Does the Fourth-Grade Slump in Creativity Actually Exist? A Meta-Analysis of the Development of Divergent Thinking in School-Age Children and Adolescents

Sameh Said-Metwaly [1,2,3], Belén Fernández-Castilla [1,2], Eva Kyndt [3,4], Wim Van den Noortgate [1,2], and Baptiste Barbot [5,6]

[1] KU Leuven (University of Leuven)

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

[3] University of Antwerp

[4] Swinburne University of Technology

[5] UCLouvain (University of Louvain)

[6] Child Study Center, Yale University

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 Social Sciences, University of Antwerp & Centre for the New Workforce, Swinburne University of Technology; Wim Van den Noortgate, Faculty of Psychology and Educational Sciences & imec-itec, KU Leuven (University of Leuven); Baptiste Barbot, Psychological Sciences Research Institute, UCLouvain & Child Study Center, Yale University. This research was supported by a grant from the Ministry of Higher Education and Scientific Research in Egypt to the first author. Correspondence to Baptiste Barbot, Psychological Sciences Research Institute, UCLouvain, Place Cardinal Mercier 10/L3.05.01, 1348 Louvain-la-Neuve, Belgium. Email: baptiste.barbot@uclouvain.be.

Abstract

The development of divergent thinking (DT) in school-age children and adolescents has received considerable attention in the educational psychology literature since the 1970's. A body of research has outlined the existence of slumps (i.e., temporary declines) in this development with, however, conflicting findings concerning the magnitude and timing of these slumps. This study is the first to meta-analyze prior research findings regarding DT development from Grades 1 to 12, with a particular emphasis on the widely controversial fourth-grade slump. A total of 2,139 standardized means from 41 studies involving 40,918 subjects were analyzed using a meta-analytic three-level model. The findings showed an overall upward developmental trend of DT across grade levels, with some discontinuities. Specifically, there was no evidence of a general fourth-grade slump; rather, evidences for a seventh-grade slump were found. Moderator analyses indicated that a fourth-grade slump might be observed depending on DT test, task content domain, intellectual giftedness, and country of study. The existence of the seventh-grade slump was also moderated by DT test, task content domain, and gender. Together, this study deciphers a longstanding debate regarding DT development, a prerequisite knowledge to support age-appropriate educational strategies that encourage creativity development. Implications of these findings for creativity research and practice are discussed.

Keywords: divergent thinking, creativity, development, slumps, meta-analysis

Does the Fourth-Grade Slump in Creativity Actually Exist? A Meta-Analysis of the Development of Divergent Thinking in School-Age Children and Adolescents

Creativity is often described as the ability to generate a novel and useful product which results from an interaction between individuals' aptitude, process, and their environment (Plucker & Beghetto, 2004). Although the field of educational psychology has for decades been at the forefront of research efforts to understand the nature and development of creativity (e.g., Glover, Ronning, & Reynolds, 1989; Lau & Cheung, 2010; Lin & Shih, 2016; Smith & Smith, 2010; Torrance, 1967), enthusiasm for this line of work has then diminished given enduring conceptual and methodological issues (Plucker, Beghetto, & Dow, 2004). More recently, creativity re-surfaces as a key ability for individuals, schools, and societies to keep pace with evolutionary changes and to meet stiff challenges in the modern world (Rubenstein, Callan, & Ridgley, 2018; Runco, 2004; Wong, Chow, Lau, & Gong, 2018). This increased attention to creativity is evident from continued efforts to encourage it in the classroom (Jeffrey & Craft, 2004), such as with worldwide initiatives to focus K–12 education on 21st century skills (Plucker, Kaufman, & Beghetto, 2015). Renewed attention to creativity in education also transpires from the inclusion of *creative thinking* as a new domain for OECD's PISA 2021, which will assess creativity among hundreds of thousand 15-years old students across the world, and may trigger an extensive policy debate on the need to foster creativity in school curricula.

This endeavor is further justified by an important line of educational research examining the contribution of creativity on academic achievement (e.g., Dowd, 1966; Hansenne & Legrand, 2012). A recent meta-analysis of 120 studies reported a small to moderate link between aspects of creativity and academic achievement (Gajda, Karwowski, & Beghetto, 2017). This effect, however, was partly moderated by education stage, with a stronger relationship between creativity and academic achievement in middle school, compared to elementary or high school. This developmental effect calls for the need to better understand creativity as it develops in school-age students (Gajda et al., 2017).

More broadly, a comprehensive view of creativity development in school-age children is needed in order to ultimately inform grade-level appropriate curriculum planning and pedagogical practices aimed at improving creative potential (Torrance, 1977). However, despite decades of research on this issue, there is still a lack of clarity regarding the developmental trajectory of creativity (Barbot, Lubart, & Besançon, 2016; Lau & Cheung, 2010) as well as the factors that contribute to individual differences in this trajectory (Barbot, 2019). This lack of clarity is not surprising given the multidimensional nature of the creativity phenomenon, involving cognitive, personality-related, and environmental factors (Amabile & Mueller, 2008; Guilford, 1966; Said-Metwaly, Kyndt, & Van den Noortgate, 2017a; Sternberg, 2006). Accordingly, developmental trends of creativity in childhood and adolescence may vary dramatically according to which aspect of creativity is accounted for and how it is specifically operationalized (Barbot, Hass, & Reiter-Palmon, 2019).

Until recently, divergent thinking (DT) has been the most common way to operationalize creativity in psychology and educational sciences. Guilford (1950, 1975) introduced the term DT as one mode of creative problem solving. Indeed, in DT, a broad search is employed to generate multiple relevant and original alternative answers in response to a single problem or stimulus (Guilford, 1975). While DT has long been recognized as a central cognitive component of the creative potential, this does not make it strictly equivalent to creativity *per se* (Barbot et al., 2019; Runco & Acar, 2012). In most common operationalization, such as in the classic Wallach-Kogan Creativity Tests (WKCT; Wallach & Kogan, 1965) or the Torrance Tests of Creative Thinking (TTCT; Torrance, 1966), DT tasks involve open-ended problems in different modalities of response (i.e., "task content domain"), including for example, generating questions about a picture, proposing unusual uses for a common object, suggesting ways of improving a product, or completing an incomplete drawing in alternative ways. Subjects' responses are then typically scored for *fluency* (number of adequate and non-redundant responses), *flexibility* (diversity of responses), *originality* (novelty of responses), and *elaboration* (richness of details; e.g., Houtz & Krug, 1995; Reiter-Palmon, Forthmann, & Barbot, 2019; Said-Metwaly, Fernández-Castilla, Kyndt, & Van den Noortgate, 2018). Although the study of DT development faces unique challenges (Barbot, 2019), a long tradition of research in psychology and education has focused on DT development, mainly as a proxy to understand creativity development. While DT doesn't fully represent the creativity phenomenon, this operationalization represents a reasonable predictor of successful creative performance (Guilford, 1975), which has been widely established empirically (Acar & Runco, 2019).

DT Development and the Fourth-Grade Slump

One could expect that children and adolescents increase their DT abilities as they get older because, beyond cognitive maturation, their educational and social experiences become richer (Domínguez, Díaz-Pereira, & Martínez-Vidal, 2015; Klausmeier & Wiersma, 1964; Lau & Cheung, 2010). However, empirical findings regarding DT development have been equivocal. Some studies have provided evidence for a continuous upward trajectory of DT across grade levels (e.g., Hong & Milgram, 2010; Lopez, Esquivel, & Houtz, 1993; Sak & Maker, 2006). Other studies have reported that DT follows an irregular developmental trajectory, including significant drops in one or more periods along this trajectory (e.g., Camp, 1994; Charles & Runco, 2001; Kim, 2011). The most popular of these drops takes place in Grade 4, at around age 9, as first identified by Torrance (1967, 1968) across seven cultures (including U.S., Australia, and India), and commonly coined the "fourth-grade slump".

This slump phenomenon has fueled and enthusiastic line of research in educational psychology, and a number of subsequent studies have confirmed the existence of the fourth-grade slump. For instance, Fishkin (1989) followed-up 45 students in Grades 2 to 5 and showed a significant decline in DT scores in Grade 4 when compared to Grades 2 and 3. Similarly, studies by Lubart and Georgsdottir (2004) and Darvishi and Pakdaman (2012) indicated that DT declined temporarily in Grade 4. However, other studies have revealed an increase or no decline in Grade 4. For example, in a sample of 841 students, Sak and Maker (2006) observed that DT scores showed a progressive increase from Grades 1 through 5, whereas Lin and Shih (2016) found that the DT of students in Grade 4 did not differ significantly from those of Grades 3 and 5.

Other studies have found a slump in other grades. Long and Henderson (1965) examined the grade level differences in DT among 288 students and showed that DT increased from Grade 2 to Grade 4, dropped in Grade 5, and rose again in Grades 6 and 7. A significant decline in DT in Grade 5 was also reported by Charles and Runco (2001) and Besançon and Lubart (2008). In contrast, Kim, Cramond, and Bandalos (2006) and Kim (2011) identified a slump in DT in Grade 6. Yet, in another study conducted by Sherwood and Strahan (1985) on a sample of 296 gifted students from Grades 4 to 8, DT reached its peak in Grade 6 and then dropped in Grade 7. A slump in DT in Grade 7 was also observed in other studies (e.g., Jastrzębska & Limont 2017; Lau & Cheung, 2010). Further, in a longitudinal study of 89 students from Grades 1 to 12, Camp (1994) reported a decline in DT in Grade 9. Similarly, a study comparing the DT of 1,367 students from Grades 3 to 11 (Hu, Shi, Han, Wang, & Adey, 2010) showed that DT increased starting from Grade 3 until it peaked in Grade 8, and then decreased dramatically in Grades 9 through 11. In sum, the extant literature on DT development is characterized by inconsistencies regarding the existence and timing of slumps, which calls for comprehensive integration of past research focused on DT slumps in school-age children and adolescents.

Possible Causes of Slumps in DT Development

Regardless of the variability in study findings with respect to the existence and timing of DT slumps, several explanations for this phenomenon have been proposed. The first line of explanation, historically proposed by Torrance (1967), is environmental: slumps may occur at times of school transitions (e.g., when children move from primary to secondary), a time when children and adolescents have to cope with the stress and demands imposed by the new school-related system they have transited into. This adjustment would exacerbate a pressure to conformity in addition to an increased awareness of social rules and need for acceptance, which would implicitly prompt children to provide appropriate responses that meet social norms or expectations, rather than providing unique or unusual ones (e.g., Gralewski, Lebuda, Gajda, Jankowska, & Wiśniewska, 2016; Kim, 2011). While this explanation is tenable, many episodes of slumps are also observed in grades that are not associated with school transition (Barbot, Besançon, & Lubart, 2015).

