ScienceDirect

Thinking Skills and Creativity
Available online 3 July 2021, 100894
In Press, Journal Pre-proof

Is More Time Better for Divergent Thinking? A Meta-Analysis of the Time-On-Task Effect on Divergent Thinking

Sue Hyeon Paek [a], Ahmed Abdulla [b], Selcuk Acar [c], Mark A. Runco [d]

Show more

Outline | Share | Cite

https://doi.org/10.1016/j.tsc.2021.100894 Get rights and content

Highlightts

- The current study aimed to examine how time affects assessing maximal performance in divergent thinking tests. The results indicate the inverted J-shaped pattern in divergent thinking performance over time. The findings will contribute to accurately assess creative potential and provide implications in manipulating time conditions to maximize creative thinking in various teaching and learning contexts.

Abstract

Performance on divergent thinking (DT) tests varies by different testing conditions. Although many studies documented that DT performance increases when more time is provided, it still remains unanswered whether DT performance continues to increase linearly or follows a rather non-linear pattern as more time is allowed. The present study examined a potential curvilinear pattern in DT performance by synthesizing 237 effect size in 22 studies using a three-level approach to account for the nested structure of data. Results indicated that liberal time conditions provided significantly better DT performance than the restricted time conditions with the mean effect size (g=0.666) being considered moderate to large. The effect size was small to moderate (g = 0.493) in a group of studies that merely compared "timed" and "untimed" conditions. The effect was larger (g = 0.901) in a group of studies comparing varying amounts of time. The quadratic term of time difference between the shorter and longer time conditions turned out significantly negative indicating the *inverted J-shaped* relationship between time-on-task and DT performance: the DT performance increases with more time, and the growth slows down at some point. Implications for testing DT are explored.

Keywords

Time-on-task; time limit; divergent thinking test; meta-analysis; non-linear relation

---

1. Introduction

Divergent thinking (DT) is a core process of creative thinking and is widely assessed as a proxy of creative potential (Runco & Acar, 2012). While creative potential has been defined in many different ways, DT is defined straightforwardly as the capacity of thinking in different directions (Taylor, 1988). Because DT is scored based on the responses to open-ended tasks, testing conditions may impact the outcomes from those tasks. Indeed, performance on creativity tasks varies by different testing conditions, as was indicated in a systematic review on measurement issues in creativity assessment (Said-Metwaly et al., 2017). Furthermore, testing conditions contribute to or reduce construct-irrelevant variance—which is defined as variance attributed to factors other than DT performance (i.e., the construct of interest)—and influence the construct validity of the scores. Therefore, understanding how testing conditions impact the estimates of DT is important.

Outcomes from DT tasks are influenced by a number of factors. Some factors are internal such as test structure, content, and modality. DT outcomes are also influenced by external testing conditions such as explicitness of the test instructions (Acar et al., 2020; Said-Metwaly et al., 2020), playfulness of the testing environment (Wallach & Kogan, 1965), and the amount of time given when taking the tests (Hattie, 1977, 1980; Said-Metwaly et al., 2020). Among these external factors, time-on-task (TOT) plays a particularly critical role in DT performance not only because TOT changes the duration of the task but also TOT may have examinees' ability, understanding about the tasks, attention, motivation, and mental/physical exhaustion (Briggs & Reining, 2020) bounded along with its impact on cognitive functioning and emotional state (Moore & Tenny, 2012). Thus, it is not surprising studies show contradictory results of that DT performance varies by time.

Many studies have documented the more time provided, the larger the number of total and original responses generated (e.g., Akinboye, 1982; Benedek et al., 2013; Johns & Morse, 1997; Morse et al., 2001; Plucker et al., 2006; Preckel et al., 2011). Interestingly, not all studies support this positive TOT effect. Rather, both positive and negative TOT effects are simultaneously evidenced in several empirical studies (e.g., Foos & Boone, 2008; Hass, 2015; Hattie, 1980; Khatena, 1972; Sajjadi-Bafghi, 1986).

Despite a considerable deal of variation in findings across studies, Said-Metwaly et al. (2020) synthesized 12 studies using a meta-analytic three-level model into a large TOT effect and concluded that creative performance is significantly enhanced when more time is allowed. More specifically, originality benefits more from longer time conditions than fluency and flexibility. The positive TOT effect in originality supports the serial order effect (Mednick, 1962) that people tend to produce more original responses later as the tasks advance because they probably need to consume mundane responses to come up with unique associations across different thoughts. Said-Metwaly et al. (2020) explained the positive TOT effect using Amabile's (1983; Amabile et al., 2002) componential theory of creativity and Benedek et al.'s (2014) controlled-attention theory of creativity. Throughout several works, Amabile and colleagues discussed the negative effect of time pressure on creative thinking (Amabile, 1996; Amabile et al., 2002). In fact, Amabile et al. (2002) asserted that time pressure could lead individuals to be less creative. As Amabile et al. (2002, pp. 52) put it, "when creativity is under the gun, it usually ends up getting killed." They concluded "complex cognitive processing takes time, and, without some reasonable time for that processing, creativity is almost impossible (Amabile et al., 2012, pp. 61)." Also, people can dedicate their mental resources to processing the task without having to deal with the stress of managing time (Benedek et al., 2014). Despite the interesting finding of Said-Metwaly et al.'s (2020) meta-analysis, the exact functional form (i.e., linear vs. non-linear) of how DT performance increases as a function of time remains unanswered. Interestingly, their meta-analysis found that people tend to show better performance with more time, if tests are timed compared to untimed. In other words, DT performance benefits from some form of a time constraint as far as people have enough time to process during DT tests.

Perhaps, people make more cognitive efforts on the tasks with a time constraint (Moore & Tenny, 2012). A time constraint naturally augments the pace in the tasks because the overall window of time that people can work on the tasks is fixed (McGrath & Kelly, 1986). Besides, people may set their achievement goals higher because a time constraint makes the tasks more difficult, and the higher achievement goals promotes their performance (Bryan & Locke, 1967). The challenges added by a time constraint may serve to cognitively stimulate people on the tasks which in turn motivates them to be more creative. Andrew and Farris (1972) found that scientists who experienced optimal time pressure outperformed the counterpart scientists experiencing less time pressure in their creative and innovative outcomes. More recently, this positive relationship was supported in Ohly et al.'s (2006) study in which 278 employees in a high-tech company showed more creative and innovative ideas at work when they were tested up to a certain point of time pressure. This finding was repeatedly supported even when daily creativity was examined in an experience sampling method which rigorously illustrated day-to-day dynamics (Ohly & Fritz, 2010). People tend to perceive the task as more stimulating and this may promote cognitive activation (i.e., neural activity), which is central in processing information (Gardner, 1986). The cognitive activation would consequently elevate task performance, and creative performance is not an exception (Baer & Oldham, 2006).

Furthermore, excessive time could impede or no longer augment performance after a certain point. This inflected curve makes most sense according to Briggs and Reining's (2010) bounded ideational theory suggesting that an ideation curve is likely to slow representing an ogive. The ogive is likely because the relationship between TOT and DT performance is moderated by several boundaries including examinees' mental ability and attentions that are limited by nature, mental/physical exhaustion, and understanding of the tasks that may change over time. As a result of prolonged engagement in certain tasks, for instance, people may run out of their cognitive and motivational energy and focus. With limited mental resources, people may feel more fatigue as the tasks advance that would slow down their performance (Moore & Tenny, 2012). Put differently, additional time would only work up to a certain point but may not proportionately promote DT performance.

Presence of this kind of a relationship implies a curvilinear relationship in TOT effects. The curvilinear relationship appears in creativity as found in that people are most productive in creativity tests when tasks are moderately stimulating (Voss, 1977). Baer and Oldham (2006) particularly affirmed the curvilinear relationship among employees at the workplace indicating that employee's creativity, rated by their supervisors, increased up to a certain point and declined after that point as they perceived more time pressure at their work. Simply speaking, creative performance can be most elevated under the optimal time conditions then stop increasing after that point, which may imply the *inverted J-shaped* relationship. However, it is worth to note that perceived time pressure is not the same as the actual amount of time given, although perceived time pressure is relevant to the shorter time condition examined in the present study as people are more likely to feel pressured when they are under strict time conditions..

2. The Present Study

The present study aimed to synthesize research examining whether people's performance curvilinearly increased in DT tests as they were allowed with more time to take the DT tests, and what characteristics of study and sample moderate the TOT effect. To this end, five research questions were specifically examined.

1. What is the overall relationship between TOT and DT performance?

a. Does DT performance increase as people are allowed more time to take DT tests?

b. Does the relationship vary by which method of time manipulation (i.e., untimed vs. timed conditions; shorter time vs. longer time conditions) is chosen?

2. Does the relationship vary by the moderators such as people's demographic backgrounds (i.e., gender, age, nationality), study characteristic (i.e., year of publication, within-subject vs. between-subject design), characteristics of DT tests (i.e., DT tests, test modality, subscales, scoring methods of originality)?

3. Is there a curvilinear relationship between TOT and DT performance?

a. Does the relationship continue to increase at the same rate as more time is allowed?

b. Does the relationship change by how much of time is allowed in the reference group?

This is not the first investigation on the impact of TOT on DT performance. As mentioned above, Said-Metwaly et al.'s (2020) meta-analysis provided an overall perspective on timed and untimed testing. However, their study did not explore if the nature of the relationship could be non-linear. The present study expands Said-Metwaly et al.'s (2020) meta-analysis aiming to test whether DT performance curvilinearly increases with longer time provided for DT tasks. It is likely that more time will bring superior DT outcomes but the gains will not be proportional to the amount of time provided and will stop to increase after a certain threshold point of time. Thus, we added a few moderators for testing a curvilinear relationship between TOT and DT performance. In spite of the similarity in the focus, the present study differs from Said-Metwaly et al.'s (2020) study in that it included a larger number of studies ($k = 22$). This is nearly twice as many investigations as examined in the earlier meta-analysis ($k = 12$). The present study includes studies testing within-subject comparisons with different time intervals in the same samples in addition to between-subject comparisons. For instance, we included all possible comparisons of how DT scores changed when the same set of participants were given with 1, 3, and 5 minutes in a study (e.g., Benedek et al., 2013). We also considered how originality is scored as another moderator.

