remain comparable to the initial ones with all effect sizes included, suggesting that the estimates obtained were fairly robust and not significantly dependent upon any particular study.

Publication bias. Figure 4 shows the funnel plots for the overall analysis and each sub-analysis in the instructions meta-analysis. A visual examination of the overall analysis plot shows that effect sizes are rather symmetrically distributed around the overall mean effect size. Regarding the indicators, the funnel plot of fluency is somewhat symmetrical; however, the funnel plots of flexibility and originality appear asymmetric, suggesting that publication bias might be present. This was confirmed by the Egger's test that indicated that publication bias was not significant for the overall analysis (t = 0.33, df = 58.2, p = .74) and fluency (t = 1.64, df = 41, p = .11), but was significant for flexibility (t = -4.75, df = 15, p = .0003) and originality (t = 8.81, df = 49, p < .0001). This indicates that the true mean effect size for flexibility and originality might be biased. It is worth mentioning here that the three extreme effect sizes in the case of flexibility were obtained from Runco's (1986) study. Moreover, two of the extreme effect sizes of originality were related to the same study (Runco, 1986). Therefore, we cannot rule out the possibility that the asymmetric pattern for either flexibility or originality can be explained by a larger population effect in this study.

Moderator analyses. For domain of creative performance, there were insufficient data available to undertake the moderator analyses for divergent thinking indicators because figural tasks were only employed in product studies, not in divergent thinking studies. Thus, this moderator variable was only tested for the overall analysis. Furthermore, except for quality, we reduced the number of categories for each moderator to two categories due to the insufficient data for some categories. It should also be mentioned that there was a high association between measurement method and domain of creative performance moderators (see Supplementary Table S2 for the associations between the moderator variables in the instructions meta-analysis). Table 3 reports the results of the analyses of the moderator variables. Among the tested moderator variables, only educational level and time limits were significant for originality (p = .007 and .03, respectively). For educational level, the mean effect size obtained for non-college subjects (2.47, p = .0005) was significantly higher compared with that of college subjects (0.50, p = .07). For time limits moderator, the mean effect size was greater for untimed conditions (1.84, p = .002) than for timed conditions (0.49, p = .11). Educational level and time limits moderators explained 29.60% and 17.76% respectively of the between-study variance for originality.

Discussion

The aim of the current study was to clear up the somewhat mixed results of previous research regarding the effects of testing conditions on creative performance. Specifically, we sought to meta-analyze the impact of variant time limits (short vs. long) and instructions (standard vs. explicit) on creative performance in divergent thinking and product-based tasks and to explore possible sources of heterogeneity. For time limits, 12 studies were analyzed, resulting in a rather large overall effect size in favor of long time limits. This matches Amabile et al.'s (2002) premise that having enough time could enhance individuals' involvement in an extensive combinatorial process that could improve their creative performance. On the other hand, the detrimental effect of short time limits might be attributed to the increasing stress that could consume part of one's mental resources for the sake of monitoring progress over time, decreasing one's cognitive capacity to process the task at hand (Roskes et al., 2013).

Analyses on divergent thinking indicators showed that the effect of time limits was only significant for originality. In particular, we found that long time limits were associated with higher originality scores compared to short time limits. This finding is in line with the "serial order effect", that is more original responses tend to appear later in one's series of responses (Christensen et al., 1957; Parnes, 1961). It seems possible that subjects initiate their creative production with more common responses as they more easily and quickly come to mind (Parnes, 1961; Ward et al., 2002). As time passes, subjects might find themselves completely exhausted of common responses, and would then start searching for unique or original ones (Parnes, 1961). Therefore, one could expect that the longer the time given to subjects, the greater the possibility that they might care for original responses, leaving out other aspects of creative performance. This could be seen in line with Beaty and Silvia's (2012) study which showed that while originality rate went up with time, fluency rate reached its highest point in the first few minutes and rapidly dropped.

For instructions, 38 studies were analyzed, resulting in a rather small effect size in favor of explicit instructions. An examination of the subscales of creative performance showed that only the mean effect size for originality was significant. Specifically, giving explicit instructions was found to lead to increased originality, without significant effects on fluency or flexibility. The explicit instructions promoting effect on originality could be interpreted as being a result of a reorientation of one's achievement goal toward the given task. By having a creativity goal, an individual might be more likely to abandon easier or less clever paths and engage in an extended exploration strategy because the goal here is to reach a response that is not only relevant but also unique (Chua & Iyengar, 2008; Ward, 1994). In other words, explicit instructions direct an individual's attention from common responses toward original ones (Chen et al., 2005). On the contrary, without a creativity goal, individuals might become less concerned about the originality of the responses, as long as they work well for the task at hand (Chua & Iyengar, 2008). In this way, it can be contended that explicit instructions might not influence the quantity (i.e., fluency and flexibility) of the responses, but instead act upon the strategies adopted to generate these responses, leading to a remarkable improvement in their quality (i.e., originality).

It is worth mentioning here that the differential effects of testing conditions (time limits or instructions) on the indicators of creative performance might have resulted from the dissimilar executive functions and mechanisms underlying these indicators. Benedek, Franz, Heene, and Neubauer (2012) revealed that both fluency and flexibility were primarily driven by cognitive inhibition, while originality was primarily driven by intelligence. Besides, it was suggested that particular brain regions were more associated with originality than fluency or flexibility (Chávez-Eakle, Graff-Guerrero, García-Reyna, Vaugier, & Cruz-Fuentes, 2007; Fink, Graif, & Neubauer, 2009). As a result, they might act differently when affected by external factors such as testing conditions. This interpretation, however, still awaits investigation.

The results of the moderator analyses for time limits revealed that only study's country of origin was found to be significant for the overall analysis. A significantly lower mean effect size was found for studies from the USA than for studies from other countries. However, this finding may need to be viewed with caution because there were few studies conducted outside the USA. Moreover, long time setting was significant for fluency and originality indicators. The mean effect size for fluency and flexibility was greater for studies using timed long periods than those using untimed long periods. This finding is in accordance with the proposition by Van Mondfrans et al. (1971) that experience of time pressure might be necessary to keep subjects more focused and actively engaged in creativity tasks. It is also probably that the presence of time constraints arouses stress and spurs subjects to act more quickly. As a result, they are more likely to limit the act of refining generated responses and rely mainly on stereotypical or ordinary responses, which can lead to gains in fluency or flexibility. Further research could shed light on this matter.

For instructions, only educational level and time limits did significantly moderate the effects of instructions on originality. With regard to educational level, non-college subjects had a significantly larger mean effect size than college subjects. Probably, this finding is due to stresses and demands that characterize pre-college educational contexts against autonomy in college contexts. It is possible that pressure to conformity in pre-college contexts in addition to individuals' greater care at this age about social rules and their need for acceptance would force them to provide responses that exactly meet assigned norms or expectations. Drawing on this argument, when non-college subjects are instructed to be creative, they would be predominantly concerned with unique responses that meet the given task requirements. This might result in more original responses compared to college subjects. Regarding time limits moderator, untimed conditions were found to have a significantly larger mean effect size than timed conditions. Probably, timed conditions elicit anxiety that can place limits on cognitive exploration and increase reliance on simple processing strategies, thereby undermining the production of original responses (Amabile et al., 2002; Byron, Khazanchi, & Nazarian, 2010; Moore & Tenney, 2012). This finding provides further support to the "serial order effect" as individuals under untimed conditions are more likely to attain more original responses emerging later in the order of flow. Lastly, failure to obtain significance for the majority of the moderator analyses in this study might be due to the small number of studies at each of the contrasting categories, which weakens the statistical power of these analyses.

Implications

The findings of our study add to the growing body of literature on creativity measurement. First, these findings inform the ongoing debate regarding the optimal conditions for administering performance-based measures of creativity, with a particular emphasis on time limits and instructions. Second, the findings of our study suggest that both the time limits and instructions of creativity tasks matter in measuring creativity. This might have significant implications for researchers and educators interested in measuring creativity. In addition, these findings, in agreement with Runco and Albert's (1985) suggestion, highlight that the results obtained from creativity measures in a particular condition might not necessarily be indicative of the results of these measures under another condition. With this in mind, it could be argued that variations in testing conditions used by different researchers might account for inconsistent results in creativity literature. Researchers thus need to pay close attention to variations in testing conditions when comparing results from different studies. Finally, the meta-analytic three level model used in our study made it possible to estimate variance at both the study and outcome level, and to explore how this variance could be interpreted by conducting moderator analyses. On that account, this model enabled us to take into account a larger scope of variations across and within studies.

