The Impact of Creativity Training on Creative Performance: A Meta-Analytic Review and Critical Evaluation of Five Decades of Creativity Training Studies

Ut Na Sio[1*] & Hugues Lortie-Forgues[2]

[1]Sheffield University Management School, The University of Sheffield, Sheffield, UK
[2]Centre for Mathematical Cognition, Loughborough University, Loughborough, UK

Author Note

Data and code are available through the Open Science Framework: https://osf.io/8z3k5

Acknowledgments

This work was partially funded by the BA/Leverhulme Small Research Grant SRG2021\210385. Hugues Lortie-Forgues is funded by a research grant from Research England via the Centre for Mathematical Cognition. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

[*]Correspondence to: Ut Na Sio (u.n.sio@sheffield.ac.uk), Sheffield University Management School, The University of Sheffield, Conduit Road, Sheffield, S10 1FL, United Kingdom.

Abstract

Creativity is widely considered a skill essential to succeeding in the modern world. Numerous creativity training programs have been developed, and several meta-analyses have attempted to summarize the effectiveness of these programs and identify the features influencing their impact. Unfortunately, previous meta-analyses share a number of limitations, most notably overlooking the potentially strong impact of publication bias and the influence of study quality on effect sizes. We undertook a meta-analysis of 169 creativity training studies across five decades (844 effect sizes, the largest meta-analysis of creativity training to date), including a substantial number of unpublished studies (48 studies; 262 effect sizes). We employed a range of statistical methods to detect and adjust for publication bias and evaluated the robustness of the evidence in the field. In line with previous meta-analyses, we found a moderate training effect (0.53 *SD*s; unadjusted for publication bias). Critically, we observed converging evidence consistent with strong publication bias. All adjustment methods considerably lowered our original estimate (adjusted estimates ranged from 0.29 to 0.32 *SD*s). This severe bias casts doubt on the representativeness of the published literature in the field and on the conclusions of previous meta-analyses. Our analysis also revealed a high prevalence of methodological shortcomings in creativity training studies (likely to have inflated our average effect), and little signs of methodological improvement over time — a situation that limits the usefulness of this body of work. We conclude by presenting implications and recommendations for researchers and practitioners, and we propose an agenda for future research.

Keywords

Creativity training, meta-analysis, publication bias, study quality, methodological improvement

Public Significance Statement

Creativity is considered an essential skill in many contexts, leading to the development of numerous training programs aimed at improving this skill. We examined five decades of creativity training studies (169 studies). Using a range of meta-analytic techniques, we found converging evidence consistent with publication bias. Moreover, our analysis revealed a large number of studies with methodological shortcomings. Both of these findings suggest that creativity training may be less effective than previously thought. In view of these findings, we proposed a set of recommendations to assist researchers and practitioners in interpreting findings from the creativity training literature, and we outlined directions for future research to increase the informativeness of creativity training studies.

The Impact of Creativity Training on Creative Performance: A Meta-Analytic Review and Critical Evaluation of Five Decades of Creativity Training Studies

Creativity is defined as the generation of novel and useful ideas (e.g., Amabile, 1996; Sawyer, 2006; Sternberg & O'Hara, 1999). It is often viewed as a foundational stage of the organizational innovation process — the process of developing and commercializing creative ideas (Amabile, 1988; West, 2002). Understandably, employers rank creativity as one of the top skills necessary to succeed both in today's workplaces and in the workplaces of the future (Bellanca & Brandt, 2010; IBM Corporation, 2010; World Economic Forum, 2018). Interest in creativity is not limited to organizations, as demonstrated by the large amount of effort that high-performing education systems (such as Finland, Australia, Singapore, and Canada) dedicate to teaching creativity (Durham Commission, 2019), as well as the recent inclusion of creativity as one of the key outcomes measured by the OECD's Program for International Student Assessment (PISA, 2022). Creativity is also listed as a key skill in solving the sustainable development challenges outlined by the United Nations (UN General Assembly, 2015).

Due to the importance of creativity in a range of contexts, enhancing creativity has been a prime concern for organizations, educators, and policymakers. In organizations, many strategies have been developed with the aim of enhancing employees' creativity, such as establishing a creative climate, offering incentives, and providing training to enhance creativity (Anderson & West, 1998; Collins & Amabile, 1999; Eisenberger & Shanock, 2003). Among these strategies, providing training is one popular approach to improve creativity (Birdi, 2016). In UK organizations, for example, 10.8% of the 14,040 businesses surveyed in the 2019 UK Innovation Survey reported investing in training for innovation activities (UK Innovation Survey, 2019)[2], a decision that involves substantial costs (on average, 1.7% of the companies' total expenditures). More broadly, the Eurostat's Community Innovation 2016 Survey showed that 44.6 % of the surveyed EU innovative enterprises engaged in innovation-related training activities. Similarly, in the countries surveyed in the UNESCO Institute for Statistics Innovation Data Collection Report (2017), the average proportion of innovative-active firms using innovation training was 44.7%.

---

[2] *Creativity* is defined as the generation of novel and useful ideas (Amabile et al., 1996). Innovation also includes the implementation of those ideas. Innovation training is often considered a mix of creativity and implementation skills (Birdi, 2016; Fischer & Afifi, 2013).

Given the large and growing interest in improving creativity, the prevalence of creativity training programs as a means to achieve this aim, and the expenses often associated with their delivery, it is crucial to ensure that creativity training programs are effective. Using rigorous meta-analytic methods, we review the existing creativity training literature to (a) assess how effective creativity training programs are, (b) identify the methodological features that influence the programs' effectiveness, and (c) evaluate how robust the evidence surrounding creativity training effectiveness is. We believe that this effort is timely. The recent development in research practices and meta-analysis techniques have cast doubt on a range of previously well-established phenomena — often suggesting that these phenomena are less impressive than previously thought (e.g., the effect of cognitive training, Gobet & Sala, 2022), and, in some cases, non-existent (e.g., ego-depletion effect, Carter et al., 2015; growth mindset, Macnamara & Burgoyne, 2023; nudging effect, Maier et al., 2022). Findings from the present study will provide a more accurate estimate of the average effect size of creativity training, help researchers identify methodological gaps in the current literature, as well as inform their effort to develop theories and effective training programs — efforts that require a large body of robust evidence (Eronen & Bringmann, 2021; Nosek et al., 2022). Our survey of creativity training research will also help practitioners make more informed decisions when selecting approaches to enhance creativity and calibrate their expectations about the impact of delivering such approaches.

Different Forms of Creativity Training

Creativity is a complex psychological process characterized by a multitude of internal and external components, including individuals' creative thinking ability, personality, motivation, and environment (for a review, see Hennessey & Amabile, 2010). Expectedly, different forms of creativity training focusing on different core components of creativity have been developed. For instance, when reviewing 67 creativity training courses, Bull et al. (1995) identified 134 distinct techniques used to enhance creativity, e.g., promoting openness to new experiences, increasing psychological understanding of creative processes, and setting a safe climate for creativity.

Numerous studies have evaluated the effectiveness of creativity training. The typical paradigm of these studies involves comparing a group receiving the training (training group) to a group not receiving the training (control group) on post-training measures of creativity, with or without the use of a pretest to adjust for pre-training between-group differences. Given the multifaceted nature of creativity, different measures have been employed to evaluate the training effectiveness. These measures can be broadly organized into two categories: non-performance- and performance-based measures. Non-performance-based measures include, for example, self-efficacy and level of motivation (e.g., Amabile et al., 1994; Tierney & Farmer, 2002). Performance-based measures often include divergent thinking, creative problem solving, and the generation of creative products (e.g., Guilford, 1967; Mednick, 1962; Torrance, 1966). It is worth noting that when assessing divergent thinking performance, researchers typically measure multiple dimensions of the construct; the three most common being fluency (i.e., quantity of the ideas generated), flexibility (i.e., variety of the ideas generated), and originality (i.e., uniqueness of the ideas generated) (see Reiter-Palmon et al., 2019 for a review).

To our knowledge, five meta-analyses[3] (Ma, 2006; Rose & Lin, 1984; Scott et al., 2004; Tsai, 2013; Yasin & Yunus, 2014) have summarized evidence relating to the effectiveness of creativity training and explored the factors impacting the effect sizes, such as which types of training are most effective in enhancing creativity, whether creativity training generates similar effects on different outcome measures, and whether intervention outcomes differ as a function of different study characteristics. We describe various features of these five meta-analyses in Table 1.

Effectiveness of Creativity Training: Findings of Previous Meta-Analyses

The five previous meta-analyses of creativity training have uniformly reported moderate to strong positive effects of creativity training on creativity. For example, Rose and Lin (1984) summarized the findings of 46 studies using the Torrance Test of Creative Thinking (TTCT) — a common measure of divergent thinking — to assess the effect of creativity training and reported an average effect size of d = 0.47. Scott et al. (2004) and Ma (2006) extended Rose and Lin's (1984) work by including studies using a wider range of creativity measures and estimated the average effect sizes to be d = 0.64 (Scott et al., 2004) and d = 0.76 (Ma, 2006)[4]. Meta-analyses examining the effect on specific populations reported even larger effect sizes: d = 1.02 in the field of technology and engineering (Yasin & Yunus, 2014) and d = 0.81 for adult learners (Tsai, 2013). Previous meta-analyses also

---

[3] We focused on meta-analyses that specifically examine the effectiveness of creativity training. As a result, we did not consider, for example, the recent meta-analysis on creativity enhancement methods in adults by Haase et al. (2023) due to its large proportion of studies (> 50%) evaluating non-training approaches (e.g., drugs).

[4] Ma (2006) reported two average effect sizes, one with and one without aggregating multiple effect sizes belonging to the same study, before combining them across studies to handle effect size dependency. The one reported here is the average effect size based on aggregated effect sizes, and the one based on unaggregated effect sizes was d = 0.77.

explored the factors moderating the impact of creativity training. See supplemental material Table S1 for a detailed description of the moderators evaluated in previous meta-analyses. Scott et al. (2004) observed that creativity training seemed to have a bigger impact on divergent thinking and creative problem solving than on other outcome measures (e.g., generation of creative products, and creative attitude and behavior). Among the different dimensions of divergent thinking, originality showed the largest improvement from creativity training compared to other dimensions such as fluency and flexibility (Rose & Lin, 1984; Scott et al., 2004). Ma (2006) and Scott et al. (2004) also explored the moderating effects of a number of training and study characteristics on training effectiveness. Scott et al. (2004) observed that training based on a cognitive framework generated, on average, larger effect sizes than training based on a non-cognitive framework. They also observed that longer training tended to be associated with larger effects. In terms of study design, Scott et al. (2004) found that studies of higher methodological quality (e.g., studies using a pretest and studies having a control group) tended to generate smaller effect sizes and that larger studies tended to be associated with smaller effect sizes. Moreover, Ma (2006) reported that adults benefited more from creativity training than younger participants (e.g., students of elementary and high schools).

As shown, previous meta-analyses provide converging evidence that creativity training is an effective way to enhance creativity and identified a number of factors associated with training effectiveness. Some of these meta-analyses are highly influential. For example, at the time of writing, Scott et al.'s (2004) meta-analysis, one of the most comprehensive meta-analyses in this field in terms of the scope and number of studies included, has been cited a total of 1604 times (195 times since 2022) according to Google Scholar. However, almost two decades have passed since the publication of this meta-analysis, and as such it may not accurately reflect the current state of the field. Possibly due to limitations in the meta-analytic methods available at the time, past meta-analyses are also subject to a number of limitations, some of which could have biased — perhaps substantially — some of their findings. It is plausible, for example, that past meta-analyses might have overestimated the overall effectiveness of creativity training.

The average effect sizes of creativity training programs reported in previous meta-analyses (which range from d = 0.47 to 1.02) are substantially larger than what is typically observed in social sciences research. For example, when reviewing 747 educational intervention trials (totaling 1942 effect sizes), Kraft (2020) observed a mean effect size of d = 0.16. In psychology, Lovakov and Agadullina (2021) extracted 6447 effect sizes from 134 meta-analyses and found a median effect size of $d = 0.36$. Although these average effect sizes are not specific to creativity training, their substantial difference from those reported in past meta-analyses of creativity training raises the question of whether the previously documented average effects of creativity training may have been overestimated. We present below a number of methodological limitations found in previous meta-analyses of creativity training that may have impacted their findings.