3. Method

3.1. Locating Previous Research Results

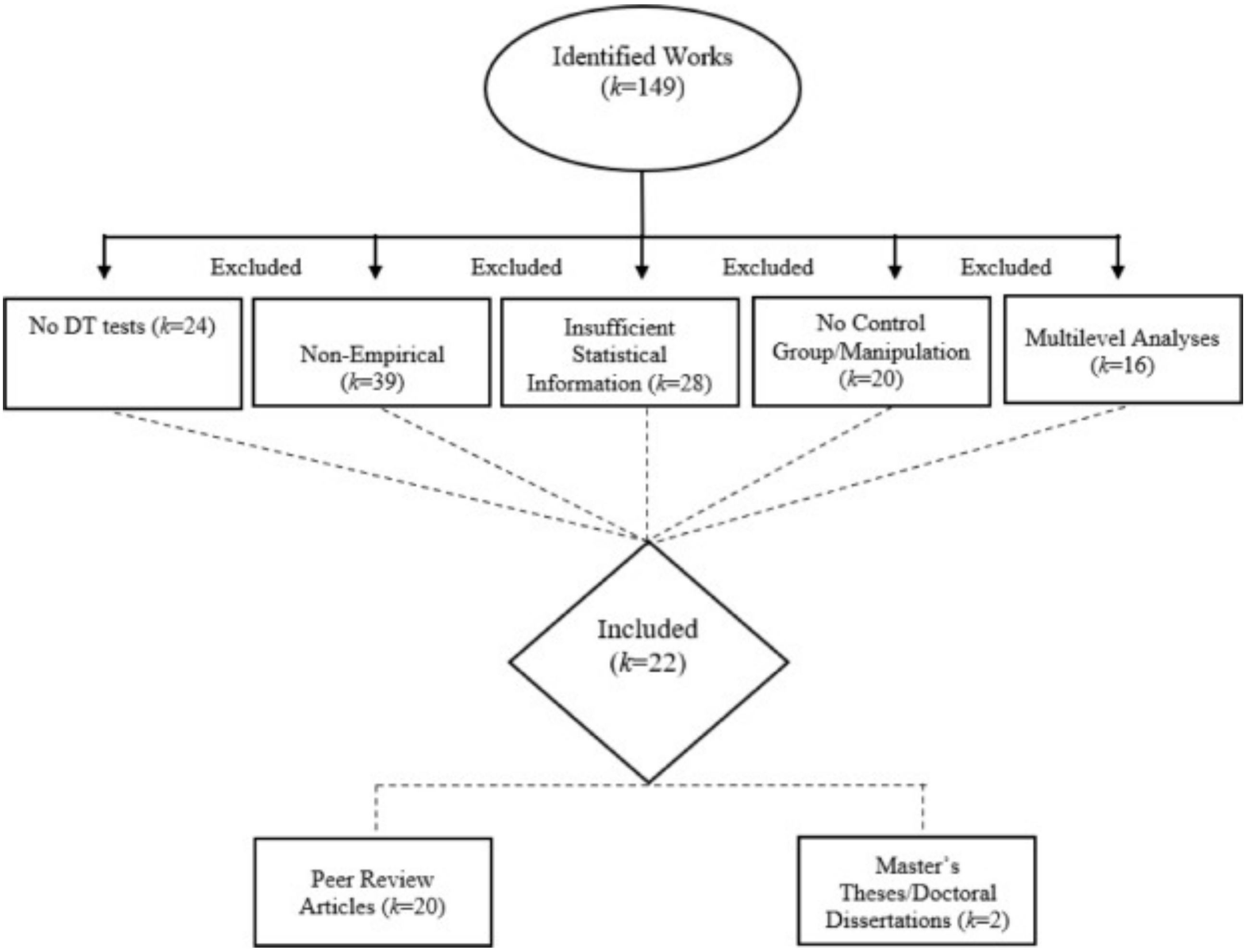
In the present study, different methods were considered in order to locate previous literature on DT and TOT. These methods included primary sources (e.g., scholarly journals, conference papers, master's theses and doctoral dissertations), and secondary sources (reference lists, reference databases, internet search, and personal contact) (Cooper, 2016).

Research articles from 1950 to 2020 were collected through a search of the following electronic databases: Academic Search Premier, PsycINFO, Educational Resources Information Center, Psychology & Behavioral Science Collection, ProQuest, and Google Scholar. In addition, reference lists were manually searched for other related studies that might be included. When the full text could not be obtained, or in case a study reported insufficient statistical information, authors were contacted (e.g., Benedek et al., 2013). Finally, the following major journals in creativity research were searched for relevant articles: *Creativity Research Journal, Journal of Creative Behavior, Psychology of Aesthetics, Creativity, and the Arts*, and *Thinking Skills and Creativity*.

The above mentioned electronic databases were searched using the advance search option. Both OR and AND options were used while searching previous literature. The following keywords were searched: *divergent thinking, creative thinking, creativity, fluency, flexibility, originality, time-on-task, timed, untimed, time, testing time, time pressure, speed, game-like*, and *test-like*. Studies that contained these keywords either in their titles or in abstracts were initially selected and reviewed to find out additional references. After removing redundancies in the results (some articles appeared twice or more because of the overlap between databases content), our literature search produced 257 publication including peer-review articles, master's theses and doctoral dissertations, reports, magazine, reviews, trade publications, and news. Among the 257 publications, 108 works were excluded because those were irrelevant to the current investigation.

3.2. Inclusion and Exclusion Criteria

Six criteria were applied to the 149 remaining publications to determine the number of works to be included in the current study. First, only studies in English were included. Second, only empirical studies, which provided quantifiable measures of DT *and* time were included. Any study that did not measure creativity through DT tests was excluded (e.g., Bar-Kochva & Hasselhorn, 2015). Third, all literature and theoretical reviews were excluded (e.g., Hattie, 1977). Fourth, for a study to be included, it should consist of an experimental group and a reference group, and time should be manipulated. Studies that did not meet this criterion were excluded (e.g., Ward, 1969a). It is worth noting that in some studies, a comparison has been made between two conditions: timed versus untimed; while other studies compared between different time intervals (e.g., 2.5 min. vs. 5 min. vs. 7.5 min.). Both types of studies were included in the present meta-analysis but were analyzed separately (see Results). Fifth, both published and unpublished studies were included to avoid publication bias (Rosenthal, 1979). Thus, the source type was not specified in the initial search resulting in publications from seven different sources as indicated above. Finally, only those studies that provided sufficient statistical information to calculate effect size were selected. The statistical information includes mean, standard deviation, sample size, product-moment correlation, and various statistics testing mean differences such as *t*-value, *F*-value, and chi-square value. Some studies were excluded because they did not report sufficient statistical information (e.g., Kang et al., 2015). Furthermore, studies that reported multivariate analyses (e.g., multiple regression and MANOVA) were excluded because their effect sizes were adjusted by other covariates (e.g., Ward, 1969b). Applying these criteria to the 149 publications resulted in including 22 publications as follows: 20 peer-review articles, 1 master's thesis and 1 doctoral dissertation (see Figure 1).



Download : Download high-res image (299KB) Download : Download full-size image

Figure 1. Flow chart for selection of studies Note. k = number of studies

---

3.3. Coding Procedure and Reliability

A coding protocol was developed by the authors, which included the following information: year of publication, country, gender, age, methods of time manipulation (i.e., timed vs. untimed conditions; shorter vs. longer time conditions), DT tests (i.e., the Torrance Test of Creative Thinking [TTCT], Wallack-Kogan test, Guilford test, and others), DT subscales (i.e., fluency, flexibility, originality, elaboration, quality, and composite), DT task modality (i.e., verbal and figural tests), study design (i.e., between- and within-subject design), methods of scoring originality, effect size, and sample size. A total of 237 effect sizes from 22 studies were independently coded by the first and second authors, and there was a very high agreement in their coding of both continuous (ICC = .99, $CI_{95\%}$ = [.998, .993]) and categorical variables (weighted kappa = 1)

3.4. Effect Size Calculation

The standardized mean difference (i.e., Hedges' *g*) in DT performance adjusted by small sample size was calculated between the experimental group and reference group. The terms "restricted" and "liberal" are relative to one another. The groups that received relatively less time to complete the DT tasks was considered "restricted" compared to the one that used more time. Likewise, the group that was "timed" was considered "restricted" compared to "untimed" group that is "liberal". Using the terminology of experimental design, we refer to "restricted" conditions as the reference group and the "liberal" conditions as the experimental group. In the case in which mean scores and standard deviations were reported, the following formula was used to estimate Hedges' *g*:

$$Cohen's\, d = \frac{M_E - M_R}{SD_{Pooled}} \quad (1)$$

$$SD_{Pooled} = \sqrt{\frac{(n_E-1)SD_E^2 + (n_R-1)SD_R^2}{n_E + n_R - 2}} \quad (2)$$

where M_E is the mean of DT performance, SD_E is the standard deviation of the mean, n_E is the sample size in the experimental group, and M_R is the mean of DT performance, SD_R is the standard deviation of the mean, n_R is the sample size in the reference group. The half of the mean differences were tested in the studies with a repeated-measure design that need to be accounted for in calculating effect sizes (Morris & DeShon, 2002). However, we should note that we calculated them using Equation 1 as the statistical information provided in most studies is not sufficient to use what Morris and DeShon (2002) suggested. We think it does not overestimate the results treating the effects tested in repeated-measure designs as if they were tested in independent samples because the effects tested in repeated-measure designs could be overly weighed if Morrison & DeShon's approach was taken. In this sense, the approach being taken in the present study is more conserved. When studies reported *F*-value, *t*-test, or Pearson product-moment correlation coefficient, these statistics were converted to Cohen's *d*, then to Hedges' *g* using the following formula:

$$g \cong d\left(1 - \frac{3}{4(n_E + n_R) - 9}\right) \tag{3}$$

3.5. Synthesizing Effect Sizes

The present study employed a random-effects model to aggregate mean differences between: a) timed and untimed conditions, and b) shorter and longer time conditions. The random-effects model assumes that there are two sources of variability, the within-study variance (i.e., sampling error) and the between-study variance (i.e., the variance of effect sizes across studies). In order to account for between-study variance, we identified several moderators that have influenced the findings of previous research on the impact of TOT on DT performance.