Another line of explanation for these slumps is cognitive: a temporary decline in DT might emerge when other facets of cognitive development are at their peak (Lubart & Georgsdottir, 2004; Rieben, 1978). Lubart and colleagues showed that DT temporarily decreases at the age of 9 to 10, when logical thinking becomes fully functional (Lubart & Georgsdottir 2004; Lubart & Lautrey, 1995). In the same vein, Charles and Runco (2001) indicated that decreases in DT would go along with increases in evaluative skills and a preference for appropriate ideas. These findings are consistent with Karmiloff-Smith's (1994) perspective on cognitive development, suggesting that acquiring new cognitive skills during the course of behavioral mastery could result in temporary decreases in related areas of cognitive development, followed by a recovery once the new skills have been consolidated. Finally, other researchers have proposed additional factors that may account for DT slumps at key developmental stages during the school-age years (e.g., puberty), such as the contribution of hormonal shifts (Karwowski & Lebuda, 2014), changes in brain structures (Barbot & Tinio, 2015; Gardner, 1982; Nelson & Guyer, 2011) or identity development (Barbot & Heuser, 2017).

Possible Moderator Variables

Inconsistencies in DT development across studies might also be explained by certain moderator variables. A review of extant literature suggests mainly five moderator variables that were examined in the present work, namely DT test, task content domain, gender, country of study, and intellectual giftedness.

DT test. Various tests have been used to measure DT. Guilford (1984) hypothesized that an individual's performance on DT tests is subject to variations depending on content and product in which DT applies. Consistently, prior studies have identified distinct developmental trends for different DT tests (e.g., Kleibeuker, De Dreu, & Crone, 2013; Lin & Shih, 2016). Hence, different DT tests might tap into distinct aspects of this construct and could account for conflicting research findings on DT development (Barbot, 2019; Barbot, Lubart, & Besançon, 2016; Said-Metwaly, Kyndt, & Van den Noortgate, 2017a, 2017b).

Task content domain. More broadly, the inconsistencies found in developmental studies of DT might also stem from the different content domains of tasks (i.e., modality of response) being employed (Barbot, Lubart, & Besançon, 2016; Domínguez et al., 2015). Similar to other facets of cognitive functioning such as curiosity which is sparking growing interest in the field of educational sciences (e.g., Alexander, 2019; Peterson & Cohen, 2019; Shin & Kim, 2019), it is increasingly acknowledged that creativity is partly a domain-specific entity. Indeed, DT performance is generally inconsistent across different modalities of response (e.g., verbal vs. figural), with only low to moderate correlations between scores obtained from tasks sampled across different content domains (Baer, 1998). Consistent with these behavioral observations, it has been evidenced that different content domains of DT recruit distinct brain regions (Gonen-Yaacovi et al., 2013), supposing dissimilar developmental trajectories. Accordingly, previous research has found differences in the developmental trend of DT according to the content domain of tasks used (e.g., Besançon & Lubart, 2008; Domínguez et al., 2015; Torrance, 1967).

Gender. Previous research has suggested that DT follows different developmental trends according to gender (e.g., Alfonso-Benlliure & Santos, 2016; Lau & Cheung, 2010). These differences might be due to gender differences in socio-environmental factors including gendered socialization practices, expectations, and educational opportunities (Baer & Kaufman, 2008). Relatedly, they might be explained by gender differences in terms of brain regions recruited during DT, and corresponding trajectories of brain development throughout childhood and adolescence (Abraham, Thybusch, Pieritz, & Hermann, 2014).

Country of study. Theoretically, variations between countries in social and cultural dimensions such as individualism vs. collectivism and autonomy vs. conformity orientations might influence the development of creative abilities such as DT (Amabile, 1983; Lubart & Georgsdottir, 2004). Consistently, few studies have reported cross-country differences in the developmental trend of DT (e.g., Jaquish & Ripple, 1984; Torrance, 1967). However, there are only limited empirical evidences examining specifically the impact of culture on DT development (Yi, Hu, Plucker, & McWilliams, 2013).

Intellectual giftedness. A large body of developmental studies of DT comes from the field of intellectual giftedness and education. This tradition stems from the fact that creativity and intelligence are considered independent in educational settings (Silvia, 2015). While only a few studies have compared developmental trends among intellectually gifted vs. non-gifted students and showed differences between groups (e.g., Lopez et al., 1993; Rosenfield & Houtz, 1978), other studies have reported no differences (e.g., Guignard & Lubart, 2007). Beyond difference in

IQ (which may partly account for dissimilar developmental trends in DT among gifted and non-gifted children; Guignard, Kermarrec, & Tordjman, 2016), gifted children are generally found to maintain superior intrinsic motivation from early childhood to adolescence (Fleith, 2016; Gottfried & Gottfried, 1996), a dimension known to affect DT development (Amabile & Mueller, 2008; Csikszentmihalyi, 1988; Hennessey, 1995; Sternberg, 2006). Therefore, inconsistent results regarding DT development might also be relative to the use of gifted vs. non-gifted participants across studies.

The Present Study

Research findings regarding DT development in school-age children and adolescents have been contradictory, with some findings suggesting a consistently increasing trend, and others indicating temporary slumps, in particular around the fourth grade. In addition to the possible influences of moderator variables outlined above, a possible reason for these mixed results is that most studies in this line of work have been carried out on small sample sizes (Barbot, Lubart, & Besançon, 2016; Lau & Cheung, 2010). This raises considerable concerns about the robustness of the findings of these individual studies, to the point that no definite conclusions on DT development can be made solely on the basis of a single study. There are indeed still recurring debates regarding whether slumps in DT do actually exist, how many they are, when they take place, and what factors explain these slumps (Barbot, Lubart, & Besançon, 2016; Gralewski et al., 2016).

This study aims to contribute to longstanding debates in this line of research through undertaking a meta-analysis of studies that addressed DT development in school-age children and adolescents (Grade 1 to Grade 12), with a focus on Grade 4, as it has sparked most controversies. Specifically, using the standardized mean (SM) level of DT as the outcome metric for the meta-analysis, this study addressed the following research questions: (1) How does the SM of DT vary across Grade 1 to Grade 12? (2) Is there a significantly lower SM of DT in Grade 4 compared to Grade 3 (i.e., a slump)? (3) Are there any moderator variables that explain the observed variability across studies regarding the SM differences in DT from Grade 3 to Grade 4? By answering these questions, this study sought to provide summary-level evidence that can establish whether there are indeed discontinuities in DT development across grades, to illuminate recurring inconsistencies in the literature on DT development. Further, with a large sample size accumulated across multiple studies, this meta-analysis offers many advantages over individual studies such as maximizing statistical power, obtaining more accurate effect estimates, enhancing generalization of findings, answering research questions not addressed in the primary studies, testing for moderator variables, and proposing hypotheses for future examination (Deeks, Higgins, & Altman, 2008; Egger & Smith, 1997; Haidich, 2010; Littell, Corcoran, & Pillai, 2008; Walker, Hernandez, & Kattan, 2008).

Method

Selection of Studies

The papers included in this meta-analysis were identified by systematically searching the creativity literature published up to April 30th, 2018, following four steps. First, searches of the following databases were conducted: ERIC, Google Scholar, JSTOR, PsycARTICLES, Scopus, and Web of Science. The following strings were used to search titles, abstracts, and keywords in the selected databases: (“divergent thinking” OR “divergent ability” OR “divergent production” OR “divergent performance” OR “creativity” OR “creative thinking” OR “creative ability” OR “creative potential”) AND (“development*” OR “growth” OR “longitudinal” OR “slump” OR “drop” OR “school” OR “grade” OR “children” OR “adolescents”). Such specific terms (over broader terms like “student*”) were chosen to avoid prolonged reviews resulting from broader queries that may lead to increased error-prone manual work (e.g., Soilemezi & Linceviciute, 2018). Second, backward searches were undertaken through screening the reference lists of the papers found in the first step for additional relevant papers. Third, forward searches were carried out by looking up papers that subsequently cited the previously identified papers in the previous two steps. Fourth, the following leading journals in creativity research were hand-searched: *Creativity Research Journal*; *Psychology of Aesthetics, Creativity, and the Arts*; *The Journal of Creative Behavior*; and *Thinking Skills and Creativity*.

The papers resulted from the search process were first checked for relevance based on a review of their titles and abstracts removing those that were clearly not relevant to the research topic. In such cases where the title and abstract were insufficient to decide on the paper's relevance, a full-text screening was undertaken to enable further evaluation. The papers deemed potentially relevant were read in detail, and were selected for inclusion if all the following criteria were met: the paper had to (1) report on an original quantitative research (qualitative and review papers were eliminated), (2) investigate differences in DT between Grade 4 and other grades (1 to 12), either cross-sectionally (comparing students of different grades) or longitudinally (comparing the same students across grades), and (3) report the statistics needed to calculate the SM and the corresponding sampling variance (i.e., mean and standard deviation). Further, we only included (4) journal articles, research reports, conference papers, or dissertations (as long as they were not also available as journal articles) that (5) were published in English (to properly interpret the findings reported), and for which (6) the full text was accessible.

In case of papers without accessible full texts or sufficient data, attempts were made to obtain them from the author(s). When a study reported analyses on more than one subsample of participants (e.g., analyses by gender, country, or intellectual giftedness), these subsamples' data were treated as independent and the data for the whole sample were excluded. In the case of studies that reported analyses on *overall* DT and also on its indicators, only indicators-level data were included in order to avoid redundancy. In cases where multiple studies reported data for the same sample, only the one with largest sample size or specificity of results (by subsample or DT indicator) was included. For studies that reported repeated measurements for a grade during the same school year, only the data from the first measurement occasion were included.

Coding of Studies

During the coding process, information regarding the year of publication, type of publication (journal article, conference paper, or dissertation), and the potential moderator variables were recorded for each of the eligible studies. The coding process was as follows: First, DT test was included as a categorical variable indicating the test used to measure DT in each study. Second, task content domain was included as a categorical variable reflecting the domain of the tasks utilized for measuring DT. Two categories for content domain were employed: verbal (generating responses to verbal prompts) and figural (generating responses to non-verbal prompts or drawing something such as figures). Third, the proportion of males in each sample was included as a continuous variable. Fourth, the country where the study was conducted was coded as a categorical variable. Fifth, intellectual giftedness was included as a categorical variable indicating whether the sample of each study was gifted or non-gifted.