Limitations

The findings of our study should be read in light of some limitations. First, our search process for both meta-analyses was limited to studies written in English, which might have resulted in cultural bias. This was readily apparent in the results of the search process as most of the eligible studies were conducted in the USA. Second, both meta-analyses only included published journal articles, and the findings therefore might be affected by publication bias because statistically significant findings might be more likely to be published than non-significant findings (Greenhouse & Iyengar, 2009). Although we searched the gray literature (dissertations, conference papers, and reports), we failed to find unpublished studies that met the inclusion criteria. Third, as no high quality studies were identified in the time limits meta-analysis and only two in the instructions meta-analysis, one cannot fully discard the possibility that our findings might be biased in one direction or the other. Finally, our findings, especially for the time limits meta-analysis, are limited by the relatively small number of studies or by the incomplete reporting of method and sample characteristics, which either hampered testing the effects of all potential moderator variables or reduced the statistical power of some of these analyses. For instance, in the moderator analyses of educational level, we had to divide the studies into only two categories (college vs. non-college). However, previous studies (Barbot et al., 2016; Gralewski et al., 2016) have indicated that there are several periods in which creativity shows serious drops, particularly during elementary school. In our moderator analyses, subjects under college age were combined together into one broad category, which might not have been sufficiently sensitive to reveal the moderator effect of educational level. Similarly, as a result of the small number of studies, only two categories were involved to test the moderator effects of study's country of origin. Therefore, it might be risky to draw comprehensive conclusions regarding the effects of these moderator variables.

Future Research

Based on our study findings, some guidelines for future research can be proposed. First, many studies were excluded from our study or were not involved in particular moderator analyses due to not reporting sufficient statistics or relevant information. Therefore, we advise future researchers to report all required data on the study's method and results that could support future relevant meta-analyses. Second, considering the limited amount of available data for some of the tested moderator variables, and accordingly the weak statistical power, we might not have been able to find out the moderator effects that actually exist. Hence, there is a need for future studies to reinvestigate the effects of these variables with further data. Third, there might be other moderator variables that could account for the differences in the effects of testing conditions among studies. For instance, previous research (e.g., Khatena, 1972, 1973) found that the effect of time limits on creative performance relies upon the creativity level (i.e., high, moderate, and low creatives). Also, the effect of instructions on creative performance was found to be dependent upon the intellectual ability (Runco, 1986). However, we did not have sufficient data to test these variables. Therefore, it would be interesting for future research to consider other potential moderator variables that were not included in our study. Fourth, significant moderating effects observed for time limits or instructions in our study await confirmation in future studies. Fifth, only three divergent thinking indicators (i.e., fluency, flexibility, and originality) were mostly reported in previous studies on testing conditions. Thus, more research is required to investigate the effects of testing conditions with regard to other indictors of creative performance. Finally, most studies in this area have been conducted on college subjects, with a limited number of studies on children. Given the importance of identifying children for gifted programs, further studies on school age children are needed.

In conclusion, our findings suggest that time limits and instructions given to subjects could significantly affect their creative performance. On that account, these testing conditions should be taken into account when measuring creativity or drawing comparisons across the results of different studies. Furthermore, future efforts should attempt to clarify these effects by exploring further moderator variables to enhance our understanding about when and how these testing conditions could affect creative performance.

References

References marked with an asterisk (*) *were included in the time limits meta-analysis, while those marked with a caret* (^) *were included in the instructions meta-analysis*

Amabile, T. M. (1979). Effects of external evaluation on artistic creativity. *Journal of Personality and Social Psychology, 37*, 221-233. doi: 10.1037/0022-3514.37.2.221

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*, 997-1013. doi: 10.1037/0022-3514.43.5.997

Amabile, T. M. (1983). *The social psychology of creativity*. New York: Springer-Verlag.

Amabile, T. M. (1996). Evidence to support the componential model of creativity: Secondary analyses of three studies. *Creativity Research Journal, 9*, 385-389.

Amabile, T., Hadley, C., & Kramer, S. (2002). Creativity under the gun. *Harvard Business Review, 80*, 52-61.

Andrews, J., & Smith, D. C. (1996). In search of the marketing imagination: Factors affecting the creativity of marketing programs for mature products. *Journal of Marketing Research, 33*, 174-187. doi: 10.2307/3152145

Armijo-Olivo, S., Stiles, C. R., Hagen, N. A., Biondo, P. D., & Cummings, G. G. (2010). Assessment of study quality for systematic reviews: A comparison of the Cochrane collaboration risk of bias tool and the effective public health practice project quality assessment tool: Methodological research. *Journal of Evaluation in Clinical Practice, 18*, 12-18. doi: 10.1111/j.1365-2753.2010.01516.x

Baer, J. (1994). Performance assessments of creativity: Do they have long-term stability? *Roeper Review, 7*, 7-11.

Baer, J., & Kaufman, J. C. (2008). Gender differences in creativity. *The Journal of Creative Behavior, 42*, 75-105. doi:10.1002/j.2162-6057.2008.tb01289.x

Barbot, B., Lubart, T. I., & Besançon, M. (2016). Peaks, slumps and bumps: Individual differences in the development of creativity in children and adolescents. *New Directions for Child and Adolescent Development, 151*, 33-45. doi: 10.1002/cad.20152

*Barrett, J. D., Vessey, W. B., & Mumford, M. D. (2011). Getting leaders to think: Effects of training, threat, and pressure on performance. *The Leadership Quarterly, 22*, 729-750. doi: 10.1016/j.leaqua.2011.05.012

Barron, F., & Harrington, D. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology, 32*, 439-476. doi: 10.1146/annurev.ps.32.020181.002255

Bart, W. M., Hokanson, B., Sahin, I., & Abdelsamea, M. A. (2015). An investigation of the gender differences in creative thinking abilities among 8th and 11th grade students. *Thinking Skills and Creativity, 17*, 17-24. doi: 10.1016/j.tsc.2015.03.003

^Bartis, S., Szymanski, K., & Harkins, S. G. (1988). Evaluation and performance: A two-edged knife. *Personality and Social Psychology Bulletin, 14*, 242-251. doi: 10.1177/0146167288142003

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

Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts, 6*, 309-319. doi: 10.1037/a0029171

Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley, & E. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 499-525). Orlando, FL: Academic Press.

Belcher, T. L., Rubovits, J. J., & Di Meo, P. A. (1981). Measurement of creativity: A factor analytic study. *Psychological Reports, 48,* 819-825. doi: 10.2466/pr0.1981.48.3.819

Benedek, M., Beaty, R., Jauk, E., Koschutnig, K., Fink, A., Silvia, P. J., Dunst, B., & Neubauer, A. C. (2014). Creating metaphors: The neural basis of figurative language production. *Neuroimage, 90*, 99-106. doi: 10.1016/j.neuroimage.2013.12.046

Benedek, M., Franz, F., Heene, M., & Neubauer, A. C. (2012). Differential effects of cognitive inhibition and intelligence on creativity. *Personality and Individual Differences, 53,* 480-485. doi: 10.1016/j.paid.2012.04.014

Benedek, M., Mühlmann, C., Jauk, E., & Neubauer, A. C. (2013). Assessment of divergent thinking by means of the subjective top-scoring method: Effects of the number of top-ideas and time-on-task on reliability and validity. *Psychology of Aesthetics, Creativity, and the Arts, 7,* 341-349. doi: 10.1037/a0033644

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 279-293). New York: Russel Sage Foundation.

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. doi: 10.1207/s15327906mbr0703_2

Buyer, L.S. (1988). Creative problem solving: A comparison of performance under different instructions. *The Journal of Creative Behavior, 22*, 55-61. doi: 10.1002/j.2162-6057.1988.tb01341.x

Byron, K., Khazanchi, S., & Nazarian, D. (2010). The relationship between stressors and creativity: A meta-analysis examining competing theoretical models. *Journal of Applied Psychology, 95*, 201-212. doi: 10.1037/a0017868

^Carson, P. P., & Carson, K. D. (1993). Managing creativity enhancement through goal-setting and feedback. *The Journal of Creative Behavior, 27*, 36-45. doi: 10.1002/j.2162-6057.1993.tb01385.x

^Chand, I., & Runco, M. A. (1993). Problem finding skills as components in the creative process. *Personality and Individual Differences, 14*, 155-162. doi: 10.1016/0191-8869(93)90185-6

Chávez-Eakle, R. A., Graff-Guerrero, A., García-Reyna, J., Vaugier, V., & Cruz-Fuentes, C. (2007). Cerebral blood flow associated with creative performance: A comparative study. *Neuroimage, 38*, 519-528. doi:10.1016/j.neuroimage.2007.07.059

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. doi: 10.1002/j.2162-6057.2006.tb01272.x

^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. doi: 10.1002/j.2162-6057.2005.tb01252.x

^Chen, C., Kasof, J., Himsel, A., Greenberger, E., Dong, Q., & Xue, G. (2002). Creativity in drawings of geometric shapes: A cross-cultural examination with the consensual assessment technique. *Journal of Cross-Cultural Psychology, 33*, 171-187. doi: 10.1177/0022022102033002004

Cheung, M. W.L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19*, 211-229. doi: 10.1037/a0032968

Christensen, P. R., & Guilford, J. P. (1963). An experimental study of verbal fluency factors. *The British Journal of Statistical Psychology, 16,* 1-26. doi: 10.1111/j.2044-8317.1963.tb00195.x

^Christensen, P. R., Guilford, J. P., & Wilson, R. C. (1957). Relations of creative responses to working time and instructions. *Journal of Experimental Psychology, 53,* 82-88. doi: 10.1037/h0045461

^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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

*Cropley, A. J. (1972). Originality scores under timed and untimed conditions. *Australian Journal of Psychology, 24*, 31-36. doi: 10.1080/00049537208255782

^Datta, L. E. (1963). Test instructions and identification of creative scientific talent. *Psychological Reports*, *13*, 495-500. doi: 10.2466/pr0.1963.13.2.495

^Di Mascio, R., Kalyuga, S., & Sweller, J. (2016). The effect of wording and placement of task instructions on problem-solving creativity. *The Journal of Creative Behavior*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/jocb.157/epdf

Donnelly, A. R. (2013). *Measuring the effects of explicit instructions and incentives on the idea generation rate of a crowd-based population*. Unpublished master thesis, Massachusetts Institute of Technology, Cambridge.