Each effect size is weighted by the inverse of its variance. The weight assigned to each study is:

$$W_i = \frac{1}{V_{Y_i}}, \tag{4}$$

$$V_{Y_i} = V_{Y_i} + T^2 \tag{5}$$

where W_i refers to weight for an effect retrieved from study i, Y_i refers to an effect in study i, V_{Y_i} refers to the within-study variance for study i plus the between-study variance (T^2). The weighted mean then can be calculated using the following formula (Borenstein et al., 2009, pp. 73):

$$M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \tag{6}$$

where k is the number of studies.

There was a dependency in our study sample due to multiple effect sizes obtained from a single study in most of the included studies. Effect sizes were nested in the obtained studies. To control for this dependency, we adopted a three-level modeling approach in the present meta-analysis. In this specific form of a three-level model, which is referred to as Level 1 Variance Known model (Bryk & Raudenbush, 1992), the first level represents a non-systematic residual variance indicating sampling error. Level 2 represents variation between individual effect sizes within a study, and Level 3 represents variation across studies that provided those effect sizes. In order to calculate the mean effect size, we ran an unconditional model with no predictors (Predictors refer to the moderators in the present study.). Following

Konstantopoulos' (2011) parameterization of the three level models, a complete unconditional three-level model is:

$$Y_{ig} = \gamma_{00} + u_{0g} + r_{ig} + e_{ig}, \quad (7)$$

where Y_{ig} represents individual effect sizes, γ_{00} represents aggregate mean, u_{0g} is the mean of the distribution of level-3 random effects that is normally distributed with a mean of zero and between publication variance, r_{ig} is the mean of the distribution of level-2 random effects that is normally distributed with a mean of zero and between effect size variance, $g = 1, 2, \ldots$, k represents the level-3 units (i.e., studies), and $i = 1, 2, \ldots$ j represents level-2 units (i.e., effect size). The full model includes all study moderators and when they were added, the final model becomes:

$$\pi_{ig} = \beta_{0g} + \beta_{1g}X_{1ig} + \cdots + \beta_{pg}X_{pig} + r_{ig}, \quad (8)$$

where β_{0g}, β_{1g},…β_{pg} represent regression coefficients, and X_{1ig},…. X_{pig} represent study moderators.

3.6. Heterogeneity and Moderator Analyses

Both Q-test and I^2 statistic were calculated to assess the heterogeneity of effect sizes. The former represents between-study variability whereas I^2 represents the proportion of variance that is due to heterogeneity between studies (Higgins & Thompson, 2002). In this meta-analysis, a high heterogeneity was expected in effects across studies as different effects were found by the characteristics of studies and samples.

In order to explain variability in the mean effect, the following moderators were considered: a) year of publication, b) country, c) gender, d) age, e) DT tests, f) DT subscales, g) DT task modality, h) study design, i) time difference between the experiment and reference groups, and j) baseline time that refers to the time given in the reference groups. The aforementioned moderators and their categories were presented in Table 1. Two characteristics of participants, age and gender, were chosen as moderators as DT performance may vary across males and females (Baer & Kaufman, 2008; Kuhn & Holling, 2009) and differ as they aged (Bijvoet-van den Berg & Hoicka, 2014). Although female students tend to outperform in DT tests over male students (Kuhn & Holling, 2009), this significant difference does not sustain in adults aged from 17 through 99 years-old (Reese et al., 2001). Thus, the mean effect size was compared between males and females, and children and adults.

Table 1. Description of the Study Moderators

Moderator	Operational Definitions
Year of Publication	
Country	Country where data were collected
Gender	
Male	Males consist of 75% or more of sample size
Female	Females consist of 75% or more of sample size
Combined	Sample combining both males and females
Other	Gender not reported
Age Groups	
G1-6	Sample with 6- to 11-year-old group
G7-12	Sample with 12- to 17-year-old group
Adults	Sample with 18-year-old or above
Other	Age not reported; or multiple age groups combined.
Time Manipulation	
Timed VS. Untimed	Study design that specified the time allowed for responding to different DT tasks (i.e., test-like) versus study design that did not specify the time for responding to different DT tasks (i.e., game-like).
Shorter VS. Longer Time	Study design that compared different time intervals (e.g., 2 minutes versus 5 minutes), or study design where a comparison has been made between untimed condition versus different time intervals.
DT Tests	
TTCT	Measuring DT using the TTCT
Guilford	Measuring DT using Guilford Structure of Intellect tests
Wallach & Kogan	Measuring DT using Wallack & Kogan's tests
Other	Measuring DT using other DT tests
DT Subscales	

Fluency	Thinking skills to produce many ideas related to the task
Flexibility	Thinking skills to produce many different types of ideas or to shift from one category of ideas to another
Originality	Thinking skills to produce uncommon and unique ideas
Elaboration	Thinking skills to work out the details of an idea
Quality	The extent to which ideas are appropriate for intended purpose
Composite	Composing two or more DT subscales into a single score
Task Modality	
Verbal	DT tasks providing verbal stimulus and/or asking participants to generate verbal responses
Figural	DT tasks providing figural or spatial stimulus and/or asking participants to generate figural or spatial responses
Other	Any modality that is neither verbal nor figural; or verbal and figural tests are collectively tested
Study design	
Between-subject design	Study design that tested the TOT effect by comparing between two independent samples that are allowed with the different amounts of time
Within-subject design	Study design that tested the TOT effect by comparing DT performance in a single sample with the different amounts of time
Originality Scoring	
Statistical Infrequency-based	Originality is scored based on the counts of responses
Subjective Rating-based	Originality is scored by judges or raters.
Other	Originality is scored in ways that are not aforementioned.
Aggregation of Originality Scores	
Summation	The counts of original responses are summed to create the score.

Snapshot	Only part of the response set is assessed to score originality
Ratio/Weighted	The proportion of original responses to the total response is counted to score originality.
Other	Originality is scored in ways that are not aforementioned.

More interestingly, the effect of the demographic characteristics of participants may also interact with the modality of DT tests. For instance, older people perform as well as younger people on DT tests when the tasks are verbally presented, while they underperform compared to younger people when the tasks are visually presented (Palmiero et al., 2014). Thus, the mean effect size was compared between verbal and figural tasks.

The impact of TOT on DT performance may vary across different types of DT tasks. This implies that different DT tests varying in other characteristics possibly elicit difference in DT performance by time. For instance, standardized instructions are required in administering TTCT, while no specific instruction is required in Wallach-Kogan and Guilford (WKG) tests. Thus, the mean effect size was compared between TTCT and WKG tests.

DT subscales represent different thinking skills in which time plays a differential role. As illustrated well in the serial order effect (Beaty & Silvia, 2012; Mednick, 1962; Runco, 1986), DT subscales representing quantity (i.e., fluency and flexibility) of mediocre responses may not necessarily favor liberal time while DT subscales representing originality of responses may favor liberal time. Thus, the mean effect size was compared between fluency/flexibility and originality.

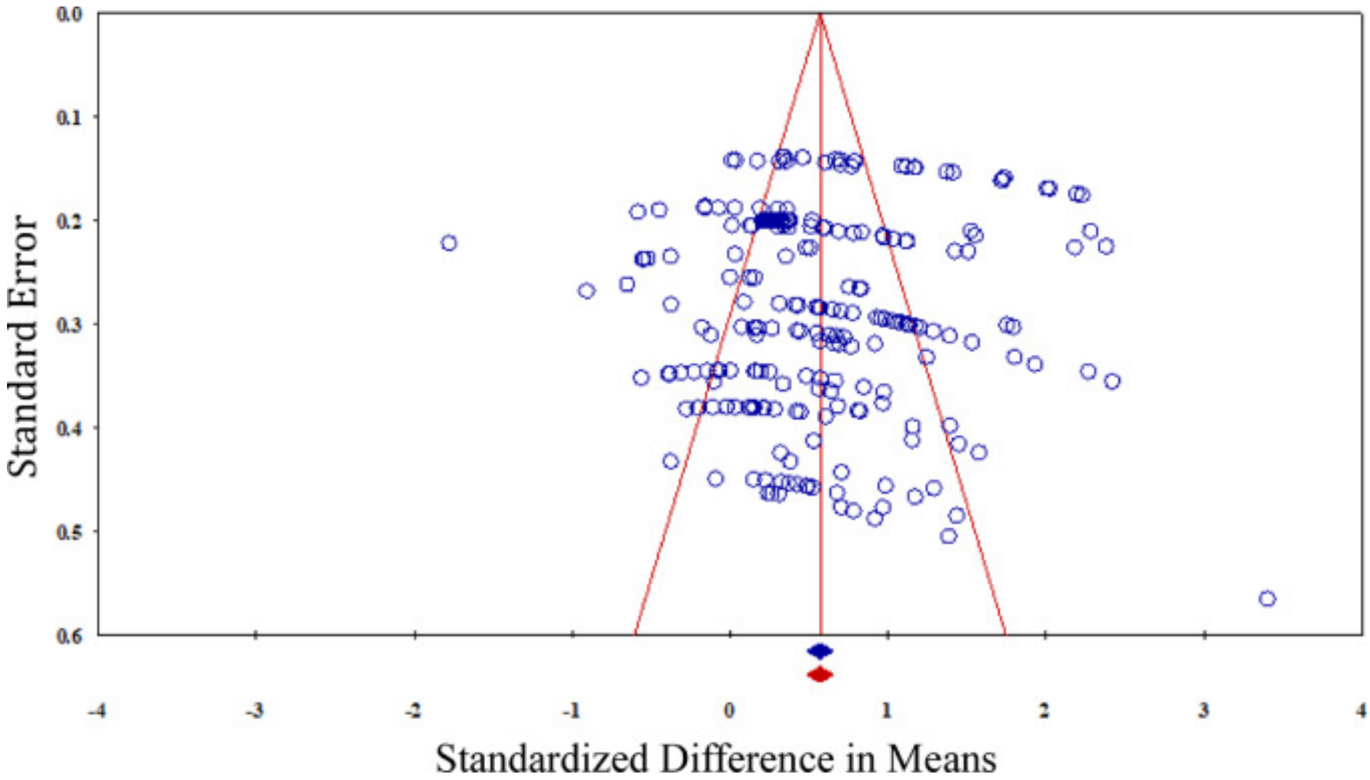
Most importantly, we tested the difference in amount of time that people were allowed to take a DT test between the experimental and reference groups and the quadratic term of time difference (time difference*time difference). Time difference is only tested in the subset comparing shorter vs. longer time conditions because it is not available in the subset comparing timed vs. untimed conditions. The time difference was coded as continuous using minutes as the unit of time. The inclusion of the quadratic term of time difference aims to test the hypothesized J-shaped relationship between TOT and DT outcomes indicating whether the impact of TOT would be proportional to the increment of TOT and get smaller and flatten after a certain period.