Limitations of previous meta-analyses of creativity training. Previous meta-analyses of creativity training present methodological limitations that could potentially impact their conclusions.

Assessment of Publication Bias. Publication bias — i.e., studies with statistically significant findings in the expected direction are more likely to be published than those with negative or null findings — can, when unaccounted for, lead meta-analyses to substantially overestimate the effect of interest (e.g., Lipsey & Wilson, 1993; Thornton & Lee, 2000). It can also distort their conclusion, for example, obscuring the influence of certain moderators (Munafò et al., 2018). Inflated effect size estimates could cause researchers and practitioners to hold unrealistic expectations as to the effectiveness of creativity training, eventually hampering the development of the field. For example, researchers may underpower future replication studies if they use the inflated estimate to determine sample size, reducing the chance of successful replication and, consequently, making it difficult to explore the mechanism underlying the effect. For practitioners, the misleading evidence would interfere with their ability to make informed decisions; it could lead to disappointment and potentially lead them to question the usefulness of this body of work. Thus, it is important to examine the presence and extent of publication bias in creativity training.

The field of creativity training could be particularly susceptible to publication bias. One key reason is that studies evaluating creativity training tend to be small. The average sample size of studies included in the meta-analysis of Scott et al. (2004)[5] was $N = 60.14$, below the sample size generally recommended in education research (e.g., the What Works Clearinghouse (2020) recommends a total sample size of at least 350 participants). In general, small (or underpowered) studies are less likely to detect a significant effect (if there is one to be found), and when they detect a significant effect, the effect found is likely an

[5] This value was computed by dividing the total number of participants ($N = 4210$) by the number of studies ($N = 70$).

overestimation of the true effect (Ioannidis, 2005, 2008). When combined with publication bias, underpowered studies can lead to a severe overestimation of the true effect size.

Also, by design, studies of creativity training tend to be noisy as they are typically conducted in clustered and hard-to-control settings (e.g., classrooms). Null findings from small and noisy studies are harder to present as convincing evidence of no effect because of the often-large uncertainty surrounding estimates of impact, reducing their chance of being published. Moreover, some creativity training programs are developed commercially and evaluated by the developers themselves, a situation that might give researchers an additional, financial incentive not to publicize null or negative findings. Together, these reasons suggest that the field of creativity training could potentially be characterized by severe publication bias.

Unfortunately, previous meta-analyses did not extensively consider the potential influence of publication bias. One key limitation of previous meta-analyses is the scarcity of data from unpublished studies. As shown in Table 1, the number of included unpublished articles is low, both as an absolute figure and as the proportion of the total number of studies included. Only Rose and Lin’s (1984) meta-analysis, published more than 35 years ago, included more than six unpublished studies. The omission of unpublished studies, in the presence of publication bias, can lead to an overestimation of the effect of interest (e.g., Lane & Dunlap, 1978). In addition to the limited inclusion of unpublished studies, none of the previous meta-analyses reported testing or correcting for publication bias (Ma, 2006; Rose & Lin, 1984; Tsai, 2013; Yasin & Yunus, 2014)[6].

Overall, given the likelihood of a substantial publication bias in creativity training research, and the potentially large influence it may have had on the conclusions of previous meta-analyses, we dedicate an important part of the present meta-analysis to estimating the influence of this bias.

Accounting for Effects’ Precision. Also noteworthy, contrary to now more standard ways of computing meta-analytic averages, previous meta-analyses did not consider the precision (i.e., sample size) of the studies in their calculation of the average estimate of the impact of creativity training. In other words, estimates from small (less precise) and large (more precise) studies were given equal weight in the overall effect sizes calculation. In

[6] Scott et al. (2004) assessed the potential for publication bias by computing the fail-safe *N* statistics (an estimation of the number of missing studies with null effects needed to nullify the observed average effect). However, this method has been widely criticized for being grounded on contentious assumptions and for understating the severity of publication bias (Becker, 2005; Ferguson et al., 2012).

addition to not using all the available information effectively (Borenstein et al., 2009), this omission could potentially exacerbate the overestimation of the average effect size in the presence of publication bias, as small studies tend to be more prone to this bias (Sterne et al., 2000).

Methodological Issues Associated with the Analysis of Moderators. Both Ma (2006) and Scott et al. (2004) conducted moderator analysis to assess the influences of study characteristics on effect size. As shown in supplemental material Table S1, Scott et al. (2004) explored an impressive number of moderators (i.e., > 40). Yet, a limitation of their analysis is that they dichotomized many of the moderators that are continuous in nature, e.g., publication year and sample size. Although dichotomization typically simplifies analyses and interpretation, it often results in a loss of information and statistical power (Altman & Royston, 2006). Findings can also vary substantially depending on the cut-off value used (Kraemer et al., 2004). These limitations may have impacted the sensitivity of Scott et al.'s (2004) analysis and the robustness of their results. Moreover, they did not evaluate the unique impact of those dichotomized moderators while controlling for the influence of the other moderators measured — as such, the effects reported may be due in part to the moderators' correlations with other factors. Given the high plausibility of some moderators being correlated (e.g., sample size and total training time as it is often more practical to evaluate long interventions on a smaller sample), the findings reported by Scott et al. (2004) may have overstated or understated the importance of certain moderators. In addition, the statistical significance of the effects of many moderators examined was not reported.

By comparison, Ma (2006) only examined a very limited number of moderators (i.e., only 6) but did report whether their effects were statistically significant. Nonetheless, many effect sizes in Ma's analysis originated from the same studies, and it is unclear how the dependency between these effect sizes was handled. This issue can dramatically impact the precision of estimates and statistical comparisons, potentially leading to misleading conclusions (Borenstein et al., 2009; Hedges, 2009). As in Scott et al. (2004), Ma (2006) also did not test the unique contribution of each individual moderator measured.

Other Methodological Limitations. A number of other limitations could impact the validity of existing meta-analyses' conclusions. One is the inclusion of studies with heterogeneous designs. Scott et al. (2004), for example, included both studies with and without a control group, the latter type generally yielding larger and potentially inaccurate effect sizes (Cheung & Slavin, 2013). Studies focusing exclusively on atypical populations were also included in previous meta-analyses. For example, both Ma (2006) and Scott et al.

(2006) included studies that involved only gifted students or only students with learning disabilities. Homogeneous populations are likely to underestimate the variation in the general population (and thus increase standardized measures of an intervention's effect), impacting the generalizability of the findings.

Because of these limitations, the results of previous meta-analyses may not reflect the effectiveness of creativity training in a way that is informative to most researchers and practitioners. In fact, some of these limitations may have led previous meta-analyses to overestimate, perhaps greatly, the impact creativity training programs are likely to achieve in practice. Practitioners and researchers using the results of these meta-analyses to inform their work may, as a result, develop inaccurate expectations and misplaced confidence in the effectiveness of creativity training. As such, it is important to update and extend previous meta-analyses of creativity training in an effort to provide a more accurate picture of the effectiveness of these programs.

Concerns About the Quality of Creativity Training Studies

A number of researchers have expressed concerns about the quality of studies evaluating creativity training programs. For example, reviewing 22 evaluations of creativity training programs from the field of organization science, Valgeirsdottir and Onarheim (2017) noted that the studies often lacked a pretest and a control group and had small sample sizes — features that limit the credibility of the causal estimates of an intervention effect (i.e., internal validity; Boot et al., 2013; Lonati et al., 2018). These concerns are not new: more than four decades ago, Mansfield et al. (1978) drew comparable observations after reviewing studies evaluating creativity training programs in wide use at the time. These review reports raise important issues, but focused only on a subset of creativity studies within a certain period. Do those concerns generalize beyond those subsets, and has research quality improved over time? Part of the present study aims at addressing these questions.

The Present Study

In this study, we conducted a meta-analysis on the effectiveness of creativity training programs, assessing publication bias. We also evaluated the methodological quality of the included studies, tracked improvements over time, and examined how quality influences effect sizes.

Table 1

Features of Previous Meta-Analyses of Creativity Training

Meta-analysis	Population	Year included	No. of Articles		Excluded studies with no control group	Test for publication bias	Correct for publication bias	Mention the use of weights in estimation	Average effect size *d*
			Unpublished	Published					
Rose and Lin, 1984	General	1960 to 1972	33 (72%)	13 (28%)	Yes	No	No	No	0.47
Scott et al., 2004	General	1965 to 2000	6 (14%)	38 (86%)	No	No[a]	No	No	0.64
Ma, 2006	General	1970 to 2003	3 (9%)	30 (91%)	Yes	No	No	No	0.76
Tsai, 2013	Adult	1989 to 2006	1 (9%)	10 (91%)	Yes	No	No	No	0.81
Yasin and Yunus, 2014	Technology and Engineering	2000 to 2012	3 (19%)	13 (81%)	Yes	No	No	No	1.02

Note. A positive effect size (*d*) means that the group receiving creativity training performed better than the control group.

[a] Compared the effect sizes between large (average *d* = 0.35) and small studies (average *d* = 1.00).

Estimation of the Average Effect Size

We first examined the effect of creativity training on creative performance and took several steps to address the limitations of previous meta-analyses. To minimize bias caused by mainly relying on published evidence, we have included a considerable number of unpublished studies (48 out of 169 studies; 262 out of 844 effect sizes). We also conducted a series of tests to investigate the presence of publication bias and correct for it if necessary. We excluded studies with no control group as having a control group is often considered a necessary condition for a study to properly estimate the causal impact of an intervention (What Works Clearinghouse, 2020). Unlike previous meta-analyses (Ma, 2006; Scott et al., 2004), we also excluded studies focusing solely on atypical populations. In terms of coding, we avoided dichotomizing continuous variables, e.g., publication year. When possible, we coded them as continuous variables, and when impractical, we categorized them into multiple levels. In terms of analysis, we used robust variance estimation (RVE; Hedges et al., 2010; Tanner-Smith & Tipton, 2014), a recently developed meta-analytic approach, to handle effect size dependency when synthesizing effect sizes and examining the influence of moderators on effect sizes. Compared to the traditional approaches used in previous meta-analyses (i.e., aggregating dependent effect sizes), RVE considers all the effect sizes in a single meta-analysis model, thus minimizing the loss of information (Hedges et al., 2010; see Method section for more details).

Moderator Analysis

In addition to estimating the average effect size, we also explored the impact of a range of study features on effect size. We included many of the moderators explored in previous meta-analyses, but excluded some, such as the depth and the difficulty of the training materials, due to their high subjectivity in coding and the frequent absence of relevant information in the included studies (see supplemental material Table S1 for excluded moderators and the rationale for their exclusion). Moreover, we included new moderators that could be coded reliably and that have been shown to influence effect size, such as the level of randomization and whether the outcome measure is researcher-made or not (Cheung & Slavin, 2016; Wilson & Lipsey, 2001). We categorized the included moderators into three groups: (a) sample characteristics, (b) training characteristics, and (c) study characteristics.

Sample Characteristics.

Participant Age. Previous meta-analyses reported inconsistent results regarding the effect of creativity training on different age groups. Scott et al. (2004) categorized participants into two groups (younger than 14 vs. 14 or older) and reported similar effect sizes between groups. Ma (2006) split participants into five age groups (kindergarten, elementary school students, high school students, college students, and adults) and found a larger training effect for adults than younger participants. Because many included studies were conducted on students across various grade levels within a school, yet only reported the aggregated effect (preventing us from coding age as a continuous variable), we followed Ma's granular categorization and re-examined how age group moderates effect size.

Training Characteristics.

Training Content. Scott et al. (2004) reported that training based on a cognitive framework generated larger effect sizes than training based on non-cognitive frameworks (e.g., motivation and personality). This meta-analysis also examined this issue, comparing the effectiveness of training targeting only cognitive skills, only non-cognitive skills, or a combination of both.

Total Training Time. It is typically expected that more training time results in a greater training effect. According to the meta-analysis of Scott et al. (2004), there was a positive bivariate correlation between training time (in minutes) and training effectiveness. However, Scott et al. (2014) did not report enough details to determine the degree to which this correlation differs from what would be expected by chance. In this meta-analysis, we examined if there was a positive link between total training time and effect size by including total training time as a potential moderator.