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629-634. doi:10.1136/bmj.315.7109.629

^Evans, R. G., & Forbach, G. B. (1983). Facilitation of performance on a divergent measure of creativity: A closer look at instructions to "be creative". *Applied Psychological Measurement, 7*, 181-187. doi: 10.1177/014662168300700206

Ezzat, H., Agogué, M., Masson, P. L., & Weil, B. (2016). *Solution-oriented versus novelty-oriented leadership instructions: Cognitive effect on creative ideation*. Design Computing and Cognition, Chicago, United States.

Farina, N., Rusted, J., & Tabet, N. (2014). The effect of exercise interventions on cognitive outcome in Alzheimer's disease: A systematic review. *International Psychogeriatrics, 26*, 9-18. doi: 10.1017/S1041610213001385

Fink, A., Graif, B., & Neubauer, A. C. (2009). Brain correlates underlying creative thinking: EEG alpha activity in professional vs. novice dancers. *Neuroimage, 46*, 854-862. doi:10.1016/j.neuroimage.2009.02.036

*Foos, P. W., & Boone, D. (2008). Adult age differences in divergent thinking: It's just a matter of time. *Educational Gerontology, 34*, 587-594. doi: 10.1080/03601270801949393

^Forthmann, B., Gerwig, A., Holling, H., Çelik, P., Storme, M., & Lubart, T. (2016). The be-creative effect in divergent thinking: The interplay of instruction and object frequency. *Intelligence, 57*, 25-32. doi: 10.1016/j.intell.2016.03.005

Forthmann, B., Holling, H., Çelik, P., Storme, M., & Lubart, T. (2017). Typing speed as a confounding variable and the measurement of quality in divergent thinking. *Creativity Research Journal, 29*, 257-269. doi:10.1080/10400419.2017.1360059

Forthmann, B., Lips, C., Szardenings, C., Scharfen, J., & Holling, H. (2018). Are speedy brains needed when divergent thinking is speeded-or unspeeded? *The Journal of Creative Behavior*. doi: 10.1002/jocb.350

Forthmann, B., Wilken, A., Doebler, P., & Holling, H. (2016). Strategy induction enhances creativity in figural divergent thinking. *The Journal of Creative Behavior*. doi: 10.1002/jocb.159

Gardner, H. (1993). *Creating minds*. New York: Basic Books.

^Gerlach, V.S., Schutz, R. E., Baker, R. L., & Mazer, G. E. (1964). Effects of variations in test directions on originality test response. *Journal of Educational Psychology, 55*, 79-83. doi: 10.1037/h0048187

^Gilchrist, M. B., & Taft, R. (1972). Originality on demand. *Psychological Reports*, *31*, 579-582. doi: 10.2466/pr0.1972.31.2.579

Goncalo, J. A., & Staw, B. M. (2006). Individualism-collectivism and group creativity. *Organizational Behavior and Human Decision Processes, 100*, 96-109. doi: 10.1016/j.obhdp.2005.11.003

Gralewski, J., Lebuda, I., Gajda, A., Jankowska, D. M., & Wiśniewska, E. (2016). Slumps and jumps: Another look at developmental changes in creative abilities. *Creativity Theories–Research-Applications, 3*, 152-177. doi: 10.1515/ctra-2016-0011

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: Pitfalls and hints. *Heart Lung Vessel, 5*, 219-225.

Greenhouse, J. B., & Iyengar, S. (2009). Sensitivity analysis and diagnostics. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 417-433). New York: Russell Sage Foundation.

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

Guilford, J. P. (1971). Some misconceptions regarding measurement of creative talents. *The Journal of Creative Behavior, 5*, 77-87. doi: 10.1002/j.2162-6057.1971.tb00877.x

Guilford, J. P. (1975). Creativity: A quarter century of progress. In I. A. Taylor, & J. W. Getzels (Eds.), *Perspectives in creativity* (pp. 37-59). Chicago: Aldine.

^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. doi: 10.1111/j.1467-6494.1975.tb00715.x

^Harris, M. B., & Evans, R. C. (1974). The effects of modeling and instructions on creative responses. *Journal of Psychology, 86,* 123-130. doi: 10.1080/00223980.1974.9923894

*Hass, R. W. (2015). Feasibility of online divergent thinking assessment. *Computers in Human Behavior 46*, 85-93. doi: 10.1016/j.chb.2014.12.056

Hattie, J. A. (1977). Conditions for administering creativity tests. *Psychological Bulletin, 84,* 1249-1260. doi: 10.1037/0033-2909.84.6.1249

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. doi: 10.1037/0022-0663.72.1.87

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,*107-128. doi: 10.2307/1164588

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, 61, 569-598. doi: 10.1146/annurev.psych.093008.100416

Hennessey, B. A., Amabile, T. M., & Mueller, J. S. (2011). Consensual Assessment. In M. A. Runco, & S. R. Pritzker (Eds.) *Encyclopedia of Creativity*, vol. 1 (2nd ed., pp. 253-260). San Diego: Academic Press.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539-1558. doi: 10.1002/sim.1186

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

^Hong, E., O'Neil, H. F., & Peng, Y. (2016). Effects of explicit instructions, metacognition, and motivation on creative performance. *Creativity Research Journal, 28,* 33-45. doi: 10.1080/10400419.2016.1125252

*^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. doi: 10.1080/02783190009554026

Johns, G. A., Morse, L. W., & Morse, D. T. (2001). An analysis of early vs. later responses on a divergent production task across three time press conditions. *The Journal of Creative Behavior, 35,* 65-72. doi: 10.1002/j.2162-6057.2001.tb01222.x

^Katz, A. N., & Poag, J. R. (1979). Sex differences in instructions to "be creative" on divergent and nondivergent test scores. *Journal of Personality, 47,* 518-530. doi: 10.1111/j.1467-6494.1979.tb00630.x

*Khatena, J. (1971). Adolescents and the meeting of time deadlines in the production of original verbal images. *Gifted Child Quarterly, 15,* 201-204. doi: 10.1177/001698627101500307

*Khatena, J. (1972). Original verbal images of children as a function of time. *Psychological Reports, 31*, 565-566. doi: 10.2466/pr0.1972.31.2.565

*Khatena, J. (1973). Production of original verbal images by college adults to variable time intervals. *Perceptual and Motor Skills, 36,* 1285-1286. doi: 10.2466/pms.1973.36.3c.1285

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.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometric*, *33*, 159-174. doi: 10.2307/2529310

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

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

^Litchfield, R. C., Fan, J., & Brown, V. R. (2011). Directing idea generation using brainstorming with specific novelty goals. *Motivation and Emotion, 35*, 135-143. doi: 10.1007/s11031-011-9203-3

Lizarraga, M. L. S., & Baquedano, M. T. S. (2008). Instrucciones explícitas para la ejecución creativa según dos tests de creatividad teniendo en cuenta la inteligencia. *Anales de Psicología*, *24*, 1695-2294.

Madjar, N., & Oldham, G. R. (2002). Preliminary tasks and creative performance on a subsequent task: Effects of time on preliminary tasks and amount of information about the subsequent task. *Creativity Research Journal*, *14*, 239-251. doi: 10.1207/S15326934CRJ1402_10

Madjar, N., & Shalley, C. E. (2008). Multiple tasks' and multiple goals' effect on creativity: Forced incubation or just a distraction? *Journal of Management, 34*, 786-805. doi: 10.1177/0149206308318611

Maltzman, I., Bogartz, W., & Breger, L. (1958). A procedure for increasing word association originality and its transfer effects. *Journal of Experimental Psychology, 56*, 392-398. doi: 10.1037/h0045940

^Manske, M. E., & Davis, G. A. (1968). Effects of simple instructional biases upon performance in the unusual uses test. *Journal of General Psychology, 79*, 25-33. doi: 10.1080/00221309.1968.9710449

^Meadow, A., Parnes, S. J., & Reese, H. (1959). Influence of brainstorming instructions and problem sequence on a creative problem solving test. *Journal of Applied Psychology, 43*, 413-416. doi: 10.1037/h0043917

Moher, D., Pham, B., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., Tugwell, P., & Klassen, T. P. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analysis? *Lancet, 352*, 609-13.