Besides, we tested baseline time that refers to the amount of time that was allowed in the reference group. We hypothesized that if the effect of baseline time is negative, it would indicate that the effect size is expected to diminish with higher baseline time. Univariate analyses were conducted to see the mean effect size in each category of individual moderators.

After obtaining the mean effect size from the unconditional model, the moderators were added into a full model in which the effects of the individual moderators were tested in a multilevel regression analysis in order to control for shared variance among various moderators. The dummy codes were created and entered in the model if the moderators were categorical (see corresponding footnotes in Table 3 and 5).

3.7. Assessing Publication Bias

Four indicators of publication bias were tested: a) Rosenthal's fail-safe *N*, b) the funnel plot test, and c) Egger's test. Rosenthal's fail-safe *N* represents the number of missing studies that could annul the reported average effect size. Figure 2 illustrates the funnel plot of standard error. Egger's test assesses asymmetry of the funnel plot using regression analysis, while the funnel plot test examines the impact of sample size on the effect sizes (Begg & Mazumdar, 1994; Egger et al., 1997). Non-significance of both indicate funnel plot symmetry implying that publication bias is unlikely to impact the results. In the present study, we tested the funnel plot test and Egger's regression test within a three-level model as suggested in Fernández-Castilla et al. (2021).



Download : Download high-res image (378KB) Download : Download full-size image

Figure 2. Funnel plot of standard error by standardized difference in means

4. Results

4.1. Preliminary Analyses

We examined the homogeneity of the data. Data were extremely heterogeneous, $Q_T(236) = 1712.93$, $p < .001$, $I^2 = 86.22\%$. We also tested the publication bias in several ways. Rosenthal's (1979) fail-safe *N* was 6,496, which represents the number of missing studies that could annul the reported average effect size in the present study. Given pretty large fail-safe *N*, it seems that the present study's findings are safe from a file-drawer issue. However, Egger's (Egger et al., 1997) regression intercept test was significant, $t(235) = 2.422$, $p = .016$, whereas the funnel plot test ,$t(235) = 0.95$, $p = .340$, was not significant. Therefore, the evidence regarding the funnel plot asymmetry (see Figure 2) and therefore publication bias is mixed.

4.2. Overall Relationships

The first analysis used the entire dataset ($j = 237$) to compare the difference between restricted and liberal time conditions. The entire dataset was then used to calculate a general mean effect size. The unconditional model compared the liberal and restricted time conditions using 237 effect sizes from 22 studies. The mean effect size was $g = 0.666$, 95% confidence interval (CI) [.397, .935], $p < .001$. The details of the unconditional models are presented in Table 2. Liberal time conditions resulted in significantly higher DT performance than the restricted time conditions, and this effect size is considered moderate to large (Cohen, 1988). This analysis provided an overall picture of comparisons with an inclusive definition of "restricted" and "liberal" time conditions. To be more specific about our findings, we divided our dataset into two subsets.

Table 2. Unconditional Models

Unconditional Model for the Entire Dataset ($j = 237$)			
	Estimates	***SE***	***t***
Fixed effects			
Intercept	0.666**	0.137	4.86
Variance Components			*Z*
Second level	0.161**	0.022	7.12
Third level	0.351**	0.134	2.63

Unconditional Model for Subset Comparing Timed VS. Untimed Conditions (*j* = 109)			
Fixed effects			
Intercept	0.493**	0.149	3.31
Variance Components			Z
Second level	0.224**	0.046	4.88
Third level	0.279*	0.127	2.19
Unconditional Model for Subset Comparing Shorter VS. Longer Time Conditions (*j* = 128)			
Fixed effects			
Intercept	0.901**	0.206	4.37
Variance Components			Z
Second level	0.118**	0.022	5.30
Third level	0.341	0.212	1.61

Note. *$p < .05$; **$p < .01$, j = number of effect size.

4.3. Comparing Timed vs. Untimed Conditions

The first subset ($j = 109$ from 15 studies) included studies comparing "timed" to "untimed" testing conditions. The results of the unconditional model were presented in Table 2. The mean effect size was $g = 0.493$, 95% CI [.201, .785], $p = .005$. DT performance was significantly higher in untimed testing conditions than timed testing conditions, and the effect size is considered small (Cohen, 1988). The data were heterogenous, $Q_T(108) = 795.27$, $p < .001$. The following analyses focused on moderator analyses, and none of the moderators turned out significant among year of publication, country, gender, age, DT tests, DT subscales, task modality, and study design (see Table 3). The mean effects of each category in the individual moderators are presented in Table 4.

Table 3. Full Model Comparing Timed vs. Untimed Conditions (j = 109)

	Estimates	SE	*t*
Fixed effects			

Intercept	1.939	32.124	0.06
Year	0.000	0.016	-0.03
Country	-0.390	0.506	-0.77
Gender (Female)	0.038	0.285	0.13
Gender (Mixed)	-0.190	0.637	0.30
Age (Grades thru 6)	-0.006	0.510	-0.01
Age (Grades 7 thru 12)	-0.266	0.296	-0.90
Age (combined)	-0.103	0.839	-0.12
DT Test (TTCT)	-0.110	0.324	-0.55
DT Test (Other)	0.147	0.384	0.49
DT Subscale (Originality)	0.265	0.186	1.50
DT Subscale (Other)	0.150	0.243	0.47
Task Modality (Figural)	0.019	0.335	0.23
Task Modality (Other)	0.730	0.625	0.76
Study design (Dependent samples)	-0.853	0.484	-1.76
Variance Components			*Z*
Second level	0.227**	0.049	4.66
Third level	0.289*	0.196	1.47

Note. *$p < .05$; **$p < .01$, j = number of effect size.

Reference groups for dummy variables: Gender = Male; Age = College students; DT Test = Wallach-Kogan & Guilford; DT Subscale = Fluency/Flexibility; Task Modality = Verbal; Study design = Dependent samples.

Table 4. Mean Effect Sizes for Moderator Categories

	j	ES	SE	p	t	Q_T	I^2
All	237	0.666	0.137	< .001	4.86	1,712.93	86.22%

Timed vs. Untimed	109	0.493	0.149	.005	3.31	795.27	86.17%
Country							
The US	88	0.377	0.125	.012	3.01	366.87	75.74%
Non-US/mixed	21	0.834	0.459	.163	1.82	387.42	94.32%
Gender							
Male	31	0.469	0.321	.283	1.46	80.27	60.13%
Female	9	0.603	1.006	.566	0.60	8.88	0%
Mixed/Else	69	0.493	0.175	.016	2.82	688.80	89.84%
Age Groups							
Grade 1-6	15	0.417	0.474	.478	.88	175.07	90.86%
Grade 7-12	30	0.465	0.362	.252	1.29	308.96	89.97%
College	51	0.434	0.147	.267	2.94	263.19	80.24%
Combined	13	0.773	0.106	.106	7.29	11.34	0%
DT Tests							
TTCT	19	0.476	0.256	.171	1.86	157.60	87.31%
WKG	27	0.247	0.166	.203	1.49	172.20	83.74%
Others	63	0.573	0.236	.042	2.43	442.77	85.55%
DT Subscales							
Fluency/@@@@ Flexibility	51	0.505	0.244	.063	2.07	569.63	90.87%
Originality	46	0.450	0.103	.007	4.37	165.95	71.68%
Others	12	0.638	0.228	.026	2.80	53.93	75.89%
Task Modality							
Verbal	83	0.353	0.128	.020	2.75	484.34	82.66%
Figural	16	0.418	0.412	.495	1.02	43.60	61.01%
Others	10	0.959	0.586	.240	1.64	244.11	95.49%
Study design							
Independent samples	69	0.827	0.245	.015	5.97	378.20	82.02%
Dependent sample	40	0.164	0.124	.228	6.74	287.26	86.42%

Shorter vs. @@@@Longer Time	128	0.901	0.206	.004	4.37	826.18	84.63%
Country							
US	107	0.781	0.247	.024	3.16	276.22	60.9%
Non-US/mixed	21	1.209	0.135	.001	8.96	313.65	93.62%
Gender							
Male	27	0.415	0.097	.146	4.30	24.66	0%
Female	39	1.298	0.424	.057	3.06	161.69	76.51%
Other	62	0.707	0.209	.034	3.39	605.49	89.6%
Age Groups							
Grade 7-12	26	1.002	0.434	.109	2.31	89.53	70.96%
College Students	102	0.804	0.182	.012	4.42	733.15	86.22%
DT Tests							
TTCT	32	1.291	1.028	.428	1.26	71.72	56.78%
WKG	72	0.607	0.120	.013	5.08	157.83	55.01%
Other	24	1.163	0.127	.001	9.15	319.45	92.8%
DT Subscales							
Fluency/ Flexibility	71	0.893	0.256	.034	3.49	605.49	88.27%
Originality	47	0.775	0.305	.096	2.54	106.52	55.01%
Other subscale	10	1.028	0.709	.384	1.45	69.24	87%
Task Modality							
Verbal	98	0.969	0.210	.006	4.62	601.26	83.87%
Figural	30	0.278	0.037	< .001	7.60	4.02	0%
Study design							
Dependent samples	95	1.013	0.241	.011	4.21	597.65	84.27%
Independent sample	32	0.405	0.188	.276	2.16	9.23	0%

Note. j = number of effect size; TTCT = Torrance Test of Creative Thinking, WKG = Wallach-Kogan & Guilford tests

To examine if potential confound among the moderators impact the results, we detected the bivariate correlations of the moderators. The correlations were high ($r > .75$) with DT task modality. When it was removed from the full model, moderators were still non-significant.