Training Duration (i.e., The Time Interval Between the First and the Last Training Session). In addition to total training time, one could argue that a more spread-out training would allow, to a certain extent, more time for knowledge and skills to be consolidated and thus lead to a larger effect, i.e., the spacing effect (Ebbinghaus, 1964). We also included this variable in our moderator analysis.

Study Characteristics.

Study Precision. Small, less precise studies often report larger effect sizes (i.e., small-study effects, Sterne et al., 2000). Scott et al. (2004), for example, found that the average effect size was considerably higher for small studies (i.e., studies with a below-average sample size) than for larger studies. Multiple causes can account for this effect: small studies may be more prone to publication bias, or they may share methodological features that tend to generate larger effect sizes (e.g., evaluating more intensive interventions; Sterne et al., 2001). Scott et al. (2004) attributed the effect size difference to methodological differences between small and large studies and did not test for publication bias. In the present meta-analysis, we included the standard error of the effect size (a common measure of study precision) to evaluate if there is a relationship between effect size and study precision. Furthermore, we explored whether the association between effect size and study precision remains after controlling for a number of study characteristics. We also used the standard error of the effect size to estimate the proportion of properly powdered studies in our sample when evaluating the quality of the existing evidence.

Randomization: Many creativity training studies have assigned existing clusters of individuals (e.g., classrooms) to either the training or control groups, but analyzed their data at the individual level (ignoring clustering). This approach is known to yield larger effect sizes, potentially due to uncontrolled baseline differences between clusters (e.g., Cheung & Slavin, 2016). We examined this issue by comparing such studies (which we label *non-random assignment*) with those that randomly assign individual participants to conditions (i.e., *random assignment*). Ma's (2006) meta-analysis of creativity training did not detect a noticeable impact of the type of randomization used; however, this may be attributed to the limited statistical power of this contrast in their meta-analysis. The current meta-analysis re-examined the impact of this factor on the effectiveness of creativity training. We also considered this factor when evaluating the quality of the evidence in the field.

Type of Control Group. Different types of control groups have been employed in creativity training studies. Whereas some studies offered an alternative, non-creativity, training to participants in the control group (i.e., active control group), some employed an inactive control group in which participants only maintained their usual activities. Studies with inactive control groups have the potential to overstate training effectiveness because the training effect may be conflated with the novelty effect induced by altering the participants' usual activities (Mohr et al., 2009). Accordingly, meta-analyses of interventions in other domains have shown that the control group type often moderates the training effect (e.g., Friese et al., 2017). The type of control group was included as a potential moderator and as a measure of study quality when evaluating the quality of the evidence in the field.

Use of a Pretest. Scott et al. (2004) considered the use of a pretest as an indicator of study quality because it helps disentangle treatment effects from effects due to pre-existing group differences. They found that studies using a pretest-posttest design reported smaller effect sizes than those using a posttest-only design. We included this as a potential moderator in the present meta-analysis. We also considered the use of a pretest as a measure of study quality when evaluating the quality of the evidence in the field.

Type of Outcome Measure. There are various ways to measure creative performance. Scott et al. (2004) categorized the outcome measures into divergent thinking, creative problem solving, performance (e.g., the generation of creative products), and creative attitude and behaviors. They reported that effect sizes on divergent thinking and creative problem solving were larger than on other outcome measures. In contrast, Ma (2006) did not observe notable differences between different types of outcome measures. To further examine this question, we included the type of outcome measure as a moderator.

Origin of Outcome Measure. Research has frequently shown that studies using researcher-made measures tend to yield larger effect sizes than studies using independent measures such as standardized tests (e.g., Cheung & Slavin, 2016; Wilson & Lipsey, 2001), perhaps due to researcher-made measures being more aligned with the content of the intervention (What Works Clearinghouse, 2020). We explored whether this effect is also present in the creativity training literature.

Training to Criterion. It is reasonable to expect effect sizes to be larger when the tasks used to measure performance are similar to the activities completed during the training (i.e., training to criterion). Nonetheless, Scott et al. (2004) found that studies using measures similar to training activities did not generate larger effect sizes than those using dissimilar measures. The present meta-analysis re-examined whether the similarity between training and assessment tasks moderates the effect sizes.

Posttest Timing. The interval between the end of training and the posttest varies across creativity training studies. The posttest is sometimes given immediately after training or delivered several days or weeks later. Longer time lags may result in smaller effect sizes due to learning decay (Bailey et al., 2020; Taylor et al., 2009). Nevertheless, Scott et al. (2004) did not find such an effect. We included this moderator in our analysis to re-examine this issue.

Publication Year. Effect sizes reported in studies tend to decline over time, and such decline has been observed in many fields (e.g., Medicine: Ioannidis, 2008; Psychology: Protzko & Schooler, 2017). To examine the impact of publication year, Scott et al. (2004) divided the studies into two groups (pre-1980 studies vs. 1980-and-after studies) and found comparable effect sizes between these two groups of studies. In this meta-analysis, we also included publication year as a moderator, which we treated as a continuous variable.

Publication Status. Comparing the effect size between unpublished and published studies is another way to estimate the extent of publication bias. Previous meta-analyses of creativity training did not examine if there were systematic differences between published and unpublished effect sizes. In this meta-analysis, we examined the impact of publication status on effect size.

The Effect of Creativity Training on Divergent Thinking. In addition to conducting moderator analysis on the whole sample of studies, we also carried out a separate analysis focusing on the effectiveness of creativity training on divergent thinking, given that divergent thinking is often conceptualized as a fundamental component of creative thinking, as well as being a key focus of many creativity training programs (Guilford, 1967; Scott et al., 2004). Previous meta-analyses suggested that creativity training has a strong impact on divergent thinking performance (Rose & Lin, 1984; Scott et al., 2004). Among the different dimensions of divergent thinking performance (e.g., fluency, flexibility, and originality), creativity training tended to have a larger positive impact on measures of originality than on the other outcome measures, such as fluency and flexibility (Rose & Lin, 1984; Scott et al., 2004). We re-assess this question in the present meta-analysis.

Prevalence of Studies Employing Rigorous Methodology. Apart from combining and comparing effect sizes across studies, this meta-analysis also aimed to identify the prevalence of rigorous research practices in the field of creativity training. We examined the proportion of studies in our sample that (a) performed randomization at the individual level, (b) employed an active control group, (c) used a pretest, and (d) had sufficient statistical power (to detect a range of plausible true effect sizes) — conditions known to influence the extent to which a study can provide a credible casual estimate of an intervention effect (i.e., internal validity). We also examined whether this proportion tended to increase over time and whether the average effect of creativity training changed when considering only studies of the highest quality. In addition, we looked at the number of studies that have adopted practices that can further enhance the strength of the evidence (studies reporting a priori power analysis, replication studies, and studies that preregistered their protocol).

Method

Literature Search and Inclusion Criteria

We searched five electronic databases (Business Source Premier, ERIC, ProQuest Dissertation & Theses, PsycARTICLES, PsycINFO, and Web of Science Core Collection) to identify articles containing one of the following key terms in the abstract or the title: “*creativity training,*” “*creative thinking training,*” “*creative problem-solving training,*” “*creative problem solving training,*” “*creativity intervention,*” “*creative thinking intervention,*” “*creative problem-solving intervention,*” “*creative problem solving intervention,*” “*creativity program,*” “*creative thinking program,*” “*creative problem-solving* *program*,” or “*creative problem solving program*.”[7] All databases were searched from 1970 to June 2022. We also checked the reference lists of previous meta-analytic and systematic reviews for relevant studies and used Google Scholar to search within papers citing these reviews to identify articles that contain the key terms. Additionally, key journals in creativity research: *The Creativity Research Journal, Journal of Creativity Behavior, Psychology of Aesthetics, Creativity, and the Arts,* and *Thinking Skills and Creativity*, were searched to locate articles containing any of the key terms. To reduce the impact of publication bias, we searched conference proceeding and preprint databases (ArXiv, ERIC, Procedia, OSF preprint, and Web of Science) for unpublished articles that contain the key terms; we also sent out a call for unpublished work (see OSF link: https://osf.io/8z3k5) to the professional networks of the research team (sent on 21st and 23rd of June-2022), as well as the members of APA Division 10: Society for the Psychology of Aesthetics, Creativity and the Arts (sent on 27th of June-2022).

In total, we identified 450 articles. We screened the titles and abstracts of these articles for relevance to the present work. Of these, 154 articles were excluded. The full texts of the remaining articles (n = 296) were obtained and assessed for eligibility according to the following inclusion criteria:

- Focuses on the effect of training specifically designed to enhance creativity.
- Focuses on the effect of creativity training on individual creativity. Studies examining training effectiveness on team creativity were excluded.
- Focuses on typical populations. Studies only focusing on atypical populations, e.g., gifted children and individuals with learning difficulties, were excluded.
- Includes a control group. As the inclusion of a control group is often considered a necessary feature to establish the efficacy of an intervention, studies without a control group and studies comparing different types of creativity programs were excluded.
- Measures the effect of creativity training on creative performance. Studies using only non-performance-based outcome measures (e.g., self-efficacy and motivation) were

[7] In line with other meta-analyses on creativity training, we excluded “innovation training” from our search terms. Innovation training studies, commonly conducted in organizational settings (Amabile, 1988), often exhibit distinct features not aligned with our inclusion criteria. For instance, they often rely on team performance measures and lack control groups (e.g., Nielsen & Miraglia, 2017). As such, including the term “innovation training” could have led to a non-representative set of studies. Furthermore, methodological differences between innovation and creativity training, such as differences in participants’ age, could bias our conclusions.

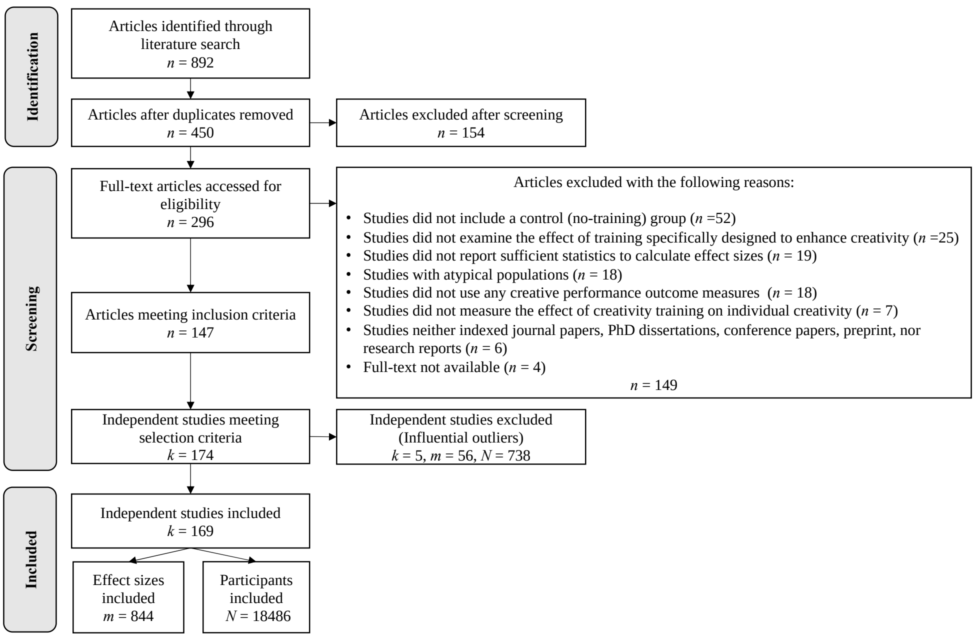
excluded. For studies employing a mix of performance- and non-performance-based creativity outcomes, only performance-based outcomes were considered.

- Reports the information needed to calculate an effect size and associated uncertainty (see section *Estimation of Effect Sizes* for details).

Based on these criteria, 149 of 296 articles were excluded. Of the remaining 147 articles, 105 were journal articles, 34 were PhD dissertations, seven were conference proceedings, and one was a grant report. For multiple versions of the same research study (e.g., dissertation and journal article), only the version published as a journal article was coded. Some articles (24 out of 147) reported multiple independent studies (e.g., Chiu, 2015). Ultimately, 169 independent studies were identified. Figure 1 presents the flowchart that outlines the search and selection process. Supplemental material Table S2 indicates which articles were excluded and provides the reasons for exclusion. In supplemental material Table S2, we also indicated which articles were included in previous meta-analytic and systematic reviews.