Moore, D. A., & Tenney, E. R. (2012). Time pressure, performance, and productivity. *Research on Managing Groups and Teams, 15*, 305-326. doi: 10.1108/S1534-0856(2012)0000015015

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105-125. doi: 10.1037/1082-989X.7.1.105

*^Morse, D. T., Morse, L. W., & Johns, J. A. (2001). Do time press, stimulus, and creative prompt influence the divergent production of undergraduate students? Yes, yes, and no, not very much. *The Journal of Creative Behavior, 35*, 102-114. doi: 10.1002/j.2162-6057.2001.tb01224.x

^Niu, W., & Liu, D. (2009). Enhancing creativity: A comparison between effects of an indicative instruction “to be creative” and a more elaborate heuristic instruction on

Chinese student creativity. *Psychology of Aesthetics, Creativity, and the Arts, 3*, 93-98. doi: 10.1037/a0013660

^Niu, W., & Sternberg, R. J. (2001). Cultural influences on artistic creativity and its evaluation. *International Journal of Psychology, 36*, 225-241. doi: 10.1080/00207590143000036

^Niu, W., & Sternberg, R. J. (2003). Societal and school influences on student creativity: The case of China. *Psychology in the Schools, 40*, 103-114. doi: 10.1002/pits.10072

^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. doi: 10.1207/S15326934CRJ1302_7

Oldham, G. R., & Cummings, A. (1996). Employee creativity, personal and contextual factors at work. *Academy of Management Journal, 39*, 607-635. doi: 10.2307/256657

Oziel, L. J., Oziel, L. D., & Cohen, R. H. (1972). Effects of instructional set on production of creative responses. *Psychological Reports, 31*, 93-94. doi : 10.2466/pr0.1972.31.1.93

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

Parnes, S. J. (1961). Effects of extended effort in creative problem solving. *Journal of Educational Psychology, 52*, 117-122. doi: 10.1037/h0044650

^Parnes, S. J., & Meadow, A. (1959). Effects of "brainstorming" instructions on creative problem solving by trained and untrained subjects. *Journal of Educational Psychology, 50*, 171-176. doi: 10.1037/h0047223

^Paulus, P. B., Kohn, N. W., & Arditti, L. E. (2011). Effects of quantity and quality instructions on brainstorming. *The Journal of Creative Behavior*, *45*, 38-46. doi: 10.1002/j.2162-6057.2011.tb01083.x

Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman, & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 48-73). New York: Cambridge University Press.

Plucker, J., Qian, M., & Wang, S. (2011). Is originality in the eye of the beholder? Comparison of scoring techniques in the assessment of divergent thinking. *The Journal of Creative Behavior, 45*, 1-22. doi: 10.1002/j.2162-6057.2011.tb01081.x

*Preckel, F., Wermer, C., & Spinath, F. M. (2011). The interrelationship between speeded and unspeeded divergent thinking and reasoning, and the role of mental speed. *Intelligence*, *39*, 378-388. doi:10.1016/j.intell.2011.06.007

Rapp, R. C., Van den Noortgate, W., Broekaert, E., & Vanderplasschen, W. (2014). The efficacy of case management with persons who have substance abuse problems: A three-level meta-analysis of outcomes. *Journal of Consulting and Clinical Psychology, 82*, 605-618. doi: 10.1037/a0036750

^Ridley, D. R., & Birney, R. C. (1967). Effects of training procedures on creativity test scores. *Journal of Educational Psychology*, *58*, 158-164. doi: 10.1037/h0024655

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2010). The selection of creative ideas after individual idea generation: Choosing between creativity and impact. *British Journal of Psychology*, *101*, 47-68. doi:10.1348/000712609X414204

^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

Rosen, D. S., Kim, Y. E., Mirman, D., & Kounios, J. (2017). All you need to do is ask? The exhortation to be creative improves creative performance more for non-expert than expert jazz musicians. *Psychology of Aesthetics, Creativity, and the Arts, 11*, 420-427. doi: 10.1037/aca0000087

Roskes, M., Elliot, A. J., Nijstad, B. A., & De Dreu, C. K. W. (2013). Time pressure undermines performance more under avoidance than approach motivation. *Personality and Social Psychology Bulletin, 39*, 803-813. doi: 10.1177/0146167213482984

Rosso, B. D. (2014). Creativity and constraints: Exploring the role of constraints in the creative processes of research and development teams. *Organization Studies, 35*, 551-585. doi: 10.1177/017084061351760

^Runco, M. A. (1986). Maximal performance on divergent thinking tests by gifted, talented, and nongifted children. *Psychology in the Schools, 23*, 308-315. doi: 10.1002/1520-6807(198607)23:3<308::AID-PITS2310230313>3.0.CO;2-V

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. doi: 10.1177/001316448504500306

^Runco, M. A., Illies, J. J., & Eisenman, R. (2005). Creativity, originality, and appropriateness: What do explicit instructions tell us about their relationships? *The Journal of Creative Behavior, 39*, 137-148. doi: 10.1002/j.2162-6057.2005.tb01255.x

^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. doi: 10.1002/acp.2350050505

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., & Van den Noortgate, W. (2018). The factor structure of the Figural Torrance Tests of Creative Thinking: A meta-confirmatory factor analysis. *Creativity Research Journal, 30*, 352-360. doi: 10.1080/10400419.2018.1530534

Said-Metwaly, S., Kyndt, E., & Van den Noortgate, W. (2017a). Approaches to measuring creativity: A systematic literature review. *Creativity. Theories - Research - Applications, 4*, 238-275. doi: 10.1515/ctra-2017-0013

Said-Metwaly, S., Kyndt, E., & Van den Noortgate, W. (2017b). Methodological issues in measuring creativity: A systematic literature review. *Creativity. Theories - Research - Applications, 4*, 276-301. doi: 10.1515/ctra-2017-0014

Sajjadi-Bafhgi, S. H. (1986). Fixed and variable time press and sex differences in production of verbal originality. *Perceptual and Motor Skills, 63*, 1011-1016. doi: 10.2466/pms.1986.63.2.1011

Sajjadi-Bafhgi, S. H., & Khatena, J. (1985). Effects of autonomy of imagery and time press on production of verbal originality. *Perceptual and Motor Skills, 61*, 787-791. doi: 10.2466/pms.1985.61.3.787

Shalley, C. E. (1991). Effects of productivity goals, creativity goals, and personal discretion on individual creativity. *Journal of Applied Psychology, 76*, 179-185. doi: 10.1037/0021-9010.76.2.179

^Shalley, C. E. (1995). Effects of coaction, expected evaluation, and goal setting on creativity and productivity. *The Academy of Management Journal, 38*, 483-503. doi: 10.2307/256689

Silvia, P. J., Martin, C., &, Nusbaum, E. C. (2009). A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills and Creati*vity, *4*, 79-85. doi: 10.1016/j.tsc.2009.06.005

Simonton, D. K. (2012). Quantifying creativity: can measures span the spectrum? *Dialogues in Clinical Neuroscience, 14,* 100-104.

^Speller, K. G., & Schumacher, G. M. (1975). Age and set in creative test performance. *Psychological Reports, 36,* 447-450. doi: 10.2466/pr0.1975.36.2.447

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing, 1*, 176-184. doi: 10.1111/j.1524-475X.2004.04006.x

^Torrance, E. P. (1961). Priming creative thinking in the primary grades. *The Elementary School Journal, 62,* 34-41. doi: 10.1086/459930

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking: Norms-technical manual*. Lexington, MA: Personal Press.

Torrance, E. P. (1969). Curiosity of gifted children and performances on timed and untimed tests of creativity. *Gifted Child Quarterly, 13,* 155-158. doi: 10.1177/001698626901300301

Treffinger, D., Renzulli, J., & Feldhusen, J. (1971). Problems in the assessment of creative thinking. *The Journal of Creative Behavior, 5*, 104-112.

Trentham, L. L. (1979). Anxiety and instruction effects on sixth-grade students in a testing situation. *Psychology in the Schools, 16,* 439-443. doi: 10.1002/1520-6807(197907)16:3%3C439::AID-PITS2310160324%3E3.0.CO;2-2

*Tsenn, J., Atilola, O., McAdams, D. A., & Linsey, J. S. (2014). The effects of time and incubation on design concept generation. *Design Studies, 35*, 500-526. doi:10.1016/j.destud.2014.02.003

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin, 135,* 452-477. doi: 10.1037/a0015329

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three- level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*, 576-594. doi: 10.3758/s13428-012-0261-6

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2014). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods, 47,* 1274-1294. doi: 10.3758/s13428-014-0527-2

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. doi: 10.1002/1520-6807(197101)8:1<65::AID-PITS2310080116>3.0.CO;2-J

Vernon, P. E. (1989). The nature-nurture problem in creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 93-110). New York: McGraw-Hill.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1-48.