4.4. Comparing Shorter vs. Longer Time Conditions

The second subset of studies involved comparisons between shorter and longer time periods. Analyses using 128 effect sizes from 9[1] studies provided a mean effect size of $g = 0.901$, 95% CI [.497, 1.305], $p = .004$. Liberal time conditions led to significantly higher performance on DT than the restricted time conditions, and this is a large effect size (Cohen, 1988). The data were heterogeneous, $Q_T(127) = 826.16$, $p < .001$. In these comparisons, the amount of time in the experimental and reference groups varied. Therefore, time difference between the experimental and reference groups and baseline time were tested as moderators along with a few others.

The full model including all tested moderators is presented in Table 5. This model included baseline time, time difference, and the quadratic term for time difference (time difference *time difference) on top of the same set of variables being tested in the subset comparing timed vs. untimed conditions. As seen in Table 5, baseline time ($\beta = -0.062$, $SE = .021$, $t = -2.91$, $p = .005$), time difference ($\beta = 0.594$, $SE = 0.065$, $t = 9.17$, $p < .001$), DT subscale-originality ($\beta = -0.190$, $SE = 0.089$, $t = -2.14$, $p = .035$), and the quadratic term of time difference ($\beta = -0.041$, $SE = 0.005$, $t = -8.73$, $p < .001$) turned out to significantly predict the effect sizes.

Table 5. Full Model Comparing Shorter VS. Longer Time Conditions ($j = 128$)

	Estimates	**SE**	***t***
Fixed effects			
Intercept	27.971	59.470	0.47
Year	-0.014	0.029	-0.48
Country	-0.338	1.342	-0.25
Gender (female)	-0.102	0.206	-0.50
Gender (mixed)	0.498	0.828	0.60
Age (Grades 7 thru 12)	0.033	0.195	0.17
Baseline Time	-0.062**	0.021	-2.91

Time Difference	0.594**	0.065	9.17
DT Test (TTCT)	0.905	1.621	.56
DT Test (other)	0.175	1.238	0.14
DT Subscale (originality)	-0.190*	0.089	-2.14
DT Subscale (other)	-0.012	0.115	-0.10
Task Modality (figural)	-2.032	1.880	-1.08
Study design (Independent samples)	-0.343	1.453	-1.08
Time Difference*Time Difference	-0.041**	0.005	-8.73
Variance Components			*Z*
Second level	0.034**	0.011	3.01
Third level	0.427	0.665	0.64

Note. Reference groups for dummy variables: Gender = Male; Age = College students; DT Test = Wallach-Kogan & Guilford; DT Subscale = Fluency/Flexibility; Task Modality = Verbal.

*$p < .05$; **$p < .01$, j = number of effect size.

Those findings indicated that the mean effect size decreases as more time provided in the reference condition and increases as the difference between the experimental and reference conditions gets larger. The significant negative quadratic term indicates that there is a non-linear relationship between time difference and the effect sizes suggesting an inverted J-shaped relationship that the relationship is positive, yet progressively decreasing. This trend was observed when the data were compared across three periods of time difference between the longer and shorter time conditions: a) less than 2 min., b) 2 min. to less than 5 min., 3) 5 to 10 min. The mean effect size was, $g = 0.545$, 95% CI [.392, .698], $p = .006$ for the time difference of less than 2 min.; $g = 1.010$, 95% CI [.440, 1.580], $p = .068$ for the time period of 2 to less than 5 min., and $g = 1.370$, 95% CI [.517, 2.223], $p = .051$ for the time period of 5 to 10 min. The amount of increase in the mean effect sizes by time difference is not proportional to the gains over time. Furthermore, the mean effects of the last two time periods are not significantly different from zero.

Another significant moderator was DT subscale where originality subscale ($g = 0.775$, 95% CI [.177, 1.373], $p = .096$) had a significantly lower mean effect size than fluency/flexibility ($g = 0.893$, 95% CI [.391, 1.395], $p = .034$). Increased time benefitted to originality less than it did to fluency and flexibility.

To make sure any confound among the study moderators does not impact the results, we examined the bivariate correlations among the moderators. The highest correlations were high ($r > .75$) among DT task type, DT modality, and the study design. Following Viechtbauer's (2007) suggestion, we removed DT task type and DT modality because they were less important than the study design. All of the main and interaction effects remained the same and the confound among study moderators did not impact the results.

4.5. Originality Scoring

Another moderator that is worthy of attention is how originality is scored. Originality is scored in a number of ways such as aggregation of frequency-based scores, average of frequency-based scores, and subjective ratings, and the impact of TOT on originality might have been influenced by the way it is scored. Since this moderator was available for only a small number of studies, the above models did not include it in order to keep statistical power high. This moderator was tested for both datasets (timed vs. untimed conditions; shorter vs. longer time conditions). This analysis was first conducted in the dataset, in which timed and untimed conditions were compared (45 effect sizes from ten studies). The mean effect size is not significantly different from the studies in which originality is scored as the average of statistically infrequent responses no matter whether originality is scored as the sum of statistically infrequent responses ($\beta = -0.090$, $SE = 0.230$, $t = -0.39$, $p = .715$) or subjective scoring ($\beta = -0.168$, $SE = 0.240$, $t = -0.70$, $p = .507$).

Likewise, in the dataset where shorter and longer time conditions are compared (48 effect sizes in six studies), neither originality scores based on the sum of statistically infrequent responses ($\beta = 0.607$, $SE = 0.927$, $t = 0.66$, $p = .562$) nor various other scoring methods ($\beta = 0.558$, $SE = 0.916$, $t = 0.69$, $p = .590$) was any different from the average of statistically infrequent responses.

6. Discussion

The present study aimed to examine whether the relationship between TOT and DT performance changed curvilinearly as more time was given. Three findings are worthy to further discussed: a) DT performance increases as more time is given, b) originality does not necessarily take an advantage of liberal time compared to fluency and flexibility, and c) the relationship is characterized in the inverted *J-shaped* relationship.

Results indicate that DT performance increased, on average, as more time was given. In other words, people tend to do better in DT tests with more time to take the tests. This finding is comparable to the positive TOT effect found in a recent meta-analysis (Said-Metwaly et al., 2020). Furthermore, the present study confirms that DT performance increases regardless of time manipulation methods although the effect is larger when comparing the shorter vs. longer time conditions (g = .901) than comparing the timed vs. untimed conditions (g = .493). This discrepancy is a nuanced result unique to the present study because it may uncover what role time plays in DT performance when accounting for the comparisons of all possible pairs of time intervals. Perhaps, it indicates that DT performance may be maximal at sufficiently enough time with a definite time limit over indefinite time because untimed conditions do not return the higher effect. One of the probable reasons is self-regulation: When people know how long they are given in the DT tests, they would be more efficient in allocating time across their cognitive processing to yield their best performance. However, this result is not enough to directly uphold the non-linear relationship between time and DT performance.

Thus, moderator analyses were conducted using meta-regression analyses to account for shared variance among the moderators and reduce the likelihood of Type I error aiming to add rigor to the analyses. In the subset comparing timed vs. untimed conditions in a meta-regression analysis, no moderator turned out to significantly explain the effect. However, the results of moderator analyses with the data from the studies comparing shorter vs. longer time revealed some interesting findings. As expected, the baseline time, time difference, and a quadratic term of time difference turned out all significant. The significant moderators provide evidence of the inverted *J-shaped* relationship between TOT and DT performance. The negative coefficient of the quadratic term indicates that the curvilinear relationship is concave implying that DT performance increases as more time allowed, but its growth slows down progressively. In other words, this result indicates the inverted *J-shaped* relationship that is marked with the progressive growth followed by some deceleration.

The negative effect of the baseline time also supports the inverted *J-shaped* relationship because the less time allowed in the reference group, the larger mean effect was found than when more time is allowed in the reference group. To exemplify this, let us think of two hypothetical testing conditions: In Study 1, the reference group is given 2 min. and the more liberal condition is allowed 4 min. In Study 2, the conditions are 4 and 6 min, respectively. In both studies, the time difference is 2 min. Although both conditions have the same time difference being added in the moderator analyses, the effect size would be larger in Study 1 than Study 2 because 2 minutes of additional time make more difference in Study 1. This is supportive of a curvilinear relationship because the additional time allowed makes less of a difference if it comes after a longer period of time rather than a shorter period of time.