Figure 1

PRISMA Flow Diagram Outlining the Stages of the Search and Selection Process



Note. *n* = number of articles; *k* = number of studies; *m* = number of effect sizes; *N* = number of participants.

Coding Procedure

We developed a coding protocol to extract the following information from the studies.

- **Identifying Information.** Authors, title, and publication year.
- **Publication Status.** This was coded categorically (Unpublished vs. Published). PhD dissertations, conference proceedings, preprints, and research reports were coded as "Unpublished." Studies published in a peer-reviewed journal were coded as "Published."
- **Age Group**. This was coded categorically based on participants' grade levels. There were five levels: (a) Preschool/Kindergarten, (b) US Grade 1-8, (c) US Grade 9-12, (d) University Students, and (e) Adults. These categories were drawn from Ma (2006).
- **Training Content.** This was coded categorically based on the content of the training. There were three levels: (a) Non-cognitive (i.e., training focusing on non-cognitive components, e.g., personality, motivation, and emotion); (b) Cognitive (i.e., training focusing on cognitive components, e.g., creative thinking and idea generation skills, memory, and attention); and (c) Combined (i.e., training focusing on both cognitive and non-cognitive components).
- **Total Training Time (in Minutes).** This was coded continuously based on the overall amount of training measured in minutes. We estimated this based on the number of training sessions and the duration of each session. If the study was a classroom study and only reported the number but not the length of each lesson, we assumed the lesson length to be 50 minutes. A small number of studies only presented participants with a set of instructions (e.g., about how to be creative), without specifying the training time. We assumed training time in these studies to be 5 minutes.
- **Training Duration (in Days).** This was coded continuously based on the days between the first and last training sessions. If the study only reported the number of months, we assumed 4.33 weeks per month (i.e., 52 weeks/12 months). Training programs lasting less than a day were coded as one day.
- **Randomization.** This was coded categorically. Studies were coded as "Random" if individuals were randomly assigned to different conditions. Studies were coded as "Non-random" if assignment to conditions was based on pre-existing clusters (e.g., classrooms). If the assignment was unspecified, it was coded as "Unspecified."

- **Type of Control Group.** This was coded categorically (Active vs. Inactive). Control groups that received non-creativity training were coded as "Active"; Control groups that did not receive any additional training and continued to maintain their usual activities were coded as "Inactive."[8]
- **Use of a Pretest.** This was coded categorically (Posttest-Only vs. Pretest-Posttest). If studies measured pretest performance, they were coded as "Pretest-Posttest." Otherwise, they were coded as "Posttest-Only."
- **Type of Outcome Measure.** This was coded categorically based on the type of measure used. There were three categories (a) Divergent thinking, (b) Creative problem solving, and (c) Creative product. This categorization was based on Scott et al. (2004).
- **Origin of Outcome Measure.** This was coded categorically (Researcher-made vs. Independent). Measures developed for the purposes of the study were coded as "Researcher-made"; Pre-existing measures were coded as "Independent"; If no external source was mentioned, the measure was coded as "Researcher-made."
- **Training to Criterion.** This was coded categorically (Yes vs. No) based on whether the training program included only exercises that were very similar to the posttest measures. For example, if the training program included only divergent thinking exercises and the divergent thinking test was the posttest measure (e.g., Franklin et al., 1977), the training program was coded as training to criterion.
- **Posttest Timing.** This was coded categorically based on the days between the end of training and the posttest. There were four levels: (a) Same day, (b) Between 1 and 7 days, (c) 8 days or longer, and (d) Unspecified. These categories were adapted from Uttal et al. (2013).
- **Modality of Divergent Thinking Measures.** This was coded categorically. For studies using a divergent thinking test as a posttest measure, the modality of the divergent thinking test was coded as (a) Verbal, (b) Figural, (c) Combined, and (d) Other/unspecified.
- **Dimension of Divergent Thinking Measures.** This was coded categorically. For studies using a divergent thinking test as a posttest measure, the dimension of divergent thinking measured was also coded as (a) Overall; (b) Fluency; (c)

[8] A small number of studies had more than one control group, i.e., active control group and inactive control group, but only one treatment group (e.g., Below, 1986). In such cases, the active control group was considered.

Flexibility; (d) Originality; and (e) Other dimensions, which included, for example, how elaborate the ideas generated were.

- **Use of a Delayed Posttest.** This was coded categorically based on whether the study included a delayed posttest. Given the very small number of studies with a delayed posttest, we did not compute the effect size associated with the delayed posttest for statistical analysis.
- **Power Analysis.** This was coded categorically based on whether the study included a priori power analysis or not.
- **Replication.** This was coded categorically based on whether the study reported replicating a previous study or not. We looked at studies that contained the term “replicat*” to assess whether it was a replication of a previous study. This method was similar to that used in Makel and Plucker (2014).
- **Preregistered.** This was coded categorically based on whether the study reported being preregistered or not.
- **Study Location.** This was coded categorically based on the continent where the study was conducted: (a) Asia, (b) Australia, (c) Europe, (d) North America, (e) South America, and (f) Unspecified. If the study location was not mentioned, but all authors were from the same continent, we assumed that continent to be the location of the study. Otherwise, we coded the study location as unspecified. For studies conducted online with no mention of restriction on the participants’ location, we coded the study location as unspecified.
- **Information Needed to Compute Effect Size Estimate and Associated Uncertainty.** Sample size of the treatment and control groups, means and standard deviations, or other statistical information (e.g., *t*-values, *F*-values, *p*-values) was needed to compute effect sizes and associated variances.

The coding was carried out by both the first and second authors. Together, they coded 50% of the studies (87 in total). For the remaining 50%, the first author initially screened the studies to identify the general sections containing information relevant to each moderator and effect size calculation. After this initial screening, both the first and second authors coded the relevant information independently. To measure interrater reliability, we calculated Cohen’s kappa for each categorical moderator and the intraclass correlation coefficient (ICC) for each continuous moderator. The interrater reliability indices ranged from 0.89 to 1, suggesting high agreement. Any disagreement was settled through discussion among the authors (see supplemental material Table S3 for the interrater reliability index for each moderator).

Estimation of Effect Sizes

Computing Effect Sizes and Variances. We computed the standardized mean difference between the training and control groups (Cohen's *d*) for each outcome measure and used it as the effect size index in the present meta-analysis. Depending on the study design (posttest-only or pretest-posttest), different equations were used to compute the effect size.

For studies with a posttest-only design, the effect size was computed by dividing the mean difference between the training and control group means by their weighted pooled standard deviation (Morris, 2008):

$$d = \frac{M_{training,\ post} - M_{control,\ post}}{\sqrt{\frac{\left(N_{training} - 1\right) \times \left(SD_{training,\ post}\right)^2 + \left(N_{control} - 1\right) \times \left(SD_{control,\ post}\right)^2}{\left(N_{training} + N_{control} - 2\right)}}}$$

If only adjusted posttest Means and *SD*s (adjusted for pretest) were provided, the effect size was computed by dividing the adjusted mean difference between the training and no-training group by their weighted pool standard deviation:

$$d = \frac{M_{training,\ adjusted\ post} - M_{control,\ adjusted\ post}}{\sqrt{\frac{\left(N_{training} - 1\right) \times \left(SD_{training,\ adjusted\ post}\right)^2 + \left(N_{control} - 1\right) \times \left(SD_{control,\ adjusted\ post}\right)^2}{\left(N_{training} + N_{control} - 2\right)}}}$$

For a pretest-posttest design, the effect size was computed by dividing the difference of mean improvement between the training group and the control group by their weighted pooled pretest standard deviation (Morris, 2008):

$$d = \frac{\left(M_{training,\ post} - M_{training,\ pre}\right) - \left(M_{control,\ post} - M_{control,\ pre}\right)}{\sqrt{\frac{\left(N_{training} - 1\right) \times \left(SD_{training,\ pre}\right)^2 + \left(N_{control} - 1\right) \times \left(SD_{control,\ pre}\right)^2}{\left(N_{training} + N_{control} - 2\right)}}}$$

If studies only reported the total sample size and did not specify the size of each group, the sample was evenly divided between groups. If means and standard deviations were not provided, effect sizes were calculated from *t*-value, *F*-value, χ^2, or *p*-value. If only a *p*-less-than value was reported, instead of an exact *p*-value, the *p*-less-than value was treated as an exact *p*-value. For the few studies that only provided statements of significant or non-significant differences between the training and control groups, we assumed $p = .05$ or $p = .50$, respectively (Cooper et al., 2009). This procedure was used only when the direction of the effect was indicated; otherwise, the studies were excluded. Once the effect sizes were estimated, we estimated their variances using the equation (Lipsey and Wilson, 2001)[9]:

$$var = \frac{N_{training} + N_{control}}{N_{training} \times N_{control}} + \frac{d^2}{2 \times (N_{training} + N_{control})}$$

Correction for Small-Sample Bias. We converted the effect sizes *d* to unbiased ones (Hedges' *g*) by multiplying each effect size *d* with a correction factor *J*. We also applied a similar correction to effect size variances, multiplying each variance by the factor J^2 (Hedges, 1981):

$$J = 1 - \frac{3}{4 \times \left(N_{training} + N_{control} - 2\right) - 1}$$

Most studies (82%) provided multiple and dependent effect sizes due to multiple outcome measures evaluated on the same participant or multiple training groups sharing the same control group. To handle effect size dependency, we conducted the meta-analysis with robust variance estimation techniques (RVE, Hedges et al., 2010; Tanner-Smith & Tipton, 2014), which we will discuss later in this section.

Outlier Detection

Prior to the main analysis, we performed a set of diagnostics to identify influential outliers[10]. Because these procedures assume independence between effect size estimates, they were performed on effect sizes aggregated at the study level. Studies identified as having

[9] Eight studies (44 effect sizes) reported only the *F*-value, or the adjusted means and standard deviations associated with ANCOVA. For these studies, we used the formula recommended by Borenstein (2009) to estimate variance.

[10] The diagnostics are the externally standardized residuals, DFFITS values, Cook's distances, covariance ratios, leave-one-out estimates of the amount of heterogeneity, leave-one-out heterogeneity test statistics, hats value, and weight.

extreme results were excluded from the analyses. We first used the *R*-package “MAd” (Del Re & Hoyt, 2014) to compute the aggregated effect size and its variance for each study while taking into account the correlation between within-studies effect sizes. Given that most of the studies did not report this correlation, we used the suggested value of $r = .50$ (Wampold et al., 1997). We then used the *R*-package “metafor” (Viechtbauer, 2010) to perform the diagnostic tests on aggregated effect sizes. We also performed the same diagnostic tests using a range of correlation values ($r = 0, .20, .40, .60, .80, 1$) as a sensitivity analysis to ensure that the results were not sensitive to the value of the assumed correlation.

Meta-Analytic Model with RVE

We used robust variance estimation techniques (RVE; Hedges et al., 2010; Tanner-Smith & Tipton, 2014) to deal with effect size dependency in our sample. These techniques assign a weight to each effect size based on its precision, taking into account effect size dependency. More specifically, RVE uses ρ, an estimate of the common correlation between dependent effect sizes, to compute effect size weights. In this study, we used the suggested value of $\rho = .80$. We also conducted a sensitivity analysis with a range of plausible correlations ($\rho = 0, .20, .40, .60, .80$, and 1) to ensure that the value of ρ did not influence the results meaningfully (see supplemental material Table S4). We used the *R*-package “robumeta” (Fisher & Tipton, 2015) to construct random-effects meta-regression models with RVE to synthesize and compare effect sizes.

Overall Effect Size

We constructed an intercept-only-random-effects meta-regression model with RVE to compute the weighted average effect size and its standard error (Fisher & Tipton, 2015). For comparison, we also computed the overall effect size using the traditional approach of aggregating effect sizes (see Borenstein et al., 2009). We followed the aggregation method detailed in the Outlier Detection section above. We used the *R*-package “metafor” to construct random-effect meta-regression models to estimate the overall effect size and its standard error (see supplemental material Table S5). The results obtained were virtually identical to those generated by the RVE approach. Thus, we only reported the results of the RVE approach in the main text.