Consistent with common practice within the field of research synthesis (e.g., Allen et al., 2016; Sailer & Homner, 2019), a random sample of 20% of the studies was independently coded by both the first and the second authors. The rest of the studies (80% of the corpus) was coded by the first author. Inter-coder agreement analyses showed that the percentage of agreement varied between 90.91% and 100% according to the coded variables. Cohen's kappa ranged between .85 and 1.00, suggesting almost perfect agreement (Landis & Koch, 1977). Disagreements between both coders were resolved through subsequent discussion, resulting in a satisfactory estimated coding reliability for the entire corpus (i.e., above the minimum 90.91% observed before resolving disagreements).

Analyses

To conduct a meta-analysis, the data of included studies need to be converted into a common metric. Since the instruments used to measure DT differed from study to study, each grade-specific mean value was standardized (SM) by diving it by the pooled standard deviation across grades, for each outcome and each study separately. The resulting SMs and their sampling variances were then incorporated across all the studies using a random effects model. The SMs were weighted by the inverse of their sampling variances, meaning that greater weight was given to more precise SMs in the analyses. Given that most studies reported more than one mean, using traditional random effects models that assume independence among these study outcomes might yield flawed inferences (Becker, 2000; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013). Therefore, a meta-analytic three-level model was employed in order to model the dependence within studies (Van den Noortgate et al., 2013). This model distinguishes three types of variance: variance between studies (σ_V^2), variance between grade-specific mean outcomes within the same study (σ_U^2), and sampling variance (σ_E^2; Van den Noortgate et al., 2013). Accordingly, this model informs differences in outcomes within studies as well as differences between studies and allows for testing potential moderator effects at the study and outcome level (Van den Noortgate et al., 2013). A likelihood ratio test was conducted to investigate whether the heterogeneity between or within studies was substantial. A significant result of the likelihood ratio test indicates that the SMs are heterogeneous, and thus calls for moderator analyses to explain this heterogeneity (Van den Bussche, Van den Noortgate, & Reynvoet, 2009).

To explore the mean differences in DT from Grade 1 to Grade 12, 11 dummy variables (i.e., number of grades considered - 1) were included as predictor variables in the model. The first dummy variable takes a value of 0 for Grade 1, and 1 otherwise. The second dummy variable takes a value of 0 for Grade 1 and Grade 2, and 1 otherwise. The remaining dummy variables were coded following the same pattern; the 11th dummy variable therefore is coded as 1 for Grade 12, and 0 otherwise. Thus, the coefficient of the first dummy variable captures the change in DT from Grade 1 to Grade 2 (i.e., the standardized mean difference (SMD) between Grade 1 and Grade 2), the coefficient of the second dummy variable captures the change in DT from Grade 2 to Grade 3, and so on. The model's intercept as well as the dummy variable capturing the change in DT from Grade 3 to Grade 4 were allowed to randomly vary across the study and outcome levels. To avoid an overly complex model and given the focus of the present meta-analysis on the fourth-grade slump, the effects of the moderator variables were tested for the difference in DT scores between Grade 3 to Grade 4. To estimate the effect of each moderator variable, an additional term was included in the model representing the interaction between the dummy variable standing for the change in DT from Grade 3 to Grade 4 and the proposed moderator variable. A separate regression equation was fitted to each moderator variable.

To check for the presence of outliers, a sensitivity analysis was conducted by leaving out extreme SMs (deviating at least 2.5 standard deviations from the mean) one by one and calculating the resulting SM. Finally, exploring potential publication biases is critical in meta-analyses in psychology and education, especially given that they are often used to make high-stake decisions in policy and practice (Chow & Ekholm, 2018). The existence of publication bias was investigated using the visual inspection of the symmetry of the funnel plots (Light & Pillemer, 1984). This provides a general idea of potential publication bias, yet it does not account for dependent data within studies. To account for such dependency, a three-level extension of

Egger’s regression test (Egger, Smith, Schneider, & Minder, 1997) was used. This test explores, through a three-level approach, whether a significant relationship exists between the SMDs and their standard errors (Fernández-Castilla et al., 2019). Statistical analyses were carried out using SAS software version 9.4 (SAS Institute, Cary, NC).

Results

Papers Meeting the Inclusion Criteria

The search process resulted in a corpus of 2,742 papers which was narrowed down to 107 pre-screened papers based on an initial review of the titles and abstracts. Of these, a total of 41 papers were eligible for inclusion in the present meta-analysis (see Figure 1). This final corpus consisted of 34 journal articles, five dissertations, one research report, and one conference paper, involving a total of 40,918 subjects (see Supplementary Table S1 for the list of the included studies). These studies were published between 1967 and 2017, including 19 (46.34%) involving data from U.S. samples, 11 (26.83%) from Asia, eight (19.51%) from Europe, one (2.44%) from Africa, and two (4.88%) from more than one continent. Fifteen (36.59%) studies measured DT using the TTCT, five (12.20%) used the WKCT, and the remaining studies used other tests such as the Creativity Assessment Packet and the Tel Aviv Creativity Test. Fourteen (34.14%) studies used verbal tasks, 12 (29.27%) used figural tasks, and 12 (29.27%) used both. The remaining three (7.32%) studies used tasks in other content domains (e.g., mathematical or musical). From these 41 studies, 2,139 SMs were calculated; many individual studies provided more than one SM for the same grade from multiple DT indicators or from independent groups. The distribution of the SMs among grades was as follows: Grade 1 (n = 168), Grade 2 (n = 186), Grade 3 (n = 290), Grade 4 (n = 506), Grade 5 (n = 392), Grade 6 (n = 338), Grade 7 (n = 65), Grade 8 (n = 67), Grade 9 (n = 55), Grade 10 (n = 26), Grade 11 (n = 26), and Grade 12 (n = 20).

Meta-Analyses of Overall DT

Table 1 summarizes the results of the meta-analysis and Figure 2 shows the resulting average developmental trends of DT (*overall* and for each DT indicator) by grade with corresponding standard errors. As illustrated, the initial *overall* three-level analysis incorporating all the SMDs revealed an irregular DT development with grade level. In essence, DT was significantly higher at each successive grade level, and reached its peak in Grade 9. No overall drop in Grade 4 was observed; rather, there was a plateau (i.e., non-significant difference between Grade 3 and Grade 4). Further, significant drops were obtained for Grades 7, 10, and 11.

Meta-Analyses of DT Indicators

As represented in Figure 2, the three DT indicators (f*luency*, *flexibility*, and *originality*) did appear to follow a developmental trend similar to the *overall* DT trend. A drop in *originality* was observed in Grade 3, whereas *fluency* and *flexibility* scores were significantly higher than those obtained in Grade 2. The plateau in Grade 4 and the drop in Grade 7 described above for the *overall* analysis were also observed at the DT indicator level. Moreover, the three indicators showed a plateau in Grade 9. In Grade 10, *flexibility* and *originality* were significantly higher than those in Grade 9, but *fluency* was comparable in both grade levels. A significant drop in *fluency* appeared in Grade 11, but *flexibility* and *originality* scores were not significantly different than those in Grade 10. In grade 12, *originality* was significantly higher than that in Grade 11, but *flexibility* was not. There was no data available to analyze the differences in *fluency* in Grade 12.

Heterogeneity Analysis

As presented in Table 1, heterogeneity analyses showed that 46.43% of the total variance observed in the SMDs of *overall* DT between Grades 3 and 4 was systematic variance between studies, 50% systematic variance between grade-specific *overall* DT means within the same study, and 3.57% sampling variance. The likelihood ratio test indicated that both systematic variances were significant (χ^2 = 16,073.4 and 427,162.2, respectively, $df = 1$, $p < .001$). At the indicator level, the likelihood ratio test showed that significant variance was present for *fluency*, *flexibility*, and *originality* at both the between-study level (χ^2 = 130.4, 96.8, and 351.0, respectively, $df = 1$, $p < .001$) and within-study level (χ^2 = 14,620.5, 17,809.0, and 12,491.0, respectively, $df = 1$, $p < .001$). Together, these results suggest that moderator analyses were warranted.

Sensitivity Analyses

After removing one outlier SMD (belonging to Besançon & Lubart, 2008), the obtained SMD for *overall* DT between Grade 3 and Grade 4 was 0.07 ($p = .38$). For *originality*, removing one potential outlier (Besançon & Lubart, 2008) yielded a SMD of - 0.10 ($p = .41$). No outliers were identified with respect to *fluency* or *flexibility*. In addition, the extent to which the results are affected by removing the studies of Torrance (1967) and Besançon and Lubart (2008) was examined as they yielded a greater number of SMs. Removing Torrance's (1967) study yielded a SMD of 0.07 ($p = .41$), 0.03 ($p = .87$), 0.03 ($p = .89$), and -0.07 ($p = .58$) for the *overall*, *fluency*, *flexibility*, and *originality* analyses, respectively. Removing the whole Besançon and Lubart's (2008) study yielded a SMD of 0.06 ($p = .50$), 0.007 ($p = .96$), 0.03 ($p = .84$), and -0.11 ($p = .43$) for the *overall*, *fluency*, *flexibility*, and *originality* analyses, respectively. Therefore, the adjusted SMDs remain comparable to those with all SMDs incorporated, indicating that the results obtained are robust and not significantly influenced by any particular study.

Publication Bias

Figure 3 displays the funnel plots for the *overall* and indicator level analyses. A visual examination of these plots shows that the distributions of the SMDs are relatively symmetrical around their means. Yet, there are some data points on the lower left portion of the *overall* and *originality* plots with no counterparts on the opposite side. On the contrary, the lower left portion of *fluency* and *flexibility* plots has some missing data points. The Egger's test showed that publication bias was significant for the *overall* analysis (t = -2.33, df = 140, p = .02), but not for *fluency* (t = -0.91, df = 39.10, p = .37), *flexibility* (t = -1.48, df = 36.90, p = .15), or *originality* (t = -0.57, df = 40.50, p = .57) taken separately.