We further examined whether the gain proportionally increased as more time is given. We hypothesized that if the linear relationship was present between time and DT performance, the gain in the mean effect would appear proportional to the amount of time difference between the two groups being tested. In the results, the gain turned out all positive no matter how long the time difference was tested. Yet, the amount of gain tends to get smaller as the time difference gets larger. For instance, the mean effect did not triple up although the time difference is more than three times ($g = 0.545$ for less than 2 min., $g = 1.370$ for 5 to 10 min.) suggesting that the linear relationship is not supported.

The curvilinear relationship is repeatedly affirmed including studies examining the inverted U-shaped relation between time pressure and creativity (Andrews & Farris, 1972; Baer & Oldham, 2006; Bechtoldt et al., 2010; Ohly et al., 2006) implying that there seems to be an optimal time for DT performance. Assessing maximal performance is more critical in ability assessment (Bandalos, 2018) and DT is defined as an ability, not an attitude (Harrington, 1975). The inverted *J-shaped* relationship suggests that sufficiently long time period contributes to maximal performance in DT tests, but excessive time does not proportionally enhance it.

One of the probable reasons is that providing ample time may contribute to a psychologically safe environment and encourage people to follow their own pace in ideation which eventually elevates their creative performance. People tend to express their maximal creativity in a psychologically safe environment where they are not evaluated and there are few constraints, including time constraints (Dentler & Mackler, 1964; Harrington et al., 1987). For instance, people would be less stressed about time with less time constraints as they do not need to deal with the pressure of being timed. Under excessive time pressure, people are more likely to experience stress in which their epistemic motivation diminishes and correspondingly hazards creative problem solving (Schultz & Searleman, 1998). People may allocate their cognitive resource (i.e., attention) in managing time in addition to thinking about original responses. Because they are cognitively distracted by maintenance work (e.g., time management), they might lack their cognitive resource in advancing their performance in cognitively complex tasks requiring more cognitive flexibility (Baer & Oldham, 2006; Bechtoldt et al., 2010; Carnevale & Probst, 1998; De Dreu & Nijstad, 2008; Van Hiel & Mervielde, 2003). In a nutshell, ample time would promote DT performance through creating a psychologically safe environment that frees up cognitive resources.

Ample time serves to uphold the construct validity of DT scores not only through creating a psychologically safe environment but also eliminating construct-irrelevant variance in DT scores (Messick, 1995). Extreme time pressure is likely to threat the construct validity of DT scores by yielding construct-irrelevant variance. Preckel et al. (2011), for instance, found that speediness of DT tests significantly contributed to mental speed variance in students' creative performance. That being said, students who quickly processed cognitive reasoning were likely to outperform in speeded DT tests. The threat of time pressure to the construct-irrelevant variance also appeared in the result that those who were good at working memory were faster in judging relatedness which is key in generating remote (i.e., original) responses (Vartanian et al., 2009). Furthermore, people who have superior executive functions also outperformed as DT tests go (Beaty & Silvia, 2012). The findings imply that people with higher mental ability may take advantages in their creative performance compared to those with lower mental ability if they are rushed.

Nonetheless, it is noteworthy that this increment does not continue proportionally over time. The inverted *J-shaped* relationship implies that time does not need to be longer than is necessary to yield optimal performance. One of the probable reasons for the decelerated growth in DT performance over time is that excessive time may alter their use of cognitive resources. At the beginning of the tasks, people should be able to use their cognitive resources the most effectively being inspired by high achievement goals (Moore & Tenny, 2012). However, people may perceive the tasks less stimulating if time is more than needed. If people perceive tasks less stimulating, they are unlikely to engage in the tasks which would result in less enjoyment and correspondingly lower task performance (Freedman & Edwards, 1988; Gardner, 1990). Besides, their performance stops promoting at a certain point when they feel fatigue. With excessive time, they may be productive at the beginning but slow down as their mental resources are consumed (Briggs & Reinig, 2010). Furthermore, excessive time may leave room for distraction by unrelated tasks (Wilson et al., 1995). In the absence of a definitive end point, people may resort to other activities that may eventually undermine their performance. In this sense, the inverted J-shaped relationship is reasonable.

Another interesting finding is that the more gains are found in fluency and flexibility over originality with more time. This finding contradicts to the previous study (Said-Metwaly et al., 2020) that reported larger originality gains than fluency and flexibility over time. Here two things should be noted about the present study. First, we adopted a meta-regression approach, rather than univariate ANOVA approach to control for shared variance among study moderators. Second, the present study analyzed a larger number of studies. Indeed, the univariate analysis of the DT subscale found that the mean effect is significantly larger for fluency or flexibility ($g = .505$) than originality ($g = .450$). However, we should note that the finding does not contraindicate the serial order effect (Beaty & Silvia, 2012) because originality is not demoted over time, but its gain is just not as big as in fluency and flexibility. Put differently, originality continues to increase as it advances just at a slower rate than fluency and flexibility. More convincingly, long time conditions may promote originality through expanding people's repertoire of associations as explained by Mednick (1962). He underlined the importance of that requisite elements should suffice in people's response repertoire in order to be original in the creative process.

This finding is particularly rigorous because how responses were scored was also controlled for in meta-regression. When originality is scoring, a few different scoring methods can be used: For instance, originality is measured as statistical infrequency (e.g., Buteau, 1988) or judges' ratings (Hattie, 1980) for scoring the responses, and as summative counts (e.g., Plucker et al., 2006) or average ratings per response (e.g., Hass, 2015) for aggregating the responses. If

DT performance is solely measured by a summative method, the progressive growth may not be surprising (Briggs & Reinig, 2010). However, several different scoring methods being used in the studies are accounted for using meta-regression in the present study. This finding supports that the relationship between DT performance and time is moderated by the mix of originality scores computed in several methods accounting for the summed counts of original responses or the originality of the best responses as suggested in Briggs and Reinig's bounded ideation theory (2010). Though, we should note that we were not able to test how effect size varied by which method was chosen to score originality due to the small number of studies.

7. Limitations and Future Research

The relationship between TOT and DT performance found in the present meta-analysis may still vary by external and internal factors. For instance, the inverted *J-shaped* relationship may not be true for everyone in some DT subscales. Originality is perhaps more affected by people's executive process and fluid intelligence, but it is not merely by a function of time (Beaty & Silvia, 2012; Nusbaum & Silvia, 2011; Silvia & Beaty, 2012). Thus, the inverted *J-shaped* relationship may need to be examined across full spectrum of moderating factors such as people's mental ability and motivation in order to obtain a complete picture of the relationship. More importantly, perceived time limit may matter in the TOT effect in addition to how much of time people are actually given. Although tests are actually timed, people might not be affected by TOT if they are not aware of time limit while taking the tests. Whether people were informed of time limit or not was not specifically tested in the present study because of lack of information on this in the studies. Thus, future research should further investigate whether the TOT effect found in the present study remains generalizable even when the external and internal factors are accounted for.

More importantly, the finding does not tell how much of time is optimal for creative performance but only shows that DT performance may slow down at some point as the DT test advances. Ideally, if we synthesized effects testing how DT performance changed in each time segment (e.g., per minute) over time, we could have provided the most accurate picture of DT performance proportional to time. For instance, it may show whether DT subscales increase or decrease over time and at which point this change exactly decelerates or accelerates. Besides, the present study does not uncover why the growth in DT performance decelerates overtime. If the think-aloud data illustrating people's thought process during DT tests was available, we could have identified what may determine its deceleration. Unfortunately, these analyses were not available in the present study because studies rarely test DT performance in each time segment and collect think-aloud data. Future research should tackle the unresolved questions including the DT changes proportional to time as well as why DT performance slows down as time progresses.

Lastly, we converted some effect sizes that were estimated in studies with a repeated-measure design without accounting for within-subject variance according to the best practice suggested in the field (Morris & DeShon, 2002) due to the lack of statistical information provided in the articles. Thus, more rigor should be added to this conversion in future research.

Author statement

Here are how each author contributes to this project.

Sue Hyeon Paek: Conceptualization, data curation, writing, review & editing, and overseeing all steps involved in the project.

Ahmed Abdulla: Conceptualization, data curation, methodology, visualization, writing, and review & editing

Selcuk Acar: Conceptualization, formal analysis, writing, and review & editing

Mark A. Runco: Conceptualization, supervision, and review

No part of this paper, including the research presented therein, has been published elsewhere, and the manuscript is not under consideration for publication in any other journal.

References marked with an asterisk indicate studies included in the meta-analysis.

Uncited References:

Bleedorn, 1980, Forthmann et al., 2020, Johns et al., 2000, Khatena, 1971, Khatena, 1973, Richardson, 1981, Antes and Mumford, 2009, Sajjadi-Bafghi and Khatena, 1985, Torrance, 1969, Torrance and Ball, 1978

Recommended articles

References

Acar et al., 2020 S. Acar, M.A. Runco, H. Park
What should people be told when they take a divergent thinking test? A meta-analytic review of explicit instructions for divergent thinking
Psychology of Aesthetics, Creativity, and the Arts, 14 (2020), pp. 39-49, 10.1037/aca0000256
CrossRef View Record in Scopus Google Scholar

Akinboye, 1982 J.O. Akinboye
Correlates of testing time, age and sex in the Nigerians' performance on the Torrance test of creativity
Journal of Psychological Researches, 26 (1982), pp. 1-5
*
View Record in Scopus Google Scholar

Amabile, 1983 T.M. Amabile
The social psychology of creativity
Springer-Verlag, New York, NY (1983), 10.1007/978-1-4612-5533-8
Google Scholar

Amabile, 1996 T.M. Amabile
Creativity in context: Update to "The social psychology of creativity
Westview Press, Boulder, CO (1996)
Google Scholar

Amabile et al., 2002 T.M. Amabile, C.N. Hadley, S.J. Kramer
Creativity under the gun
Harvard Business Review, 80 (2002), pp. 52-63
View Record in Scopus Google Scholar