Publication Bias

Publication bias refers to situations where the published literature is not a representative reflection of the overall research evidence (Rothstein et al., 2005). It occurs when studies are more likely to be published if they report favorable and statistically significant findings, such that published effect sizes are systematically different from the unpublished ones; and this bias tends to be more likely to affect small studies (Sterne et al., 2000). If publication bias exists, the averaged effect size estimate largely based on published studies is likely to be inflated because of the underrepresentation of non-significant or negative findings. To reduce the impact of publication bias on the meta-analytic results, we conducted a thorough search to identify relevant unpublished studies. We also undertook a series of analyses to evaluate and, if necessary, correct for publication bias in our sample.

Detecting Publication Bias. We used three methods to evaluate the presence of publication bias quantitatively. First, we tested if there was a systematic difference between published and unpublished effect sizes. To do so, we constructed a random-effects meta-regression using the RVE weighted effect size as the dependent variable and the publication status (i.e., published vs. unpublished) as the predictor variable to examine if there was a significant association between the two.

Second, we examined if small studies with negative or null findings were underrepresented in our sample, an observation that is often taken as suggestive evidence of publication bias (Egger et al., 1997; Sterne et al., 2001). To examine this, we created a funnel plot of the unbiased effect size estimates against their precision (i.e., standard error). We then assessed its asymmetry, focusing on the potential absence of small studies with negative or null findings (the lower-left corner of the plot). More specifically, we performed Egger's test to formally evaluate funnel plot asymmetry by regressing the weighted effect sizes on their standard errors (Egger et al., 1997). If publication bias exists and causes the omission of small studies with negative or null findings, effect sizes from small studies (i.e., less precise, with larger standard error) would tend to be larger than those from large studies, and a positive relation between effect size and standard error would be expected. We conducted Egger's test within the RVE framework for handling effect size dependency (Rodgers & Pustejovsky, 2021).

It should be noted that Egger's test assumes that publication bias operates based on the size of the effect rather than its statistical significance. The test also only offers an all-or-none decision on whether publication bias exists and does not quantify the evidence for the presence vs. absence of publication bias. To address these issues, we used a third approach —

robust Bayesian meta-analysis method (RoBMA, Maier et al., 2022) — to assess and quantify the evidence for publication bias. RoBMA draws inference by simultaneously considering multiple meta-analytic models based on different assumptions (e.g., publication bias acting on the magnitude or on the significance level of the effect) and computes a model-averaged Bayes factor to quantify the evidence for the presence vs. absence of publication bias. We used the *R*-package "RoBMA" (Bartoš & Maier, 2020) to conduct the analysis. As RoBMA does not account for the dependency of within-study effect sizes, we conducted the analysis on aggregated effect sizes. As a test of robustness, we also performed the analysis on randomly selected effect sizes (one per study). The results were similar to those on aggregated effect sizes (see supplemental material Table S6). We only reported the results of RoBMA conducted on aggregated effect sizes here.

Correcting for Publication Bias. If our analyses suggest the presence of a substantial publication bias, the overall effect size estimate will be corrected for publication bias. There are different correction methods, each making different assumptions and having different limitations (Carter et al., 2019; Inzlicht et al., 2015). Therefore, as a test of robustness, we employed four different correction methods and evaluated whether they yielded similar estimates. The correction methods used were PET-PEESE (Stanley & Doucouliagos, 2007), trim-and-fill (Duval & Tweedie, 2000), RoBMA (Maier et al, 2022), and Top10% (Stanley et al., 2010).

The first method, PET-PEESE, estimates the corrected overall effect size by regressing the effect sizes on their standard errors (PET) and variances (PEESE). The intercept of the regression model, which represents the effect size when standard error or variance is equal to zero (akin to an infinite sample size), could be interpreted as the corrected effect size estimate (Stanley & Doucouliagos, 2007). The PEESE intercept is often viewed as a more reliable estimate when the PET intercept is statistically significant at $\alpha = 0.10$; otherwise, the PET is considered the more reliable estimate (Stanley, 2008; Stanley & Doucouliagos, 2014; although see e.g., Inzlicht, 2015; Reed, 2015 for different views). We used the *R*-package "metafor" (Viechtbauer, 2010) to perform the PET-PEESE analysis. The second method, trim-and-fill, estimates the number of missing studies that might exist in the meta-analysis based on the funnel-plot asymmetry and estimates a new overall effect size with the imputed missing studies (Duval & Tweedie, 2000). We used the *R*-package "meta" (Balduzzi et al., 2019) to conduct the trim-and-fill analysis. Note that both PET-PEESE and trim-and-fill correct for publication bias by modeling the relationship between effect sizes and standard errors. There are other bias-correction methods based on the distribution of *p*-

values (Simonsohn et al., 2014). To take both models into consideration, the third method, RoBMA, combines these different bias-correction models to form a model-averaged adjusted estimate. The last correction method, Top10%, estimates the overall effect size based on the top 10% of most precise studies in terms of standard error (Stanley et al., 2010; van Aert et al., 2019).

As PET-PEESE, trim-and-fill, and RoBMA assume independence among effect sizes, we performed these bias-correction methods on aggregated effect sizes. We also performed them on randomly selected effect sizes (one per study) as a test of robustness and obtained similar results. We presented the results based on aggregated effect sizes in the main text. For the results based on randomly selected effect sizes, see supplemental material Table S7. For the Top10% method, we selected the 10% most precise studies in terms of the standard errors of the aggregated effect sizes. We then constructed a meta-regression model with RVE to estimate the average effect size of these most precise studies.

Moderator Analysis

The Influence of Study Characteristics on Effect Size. We conducted a moderator analysis to investigate the potential source of between-studies heterogeneity of the effect sizes. The following moderators were examined: *Participants' Age Group, Training Content, Total Training Time, Training Duration, Study Precision (i.e., standard error of the effect size), Randomization, Type of Control Group, Use of a Pretest, Type* and *Origin of Outcome Measure, Training to Criterion, Posttest Timing, Publication Year* and *Status* (see the section Coding Procedure for the description of these moderators). To examine the relation between each moderator and effect size, we constructed multiple meta-regression models, each regressing the RVE weighted effect size on one single moderator. We then included all the moderators in a single meta-regression with RVE to examine the unique influence of each moderator.

The moderators *Total Training Time* and *Training Duration* were log-transformed (base 2) to reduce skewness. All the categorical moderators were dummy coded. For categorical moderators with more than two levels (e.g., *Type of Outcome Measure*), we used the *R*-package "clubSandwich" (Pustejovsky, 2016) to conduct an omnibus *F*-test to determine if there was a statistically significant difference among all levels of the moderator.

The Effect of Creativity Training on Divergent Thinking. Given that divergent thinking is the basis of many creativity training activities and is often considered a fundamental aspect of creativity, we constructed a meta-regression model examining the effect of creativity training on divergent thinking performance. We tested if creativity training would have differential effects for different dimensions of divergent thinking. Only studies measuring divergent thinking performance were included in this analysis. We focused on the fluency, flexibility, and originality of the ideas generated because they are the core dimensions of divergent thinking. We constructed a meta-analysis model with the RVE weighted effect size as the dependent variable and the dimension of divergent thinking as the predictor variable. The modality of the measure (verbal vs. figural), along with the significant moderators identified by the moderator analyses on the whole sample, were included as predictor variables to control for their potential influence on effect size.

Estimating the Proportion of Studies Employing Rigorous Methodology

To assess the methodological quality (i.e., internal validity) of the studies in our sample, we examined the proportion of studies employing the following rigorous methodological features: (a) randomization at the individual level[11], (b) use of an active control group, (c) use of a pretest, and (d) adequate statistical power. Whether a study has adequate statistical power depends on the size of the effect it plans to examine. Given that the true effect of creativity training is unknown, we calculated the proportion of studies that have at least 80% power for detecting plausible effect size estimates (e.g., the overall effect size estimated by this meta-analysis and the average effect corrected for publication bias). This calculation was achieved by using the standard error of the aggregated effect size estimate from each study.

To determine whether there is an improvement in study design over time, we performed logistic regressions to examine if publication year predicts whether studies (a) are randomized at the individual level, (b) use an active control group, and (c) use a pretest. We also conducted a correlation analysis to examine whether publication year predicts the standard error of the effect size (a measure of sample size) and performed a sensitivity analysis to determine if the average effect size estimate changes when only considering the highest-quality studies in our sample. We also calculated the proportion of replication and preregistered studies so as to evaluate the progress of the field in response to the recent call for more rigorous research practices.

[11] Given that all included studies analyzed the data at the individual level, without considering clustering, the absence of individual-level randomization presents an issue.

Transparency and Openness

This meta-analysis was not preregistered. All data, analysis code, and the open call for unpublished data sent to professional networks and members of APA Division 10 can be found here: https://osf.io/8z3k5. The list of all articles accessed for eligibility and reasons for exclusion, interrater reliability score for each moderator, and sensitivity analyses are available in the supplemental material.

Results

Characteristics of Included Studies

A total of 174 studies met the inclusion criteria. Among them, five studies (totaling 56 effect sizes) were excluded because they were identified as influential outliers. The final sample included 169 studies and 844 effect size estimates, an average of 4.99 effect sizes per study. Eighty-four percent of these effect sizes (712 of out the 844 effect sizes) were positive. The total number of participants was 18486, and the median number of participants per effect size was 53. Tables 2 and 3 present the descriptive statistics for the included studies.

Table 2

Descriptive Statistics for Included Studies

Categorical Characteristics	Number of Studies	Number of Effect Sizes
Overall	169	844
Sample Characteristics		
Age Group		
Preschool/Kindergarten	15	73
US Grade 1-8	59	355
US Grade 9-12	9	31
University Students	68	313
Adults	18	72
Training Characteristics		
Training Content[a]		
Non-Cognitive	8	20
Cognitive	122	590
Combined	52	234
Study Characteristics		
Randomization		
Non-Random	74	351
Random	82	445
Unspecified	13	48

Table 2 (*continued*)

Type of Control Group		
Inactive	110	548
Active	59	296
Use of a Pretest[b]		
Posttest Only	61	288
Pretest-Posttest	108	556
Type of Outcome Measure[a, b]		
Divergent Thinking	137	685
Creative Problem Solving	30	76
Creative Product	29	74
Origin of Outcome Measure[a]		
Researcher-Made	43	125
Independent	136	719
Training to Criterion[a]		
No	150	725
Yes	25	119
Posttest Timing		
Same Day	68	314
Between 1-7 Days	42	229
8 Days or Longer	8	49
Unspecified	51	252
Publication Year[c]		
1970-1979	29	203
1980-1989	22	128
1990-1999	26	159
2000-2009	25	74
2010-2022	67	280
Publication Status		
Unpublished	48	262
Published	121	582
Use of a Delayed Posttest		
No	146	710
Yes	23	134
Power Analysis		
No	158	779
Yes	11	65

Table 2 (continued)

Replication		
No	160	786
Yes	9	58
Preregister		
No	168	831
Yes	1	13
Study Location		
North America	97	497
Asia	39	132
Europe	26	169
Australia	2	5
South America	1	24
Unspecified	4	17

Continuous Characteristic	Median	*M* (*SD*)
Training Characteristics		
Total Training Time (in Minutes)	360	730.63 (1182.70)
Training Duration (in Days)	21	52.62 (83.99)

[a] Some studies appeared in multiple categories. Thus, the sum of the number of studies from all the categories was larger than 169.
[b] Two studies used multiple types of outcome measures (Divergent thinking and creative problem solving) but reported only the overall effect (2 studies; 9 effect sizes). They were excluded when counting the number of effect sizes for each outcome measure.
[c] *Publication Year* was organized into 5 categories only for ease of presentation. It was kept as a continuous variable in the moderator analysis.

Table 3

Number of Studies by Outcome Measure Combinations

Outcome Measure Combinations	Number of Studies (%)
One outcome measure	
Divergent Thinking only	111 (66%)
Creative Problem Solving only	13 (8%)
Creative Product only	19 (11%)
Two outcome measures	
Divergent Thinking + Creative Problem Solving	16 (9%)
Divergent Thinking + Creative Product	9 (5%)
Creative Problem Solving + Creative Product	0 (0%)
Three outcome measures	
Divergent Thinking + Creative Problem Solving + Creative Product	1 (1%)

Overall Effect

Using meta-regression with RVE, the overall effect size estimate was $g = 0.53$, $p < .001$, 95% *CI* [0.47, 0.59], 95% *PI* [-0.58, 1.63], suggesting a positive effect of creativity training on creative thinking. This estimate is slightly smaller but still similar to the average effect sizes reported in previous meta-analyses on the same topic (Scott et al., 2004: $d = 0.64$; Ma, 2006: $d = 0.76$), all falling into the range of medium-to-strong effect sizes according to Cohen's benchmark criteria (Cohen, 1988). Keeping the outliers in our analysis only had a marginal impact on the average effect size (with outliers: average effect size g = 0.58, $p < .001$, 95% *CI* [0.51, 0.66], 95% *PI* [-0.68, 1.85]). Our meta-analysis reported a substantial amount of heterogeneity in the effect sizes ($I^2 = 87.34\%$, $\tau = 0.56$), suggesting systematic differences between the studies in our sample. We conducted a moderator analysis to explore the potential sources of the between-studies variances. Prior to the moderator analysis, we undertook a series of analyses to detect and, if necessary, correct for publication bias.