Moderator Analyses

The effect of each of the five candidate moderator variables on the differences in DT between Grade 3 and Grade 4 was investigated at the *overall* and indicator level, to further examine under which condition a fourth-grade slump may be observed. To examine the DT test moderator, the studies were split into three categories (TTCT, WKCT, and others; because of insufficient data for specific other tests). In addition, the studies were divided according to country into eastern (including Asian countries), middle-eastern (including Arab countries), and western (including the U.S., European countries, and Australia) countries[1].

Of the five moderator variables, only task content domain was significant for the *overall* analysis (Table 2). The obtained SMD was positive when using figural tasks (0.10) and negative when using verbal tasks (-0.02). However, both SMDs were not significantly different from zero (p = .36 and .84 for figural and verbal content domains, respectively), suggesting a plateau from Grade 3 to Grade 4. Task content domain explained 19.61% of the between-study variance and 56.30% of the within-study variance for the *overall* analysis. At the indicator level, intellectual giftedness was a significant moderator variable for *fluency*. A positive SMD was obtained for gifted subjects (0.46) compared to a negative value for non-gifted subjects (-0.20). Yet, both values were not significantly different from zero (p = .10 and .19 for gifted and non-gifted subjects, respectively), suggesting a plateau in *fluency* from Grade 3 to Grade 4. Intellectual

[1] The middle eastern category was only included in the *overall* analysis due to insufficient data for this category at the indicator level.

giftedness explained 68.96% of the between-study variance and 43.71% of the within-study variance for *fluency*. DT test and country of study were also found to be significant moderator variables for *originality*. For DT test, a negative SMD was obtained for WKCT and *others* compared to a positive SMD for TTCT. Yet, the SMD was significant for WKCT (-0.54, p = .02), but not for TTCT (0.09, p = .49) or *others* (-0.23, p = .22). DT test explained 74.93% of the between-study variance and 52.42% of the within-study variance for *originality*. For country, studies involving subjects from eastern countries had a significant negative SMD (-0.29, p = .02), whereas those involving subjects from western countries had a non-significant positive SMD (0.05, p = .62). Country of study explained 72.50% of the between-study variance and 42.43% of the within-study variance for *originality*.

Given the slump identified in Grade 7, post-hoc analyses were conducted to test the potential influence of the moderator variables on the differences in DT between Grade 6 and Grade 7. DT test, task content domain, and gender were found to be significant for the *overall* analysis (see Table 2). For DT test, a negative SMD was found for TTCT and WKCT compared to a positive value for *other*s. However, the SMD was significant for WKCT (-0.39, p = .02), but not for TTCT (-0.26, p = .12) or *others* (0.08, p = .47). DT test explained 65.35% of the between-study variance and 49.68% of the within-study variance for the *overal*l analysis. For task content domain, a significant negative SMD was obtained when using figural tasks (-0.30, p = .003) compared to a non-significant negative SMD when using verbal tasks (-0.14, p = .15). Task content domain explained 65.61% of the between-study variance and 47.22% of the within-study variance for the *overall* analysis. For gender, a larger proportion of males was associated with a larger negative SMD (B = -0.002, p = .02). Gender explained 28.61% of the between-study variance and 59.51% of the within-study variance for the *overall* analysis. At the indicator level, task content domain was significant for *fluency* and *flexibility*. Specifically, a larger negative

SMD was found for figural tasks (-0.65, $p < .001$ for *fluency* and -0.77, $p < .001$ for *flexibility*) than for verbal tasks (-0.19, $p = .046$ for *fluency* and -0.28, $p = .006$ for *flexibility*). Task content domain explained 100% and 80.03% of the between-study variance and 41.33% and 69.31% of the within-study variance for *fluency* and *flexibility*, respectively. Gender was significantly moderating the results relative to *fluency*, *flexibility*, and *originality*. A larger proportion of males was associated with a greater negative SMD (B = -0.002, $p < .001$ for *fluency*, B = -0.003, $p = .02$ for *flexibility*, and B = -0.002, $p < .001$ for *originality*). Gender explained 100%, 91.59%, and 100% of the within-study variance for *fluency*, *flexibility*, and *originality*, respectively. The other moderator variables investigated did not yield statistically significant effects.

Discussion

The aim of the present study was to decipher a controversial line of empirical findings regarding DT development in children and adolescents from Grade 1 to Grade 12, with a focus on the fourth-grade slump (Torrance, 1967, 1968) that has historically stimulated this line of work. To address this question, this study was the first to use a three-level meta-analytic technique which involved 41 eligible studies and a total of 40,918 participants, allowing to derive findings that could not be obtained with such robustness by any individual study to date. Together, this effort showed that (1) although DT rose progressively from Grade 1 to Grade 12, irregularities in this trend were also observed. Further, and of particular importance for our main research purpose, (2) no evidence for a DT slump in Grade 4 was found at either the *overall* or indicator level, suggesting instead an average, temporary plateau. However, (3) this effect was mitigated by a number of relevant moderator variables identified from suggestions in past literature on DT development. Finally, (4) a DT slump in Grade 7 was observed at the *overall* as well as indicator-level.

Developmental Trends and the Fourth-Grade Slump

The overall upward trend in DT observed in this meta-analysis is consistent with general conclusions of studies focused on school-age children and adolescents (McCrae, Arenberg, & Costa, 1987). However, this study goes further by explaining how selected factors may explain whether episodes of irregularity in DT development are observed or not across studies. Specifically, moderator analyses revealed that only task content domain significantly moderated the presence of a slump in Grade 4 for the *overall* DT. Although there was not significant difference in DT between Grade 3 and Grade 4 across content domains, the SMD obtained for figural tasks was positive whereas it was negative for verbal tasks. This finding is consistent with an increasing body of research pointing to the content domain and task specificity of creativity and its development (Baer, 1998; Barbot, 2019; Barbot, Besançon, & Lubart, 2016; Runco & Albert, 1985). This line of work is supported by neuroscience of creativity studies establishing domain-specificity (e.g., Gonen-Yaacovi et al., 2013), which could easily extend to neurodevelopmental findings (e.g., Paterson, Heim, Friedman, Choudhury, & Benasich, 2006) and explain a differential DT development according to different content domains of production. Future research is warranted to explore more closely the domain- and task-specificity of DT as it develops, by examining competing causes of domain- and task-specific DT development (e.g., whether tied to neurological or educational underpinning).

A focus on moderator analyses at the DT indicator level showed that multiple relevant factors moderated the presence of the observed fourth-grade slump, according to the DT-indicator considered. In other words, it revealed that some key factors interact distinctly with the various facet of DT as they develop. In keeping with the issue of domain- and task-specificity of DT development outlined above for the *overall* analysis, the present work revealed that the specific DT test used across studies was a significant moderator of the observed fourth-grade slump in *originality*: a drop was found when using the WKCT, while a plateau was obtained when using the TTCT or other tests. This finding is contradictory with the classic, yet poorly replicated study by Torrance (1967) with the TTCT that has initially stimulated this line of work. In addition to the hypotheses regarding the domain- and task-specificity of DT development outlined above, it is also possible that the aspects of DT captured by the WKCT are typically less emphasized in the fourth-grade curriculum compared to those measured by other tests. Beyond content domains, test-specific differences might also be attributed to differences in testing procedures. WKCT is administered in a game-like context, while other DT tests are generally administered in a test-like context. Such variations, together with other differences regarding instructions or time on tasks, have shown to impact the *originality* of DT production (Forthmann et al., 2016; Forthmann, Lips, Szardenings, Scharfen, & Holling, in press; Gerlach, Schutz, Baker, & Mazer, 1964; Hattie, 1980; Said-Metwaly, Fernández-Castilla, Kyndt, & Van den Noortgate, 2019; Wallach & Kogan, 1965) and could account for the observed variations in developmental trends.

Country of study was also a significant moderator variable of *originality*: while a fourth-grade drop was found for subjects from eastern countries, such trend was not observed for subjects from western countries. This finding aligns with the hypothesis that the development of original ideas is likely to be lower in social contexts emphasizing collectivism and conformity and higher in those emphasizing individualism and autonomy (Mainemelis, 2010). Given that collectivism is more dominant in eastern countries as opposed to individualism in western countries (Brewer & Chen, 2007), eastern subjects might be more likely than western subjects to experience a slump in original thinking. Relatedly, different countries may be related to different school curricula and values weighted on academic achievement, which could all account for the moderating effect of country of study on DT *originality* observed here.

Regarding *fluency*, intellectual giftedness was found to be a significant moderator variable of the fourth-grade slump, which is consistent with numerous studies showing a differential DT development for gifted vs. non-gifted subjects (e.g., Dai, 2019; Guignard et al., 2016; Hopp, Zhang, Hinch, O'Reilly, & Ziegler, 2019). In the present meta-analysis, there was on average a third to fourth-grade plateau in *fluency* when considering both gifted and non-gifted subjects. However, the corresponding SMD was positive for gifted subjects and negative for non-gifted subjects. That is, while gifted subjects tended to increase *fluency*, non-gifted subjects tended to decline from Grade 3 to Grade 4. In addition to the potential effect of intellectual precocity in DT development (Guignard et al., 2016), this finding might be linked to the differences between gifted and non-gifted students in intrinsic motivation. Indeed, Gottfried and Gottfried (1996) reported that gifted students at ages 9 through 13 years exhibit superior intrinsic motivation compared with non-gifted students. Considering the fundamental role of intrinsic motivation inDT (Amabile & Mueller, 2008; Hennessey, 1995; Sternberg, 2006), it is possible that, compared to non-gifted students, gifted-students overcome the slump by a form of motivational compensation, a mechanism which has been conceptualized in contemporary models of creativity (e.g., Lubart, 2001).

Overall, the focus on the fourth-grade slump confirmed the contribution of multiple factors (i.e., moderator variables) that interact differentially with the DT indicators. While these effects help to formulate important developmental hypotheses, further investigations are also needed to address why these moderator variables may not impact consistently DT developmental patterns across all grades, and why they may be particularly salient in the transition from Grade 3 to Grade 4. In an attempt to preliminary address this important question, the present work has also investigated the effect of the selected moderator variables on the average seventh-grade slump identified herein.