Andrews and Farris, 1972 F.M. Andrews, G.F. Farris
Time pressure and performance of scientists and engineers: A five-year panel study
Organizational Behavior & Human Performance, 8 (1972), pp. 185-200
Article Download PDF View Record in Scopus Google Scholar

Antes and Mumford, 2009 *
A. Antes, M. Mumford
Effects of time frame on creative thought: Process versus problem-solving effects
Creativity Research Journal, 21 (2009), pp. 166-182, 10.1080/10400410902855267
CrossRef View Record in Scopus Google Scholar

Baer and Kaufman, 2008 J. Baer, J.C. Kaufman
Gender differences in creativity
Journal of Creative Behavior, 42 (2008), pp. 75-105
CrossRef View Record in Scopus Google Scholar

Baer and Oldham, 2006 M. Baer, G.R. Oldham
The curvilinear relation between experienced creative time pressure and creativity: moderating effects of openness to experience and support for creativity

The Journal of Applied Psychology, 91 (2006), pp. 963-970
CrossRef View Record in Scopus Google Scholar

Bandalos, 2018 D.L. Bandalos
Measurement theory and applications for the social sciences
Guilford Publications, New York, NY (2018)
Google Scholar

Bar-Kochva and Hasselhorn, 2015 I. Bar-Kochva, M. Hasselhorn
In search of methods enhancing fluency in reading: An examination of the relations between time constraints and processes of reading in readers of German
Journal of Experimental Child Psychology, 140 (2015), pp. 140-157
Article Download PDF View Record in Scopus Google Scholar

Beaty and Silvia, 2012 R.E. Beaty, P.J. Silvia
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 (2012), pp. 309-319
CrossRef View Record in Scopus Google Scholar

Bechtoldt et al., 2010 M.N. Bechtoldt, C.K. De Dreu, B.A. Nijstad, H.S. Choi
Motivated information processing, social tuning, and group creativity
Journal of Personality and Social Psychology, 99 (2010), pp. 622-637, 10.1037/a0019386
CrossRef View Record in Scopus Google Scholar

Begg and Mazumdar, 1994 C.B. Begg, M. Mazumdar
Operating characteristics of a rank correlation test for publication bias
Biometrics, 50 (1994), pp. 1088-1101
CrossRef View Record in Scopus Google Scholar

Benedek et al., 2014 . .
M. Benedek, R. Beaty, E. Jauk, K. Koschutnig, A. Fink, P.J. Silvia, A.C. Neubauer
Creating metaphors: The neural basis of figurative language production
NeuroImage, 90 (2014), pp. 99-106, 10.1016/j.neuroimage.2013.12.046
Article Download PDF View Record in Scopus Google Scholar

Benedek et al., 2013 * 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. https://doi.org/10.1037/a0033644

Google Scholar

Bijvoet-van den Berg and Hoicka, 2014 S. Bijvoet-van den Berg, E Hoicka
Individual differences and age-related changes in divergent thinking in toddlers and preschoolers
Developmental Psychology, 50 (2014), pp. 1629-1639
CrossRef View Record in Scopus Google Scholar

Bleedorn, 1980 *
B.D. Bleedorn
Timed and untimed tests of creative thinking: A timely comparison
The General College Studies, 6 (1980), pp. 1-14
Google Scholar

Borenstein et al., 2009 M. Borenstein, L.V. Hedges, J.P.T. Higgins, H.R. Rothstein
Introduction to meta-analysis
Wiley, Chichester, U.K. (2009)
Google Scholar

Briggs and Reinig, 2010 R.O. Briggs, B.A. Reinig
Bounded Ideation Theory
Journal of Management Information Systems, 27 (2010), pp. 123-144, 10.2753/MIS0742-1222270106
View Record in Scopus Google Scholar

Bryk and Raudenbush, 1992 A.S. Bryk, S.W. Raudenbush
Hierarchical linear models: Applications and data analysis methods
Sage, Newbury Park, CA (1992)
Google Scholar

Bryan and Locke, 1967 J.F. Bryan, E.A. Locke
Parkinson's Law as a goal-setting phenomenon
Organizational Behavior and Human Decision Processes, 2 (1967), pp. 258-275
Article Download PDF View Record in Scopus Google Scholar

Buteau, 1988 *
A. Buteau
Effects of time press on figural creative performance
Dissertation Abstracts International (1988)
48-10A, 2575.
Google Scholar

Carnevale and Probst, 1998 P.J. Carnevale, T.M. Probst
Social values and social conflict in creative problem solving and categorization
Journal of Personality and Social Psychology, 74 (1998), pp. 1300-1309, 10.1037/0022-3514.74.5.1300
View Record in Scopus Google Scholar

Cohen, 1988 J. Cohen
Statistical power analysis for the behavioral sciences
(2nd ed.), Lawrence Earlbaum Associates, Hillsdale, NJ (1988)
Google Scholar

Cooper, 2016 H. Cooper
Research synthesis and meta-analysis: A step-by-step approach
(6th ed.), Sage Publications, Inc (2016)
Google Scholar

De Dreu and Nijstad, 2008 C.K.W. De Dreu, B.A. Nijstad
Mental set and creative thought in social conflict: Threat rigidity versus motivated focus
Journal of Personality and Social Psychology, 95 (2008), pp. 648-661, 10.1037/0022-3514.95.3.648
CrossRef View Record in Scopus Google Scholar

Dentler and Mackler, 1964 R.A. Dentler, B. Mackler
Originality: Some social and personal determinants
Behavioral Science, 9 (1964), pp. 1-7
https://doi.org/10.1002/bs.383009010210.1136/bmj.315.7109.629
CrossRef View Record in Scopus Google Scholar

Morris and DeShon, 2002 S.B. Morris, R.P. DeShon
Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs
Psychological Methods, 7 (2002), pp. 105-125
View Record in Scopus Google Scholar

Egger et al., 1997 M. Egger, G.D. Smith, M. Schneider, C. Minder
Bias in meta-analysis detected by a simple, graphical test
British Medical Journal, 315 (1997), pp. 629-634, 10.1136/bmj.315.7109.629
CrossRef View Record in Scopus Google Scholar

Fernández-Castilla et al., 2021 B. Fernández-Castilla, L. Declercq, L. Jamshidi, S.N. Beretvas, P. Onghena, W. Van den Noortgate
Detecting selection bias in meta-analyses with multiple outcomes: a simulation study
The Journal of Experimental Education, 89 (2021), pp. 125-144, 10.1080/00220973.2019.1582470
CrossRef View Record in Scopus Google Scholar

Foos and Boone, 2008 *
P.W. Foos, D. Boone
Adult age differences in divergent thinking: It's just a matter of time
Educational Gerontology, 34 (2008), pp. 587-594
CrossRef View Record in Scopus Google Scholar

Forthmann et al., 2020 *
B. Forthmann, C. Lips, C. Szardenings, J. Scharfen, H. Holling
Are speedy brains needed when divergent thinking is speeded—or unspeeded?
The Journal of Creative Behavior, 54 (2020), pp. 123-133, 10.1002/jocb.350
CrossRef View Record in Scopus Google Scholar

Freedman and Edwards, 1988 Freedman, J. L., & Edwards, D. R. (1988). Time pressure, task performance, and enjoyment. In J. McGrath (Ed.), *The social psychology of time: New perspectives* (Vol. 91, pp. 113–133). Newbury Park, CA: Sage.
Google Scholar

Gardner, 1986 D.G. Gardner
Activation theory and task design: an empirical test of several new predictions
Journal of Applied Psychology, 71 (1986), pp. 411-418
View Record in Scopus Google Scholar

Gardner, 1990 Gardner, D. G. (1990). Task complexity effects on non-task-related movements: A test of activation theory. *Organizational Behavior and Human Decision Processes*, 45, 209-231. https://doi.org/10.1016/0749-5978(90)90012-X
Google Scholar

Harrington, 1975 D.M. Harrington
Effects of explicit instructions to 'be creative' on the psychological meaning of divergent thinking test scores
Journal of Personality, 43 (1975), pp. 434-454, 10.1111/jopy.1975.43.issue-3
CrossRef View Record in Scopus Google Scholar

Harrington et al., 1987 D.M. Harrington, J.H. Block, J. Block

Testing aspects of Carl Rogers's theory of creative environments: child-rearing antecedents of creative potential in young adolescents
Journal of Personality and Social Psychology, 52 (1987), pp. 851-856
View Record in Scopus Google Scholar

Hass, 2015 *
R.W. Hass
Feasibility of online divergent thinking assessment
Computers in Human Behavior, 46 (2015), pp. 85-93, 10.1016/j.chb.2014.12.056
Article Download PDF View Record in Scopus Google Scholar

Hattie, 1977 J.A. Hattie
Conditions for administering creativity tests
Psychological Bulletin, 84 (1977), pp. 1249-1260, 10.1037/0033-2909.84.6.1249
View Record in Scopus Google Scholar

Hattie, 1980 *
J.A. Hattie
Should creativity tests be administered under test like conditions? An empirical study of three alternative conditions
Journal of Educational Psychology, 72 (1980), p. 87, 10.1037/0022-0663.72.1.87
View Record in Scopus Google Scholar

Higgins and Thompson, 2002 J.P.T. Higgins, S.G. Thompson
Quantifying heterogeneity in a meta-analysis
Statistics in Medicine, 21 (2002), pp. 1539-1558
CrossRef View Record in Scopus Google Scholar

Johns and Morse, 1997 *
G.A. Johns, L.W. Morse
Research note: Divergent thinking as a function of time and prompting to "be creative" in undergraduates
The Journal of Creative Behavior, 31 (1997), pp. 156-165, 10.1002/j.2162-6057.1997.tb00788.x
CrossRef Google Scholar

Johns et al., 2000 *
G.A. Johns, L.W. Morse, D.T. Morse
Divergent production in gifted adolescents using timed vs. untimed stimuli with creative prompting
Roeper Review, 22 (2000), pp. 165-166, 10.1080/02783190009554026