Detecting Publication Bias

We evaluated the presence of publication bias in three ways. First, we examined if the published effect sizes differed systematically from the unpublished ones. Second, we performed Egger's regression test to examine funnel plot asymmetry. Third, we used RoBMA to compute a Bayes factor to quantity the evidence for publication bias.

Published vs. Unpublished Studies. We constructed a meta-regression with the RVE weighted effect size as the dependent variable and publication status (0: unpublished, 1: published) as the predictor variable. A significant effect of publication status was found, estimated coefficient, $B = 0.19$, $p = .006$, indicating that published studies yield significantly larger effects than unpublished studies (published studies: $g = 0.58$, $SE = 0.04$; unpublished studies: $g = 0.38$, $SE = 0.06$; see Table 4). The published and unpublished studies had similar sample sizes and comparable proportions of studies (a) randomizing at the individual level, (b) having an active control group, and (c) employing a pretest (see Table 5). Thus, observed differences in effect size seem unlikely due to differences in these methodological features, which have been suggested to impact effect size.

Table 4

Overall Effect Size Estimate by Publication Status

	k	*m*	I^2	τ	*g*	*SE*	95% *CI*	95% *PI*	*p*
All Studies	169	844	87.34%	0.56	0.53	0.03	0.47, 0.59	-0.58, 1.63	< .001
Unpublished	48	262	72.77%	0.40	0.38	0.06	0.27, 0.50	-0.42, 1.19	< .001
Published	121	582	89.53%	0.60	0.58	0.04	0.51, 0.65	-0.61, 1.77	< .001

Note. k = number of studies; *m* = number of effect size estimates. A positive effect size (*g*) means that the group receiving creativity training performed better than the control group.

Table 5

Study Features of Published and Unpublished Studies

	Unpublished studies			Published studies			Diff. between the two groups
Design Features	Proportion of studies			Proportion of studies			χ^2 test
Randomization	44%			50%			$\chi^2(1, 169) = 0.373$, $p = .541$
Active control group	35%			35%			$\chi^2(1, 169) < 0.001$, $p = 1.000$
Pretest	67%			63%			$\chi^2(1, 169) = 0.09$, $p = .769$
	Median	*M*	*SD*	Median	*M*	*SD*	Wilcoxon rank sum test
Participants per effect size	49.08	71.84	61.76	56	107.2	209	$W = 2468$, $p = 0.129$

Funnel Plot and Egger's Regression Test. We also examined if small studies with small or negative effects were underrepresented in our sample. Figures 2a and 2b present the funnel plot of the effect sizes by their standard errors. Consistent with publication bias, there was a lower number of small or negative effect sizes from small studies (the lower-left corner of the plot). To formally test for funnel plot asymmetry, we conducted Egger's test by constructing a meta-regression with the RVE weighted effect size as the dependent variable and the standard error of the effect size as the predictor variable. If publication bias exists such that small studies with null or negative findings are often omitted, effect sizes found in small studies should tend to be larger than those found in large studies. In line with publication bias, there was a positive association between effect size and standard error, $B = 1.76$, $p < .001$, suggesting that small studies (i.e., larger standard error) tended to yield larger effects (see Table 6).

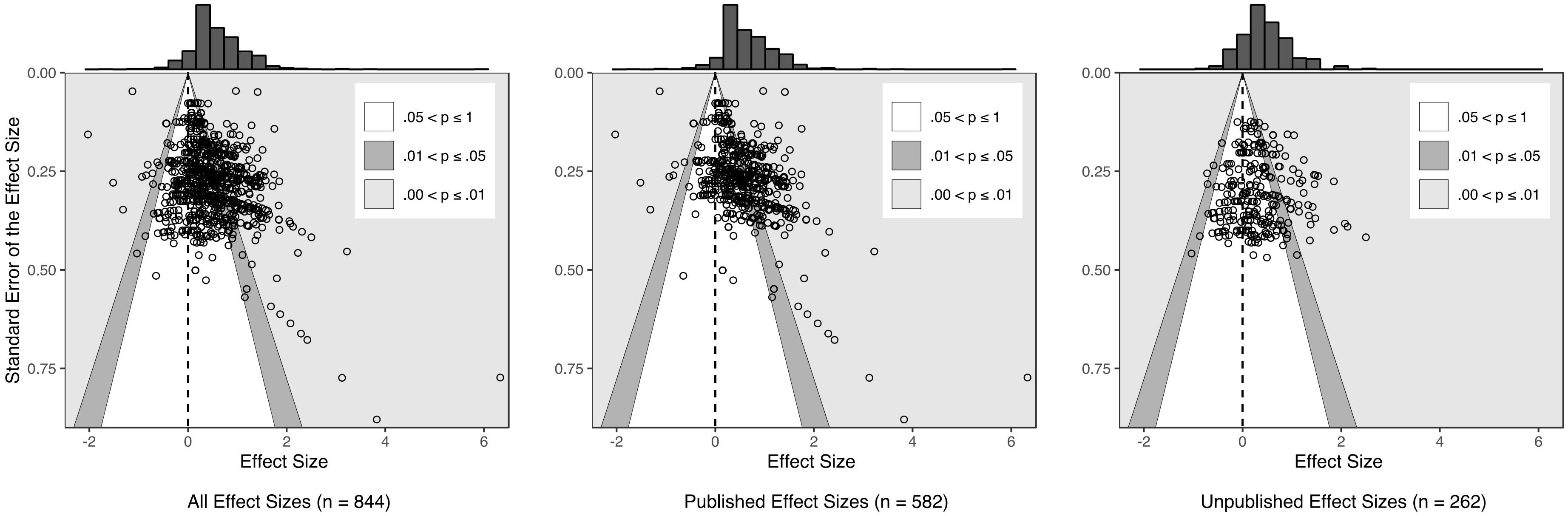
One might argue that the relation between effect size and standard error does not necessarily indicate publication bias. For instance, the asymmetry could result from differences between small and large studies due to inherent heterogeneity among the studies (e.g., Lau et al., 2006). Larger studies may, for example, evaluate shorter training programs (due to the cost of implementing long programs in large samples), which in turn could yield smaller effects. To examine this possibility, we included the standard error of the effect size, along with a range of study features (e.g., training duration and total training time), in the moderator analysis. After controlling for these features, the association between effect size and standard error remained statistically significant (see the "Moderator Analysis" section for details). Similarly, researchers conducting studies that are, by design, likely to yield small effects may have recruited larger samples to achieve sufficient statistical power (Kühberger et al., 2014). This situation could also lead to a relationship between effect size and standard error. Although possible, this explanation appears unlikely in the current context because only a small proportion of the studies (11 out of 169; less than 7%) justified their sample size on the basis of a power analysis (see Table 2). Perhaps more importantly, these two alternative explanations are difficult to reconcile with the fact that the relationship between standard error and effect size was observed only in published studies (published studies: B = 2.33, p < .001; unpublished studies: B = 0.88, p = .111; see Table 6). This situation aligns more closely with what would be expected if the relationship was the result of publication bias.

RoBMA. We also quantified the evidence for the presence of publication bias using robust Bayesian meta-analysis (RoBMA, Maier et al., 2022), and the results indicated that our data were 64.56 times (BF_{10} = 64.56) more likely under the model assuming the presence of publication bias than under the model assuming no publication bias. We conducted the same analyses, one for published and one for unpublished effect sizes. Consistent with the presence of publication bias, we found very strong evidence of publication bias but only for published studies (published studies: BF_{10} =575.84; unpublished studies: BF_{10} = 0.97).

In sum, we found that published studies reported larger effect sizes than unpublished studies, and the results of Egger's test suggest funnel plot asymmetry. RoBMA also revealed very strong evidence for the presence of publication bias. These different approaches present converging evidence to suggest the presence of publication bias in our sample of studies.

Figure 2a

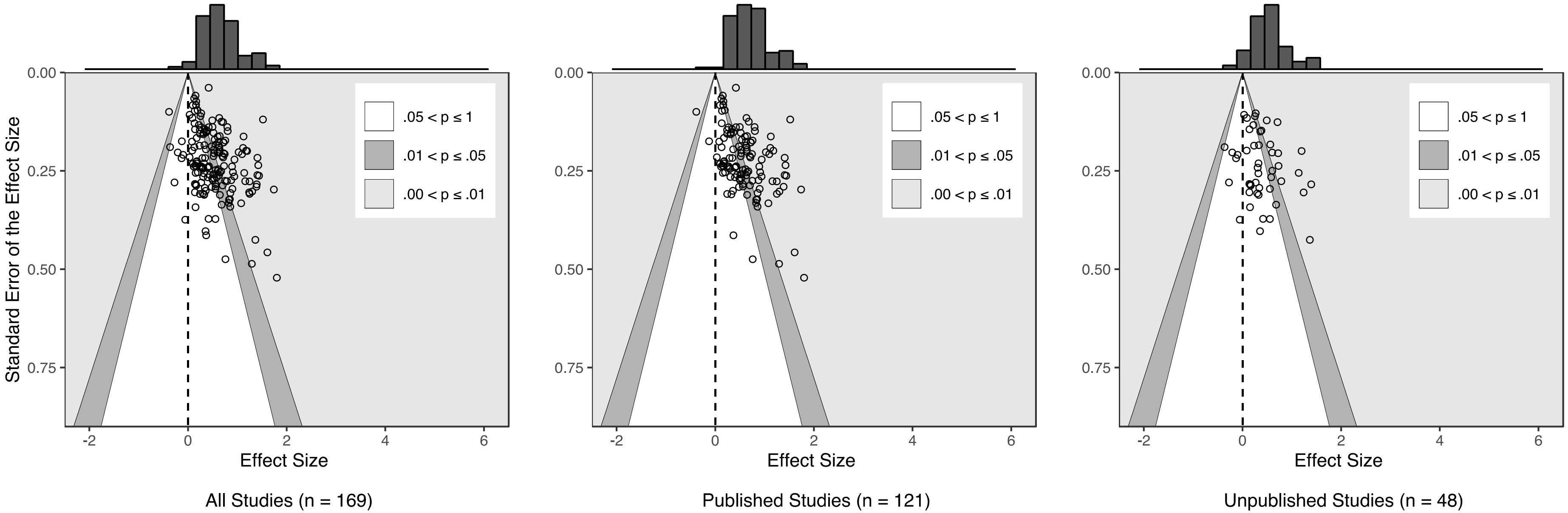
Funnel Plot (Non-Aggregated Effect Sizes)



Note. Each circle represents a non-aggregated effect size.

Figure 2b

Funnel Plot (Effect Sizes Aggregated at the Study Level)



Note. Each circle represents an effect size aggregated at the study level.

Table 6

RVE Meta-Regression Model with SE as the Predictor

			RVE Meta-Regression Model with SE as the Predictor Variable						
Studies Included	*k*	*m*	I^2	τ	*B*	*SE B*	*df*	*t*	*p*
All Studies	169	844	86.65	0.55	1.76	0.41	55.6	4.34	< .001
Unpublished	48	262	72.03	0.39	0.88	0.53	23.5	1.66	.111
Published	121	582	88.81	0.59	2.33	0.47	33.2	4.93	< .001

Note. k: number of studies; *m* = number of effect size estimates; *B*: unstandardized coefficients of the predictors.

Correcting for Publication Bias

Given that publication bias can inflate the overall effect size estimate in a meta-analysis, four methods — PET-PEESE, trim-and-fill, RoBMA, and Top10% — were used to correct for the potential overestimation (see Table 7 for the summary).