Beyond the Fourth-Grade Slump

The post-hoc analyses investigating the seventh-grade slump give an opportunity to gauge whether other slumps observed in school-age children and adolescents resemble the fourth-grade slump, including regarding the role of the moderator variables considered here. First, as reflected in the present meta-analysis, the seventh-grade slump has been observed in several studies (e.g., Jastrzębska & Limont 2017; Lau & Cheung, 2010) and a number of hypotheses regarding the reasons for this slump have been proposed. These reasons include the possible effect of school transitions during that period (Lau & Cheung, 2010), a hypothesis that was initially proposed by Torrance (1967). Although transition to middle-school happens at different grade in different countries (Barbot et al., 2015), and even within the same country like in the U.S., we cannot rule out that this factor may have contributed to the observed results (notably given the large proportion of studies from the U.S.). Other reasons include neurobiological changes associated with puberty including the dramatic maturation of the prefrontal cortex which impacts higher cognitive functions such as DT (Barbot & Tinio, 2015), or similarly, the onset of the formal operational stage, marked by the emergence of hypothetical reasoning and abstract thinking. As suggested by pioneer (Rieben, 1978) and more contemporary work (Lubart & Georgsdottir, 2004) in reference to the fourth-grade slump, DT development might be in a slump when other facets of the cognitive development are in a peak, and this process could be at play too during other critical periods of a slump in DT development.

Moderator analyses focused on the seventh-grade slump indicated that DT test, task content domain, and gender were significant moderator variables for the *overall* DT analysis. A slump was likely more observed when the study involved the WKCT, figural tasks, and a larger proportion of males. These effects were confirmed at the indicator level for *fluency* and *flexibility*, as moderated by task content domain, and for all three DT indicators as moderated by gender (with again, a higher proportion of males associated with greater seventh-grade slump across all DT indicators). In other words, the slump in Grade 7 appears more evident for males than for females. The effect of domain- and task-specific factors on DT development seems therefore consistent across the fourth- and seventh-grade slumps. However, the moderating role of gender on the seventh-grade slump appears particularly robust as it is observed across DT indicators, while gender was not a significant moderator in the fourth grade. Gender differences in the susceptibility to slumps in Grade 7 might be attributable to an increasingly salient gendered socialization (Hill & Lynch, 1983) that impacts creative performance (Baer & Kaufman, 2008), which also coincides with gendered differences in brain activity patterns and corresponding DT developmental trajectories (Abraham et al., 2014).

Concerning the other moderator variables, the seventh-grade slump analyses suggested that, contrary to the fourth-grade slump, culture and intellectual giftedness were not significant moderator variables of DT development during that period. Together, these findings suggest that some factors consistently moderate DT developmental trajectories (such as those relative to domain- and task-specificity), while other factors are more specific to the developmental period considered (e.g., gender, culture, or intellectual giftedness may interact with DT development during periods that are more “sensitive” to these factors).

Implications, Limitations, and Future Research

This study contributes to the existing body of the literature on DT development in many ways. As the first meta-analysis on a research topic which has proved largely inconsistent, this study represents a significant step toward a better understanding of DT as it develops in school-aged children and adolescents and adds new insights to the ongoing debate over DT slumps in the literature. Another strength lies in the scope of this study. This meta-analysis covered a wide range of grades (from Grades 1 to 12) compared to most studies in this line of work that are typically limited to three or four grades. Together, the findings outlined in this work could have significant implications for development and education scientists, educators, and policymakers and might be used to guide curriculum development and implementation to promote DT growth. These findings could also inform the development of programs and interventions for promoting creative abilities in line with grade-level and developmental specificities, while accounting for the relevant factors that have been found to moderate this development. For instance, gender differences evidenced in this meta-analysis around the seventh-grade suggest that training programs should account for gendered developmental specificities to help all genders reach their optimal DT growth in specific periods.

Lastly, the present meta-analysis helps understand the extent to which relevant variables (e.g., task at hand, content domain, country of study) account for the inconsistent developmental trends in DT outlined in the literature. Specifically, some factors seem to have a systematic effect on the observed developmental trends, regardless of DT indicator or grade level considered. In particular, the present findings suggest that DT tests are not equivalent and that the resulting conclusions in the developmental study of DT might be test- and domain-dependent (e.g., Baer, 1998; Barbot, Lubart, & Besançon, 2016; Plucker & Beghetto, 2004). Further, the effect of other factors (gender, intellectual giftedness, or culture) seems more complex, as they interact with the various DT indicators and developmental periods considered. Accordingly, researchers need to carefully consider these important factors when studying DT development, adding to a recent set of methodological recommendations for the study of creativity as it develops (Barbot, 2019).

Several limitations of this study must also be acknowledged. First, the literature search process was limited to English language literature. Although English remains the most common language for the diffusion of research findings, the study selection might have been somewhat skewed, as a large majority of studies eligible involved samples from the U.S., and possible publication biases. Thus, and given that one of the findings presented here outlined the moderating role of the country of study, caution should be used when generalizing these findings to inference on "general" developmental trends in DT. Second, as any meta-analysis, our study was itself restrained by the limitations of the corresponding primary studies selected (e.g., regarding sampling or other methodological limitation). Third, given the small set of studies included for some grades, particularly Grades 10 to 12, our meta-analysis might have had a lower power to identify effects of smaller magnitude. Finally, the findings regarding the moderator variables of the seventh-grade slump should be considered preliminary, as, by design, we have included only studies that covered at least the Grade 4 in addition to any other grade from 1 to 12. Therefore, while the effect obtained when exploring the seventh-grade slump can be considered fairly robust given the number of studies included, this set of studies is not equivalent to the whole body of existing studies potentially eligible if the focus was specifically on the seventh-grade slump.

As a general recommendation for future research, we shall note that several studies were excluded from the present meta-analysis or particular moderator analyses due to insufficient statistics reported or missing information (most of which were older studies with authors unavailable to provide missing information). Hence, we urge researchers to provide enough data on the study's method and outcomes so that their findings can be integrated into subsequent meta-analyses meaningfully.

Conclusion

The present study represents the first attempt to meta-analyze a controversial line of research findings on DT development from Grades 1 to 12 with a focus on the fourth-grade slump. Together, the general developmental pattern of DT is consistent with an overall upward trend punctuated with patterns of irregularity observed throughout childhood and adolescence

(Barbot, Lubart, & Besançon, 2016), although we didn't detect an average fourth-grade slump. Our findings also suggest some differences in the developmental trends of DT according to DT test, task content domain, gender, country of study, and intellectual giftedness of study samples. Further, these factors may have differential effects on the observed DT trends according to the period considered (e.g., fourth vs. seventh grade) and the DT indicator considered. Together, these new findings open important directions for future research aiming at (1) further validating and replicating these findings (the present work was somewhat limited by study selectivity and, at times, limited availability of data for some moderator variables), (2) better understand the role of the moderator variables outlined here (for instance, what specifically explains the differential trajectories of gifted vs. non-gifted children: intellectual precocity, motivation, educational opportunities or a combination of these factors?), and more broadly, (3) better decipher the factors that explain the slump phenomenon, over and beyond the contribution of the identified moderator variables (e.g., are slumps associated with change in brain maturation, cognitive functioning, psychosocial development, educational opportunities, or a combination of these factors?). The purpose of the present meta-analysis was not to precisely address this question, but it offers a solid basis to formulate more specific hypotheses to study DT development and the slump phenomenon. A better understanding of this phenomenon is critical in order to accurately intervene and promote creativity development in school-aged children and adolescents, as well as other psychological dimensions that rely on DT, such as exploratory behaviors involved in curiosity (e.g., Peterson & Cohen, 2019), or self-regulated learning (Rubenstein et al., 2018).

References

Abraham, A., Thybusch, K., Pieritz, K., & Hermann, C. (2014). Gender differences in creative thinking: Behavioral and fMRI findings. *Brain Imaging and Behavior, 8,* 39-51. doi: 10.1007/s11682-013-9241-4

Acar, S., & Runco, M. A. (2019). Divergent thinking: New methods, recent research, and extended theory. *Psychology of Aesthetics, Creativity, and the Arts, 13,* 153. doi: 10.1037/aca0000231

Alexander, P. A. (2019). Seeking common ground: Surveying the theoretical and empirical landscapes for curiosity and interest. *Educational Psychology Review*. doi: 10.1007/s10648-019-09508-x

Alfonso-Benlliure, V., & Santos, M. R. (2016). Creativity development trajectories in elementary education: Differences in divergent and evaluative skills. *Thinking Skills and Creativity, 19*, 160-174. doi: 10.1016/j.tsc.2015.11.003

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology, 45*, 357-376. doi: 10.1037/0022-3514.45.2.357

Amabile, T. M., & Mueller, J. S. (2008). Studying creativity, its processes, and its antecedents: An exploration of the componential theory of creativity. In J. Zhou, & C. E. Shalley (Eds.), *Handbook of organizational creativity* (pp. 33-64). New York, NY: Lawrence Erlbaum Associates.

Baer, J. (1998). The case for domain specificity in creativity. *Creativity Research Journal, 11*, 173-177. doi: 10.1207/s15326934crj1102_7

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. (2019). Measuring creativity change and development. *Psychology of Aesthetics, Creativity, and the Arts, 13*, 203-210. doi: 10.1037/aca0000232

Barbot, B., Besançon, M., & Lubart, T. (2015). Creative potential in educational settings: Its nature, measure, and nurture. *Education 3-13, 43*, 371-381. doi: 10.1080/03004279.2015.1020643

Barbot, B., Besançon, M., & Lubart, T. I. (2016). The generality-specificity of creativity: Exploring the structure of creative potential with EPoC. *Learning and Individual Differences, 52*, 178-187. doi: 10.1016/j.lindif.2016.06.005

Barbot, B., Hass, R. W., & Reiter-Palmon, R. (2019). Creativity assessment in psychological research: (Re)setting the standards. *Psychology of Aesthetics, Creativity, and the Arts, 13*, 233-240. doi: 10.1037/aca0000233

Barbot, B., & Heuser, B. (2017). Creativity and the identity formation in adolescence: A developmental perspective. In M. Karwowski, & J. C. Kaufman (Eds.), *The creative self: Effect of beliefs, self-efficacy, mindset, and identity* (pp. 87-98). San Diego, CA: Academic Press.

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

Barbot, B., & Tinio, P. P. L. (2015). Where is the “g” in “creativity”? A specialization-differentiation hypothesis. *Frontiers in Human Neuroscience, 8*, 1041. doi: 10.3389/fnhum.2014.01041

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.