CrossRef View Record in Scopus Google Scholar

Kang et al., 2015 D. Kang, J. Park, H. Hong
Changes in the number of ideas depending on time when conducting scientific creativity activities
Journal of Baltic Science Education, 14 (2015), pp. 448-459
View Record in Scopus Google Scholar

Khatena, 1971 *
J. Khatena
Adolescents and the meeting of time deadlines in the production of original verbal images
Gifted Child Quarterly, 15 (1971), pp. 201-204
CrossRef View Record in Scopus Google Scholar

Khatena, 1972 *
J. Khatena
Original verbal images of children as a function of time
Psychological Reports, 31 (1972), pp. 565-566
CrossRef View Record in Scopus Google Scholar

Khatena, 1973 *
J. Khatena
Production of original verbal images by college adults to variable time intervals
Perceptual and Motor Skills, 36 (1973), pp. 1285-1286
CrossRef View Record in Scopus Google Scholar

Konstantopoulos, 2011 S. Konstantopoulos
Fixed effects and variance components estimation in three-level meta-analysis?
Research Synthesis Methods, 2 (2011), pp. 61-76
CrossRef View Record in Scopus Google Scholar

Kuhn and Holling, 2009 J.-T. Kuhn, H. Holling
Measurement invariance of divergent thinking across gender, age, and school forms
European Journal of Psychological Assessment, 25 (2009), pp. 1-7
CrossRef View Record in Scopus Google Scholar

McGrath and Kelly, 1986 J.E. McGrath, J.R. Kelly
Time and human interaction: Toward a social psychologyof time
Guilford Press, New York, NY, USA (1986)
Google Scholar

Mednick, 1962 S.A. Mednick
The associative basis of the creative process
Psychological Review, 69 (1962), pp. 220-232
CrossRef View Record in Scopus Google Scholar

Messick, 1995 S. Messick
Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning
American Psychologist, 50 (1995), pp. 741-749, 10.1037/0003-066X.50.9.741
View Record in Scopus Google Scholar

Moore and Tenney, 2012 D.A. Moore, E.R. Tenney
Time pressure, performance, and productivity
Research on Managing Groups and Teams, 15 (2012), pp. 305-326
View Record in Scopus Google Scholar

Morse et al., 2001 *
D.T. Morse, L.W. Morse, G.A. Johns
Download
Journal of Creative Behavior, 35 (2001), pp. 102-114
CrossRef View Record in Scopus Google Scholar

Nusbaum and Silvia, 2011 E.C. Nusbaum, P.J. Silvia
Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking
Intelligence, 39 (2011), pp. 36-45, 10.1016/j.intell.2010.11.002
Article Download PDF View Record in Scopus Google Scholar

Ohly and Fritz, 2010 S. Ohly, C. Fritz
Work characteristics, challenge appraisal, creativity, and proactive behavior: A multilevel study
Journal of Organizational Behavior, 31 (2010), pp. 543-565, 10.1002/job.633
CrossRef View Record in Scopus Google Scholar

Ohly et al., 2006 S. Ohly, S. Sonnentag, F. Pluntke
Routinization, work characteristics and their relationships with creative and proactive behaviors
Journal of Organizational Behavior, 27 (2006), pp. 257-279, 10.1002/job.376
CrossRef View Record in Scopus Google Scholar

Palmiero et al., 2014 M. Palmiero, D. Di Giacomo, D. Passafiume
Divergent thinking and age-related changes
Creativity Research Journal, 26 (2014), pp. 456-460
CrossRef View Record in Scopus Google Scholar

Plucker et al., 2006 *
J.A. Plucker, M.A. Runco, W. Lim
Predicting ideational behavior from divergent thinking and discretionary time on task
Creativity Research Journal, 18 (2006), pp. 55-63
CrossRef View Record in Scopus Google Scholar

Preckel et al., 2011 *
F. Preckel, C. Wermer, F.M. Spinath
The interrelationship between speeded and unspeeded divergent thinking and reasoning, and the role of mental speed
Intelligence, 39 (2011), pp. 378-388
Article Download PDF View Record in Scopus Google Scholar

Reese et al., 2001 H.W. Reese, L.-J. Lee, S.H. Cohen, Puckett Jr, M J.
Effects of intellectual variables, age, and gender on divergent thinking in adulthood
International Journal of Behavioral Development, 25 (2001), pp. 491-500, 10.1080/01650250042000483
View Record in Scopus Google Scholar

Richardson, 1981 *
M.C. Richardson
***The effects of rewards and timed versus untimed conditions on creativity scores for seventh and eleventh grade students* (unpublished master's thesis)**
Kean College of New Jersey, NJ, United States of America. (1981)
Google Scholar

Rosenthal, 1979 R. Rosenthal
The "file drawer problem" and tolerance for null results
Psychological Bulletin, 5 (1979), pp. 638-641
Article Download PDF CrossRef View Record in Scopus Google Scholar

Runco and Acar, 2012 M.A. Runco, S. Acar
Divergent thinking as an indicator of creative potential
Creativity Research Journal, 24 (2012), pp. 66-75
CrossRef View Record in Scopus Google Scholar

FEEDBACK

Runco, 1986 M.A. Runco
Flexibility and originality in children's divergent thinking
The Journal of Psychology, 120 (1986), pp. 345-352
CrossRef View Record in Scopus Google Scholar

Said-Metwaly et al., 2020 S. Said-Metwaly, B. Fernández-Castilla, E. Kyndt, W. Van den Noortgate
Testing conditions and creative performance: Meta-analyses of the impact of time limits and instructions
Psychology of Aesthetics, Creativity, and the Arts, 14 (2020), pp. 15-38
CrossRef View Record in Scopus Google Scholar

Said-Metwaly et al., 2017 S. Said-Metwaly, W. Van den Noortgate, E. Kyndt
Approaches to measuring creativity: A systematic literature review
Creativity Theories - Research - Applications, 4 (2017), pp. 238-275
CrossRef View Record in Scopus Google Scholar

Sajjadi-Bafghi, 1986 *
S.H. Sajjadi-Bafghi
Fixed and variable time press and sex differences in production of verbal originality
Perceptual and Motor Skills, 63 (1986), pp. 1011-1016
CrossRef Google Scholar

Sajjadi-Bafghi and Khatena, 1985 *
S.H. Sajjadi-Bafghi, J. Khatena
Effects of autonomy of imagery and time press on production of verbal originality
Perceptual and Motor Skills, 61 (1985), pp. 787-791, 10.2466/pms.1985.61.3.787
CrossRef Google Scholar

Schultz and Searleman, 1998 P.W. Schultz, A. Searleman
Personal need for structure, the Einstellung task, and the effects of stress
Personality and Individual Differences, 24 (1998), pp. 305-310, 10.1016/S0191-8869(97)00179-7
Article Download PDF View Record in Scopus Google Scholar

Silvia and Beaty, 2012 P.J. Silvia, R.E. Beaty
Making creative metaphors: The importance of fluid intelligence for creative thought
Intelligence, 40 (2012), pp. 343-351
Article Download PDF View Record in Scopus Google Scholar

Taylor, 1988 C.W. Taylor

Various approaches to and definitions of creativity
R.J. Sternberg (Ed.), The nature of creativity: Contemporary psychological perspectives, Cambridge University Press, New York, NY (1988), pp. 99-121
CrossRef Google Scholar

Torrance, 1969 *
E.P. Torrance
Curiosity of gifted children and performance on timed and untimed tests of creativity
Gifted Child Quarterly, 13 (1969), pp. 155-159
CrossRef View Record in Scopus Google Scholar

Torrance and Ball, 1978 *
E.P. Torrance, O. Ball
Effects of increasing the time limits of the Just suppose test
Journal of Creative Behavior, 12 (1978), p. 281
CrossRef View Record in Scopus Google Scholar

Van Hiel and Mervielde, 2003 A. Van Hiel, I. Mervielde
The measurement of cognitive complexity and its relationship with political extremism
Political Psychology, 24 (2003), pp. 781-801, 10.1046/j.1467-9221.2003.00354.x
View Record in Scopus Google Scholar

Vartanian et al., 2009 O. Vartanian, C. Martindale, J. Matthews
Divergent thinking ability is related to faster relatedness judgements
Psychology of Aesthetics, Creativity, and the Arts, 3 (2009), pp. 99-103
CrossRef View Record in Scopus Google Scholar

Viechtbauer, 2007 W. Viechtbauer
Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis
Zeitschrift für Psychologie/Journal of Psychology, 215 (2007), pp. 104-121
CrossRef View Record in Scopus Google Scholar

Voss, 1977 H.G. Voss
The effect of experimentally induced activation on creativity
Journal of Psychology, 96 (1977), pp. 3-9
CrossRef View Record in Scopus Google Scholar

Wallach and Kogan, 1965 M.A. Wallach, N. Kogan
Modes of thinking in young children

Holt, Rinehart and Winston, New York (1965)
Google Scholar

Ward, 1969a W.C. Ward
Rate and uniqueness in children's creative responding
Child Development, 40 (1969), pp. 869-878, 10.2307/1127195
View Record in Scopus Google Scholar

Ward, 1969b W.C. Ward
Creativity and environmental cues in nursery school children
Developmental Psychology, 1 (1969), pp. 543-547
https://doi.org/10.1037/h0027977
CrossRef View Record in Scopus Google Scholar

Wilson et al., 1995 T.D. Wilson, S.D. Hodges, S.J. LaFleur
Effects of introspecting about reasons: Inferring attitudes from accessible thoughts
Journal of Personality and Social Psychology, 69 (1995), pp. 16-28, 10.1037/0022-3514.69.1.16
CrossRef View Record in Scopus Google Scholar

1 There are two studies (e.g., Khatena, 1972) testing both timed vs. untimed and shorter vs. longer time conditions.

View Abstract