Table 7

Effect Size Estimate Corrected for Publication Bias

	Estimated Overall Effect Size				
Method used	*g*	*SE*	τ	95% *CI*	*p*-value or Bayes Factor
No correction	0.53	0.03	0.56	0.47, 0.59	$p < .001$
PET	0.14	0.09	0.31	-0.03, 0.31	$p = .103$
PEESE	0.32	0.05	0.31	0.23, 0.42	$p < .001$
Trim-and-fill	0.30	0.04	0.50	0.23, 0.37	$p < .001$
RoBMA	0.29	/	0.29	0.00, 0.48[a]	$BF_{10} = 11.70$[b]
Top10%	0.30	0.10	0.71	0.09, 0.50	$p = .007$

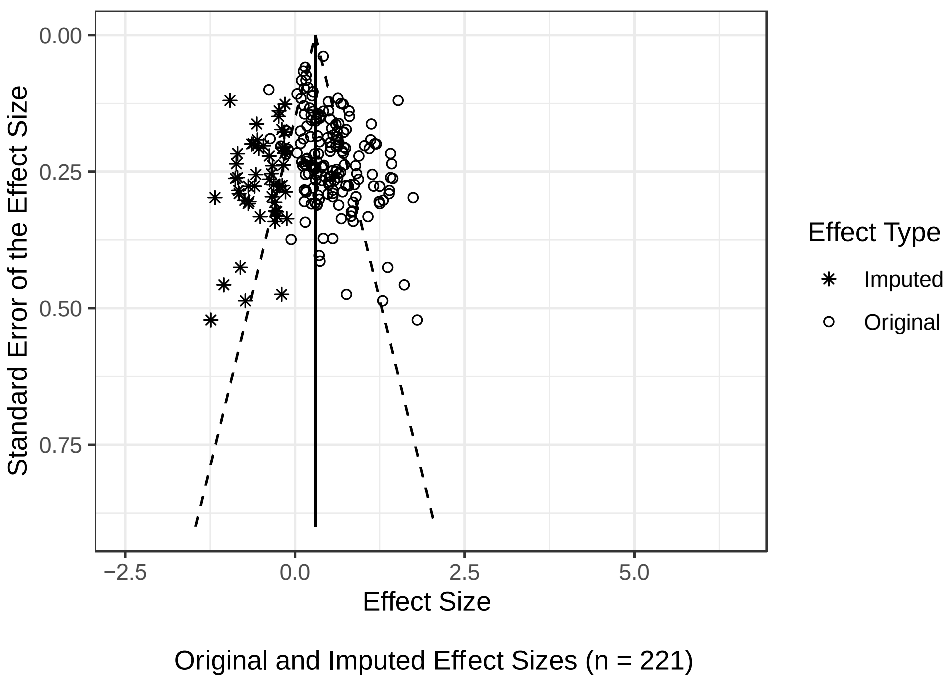
Note. A positive effect size (*g*) means that the group receiving creativity training performed better than the control group.
[a] 95% credible interval.
[b] our data were 11.7 times more likely under the model assuming the presence of an effect than under the model assuming no effect.

The first method produces two corrected estimates: PET and PEESE. Typically, the PET estimate is favored over PEESE when its *p*-value is larger than 0.10. Here, we chose to focus on the PEESE estimate considering: (a) the notable difference between the PET and our other correction method estimates, (b) the fact that the PET's *p*-value barely passed the threshold of 0.10, and (c) the PET's known tendency to underestimate effect sizes when heterogeneity is high—as is the case in our meta-analysis. Our decision aligns with the recommendations of Inzlicht et al. (2015) and Reed (2015).

The second method, trim-and-fill, imputed 52 additional studies. In accordance with a publication bias towards positive results, all imputed studies fall into the category of negative effects. Figure 3 presents the funnel plot of study-level aggregated effect sizes by their standard errors, including both observed studies and imputed studies. The adjusted overall effect size was $g_{\text{adjusted}} = 0.30$, $SE = 0.04$, $p < .001$, 95% CI [0.23, 0.37], which was similar to the PEESE estimate. RoBMA estimated the adjusted effect size to be $g_{\text{adjusted}} = 0.29$, 95% credible interval [0.00, 0.48], BF_{10} for the presence of the effect = 11.70. Last, we constructed a meta-regression with RVE to estimate the overall effect size from the 10% most precise studies (Top10%). Seventeen studies (number of effect sizes = 115) were identified as the 10% most precise studies. The average effect size estimate was $g_{\text{adjusted}} = 0.30$, $p = .007$, 95% CI [0.09 0.50]. It is important to note that the top 10% most precise studies all examined the effect of creativity training with cognitive components. Hence, the adjusted estimate may not be as representative as those generated by other methods. Still, the Top10% estimate was very similar to the estimates derived from the other methods, also within the range of small effect sizes.

Figure 3

Funnel Plot (Study-Level Aggregated Effect Sizes), Including Both Original and Imputed Effect Sizes



Note. The vertical solid line represents the corrected average effect size, and the two dashed lines represent the 95% confidence intervals around the corrected average effect size.

Moderator Analysis

There was a large amount of heterogeneity among the effect sizes (I^2 = 87.34, τ = 0.56), suggesting the presence of moderators. The publication bias analysis identified the standard error of the effect size and publication status as two significant moderators of effect size. We conducted the following analyses to explore if certain study features could further explain the heterogeneity among effect sizes.

Meta-Regression with a Single Moderator. First, we examined the impact of the moderators one at a time by constructing separate meta-regression models with RVE for each moderator (see supplemental material Table S8). As mentioned earlier, *Standard Error* and *Publication Status* were significant moderators of effect sizes (*Standard Error*: B = 1.76, p < .001; *Publication Status*: B = 0.19, p =.006). Our results also show a statistically significant and positive relationship between *Total Training Time* and effect size, B = 0.03, p = .035. None of the remaining moderators were significant.

Meta-Regression with Multiple Moderators. Because some of the moderators may be correlated, their effects (or lack thereof) may be due to their correlations with other moderators. To examine their unique contribution, we included all the moderators in a single RVE meta-regression model (see Table 8). The positive effect of *Total Training Time* observed earlier was attenuated after controlling for the other factors, p = .556. In contrast, the effect of *Training Content* became larger and significant, p = .024, with combined training and cognitive training studies reporting larger effect sizes than non-cognitive training studies (combined vs. non-cognitive, p = .011; cognitive vs. non-cognitive: p = .022). There was no significant effect size difference between the cognitive and the combined training (p = .241). *Standard Error* and *Publication Status* were still statistically significant. Again, both moderators were positively associated with effect size (*Standard Error:* B = 2.55, p < .001; *Publication Status:* B = 0.29, p = .002). These relationships, which are consistent with publication bias, remained robust even after controlling for all the other moderators.

Table 8

Results of the RVE Mixed-Effects Meta-Regression Model with Multiple Moderators

Moderators *(Reference Group)*	*B*	*SE B*	Statistics	*p*
Intercept	-1.26	5.14	*t*(50.05) = -0.25	.807
Sample Characteristics				
Age Group *(Preschool/Kindergarten)*			*F*(4, 34.1) = 0.44	.780
US Grade 1-8	0.08	0.12	*t*(22.69) = 0.69	.495
US Grade 9-12	0.05	0.26	*t*(15.95) = 0.19	.854
University Students	-0.05	0.14	*t*(25.29) = -0.40	.689
Adults	-0.13	0.18	*t*(27.35) = -0.72	.476
Training Characteristics				
Training Content *(Non-Cognitive)*			*F*(2, 21.80) = 4.43	.024
Cognitive	0.40	0.15	*t*(9.88) = 2.72	.022
Combined	0.49	0.16	*t*(10.72) = 3.04	.012
Log-transformed Total Training Time	0.02	0.03	*t*(33.77) = 0.59	.558
Log-transformed Training Duration	-0.01	0.04	*t*(41.98) = -0.33	.745
Study Characteristics				
Standard Error	2.55	0.64	*t*(35.19) = 3.95	< .001
Randomization *(Non-Random)*			*F*(2, 27.60) = 0.69	.512
Random	-0.11	0.10	*t*(56.89) = -1.02	.313
Unspecified	0.03	0.17	*t*(16.33) = 0.19	.854
Type of Control Group (*Inactive)*				
Active	-0.06	0.09	*t*(52.80) = -0.70	.489
Use of a Pretest *(Posttest only)*				
Pretest-Posttest	-0.07	0.10	*t*(49.04) = -0.73	.468
Type of Outcome Measure *(Divergent Thinking)*			*F*(2, 40.60) = 0.60	.553
Creative Problem Solving	-0.12	0.11	*t*(28.83) = -1.10	.280
Creative Product	-0.07	0.14	*t*(39.61) = -0.53	.602
Origin of Outcome Measure (*Researcher-Made)*				
Independent	-0.17	0.11	*t*(36.35) = -1.60	.119
Training to Criterion (*No)*				
Yes	0.04	0.10	*t*(35.74) = 0.40	.694
Posttest Timing *(Same day)*			*F*(3, 33.1) = 0.54	.659
Between 1-7 Days	-0.06	0.11	*t*(46.41) = -0.53	.597
8 Days or Longer	-0.13	0.15	*t*(14.24) = -0.87	.397
Unspecified	0.04	0.11	*t*(49.90) = 0.33	.741
Publication Year	< 0.01	< 0.01	*t*(50.92) = 0.14	.893
Publication Status *(Unpublished)*				
Published	0.29	0.09	*t*(52.28) = 3.22	.002

Note. Due to missing values only 725 effect sizes from 142 studies were included. I^2= 88.34; τ = 0.69.

Effects of Creativity Training on Divergent Thinking Measures

We also examined the impact of creativity training on three core dimensions of divergent thinking (i.e., fluency, flexibility, and originality of the ideas generated). Eighty-six studies — totaling 407 effect sizes — were included in this analysis. We constructed a meta-regression model with the RVE weighted effect size as the dependent variable and the dimension of divergent thinking as the predictor variable. Apart from the modality of divergent thinking (i.e., verbal vs. figural), we also included the standard error of the effect size, the publication status, and the total training time as predictor variables because they were identified as significant moderators in the single or multiple-moderator analyses for the whole sample. We did not include training content as a predictor variable, even though it was a significant moderator in the multiple-moderator analysis, because only a sparse number of studies that used divergent thinking tests as outcome measures examined non-cognitive training in this sub-sample (2 out of 86 studies). We present the results of our moderator analyses in supplemental material Table S9 where moderators were evaluated one at a time and in Table 9 where they were evaluated all together. As shown in Table 9, the effect sizes for fluency, flexibility, and originality were not statistically significantly different from each other, $F = 0.62$, $p = .540$. This finding does not support the previously made suggestion that creativity training has a larger impact on originality than on the other dimensions of divergent thinking (Rose & Lin, 1984; Scott et al., 2004).

Table 9

Results of the RVE Mixed-Effects Meta-Regression Model (Divergent Thinking Measure Only)

Moderator *(Reference category)*	*B*	*SE B*	Statistics	*p*
Intercept	-0.33	0.23	$t(30.4) = -1.41$	.170
Standard Error	1.68	0.54	$t(22.0) = 3.10$	.005
Publication Status *(Unpublished)*				
Published	0.27	0.11	$t(36.1) = 2.40$	.022
Total Training Time (log-transformed)	0.03	0.01	$t(24.7) = 2.03$	.053
Dimension (*Fluency*)			$F(2, 68.2) = 0.62$	.540
Flexibility	0.07	0.09	$t(57.0) = 0.85$	.397
Originality	0.05	0.05	$t(79.2) = 0.92$	.361
Modality (*Verbal*)				
Figural	-0.14	0.08	$t(62.1) = -1.72$	.091

Note. Number of studies = 86. Number of effect sizes = 407. $I^2 = 90.74$; $\tau = 0.71$.

Prevalence of Studies Employing Rigorous Methodology

We evaluated the quality of the studies on a range of methodological features and identified several issues. We describe each of them in turn.