Besançon, M., & Lubart, T. I. (2008). Differences in the development of creative competencies in children schooled in diverse learning environments. *Learning and Individual Differences, 18*, 381-389. doi: 10.1016/j.lindif.2007.11.009

Brewer, M. B., & Chen, Y.-R. (2007). Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological Review, 114*, 133-151. doi: 10.1037/0033-295X.114.1.133

Camp, G. C. (1994). A longitudinal study of psychological correlates of creativity. *Creativity Research Journal*, *7*, 125-144. doi: 10.1080/10400419409534519

Charles, R. E., & Runco, M. A. (2001). Developmental trends in the evaluative and divergent thinking of children. *Creativity Research Journal*, *13*, 417-437. doi: 10.1207/S15326934CRJ1334_19

Chow, J. C., & Ekholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review*, *30*, 727-744. doi: 10.1007/s10648-018-9437-7

Csikszentmihalyi, M. (1988). Motivation and creativity: Towards a synthesis of structural and energistic approaches to cognition. *New Ideas in Psychology, 6*, 159-176.

Dai, D. Y. (2019). New directions in talent development research: A developmental systems perspective. *New Directions for Child and Adolescent Development, 168*, 177-197. doi: 10.1002/cad.20322

Darvishi, Z., & Pakdaman, S. (2012). Fourth grade slump in creativity: Development of creativity in primary school children. *GSTF International Journal of Law and Social Sciences, 1*, 40-48.

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2008). Analysing data and undertaking meta-analyses. In I. P. T. Higgins, & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 243-296). Chichester, England: John Wiley & Sons.

Domínguez, A., Díaz-Pereira, M. P., & Martínez-Vidal, A. (2015). The evolution of motor creativity during primary education. *Journal of Human Sport and Exercise, 10*, 583-591. doi: 10.14198/jhse.2015.102.05

Dowd, R. J. (1966). *The relationship of creative thinking to college success* (Final report of Small Contract Project No. 5–8202). Washington, DC: U. S. Office of Education.

Egger, M., & Smith, G. D. (1997). Meta-analysis: Potentials and promise. *British Medical Journal, 315*,1371-1374. doi: 10.1136/bmj.315.7119.1371

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.

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2019). Detecting selection bias in meta-analyses with multiple outcomes: A simulation study. *The Journal of Experimental Education*. doi: 10.1080/00220973.2019.1582470

Fishkin, A. S. (1989). *Effort of odyssey of the mind creative problem solving teams: Effects on creativity, creative self-concept, locus of control and general self-concept in gifted children* (Doctoral dissertation). Retrieved from https://shareok.org/bitstream/handle/11244/20996/Thesis-1989D-F537e.pdf

Fleith, D. S. (2016). Creativity, motivation to learn, family environment, and giftedness: A comparative study. *Psicologia: Teoria e Pesquisa, 32*, 1-9. doi: 10.1590/0102-3772e32ne211

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.

Forthmann, B., Lips, C., Szardenings, C., Scharfen, J., & Holling, H. (in press). Are Speedy Brains Needed when Divergent Thinking is Speeded—Or Unspeeded? *The Journal of Creative Behavior*. https://doi.org/10.1002/jocb.350

Gajda, A., Karwowski, M., & Beghetto, R. A. (2017). Creativity and academic achievement: A meta-Analysis. *Journal of Educational Psychology, 109,* 269-299. doi: 10.1037/edu0000133

Gardner, H. (1982). *Art, mind, and brain: A cognitive approach to creativity.* New York, NY: 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

Glover, J. A., Ronning, R. R., & Reynolds, C. (Eds). (1989). *Handbook of creativity*. New York, NY: Springer Science & Business Media.

Gonen-Yaacovi, G., de Souza, L. C., Levy, R., Urbanski, M., Josse, G., & Volle, E. (2013). Rostral and caudal prefrontal contribution to creativity: A meta-analysis of functional imaging data. *Frontiers in Human Neuroscience, 7,* 465. doi: 10.3389/fnhum.2013.00465

Gottfried, A. E., & Gottfried, A. W. (1996). A longitudinal study of academic intrinsic motivation in intellectually gifted children: Childhood through early adolescence. *Gifted Child Quarterly, 40*, 179-183. doi: 10.1177/001698629604000402

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

Guignard, J.-H., Kermarrec, S., & Tordjman, S. (2016). Relationships between intelligence and creativity in gifted and non-gifted children. *Learning and Individual Differences, 52,* 209-215. doi: 10.1016/j.lindif.2015.07.006

Guignard, J.-H., & Lubart, T. I. (2007). A comparative study of convergent and divergent thinking in intellectually gifted children. *Gifted and Talented International, 22,* 9-15. doi: 10.1080/15332276.2007.11673481

Guilford, J. P. (1950). Creativity. *American Psychologist, 5,* 444-454.

Guilford, J. P. (1966). Measurement and creativity. *Theory into Practice, 5,* 186-202.

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.

Guilford, J. P. (1984). Varieties of divergent production. *The Journal of Creative Behavior, 18,* 1-10. doi:10.1002/j.2162-6057.1984.tb00984.x

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia, 14,* 29-37.

Hansenne, M., & Legrand, J. (2012). Creativity, emotional intelligence, and school performance in children. *International Journal of Educational Research, 53,* 264-268. doi: 10.1016/j.ijer.2012.03.015

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

Hennessey, B. A. (1995). Social, environmental, and developmental issues and creativity. *Educational Psychology Review, 7,* 163-183.

Hill, J. P., & Lynch, M. E. (1983). The intensification of gender-related role expectations during early adolescence. In J. Brooks-Gunn, & A. C. Petersen (Eds.), *Girls at puberty:* : *Biological and psychosocial perspectives* (pp. 201-228). New York, NY: Springer Science & Business Media.

Hong, E., & Milgram, R. M. (2010). Creative thinking ability: Domain generality and specificity. *Creativity Research Journal, 22,* 272-287. doi: 10.1080/10400419.2010.503535

Hopp, M. D. S., Zhang, Z. S., Hinch, L., O'Reilly, C., & Ziegler, A. (2019). Creative, thus connected: The power of sociometric creativity on friendship formation in gifted adolescents—a longitudinal network analysis of gifted students. *New Directions for Child and Adolescent Development, 168,* 47-73. doi: 10.1002/cad.20324

Houtz, J. C., & Krug, D. (1995). Assessment of creativity: Resolving a mid-life crisis. *Educational Psychology Review, 7*, 269-300. doi:10.1007/bf02213374

Hu, W., Shi, Q. Z., Han, Q., Wang, X., & Adey, P. (2010). Creative scientific problem finding and its developmental trend. *Creativity Research Journal*, *22*, 1-7. doi: 10.1080/10400410903579551

Jaquish, G. A., & Ripple, R. E. (1984). A life-span developmental cross-cultural study of divergent thinking abilities. *International Journal of Aging and Human Development, 20*, 1-11. doi: 10.2190/RNJJ-NBD0-4A3K-0XPA

Jastrzębska, D., & Limont, W. (2017). Not only jumps, slumps, but also mini plateau. Creative potential assessed by the Test for Creative Thinking-Drawing Production. A cross-sectional study of polish students aged from 7 to 18. *Creativity Research Journal*, *29*, 337-342. doi: 10.1080/10400419.2017.1360060

Jeffrey, B., & Craft, A. (2004). Teaching creatively and teaching for creativity: Distinctions and relationships. *Educational Studies, 30*, 77-87. doi: 10.1080/0305569032000159750

Karmiloff-Smith, A. (1994). Précis of beyond modularity: A developmental perspective on cognitive science. *Behavioral and Brain Sciences, 17*, 693-745.

Karwowski, M., & Lebuda, I. (2014). Digit ratio predicts eminence of Polish actors. *Personality and Individual Differences, 64*, 30-34. doi: 10.1016/j.paid.2014.02.014

Kim, K. H. (2011). The creativity crisis: The decrease in creative thinking scores on the Torrance Tests of Creative Thinking. *Creativity Research Journal, 23*, 285-295. doi: 10.1080/10400419.2011.627805

Kim, K. H., Cramond, B., & Bandalos, D. L. (2006). The latent structure and measurement invariance of scores on the Torrance Tests of Creative Thinking – Figural. *Educational and Psychological Measurement, 66*, 459-477. doi: 10.1177/0013164405282456

Klausmeier, H. J., & Wiersma, W. (1964). Relationship of sex, grade level, and locale to performance of high IQ students on divergent thinking tests. *Journal of Educational Psychology, 55*, 114-119. doi:10.1037/h0041717

Kleibeuker, S. W., De Dreu, C. K. W., & Crone, E. A. (2013). The development of creative cognition across adolescence: Distinct trajectories for insight and divergent thinking. *Developmental Science, 16*, 2-12. doi: 10.1111/j.1467-7687.2012.01176.x

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

Lau, S., & Cheung, P. C. (2010). Developmental trends of creativity: What twists of turn do boys and girls take at different grades? *Creativity Research Journal*, *22*, 329-336. doi: 10.1080/10400419.2010.503543

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

Lin, W.-L., & Shih, Y.-L. (2016). The developmental trends of different creative potentials in relation to children's reasoning abilities: From a cognitive theoretical perspective. *Thinking Skills and Creativity, 22*, 36-47. doi: 10.1016/j.tsc.2016.08.004

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. New York, NY: Oxford University Press.

Long, B. H., & Henderson, E. H. (1965). Opinion formation and creativity in elementary school children. *Psychological Reports, 17*, 219-223. doi: 10.2466/pr0.1965.17.1.219

Lopez, E. C., Esquivel, G. B., & Houtz, J. C. (1993). The creativity skills of culturally and linguistically gifted and diverse students. *Creativity Research Journal, 6*, 401-412. doi: 10.1080/10400419309534495

Lubart, T. I. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal, 13*, 295-308. doi: 10.1207/S15326934CRJ1334_07

Lubart, T. I., & Georgsdottir, A. (2004). Creativity: Development and cross-cultural issues. In S. Lau, A. N. N. Hui, & G. Y. C. Ng (Eds.), *Creativity: When east meets west* (pp. 23-54). Singapore: World Scientific Publishing.

Lubart, T. I., & Lautrey, J. (1995, August). *Relationships between creative development and cognitive development*. Paper presented at the Seventh European Conference on Developmental Psychology, Krakow, Poland.