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

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology, 27,* 1-40. doi: 10.1006/cogp.1994.1010

Ward, T. B., Patterson, M. J., Sifonis, C. M., Dodds, R. A., & Saunders, K. N. (2002). The role of graded category structure in imaginative thought. *Memory and Cognition, 30,* 199-216. doi: 10.3758/BF03195281

^Ward, T. B., Saunders, K. N., & Dodds, R. A. (1999). Creative cognition in gifted adolescents. *Roeper Review, 21*, 260-266. doi: 10.1080/02783199909553973

^Ward, T. B., & Sifonis, C. M. (1997). Task demands and generative thinking: What changes and what remains the same? *The Journal of Creative Behavior, 31*, 245-259. doi: 10.1002/j.2162-6057.1997.tb00797.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: 10400419.2011.545713

Table 1

Studies Included in the Time Limits Meta-analysis

No.	Study	Year	Country	*n* (Male%)	Age range (*n*/*SD*)	Grade	Measurement method	Domain	Quality	Scoring method	Long time setting	Indicator	*g*	v_g
1	Khatena	1971	USA	142	/	10th-12th	Divergent thinking	Verbal	Moderate	Uniqueness	Untimed	Originality	1.59	0.05
												Originality	1.54	0.05
												Originality	1.29	0.04
2	Cropley	1972	Canada	115 (51.30%)	(17.25)	12th	Divergent thinking	Figural	Weak	Uniqueness	Untimed	Originality	1.55	0.02
3	Khatena	1972	USA	87 (55.17%)	8-11	Elementary school	Divergent thinking	Verbal	Moderate	Uniqueness	Untimed	Originality	0.50	0.08
												Originality	0.15	0.09
												Originality	0.38	0.08
												Originality	0.98	0.15
												Originality	0.00	0.07
												Originality	-0.03	0.09
4	Khatena	1973	USA	90 (31.11%)	/	Undergraduate	Divergent thinking	Verbal	Moderate	Uniqueness	Untimed	Originality	1.08	0.23
												Originality	0.73	0.07
												Originality	0.71	0.20
												Originality	1.29	0.10
												Originality	1.69	0.46
												Originality	1.00	0.08
5	Johns & Morse	1997	USA	76 (25.32%)	/	Undergraduate	Divergent thinking	Verbal	Moderate		Untimed	Fluency	0.51	0.05
												Flexibility	0.48	0.05
6	Johns et al.	2000	USA	Study 1: 22	12-17 (14.15/1.18)	Non-college	Divergent thinking	Verbal	Weak		Untimed	Fluency	0.27	0.05
				Study 2: 20	12-18 (14.35/1.60)	Non-college	Divergent thinking	Verbal	Weak		Untimed	Flexibility	0.16	0.06
7	Morse et al.	2001	USA	50 (16%)	18-38 (21/4)	Undergraduate	Divergent thinking	Verbal	Weak	Uniqueness	Timed	Fluency	1.80	0.11
												Flexibility	1.93	0.11
												Originality	1.07	0.09
												Fluency	1.39	0.10
												Flexibility	1.53	0.10
												Originality	0.98	0.09
												Fluency	1.29	0.09
												Flexibility	1.20	0.09
												Originality	0.31	0.08

Table 1 *(Continued)*

No.	Study	Year	Country	*n* (Male%)	Age range (*n*/*SD*)	Grade	Measurement method	Domain	Quality	Scoring method	Long time setting	Indicator	*G*	v_g
8	Foos & Boone	2008	USA	120 (37.50%)	(46.32)	College	Divergent thinking	Verbal	Weak		Untimed	Total	0.82	0.07
												Total	0.83	0.07
												Total	0.12	0.07
												Total	1.75	0.09
												Total	0.75	0.07
												Total	-0.65	0.07
												Total	0.16	0.07
												Total	-0.91	0.07
												Total	1.79	0.09
												Total	0.00	0.06
9	Barrett, Vessey, & Mumford	2011	USA	193 (38.86%)	(18.7)	Undergraduate	Divergent thinking	Verbal	Weak		Untimed	Total	-0.44	0.02
10	Preckel et al.	2011	Germany	261 (42.53%)	(14.84/.58)	9th Grade	Divergent thinking	Verbal, figural, and numerical	Moderate		Timed	Fluency	3.13	0.10
												Fluency	1.78	0.05
												Fluency	3.74	0.12
												Fluency	3.42	0.11
11	Tsenn, Atilola, McAdams, & Linsey	2014	USA	39	/	Undergraduate	Product	Figural	Weak		Timed	Quantity	1.07	0.12
												Quality	0.06	0.10
												Novelty	0.00	0.10
												Variety	0.62	0.11
												Quality	0.36	0.10
												Novelty	0.08	0.10
12	Hass	2015	USA	73 (46.29%)	18-31	Undergraduate	Divergent thinking	Verbal	Weak	Top	Untimed	Fluency	0.39	0.06
												Fluency	-0.56	0.06
												Fluency	-0.52	0.06
												Fluency	-0.38	0.06
												Originality	0.35	0.05
												Originality	0.03	0.05
												Originality	0.54	0.06

Note. *n* = number of participants; *g* = Hedges' *g* effect size; v_g = variance of the effect size estimate. A slash (/) indicates not reported information.

Table 2

A summary of the Results of Time Limits and Instructions Meta-analyses

Meta-analysis	Level	Effect size	SE	95% CI	*p*-value	σ_V^2	σ_U^2	σ_E^2
Time-limits	Overall	0.81	0.26	[0.24, 1.38]	.01	0.70	0.25	0.07
	Fluency	1.02	0.59	[-0.58, 2.63]	.16	1.57	0.30	0.02
	Flexibility	0.74	0.42	[-1.04, 2.52]	.22	0.47	0.02	0.03
	Originality	0.89	0.22	[0.32, 1.47]	.01	0.28	0.006	0.03
Instructions	Overall	0.26	0.11	[0.03, 0.49]	.03	0.00	1.87	0.04
	Fluency	0.02	0.24	[-0.48, 0.51]	.94	0.87	0.40	0.04
	Flexibility	-0.85	0.87	[-2.95, 1.26]	.37	4.86	0.57	0.07
	Originality	0.83	0.28	[0.25, 1.41]	.007	1.49	0.45	0.05

Note. SE = standard error of the effect size; CI = confidence interval; σ_V^2 = between-study variance; σ_U^2 = within-study variance; σ_E^2 = typical sampling variance (calculated using Higgins and Thompson's (2002) formula).

Table 3

Results of Moderator Analyses for Time Limits and Instructions Meta-analyses

Analysis	Moderator variable	Time limits						Instructions					
		k	*n*	*N*	*F*	*df*	*p*	*k*	*n*	*N*	*F*	*df*	*p*
Overall													
	Measurement method				0.27	1, 8.9	.61				0.27	1, 8.9	.61
	Divergent thinking	51	1189	27				139	3288	44			
	Product	6	39	1				26	1400	11			
	Domain				0.26	1, 9.3	.62				3.36	1, 124	.07
	Verbal	46	813	22				140	3814	45			
	Figural	7	154	2				18	1283	10			
	Gender	46	1005	22	0.01	1, 44	.93	107	3132	32	0.52	1, 32.3	.48
	Country				13.96	1, 11.4	.003				1.32	1, 121	.25
	USA	52	852	23				107	2811	41			
	Non-USA	5	376	5				47	1073	12			
	Educational level				2.75	1, 10.1	.13				1.59	1, 133	.21
	Non-college	16	647	16				38	1599	16			
	College	41	581	12				125	3061	38			
	Quality				2.48	1, 9.56	.15				2.32	2, 130	.10
	Weak	36	572	8				106	2553	36			
	Moderate	21	656	20				50	1985	16			
	Strong	0	0	0				9	170	3			
	Long time setting				3.08	1, 8.93	.11						
	Timed	19	350	6									
	Untimed	38	878	22									
	Time limits										1.36	1, 132	.25
	Timed							126	3218	43			
	Untimed							39	1470	12			
Fluency													
	Domain				-	-	-				-	-	-
	Verbal	9	221	4				45	2449	31			
	Figural	0	0	0				0	0	0			
	Gender	12	460	7	0.00	1, 2.02	.95	31	1369	17	2.55	1, 29	.12
	Country				7.79	1, 2.88	.07				1.50	1, 17.5	.24
	USA	9	221	4				30	1705	24			
	Non-USA	4	261	4				16	744	7			
	Educational level				0.9	1, 3.02	.41				2.83	1, 18.9	.11
	Non-college	5	283	5				12	1088	12			
	College	8	199	3				34	1361	19			
	Quality				1.37	1, 3.03	.33				0.91	2, 19.9	.42
	Weak	8	464	3				32	1524	20			
	Moderate	5	337	5				13	853	10			
	Strong	0	0	0				1	72	1			
	Long time setting				9.82	1, 3.15	.04						
	Timed	7	311	5									
	Untimed	6	171	3									
	Time limits										2.17	1, 18.8	.16
	Timed							34	1384	22			
	Untimed							12	1065	9			
Flexibility													
	Domain				-	-	-				-	-	-
	Verbal	5	146	3				16	1227	14			
	Figural	0	0	0				0	0	0			
	Gender	4	126	2	6.68	1, 1.76	.14	5	429	3	1.54	1, 2.09	.34
	Country				-	-	-				0.12	1, 5.38	.74
	USA	5	146	3				16	924	13			
	Non-USA	0	0	0				1	303	1			
	Educational level				0.88	1, 1.05	.51				0.98	1, 5.12	.37
	Non-college	1	20	1				11	1016	11			
	College	4	126	2				6	211	3			
	Quality				0.10	1, 1.02	.80				0.93	1, 5.06	.38
	Weak	4	70	2				11	586	9			
	Moderate	1	76	1				6	641	5			
	Strong	0	0	0				0	0	0			
	Long time setting				20.89	1, 2.47	.03						
	Timed	3	50	1									
	Untimed	2	96	2									
	Time limits										0.77	1, 4.09	.43
	Timed							10	330	8			
	Untimed							7	897	6			
Originality													
	Domain				2.01	1, 4.15	.23				-	-	-
	Verbal	21	442	17				50	2730	35			
	Figural	1	115	1				0	0	0			
	Gender	19	415	15	0.00	1, 17	.97	30	1483	17	0.03	1, 8.97	.87