Mainemelis, C. (2010). Stealing fire: Creative deviance in the evolution of new ideas. *Academy of Management Review, 35*, 558-578. doi: 10.5465/amr.35.4.zok558

McCrae, R. R., Arenberg, D., & Costa, P. T. (1987). Declines in divergent thinking with age: Cross-sectional, longitudinal, and cross-sequential analyses. *Psychology and Aging, 2*, 130- 137. doi: 10.1037/0882-7974.2.2.130

Nelson, E. E., & Guyer, A. E. (2011). The development of the ventral prefrontal cortex and social flexibility. *Developmental Cognitive Neuroscience*, *1*, 233-245. doi: 10.1016/j.dcn.2011.01.002

Paterson, S., Heim, S., Friedman, J. T., Choudhury, N., & Benasich, A. (2006). Development of structure and function in the infant brain: Implications for cognition, language and social behaviour. *Neuroscience and Biobehavioral Reviews, 30,* 1087-1105. doi:10.1016/j.neubiorev.2006.05.001

Peterson, E. G., & Cohen, J. (2019). A case for domain-specific curiosity in mathematics. *Educational Psychology Review*. doi: 10.1007/s10648-019-09501-4

Plucker, J. A., & Beghetto, R. A. (2004). Why creativity is domain general, why it looks domain specific, and why the distinction does not matter. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: From potential to realization* (pp. 153-167). Washington, DC: American Psychological Association.

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist, 39*, 83-96. doi: 10.1207/s15326985ep3902_1

Plucker, J. A., Kaufman, J. C., & Beghetto, R. A. (2015). *What we know about creativity*. Washington, DC: Partnership for 21st Century Skills.

Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts, 13*, 144-152. doi: 10.1037/aca0000227

Rieben, L. (1978). *Inteligencia global, inteligencia operatoria y creatividad*. Barcelona, España: Médica y Técnica.

Rosenfield, S., & Houtz, J. C. (1978). Developmental patterns in problem solving and divergent thinking abilities in gifted elementary school children. *Journal for the Education of the Gifted, 1,* 37-48. doi: 10.1177/016235327800100207

Rubenstein, L. D., Callan, G. L., & Ridgley, L. M. (2018). Anchoring the creative process within a self-regulated learning framework: Inspiring assessment methods and future research. *Educational Psychology Review, 30*, 921-945. doi: 10.1007/s10648-017-9431-5

Runco, M. A. (2004). Creativity. *Annual Review of Psychology, 55*, 657-687. doi: 10.1146/annurev.psych.55.090902.141502

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24,* 66-75. doi: 10.1080/10400419.2012.652929

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

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., Fernández-Castilla, B., Kyndt, E., & Van den Noortgate, W. (2019). Testing conditions and creative performance: Meta-analyses of the impact of time limits and instructions. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. doi: 10.1037/aca0000244

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

Sak, U., & Maker, C. J. (2006). Developmental variation in children's creative mathematical thinking as a function of schooling, age, and knowledge. *Creativity Research Journal*, *18*, 279-291. doi: 10.1207/s15326934crj1803_5

Sherwood, L. A., & Strahan, D. B. (1985). Developmental patterns of logical and creative thinking among gifted learners in the middle school. *Middle School Research Selected Studies*, *10*, 68-82. doi: 10.1080/08851700.1985.11670261

Shin, D. D., & Kim, S. (2019). Homo curious: Curious or interested? *Educational Psychology Review*. doi: 10.1007/s10648-019-09497-x

Silvia, P. J. (2015). Intelligence and creativity are pretty similar after all. *Educational Psychology Review*, *27*, 599-606. doi: 10.1007/s10648-015-9299-1

Smith, J. K., & Smith, L. F. (2010). Educational creativity. In J. C. Kaufman, & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 250-264). New York, NY: Cambridge University Press.

Soilemezi, D., & Linceviciute, S. (2018). Synthesizing qualitative research. *International Journal of Qualitative Methods, 17.* doi: 10.1177/1609406918768014

Sternberg, R. J. (2006). The nature of creativity. *Creativity Research Journal*, *18*, 87-98. doi: 10.1207/s15326934crj1801_10

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

Torrance, E. P. (1967). *Understanding the fourth grade slump in creative thinking* (Report No. BR-5-0508; CRP-994). Washington, DC: U. S. Office of Education.

Torrance, E. P. (1968). A longitudinal examination of the fourth-grade slump in creativity. *Gifted Child Quarterly, 12*, 195-199. doi: 10.1177/001698626801200401

Torrance, E. P. (1977). *Creativity in the classroom: What research says to the teacher*. Washington, DC: National Education Association.

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

Walker, E., Hernandez, A. V., & Kattan, M. W. (2008). Meta-analysis: Its strengths and limitations. *Cleveland Clinic Journal of Medicine, 75*, 431-439. doi: 10.3949/ccjm.75.6.431

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

Wong, Y. Y., Chow, I. H., Lau, V. P., & Gong, Y. (2018). Benefits of team participative decision making and its potential to affect individual creativity. *Journal of Applied Social Psychology, 48*, 369-376. doi: 10.1111/jasp.12517

Yi, X., Hu, W., Plucker, J. A., & McWilliams, J. (2013). Is there a developmental slump in creativity in China? The relationship between organizational climate and creativity development in Chinese adolescents. *The Journal of Creative Behavior, 47*, 22-40. doi: 10.1002/jocb.21

Table 1

A summary of the Results of the Meta-analysis

	Overall			Fluency			Flexibility			Originality		
	SM	*SE*	SMD	SM	*SE*	SMD	SM	*SE*	SMD	SM	*SE*	SMD
Grade 1	2.49	0.28		2.89	0.39		3.06	0.47		2.92	0.49	
Grade 2	2.71	0.01	0.22***	3.03	0.03	0.14***	3.38	0.03	0.32***	3.12	0.03	0.20***
Grade 3	2.76	0.01	0.05***	3.15	0.03	0.12***	3.51	0.03	0.13***	3.06	0.03	-0.06*
Grade 4	2.81	0.08	0.05	3.10	0.15	-0.05	3.46	0.16	-0.05	3.00	0.11	-0.06
Grade 5	2.97	0.01	0.16***	3.31	0.01	0.21***	3.71	0.01	0.25***	3.13	0.01	0.13***
Grade 6	3.11	0.01	0.14***	3.42	0.02	0.11***	3.81	0.02	0.10***	3.21	0.01	0.08***
Grade 7	2.97	0.01	-0.14***	3.18	0.03	-0.24***	3.63	0.03	-0.18***	3.16	0.02	-0.05**
Grade 8	3.16	0.01	0.19***	3.49	0.03	0.31***	3.93	0.03	0.30***	3.34	0.02	0.18***
Grade 9	3.19	0.01	0.03*	3.56	0.04	0.07	3.99	0.04	0.06	3.39	0.03	0.05
Grade 10	3.15	0.02	-0.04*	3.49	0.06	-0.07	4.12	0.05	0.13*	3.65	0.05	0.26***
Grade 11	3.09	0.02	-0.06***	3.26	0.08	-0.23**	4.09	0.06	-0.03	3.55	0.06	-0.10
Grade 12	3.12	0.01	0.03*	/	/	/	4.20	0.07	0.11	3.72	0.07	0.17*
σ_V^2		0.13			0.20			0.19			0.11	
σ_U^2		0.14			0.20			0.19			0.12	
σ_E^2		0.01			0.01			0.01			0.01	

Note. SM = standardized mean; *SE* = standard error; SMD = standardized mean difference; σ_V^2 = between-study variance of the standardized mean differences between Grades 3 and 4; σ_U^2 = within-study variance of the standardized mean differences between Grades 3 and 4; σ_E^2 = typical sampling variance of the standardized mean differences between Grades 3 and 4; / indicates data were not available.

**p* < .05. ***p* < .01. ****p* < .001.

Table 2

Results of the Moderator Analyses

Analysis	Moderator variable	Grade 4						Grade 7					
		n	*N*	*k*	*F*	*df*	*p*	*n*	*N*	*k*	*F*	*df*	*p*
Overall													
	DT test				0.64	2, 1297	.53				3.24	2, 1549	.04
	TTCT	3011	44	538				1355	20	212			
	WKCT	1013	6	56				1172	5	52			
	Other	8590	34	202				6067	25	139			
	Domain				4.14	1, 1159	.04				4.18	1, 1399	.04
	Verbal	7804	55	436				5387	31	215			
	Figural	5731	56	325				3302	24	145			
	Gender	4160	52	455	0.54	1, 722	.46	2657	25	193	5.47	1, 883	.02
	Country				1.58	2, 1297	.21				0.27	1, 1549	.60
	Western	4453	61	583				2430	31	235			
	Eastern	7869	21	161				6082	17	142			
	Middle Eastern	292	2	52				/	/	/			
	Intellectual giftedness				3.6	1, 1297	0.06				0.02	1, 1549	.89
	Gifted	384	10	70				252	5	31			
	Non-Gifted	12230	74	726				8342	45	372			
Fluency													
	DT test				0.72	2, 250	.49				0.4	2, 282	.67
	TTCT	2648	26	111				1355	20	67			
	WKCT	985	5	13				1147	4	13			
	Other	1886	11	31				1371	9	20			
	Domain				0.08	1, 233	.77				8.91	1, 266	.003
	Verbal	3462	32	111				2743	25	73			
	Figural	3286	26	40				2153	18	22			
	Gender	1664	19	59	0.14	1, 97	.71	1301	15	41	15.24	1, 100	< .001
	Country				1.44	1, 250	.23				0	1, 282	.97
	Western	2270	30	114				1131	22	66			
	Eastern	3249	12	41				2742	11	34			
	Intellectual giftedness				4.16	1, 250	.04				0.72	1, 282	.40
	Gifted	347	8	25				252	5	13			
	Non-gifted	5172	34	130				3621	28	87			

Table 2 (*continued*)

Analysis	Moderator variable	Grade 4 *n*	Grade 4 *N*	Grade 4 *k*	Grade 4 *F*	Grade 4 *df*	Grade 4 *p*	Grade 7 *n*	Grade 7 *N*	Grade 7 *k*	Grade 7 *F*	Grade 7 *df*	Grade 7 *p*
Flexibility													
	DT test				0.50	2, 224	.61				0.81	2, 247	.45
	TTCT	2523	25	103				1290	19	66			
	WKCT	985	5	13				1147	4	13			
	Other	2054	10	18				2055	9	14			
	Domain				0.41	1, 203	.52				10.21	1, 230	.002
	Verbal	3426	31	98				2706	24	66			
	Figural	3011	18	29				2051	16	20			
	Gender	1523	17	44	0.04	1, 73	.84	1199	13	32	5.81	1, 71	.02
	Country				0.26	1, 224	.61				0.67	1, 247	.41
	Western	1964	27	92				1029	20	57			
	Eastern	3598	13	42				3463	12	36			
	Intellectual giftedness				3.31	1, 224	.07				0.73	1, 247	.39
	Gifted	222	7	10				187	4	8			
	Non-gifted	5340	33	124				4305	28	85			
Originality													
	DT test				3.22	2, 390	.04				1.58	2, 483	.21
	TTCT	2851	42	279				1355	20	67			
	WKCT	965	4	20				1127	3	24			
	Other	2549	12	19				2377	11	16			
	Domain				1.89	1, 361	.17				1.4	1, 458	.24
	Verbal	3516	45	211				2651	22	69			
	Figural	3953	43	103				2366	19	29			
	Gender	2326	36	220	0.10	1, 220	.75	1586	16	43	25.78	1, 287	< .001
	Country				7.66	1, 390	.006				0.3	1, 483	.59
	Western	2767	45	268				1396	22	59			
	Eastern	3598	13	50				3463	12	48			
	Intellectual giftedness				1.24	1, 390	.27				0	1, 483	.97
	Gifted	254	7	13				217	4	7			
	Non-gifted	6111	51	305				4642	30	100			

Note. n = number of participants; *N* = number of samples; *k* = number of standardized means; / indicates data were not available.