Table 3 *(continued)*

Analysis	Moderator variable	Time limits						Instructions					
		k	*n*	*N*	*F*	*df*	*p*	*k*	*n*	*N*	*F*	*df*	*p*
Country					2.01	1, 4.15	.23				0.15	1, 18.2	.70
USA		21	442	17				33	1884	26			
Non-USA		1	115	1				18	846	9			
Educational level					0.77	1, 4.08	.43				9.11	1, 21	.007
Non-college		10	344	10				11	1016	11			
College		12	213	8				40	1714	24			
Quality					0.01	1, 4.04	.94				0.13	1, 19.2	.72
Weak		7	238	3				36	1710	24			
Moderate		15	319	15				15	1020	11			
Strong		0	0	0				0	0	0			
Long time setting					0.05	1, 4.19	.84						
Timed		3	50	1									
Untimed		19	507	17									
Time limits											5.14	1, 20.3	.03
Timed								40	1665	26			
Untimed								11	1065	9			
Scoring method					1.49	1, 4.03	.29				0.94	3, 17.7	.44
Uniqueness		19	484	17				19	1182	15			
Subjective		0	0	0				19	1228	16			
Snapshot		0	0	0				8	249	1			
Ratio		0	0	0				5	71	3			
Top scoring		3	73	1				0	0	0			

Note. *k* = number of effect sizes; *n* = number of participants; *N* = number of samples. A dash (-) indicates insufficient data to carry out the analysis.

Table 4
Studies Included in the Instructions Meta-analysis

No.	Study	Year	Country	*n* (Male%)	Age range (*M*/*SD*)	Grade	Measurement method	Domain	Quality	Scoring	Time limits	Indicator	*g*	v_g
1	Christensen et al.	1957	USA	52	/	College	Divergent thinking	Verbal	Weak	Subjective	Timed	Originality	1.36	0.09
												Originality	0.70	0.08
2	Meadow, Parnes, & Reese	1959	USA	32	/	College	Divergent thinking	Verbal	Weak	Subjective	Timed	Originality	-1.31	0.06
3	Parnes & Meadow	1959	USA	52	/	Undergraduate	Divergent thinking	Verbal	Moderate	Subjective	Timed	Originality	-0.27	0.08
												Originality	-1.08	0.09
4	Torrance	1961	USA	204	/	1st-3rd	Divergent thinking	Verbal	Weak	Uniqueness	Timed	Fluency	0.25	0.12
												Flexibility	-0.07	0.12
												Originality	0.21	0.12
												Fluency	0.03	0.11
												Flexibility	0.14	0.11
												Originality	0.18	0.11
												Fluency	0.83	0.11
												Flexibility	0.65	0.11
												Originality	0.96	0.12
												Fluency	0.28	0.10
												Flexibility	0.27	0.10
												Originality	0.38	0.10
												Fluency	0.73	0.15
												Flexibility	0.85	0.15
												Originality	0.53	0.14
												Fluency	-0.27	0.12
												Flexibility	0.05	0.12
												Originality	-0.37	0.13
5	Datta	1963	USA	31	/	Postgraduate	Divergent thinking	Verbal	Moderate	Subjective	Timed	Originality	0.47	0.13
6	Gerlach et al.	1964	USA	39	/	Postgraduate	Divergent thinking	Verbal	Weak	Subjective	Timed	Fluency	-0.47	0.10
												Originality	0.76	0.11
7	Ridley & Birney	1967	USA	159 (100%)	/	Undergraduate	Divergent thinking	Verbal	Moderate	Uniqueness	Timed	Originality	0.39	0.04
												Originality	0.45	0.04
												Total	0.00	0.04
												Total	0.19	0.04
8	Manske & Davis	1968	USA	30 (26.67%)	/	Undergraduate	Divergent thinking	Verbal	Weak	Subjective	Timed	Originality	1.02	0.07
												Practicality	-1.01	1.99
												Fluency	-1.01	2.97

Table 4 *(continued)*

No.	Study	Year	Country	*n* (Male%)	Age range (*M*/*SD*)	Grade	Measurement method	Domain	Quality	Scoring	Time limits	Indicator	*g*	v_g
9	Gilchrist & Taft	1972	Australia	60	/	College	Divergent thinking	Verbal	Weak	Uniqueness	Timed	Fluency	1.84	0.05
												Originality	0.51	0.02
10	Harris & Evans	1974	USA	67 (40.60)	19-54 (25.4)	College	Divergent thinking	Verbal	Weak		Timed	Total	-0.86	0.06
												Total	0.72	0.06
11	Harrington	1975	USA	105 (100%)	/	Undergraduate	Divergent thinking	Verbal	Moderate	Subjective	Timed	Originality	0.44	0.04
												Fluency	-0.29	0.04
												Uncreative uses	-0.87	0.04
12	Speller & Schumacher	1975	USA	72 (50%)	10-17	5th and 12th	Divergent thinking	Verbal	Strong		Timed	Fluency	0.95	0.06
13	Katz & Poag	1979	Canada	42 (47.62%)	/	College	Divergent thinking	Verbal	Moderate	Ratio	Timed	Fluency	1.58	0.25
												Fluency	-0.95	0.19
												Fluency	0.98	0.21
												Fluency	0.20	0.17
												Originality	0.88	0.20
												Originality	1.32	0.21
												Originality	0.78	0.20
												Originality	-0.10	0.17
14	Evans & Forbach	1983	USA	65 (46.21%)	/	College	Divergent thinking	Verbal	Moderate		Timed	Fluency	0.43	0.13
												Fluency	0.53	0.11
15	Runco	1986	USA	240	/	5th to 8th	Divergent thinking	Verbal	Moderate	Uniqueness	Untimed	Fluency	-6.47	0.24
												Fluency	-1.58	0.05
												Fluency	-3.03	0.07
												Flexibility	-8.68	0.42
												Flexibility	-4.15	0.20
												Flexibility	-5.41	0.18
												Originality	2.73	0.05
												Originality	5.54	0.34
												Originality	9.69	0.56
16	Bartis et al.	1988	USA	111	/	Undergraduate	Divergent thinking	Verbal	Weak	Subjective	Timed	Fluency	0.93	0.04
												Originality	0.77	0.04
17	Runco & Okuda	1991	USA	29 (65.52%)	15-17	Non-college	Divergent thinking	Verbal	Weak	Ratio	Untimed	Fluency	-0.82	0.05
												Flexibility	-0.70	0.05
												Originality	4.70	0.49
18	Carson & Carson	1993	USA	28	/	/	Divergent thinking	Verbal	Weak		Timed	Total	0.36	0.14
												Total	0.76	0.14

Table 4 *(continued)*

No.	Study	Year	Country	*n* (Male%)	Age range (*M*/*SD*)	Grade	Measurement method	Domain	Quality	Scoring	Time limits	Indicator	*g*	v_g
19	Chand & Runco	1993	USA	78 (37.18%)	/	College	Divergent thinking	Verbal	Weak	Uniqueness	Untimed	Fluency	0.15	0.05
												Originality	0.73	0.05
												Fluency	-0.34	0.05
												Originality	-0.12	0.05
												Fluency	0.21	0.05
20	Shalley	1995	USA	136 (48%)	(22)	Undergraduate	Divergent thinking	Verbal	Weak		Timed	Total	0.65	0.03
												Total	0.49	0.03
												Total	0.60	0.03
												Total	0.54	0.03
21	Johns & Morse	1997	USA	76 (26.32%)	/	Undergraduate	Divergent thinking	Verbal	Moderate		Timed	Fluency	0.33	0.05
												Flexibility	0.31	0.05
22	Ward & Sifonis	1997	USA	105	/	Undergraduate	Product	Figural	Weak		Untimed	Originality	0.57	0.06
23	Ward et al.	1999	USA	Study 1: 54	(13.83)	7th to 12th	Product	Figural	Weak		Timed	Originality	0.33	0.07
				Study 2: 100	/	College	Product	Figural	Weak		Timed	Originality	0.51	0.04
24	Morse et al.	2001	USA	50 (16%)	18-38 (21/4)	College	Divergent thinking	Verbal	Weak	Uniqueness	Timed	Fluency	-0.28	0.08
												Flexibility	0.00	0.08
												Originality	0.00	0.08
												Fluency	-0.49	0.08
												Flexibility	-0.42	0.08
												Originality	-0.48	0.08
												Fluency	-0.22	0.08
												Flexibility	0.00	0.08
												Originality	0.00	0.08
25	Niu & Sternberg	2001	USA and China	USA 76 (28.95%) China 63 (50.79%)	USA (18.5) China (20.2)	Undergraduate	Product	Figural	Moderate		Untimed	Total	0.30	0.01
26	Chen et al.	2002	USA and China	USA 50 (50%) China 48 (50%)	USA (23.5) China (21.8)	College	Product	Figural	Strong		Timed	Creativity	1.24	0.09
												Creativity	1.12	0.09
												Uniqueness	1.21	0.09
												Uniqueness	1.21	0.10
												Quality	1.15	0.09
												Quality	0.69	0.09
												Liking	1.19	0.09

Table 4 *(continued)*

No.	Study	Year	Country	*n* (Male%)	Age range (*M*/*SD*)	Grade	Measurement method	Domain	Quality	Scoring	Time limits	Indicator	*g*	v_g
												Liking	0.86	0.09
27	Niu & Sternberg	2003	China	62	/	High School	Product	Figural	Weak		Untimed	Total	2.81	0.13
28	Chen et al.	2005	USA and China	USA 248 (25%) China 278 (31%)	USA (22.05) China (21.69)	Undergraduate	Product	Verbal	Moderate		Timed	Total	0.18	0.01
								Verbal				Total	0.38	0.01
								Verbal				Total	0.32	0.01
								Verbal				Total	0.24	0.01
								Verbal				Total	0.22	0.01
								Figural				Total	1.27	0.01
								Figural				Total	0.42	0.01
								Figural				Total	0.18	0.01
								Mathematical				Total	0.99	0.01
								Mathematical				Total	0.65	0.01
29	Runco et al.	2005	USA	85	/	Undergraduate	Divergent thinking	Verbal and figural	Moderate	Uniqueness	Untimed	Fluency	0.30	0.05
								Verbal and figural				Flexibility	0.14	0.05
								Verbal and figural				Originality	0.20	0.05
								Verbal and figural				Appropriateness	0.39	0.05
								Verbal				Fluency	0.26	0.05
								Verbal				Flexibility	0.40	0.05
								Verbal				Originality	0.16	0.05
								Verbal				Appropriateness	0.18	0.05
30	Chua & Iyengar	2008	USA	Study 1: 100 (38%)	/	College	Product	Figural	Weak		Timed	Total	-0.24	0.08
												Total	0.38	0.08
				Study 2: 114 (48%)	(23)	College	Divergent thinking	Verbal	Weak		Timed	Total	-0.11	0.08
												Total	0.31	0.09
31	Niu & Liu	2009	China	117 (35.56%)	(16.2/0.39)	High School	Product	Verbal and Figural	Weak		Timed	Total	-0.87	0.04
32	Litchfield, Fan, & Brown	2011	USA	101	/	College	Divergent thinking	Verbal	Weak	Subjective	Timed	Fluency	0.05	0.04
												Originality	0.21	0.04
												Creativity	0.19	0.04
												Effectiveness	0.17	0.04
												Practicality	0.17	0.04

Table 4 *(continued)*

No.	Study	Year	Country	*n* (Male%)	Age range (*M*/*SD*)	Grade	Measurement method	Domain	Quality	Scoring	Time limits	Indicator	*g*	v_g
33	Paulus et al.	2011	USA	39 (26.92%)	(21)	Undergraduate	Divergent thinking	Verbal	Weak	Subjective	Timed	Fluency	0.85	0.11
												Originality	2.49	0.18
34	Nusbaum et al.	2014	USA	141 (28%)	(19.6/4.2)	College	Divergent thinking	Verbal	Weak	Subjective	Timed	Fluency	-0.24	0.03
												Fluency	-0.78	0.03
												Originality	0.90	0.03
												Originality	0.88	0.03
35	Rietzschel et al.	2014	Netherlands	102 (25.49%)	(21.1)	Undergraduate	Divergent thinking	Verbal	Weak	Subjective	Timed	Originality	0.27	0.08
												Originality	0.49	0.08
												Feasibility	-0.14	0.08
												Feasibility	-0.38	0.08
36	Di Mascio et al.	2016	Australia	Study1: 47 (40%)	/	Undergraduate	Divergent thinking	Verbal	Weak	Subjective	Untimed	Fluency	1.18	0.10
												Originality	0.97	0.09
												Workability	0.29	0.08
												Relevance	-0.42	0.08
				Study2: 43 (62%)	/	Undergraduate	Divergent thinking	Verbal	Weak	Subjective	Untimed	Fluency	0.86	0.10
												Originality	0.91	0.10
												Workability	-0.11	0.09
												Relevance	0.18	0.09
37	Forthmann, Gerwig, Holling, Çelik, Storme, & Lubart	2016	Germany	249 (20.88%)	18-60 (23.48)	Undergraduate	Divergent thinking	Verbal	Weak	Snapshot	Timed	Fluency	-0.05	0.02
												Fluency	0.15	0.02
												Fluency	-0.22	0.02
												Fluency	-0.42	0.02
												Fluency	-0.63	0.02
												Fluency	-0.52	0.02
												Fluency	-0.34	0.02
												Fluency	-0.52	0.02
												Originality	1.24	0.02
												Originality	0.42	0.02
												Originality	0.92	0.02
												Originality	0.93	0.02
												Originality	1.28	0.02
												Originality	0.68	0.02
												Originality	0.78	0.02
												Originality	0.48	0.02

Table 4 *(continued)*

No.	Study	Year	Country	*n* (Male%)	Age range (*M*/*SD*)	Grade	Measurement method	Domain	Quality	Scoring	Time limits	Indicator	*g*	ν_g
38	Hong et al.	2016	China	303 (48.18%)	/	High School	Divergent thinking	Verbal	Weak	Subjective	Untimed	Fluency	0.10	0.01
												Flexibility	-0.02	0.01
												Originality	0.41	0.01

Note. n = number of participants; *g* = Hedges' *g* effect size; ν_g = variance of the effect size estimate. A slash (/) indicates not reported information.

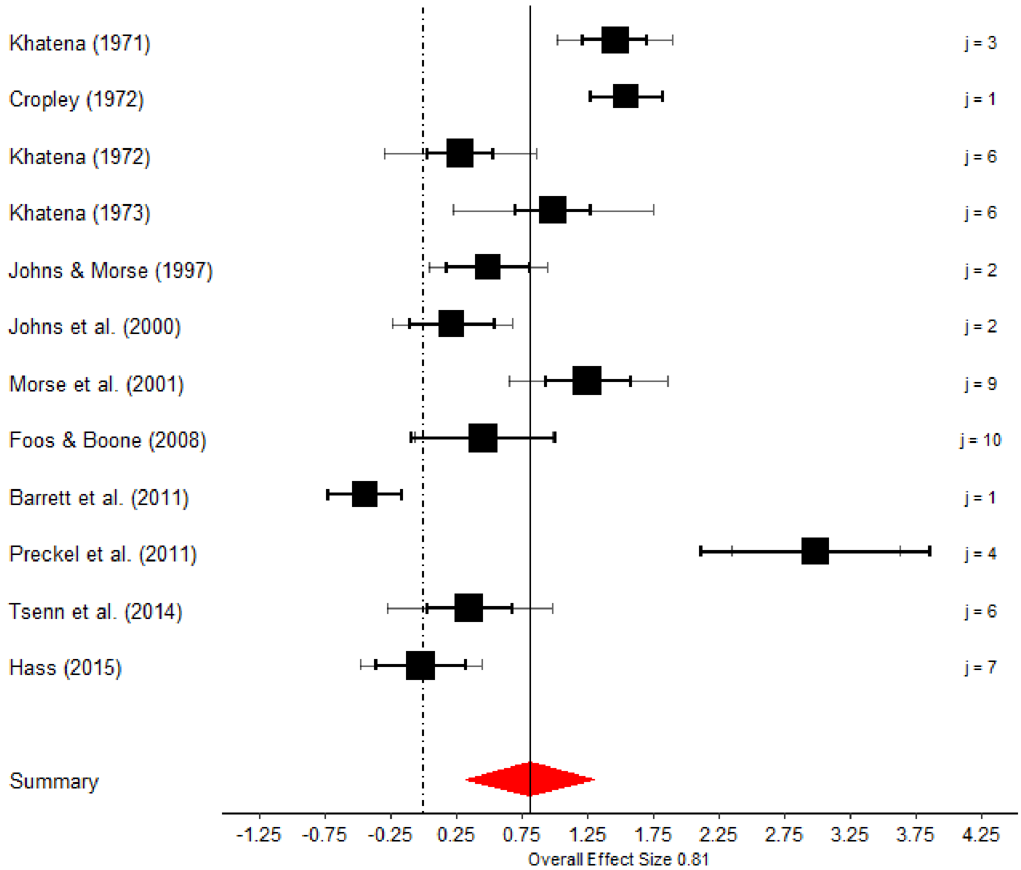

Figure 1. Forest plot of the time limits meta-analysis. j = number of effect sizes for each study. The bold confidence interval represents the precision of a study, which depends on the sampling variance as well as on the number and variability of effect sizes within that study. The non-bold confidence interval represents the precision of an individual effect size within the study, which depends on the sampling variance only.