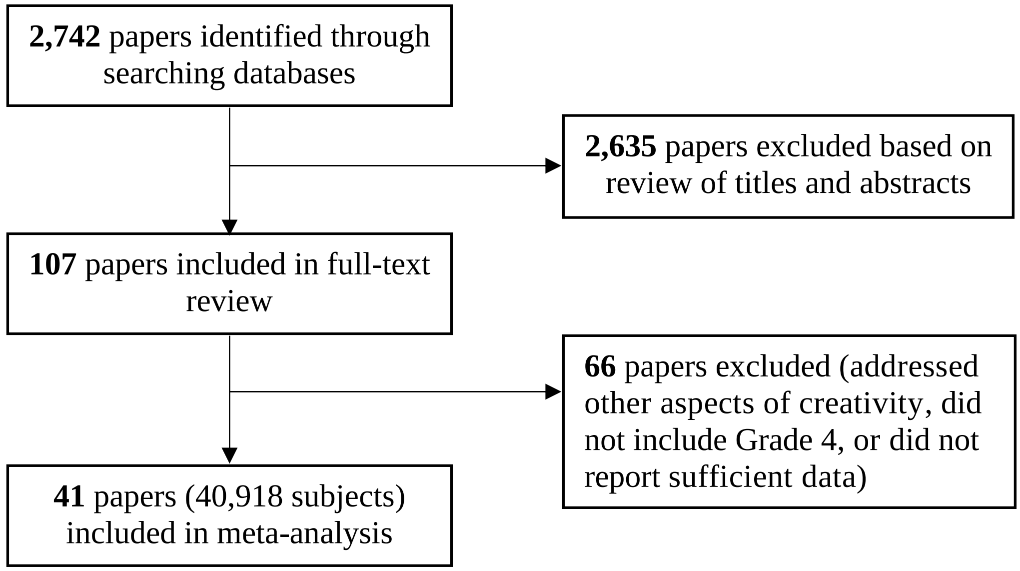

Figure 1. Flow chart of data collection in the meta-analysis

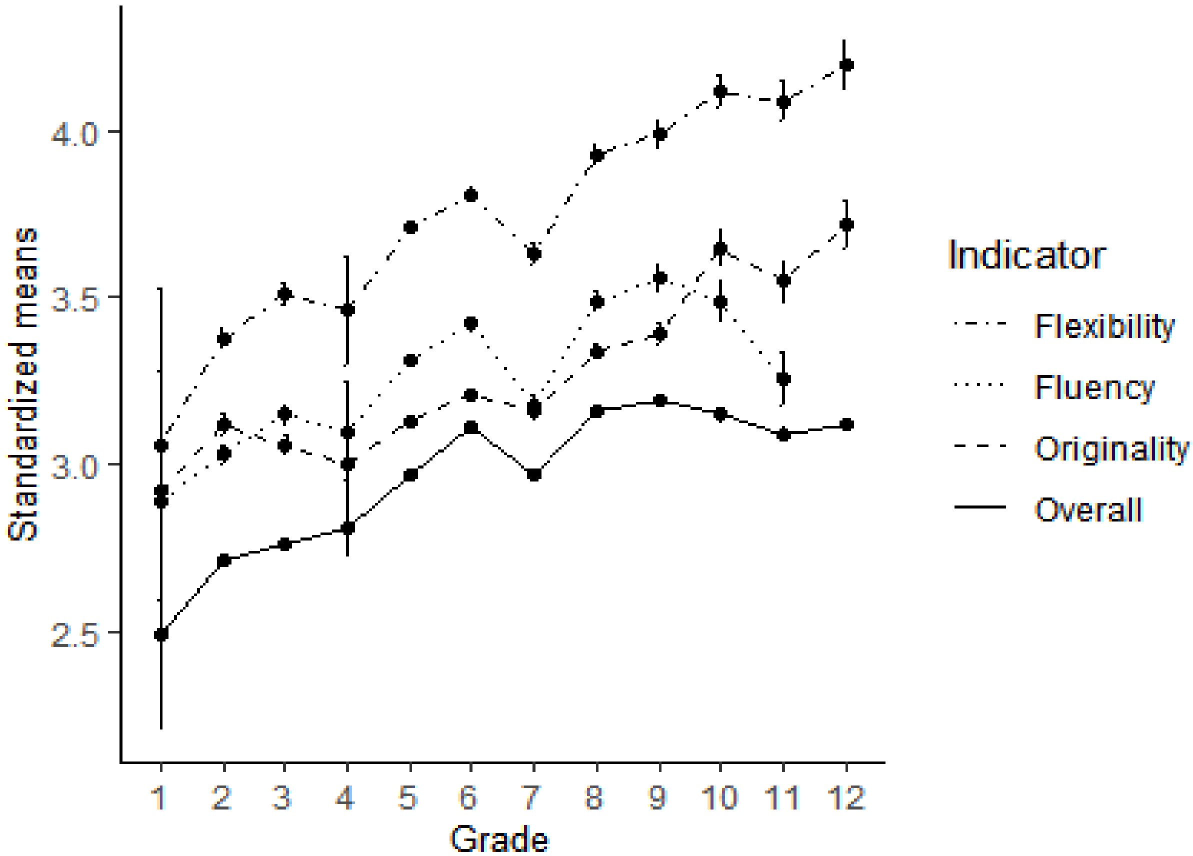

Figure 2. Developmental trends of divergent thinking indicators by grade with corresponding standard errors

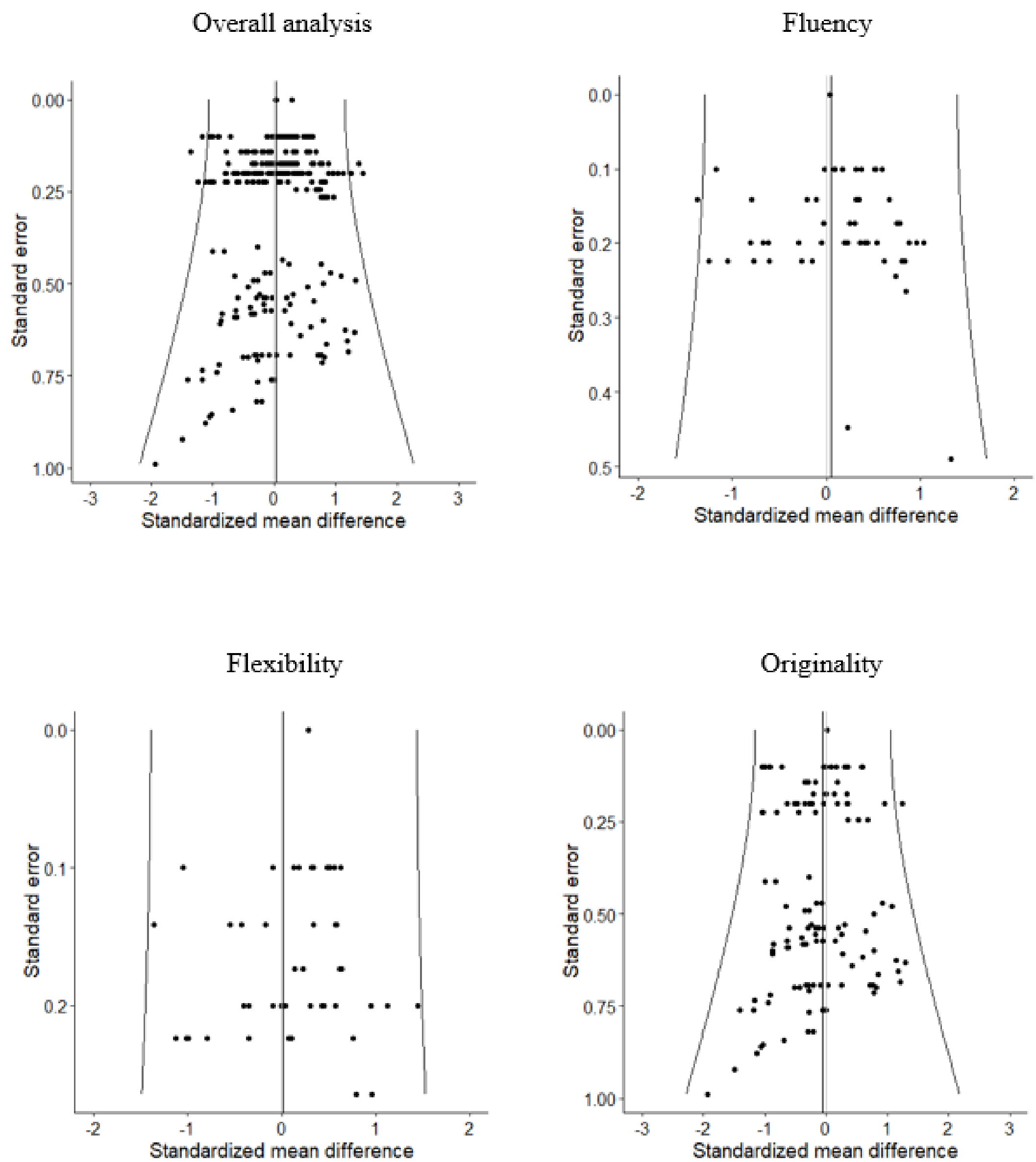

Figure 3. Funnel plots of the meta-analysis