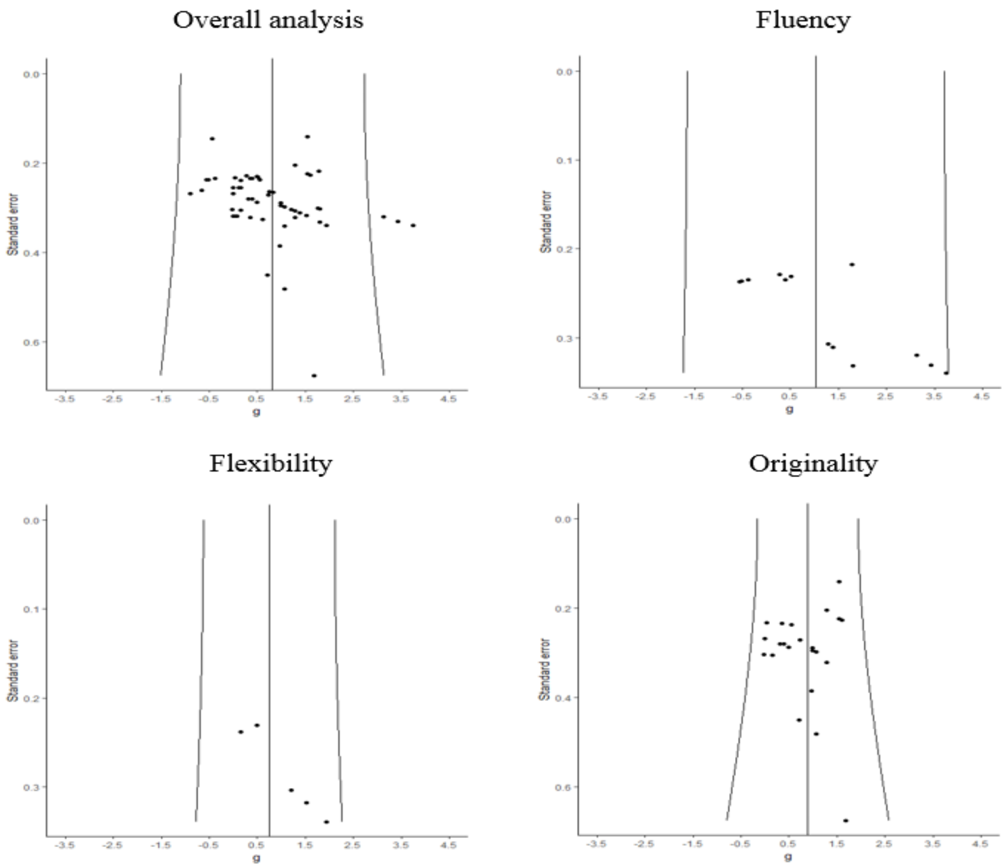

Figure 2. Funnel plots of the time limits meta-analysis

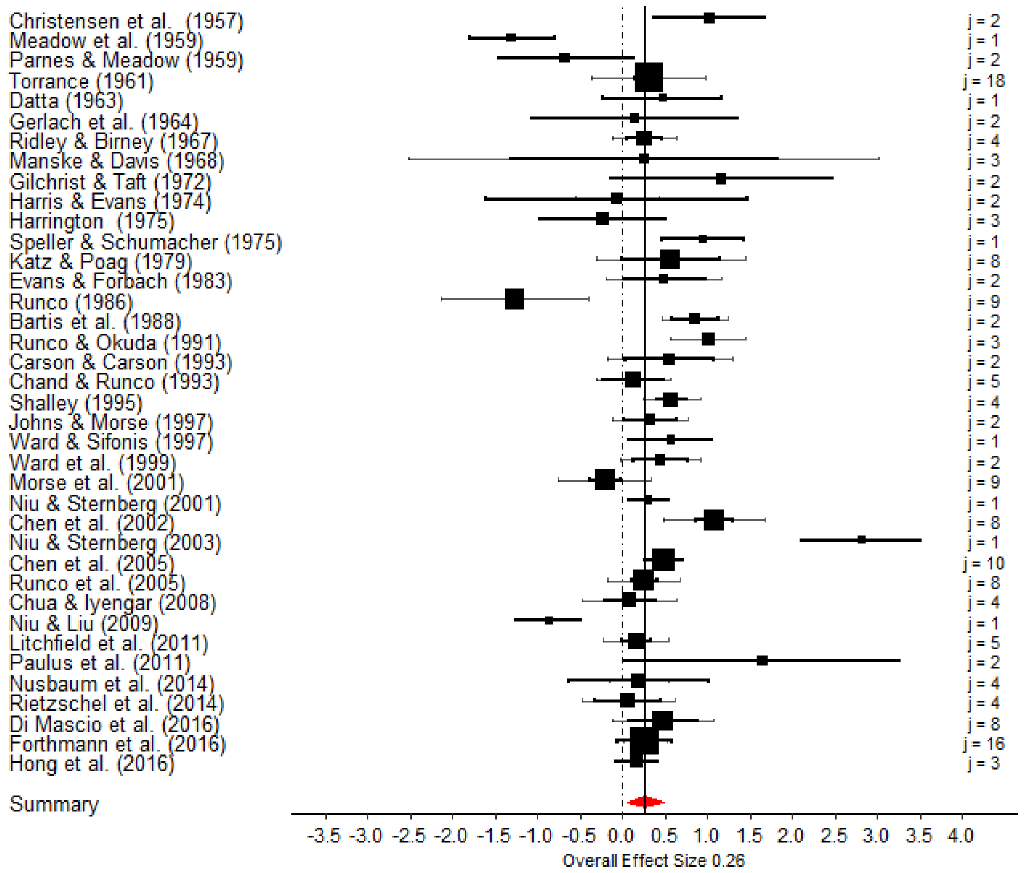

Figure 3. Forest plot of the instructions meta-analysis. j = number of effect sizes for each study. The bold confidence interval represents the precision of a study, which depends on the sampling variance as well as on the number and variability of effect sizes within that study. The non-bold confidence interval represents the precision of an individual effect size within the study, which depends on the sampling variance only.

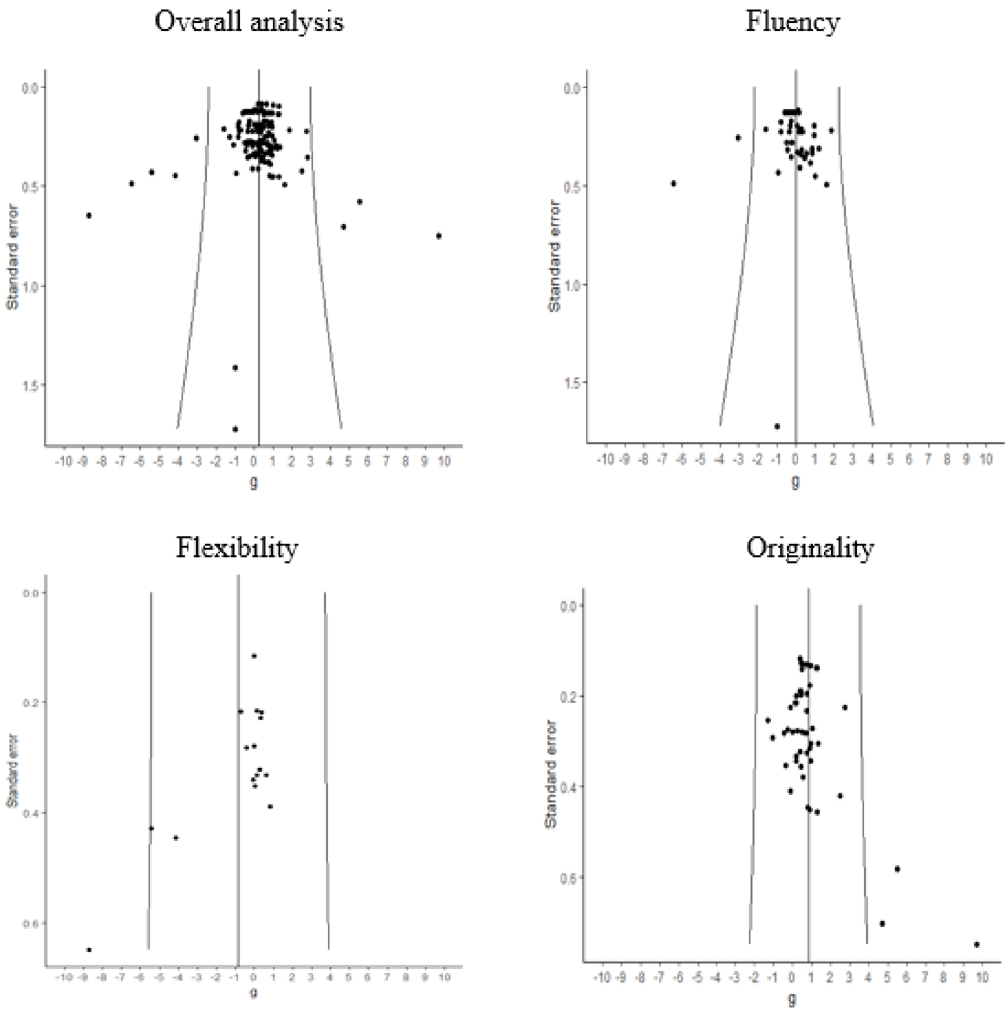

Figure 4. Funnel plots of the instructions meta-analysis