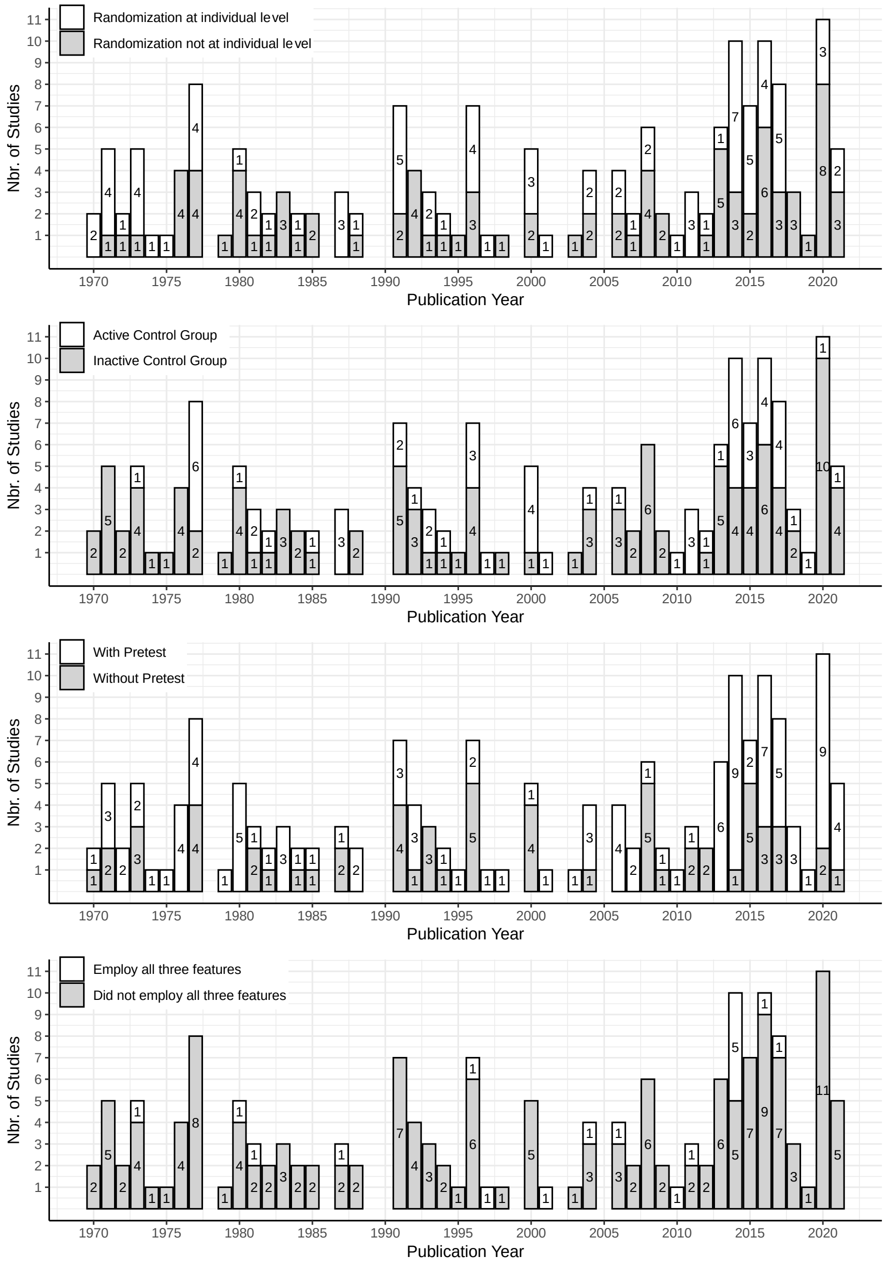
Low Proportion of Studies Randomizing at the Individual Level, Using an Active Control Group, and Using a Pretest. We examined the proportion of studies randomizing at the individual level, using an active control group, and using a pretest. Of 169 studies, 82 (49%) performed randomization at the individual level, 59 (35%) used an active control group, and 108 (64%) used a pretest. Only 18 of these 169 studies (11%) had adopted all three features. We constructed four logistic regression models with publication year as the predictor variable to examine whether the proportion of studies adopting these methodological features increased over time. None of the models were significant, all $p > .307$, suggesting that the proportion of studies employing these design features did not change over time (see supplemental material Table S10 for results). Figure 4 presents the number of studies employing these design features over time.

Low Proportion of Adequately Powered Studies. We also estimated the proportion of adequately powered (80% or above) studies for a range of effect size estimates (see Figure 5). According to our results, only 30% of the studies had enough power to detect an effect size of $d = 0.53$, the unadjusted estimate of the overall effect size. For effect sizes of $d = 0.29$ and 0.32, i.e., the lowest and highest estimate of the adjusted effect sizes, the proportion fell to 5% and 7%, respectively. These low figures imply that the vast majority of the included studies had insufficient sample sizes. When examining both sample size and randomization at the same time, we noticed a lack of large-scale randomized trials. For example, only 1 out of the 82 studies randomizing at the individual level met the minimum sample size ($N = 350$) recommended by the What Works Clearinghouse (2020) for generating moderate to strong evidence for educational interventions.

The low proportion of adequately powered studies is unsurprising given the prevalence of this issue in other fields and the fact that studies in our sample rarely (< 5%) justified their sample size on the basis of a power analysis. We also examined whether the number of adequately powered studies increased over time. We correlated the standard error of the aggregated effect size (a measure of study precision) with publication year and found a weak and non-significant relationship between them[12], $r = -.01$, $p = .860$, suggesting no improvement over time.

Figure 4

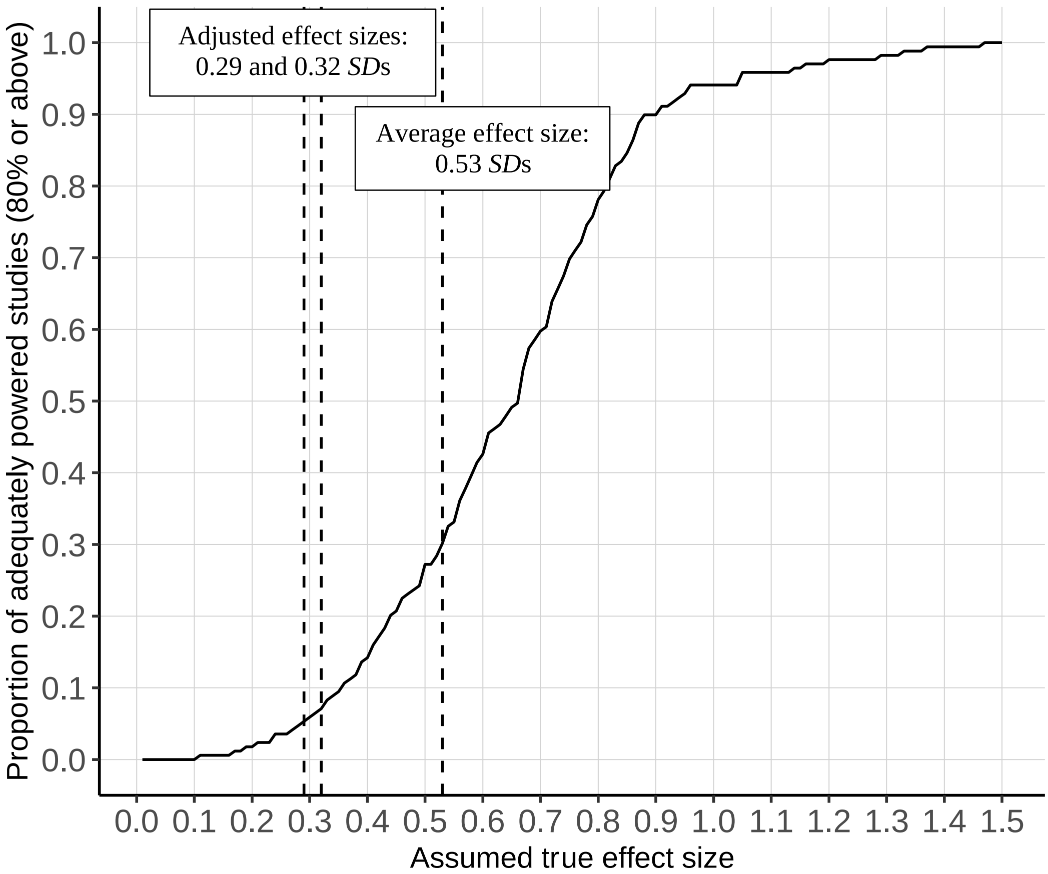
Number of Studies with Randomization at the Individual Level, Active Control Group, and Pretest Over Time



[12] To account for effect size dependency, we performed the correlation analysis on the aggregated study-level effect sizes.

Figure 5

Proportion of Adequately Powered Studies for a Range of Assumed True Effect Sizes



To further explore the relationship between study quality and the impact of creativity training, we computed the average effect size using studies meeting at least 1, 2, or 3 of the following quality criteria (no study met all four criteria):

- Randomization at the individual level.
- Use of an active control group.
- Use of a pretest.
- Sufficient statistical power to detect an effect size of 0.3 *SD*s or above—a value consistent with most of our adjusted effect size estimates (see Table 7).

As indicated in Table 10, only 19 out of 169 (11%) satisfied three criteria. When only considering the highest quality evidence in our sample (studies meeting three criteria), the average effect was notably lower ($g = 0.28$, $p < .001$). These effects were further reduced after adjusting for publication bias (ranging between $g = 0.09$ and 0.13; see supplemental material Tables S11 and S12 for details).

Table 10

Effect Size Estimates by Number of Met Quality Criteria

Number of criteria met	*k*	*m*	I^2	τ	*g*	*SE*	95% *CI*	95% *PI*	*p*
All studies	169	844	87.34%	0.56	0.53	0.03	0.47, 0.59	-0.58, 1.63	< .001
At least 1 of the 4 criteria	156	782	87.97%	0.57	0.53	0.03	0.46, 0.59	-0.59, 1.65	< .001
At least 2 of the 4 criteria	84	480	90.66%	0.67	0.47	0.04	0.39, 0.55	-0.86, 1.80	< .001
At least 3 of the 4 criteria	19	129	61.40%	0.30	0.28	0.06	0.15, 0.42	-0.36, 0.92	< .001
All 4 criteria	0	0	/	/	/	/	/	/	/

Note. *k* = number of studies; *m* = number of effect size estimates. A positive effect size (*g*) means that the group receiving creativity training performed better than the control group. The 4 criteria: (1) randomization at individual level, (2) use of an active control group, (3) use of a pretest, and (4) sufficient statistical power to detect an effect size of 0.3 or above.

Low Proportion of Direct Replication. Replication studies help determine if an observed effect is real or not. Despite the importance of this effort, only 9 of the 169 studies included in our meta-analysis reported being replication studies. Interestingly, the majority (7 out of 9) were published before 2010, suggesting no recent progress towards improving replicability.

Near Absence of Preregistered Studies. Preregistration is often viewed as an effective way to prevent selective reporting, a questionable research practice common in studies containing multiple outcome measures — a feature shared by many studies in our sample (on average, studies reported 4.99 effect sizes). Unfortunately, preregistration was uncommon: only one study in our dataset (i.e., Ritter et al., 2020) reported preregistering their protocol. Selective reporting may have occurred in some of the studies we considered. For example, although fluency, flexibility, and originality are recognized as the core dimensions of verbal divergent thinking (Runco & Jaeger, 2012), over 40% of the studies (29 out of 72) using this test to assess training effectiveness reported only a subset of these dimensions, or focused on different, less common dimensions entirely (see supplemental material Table S13). This evidence, although inconclusive, is suggestive of selective reporting.

Discussion

In this meta-analysis and critical evaluation of the creativity training literature, we considered 169 studies conducted between 1970 and 2022, totaling 844 effect sizes. Close to one-third (262 effect sizes) were extracted from unpublished studies. We used random-effects meta-analysis with robust variance estimation (RVE) to synthesize and compare these effect sizes while taking into account within-study dependency. We also quantified the extent of publication bias and surveyed the prevalence of study features that influence the credibility of creativity training studies and their usefulness for decision makers.

Overall Effect Size and Publication Bias

We found that the overall effect of creativity training, uncorrected for publication bias, was large and significantly greater than zero, g = 0.53, p < .001, 95% CI [0.47, 0.59], 95% PI [-0.58, 1.63]. This estimate is very similar to those reported in existing meta-analyses of creativity training programs. Importantly, however, we found converging evidence consistent with substantial publication bias, suggesting that this estimate is inflated. First, we observed a large (and significant) difference between effect sizes from published and unpublished studies, with published studies reporting larger effect sizes than unpublished ones (g = 0.58 vs. g = 0.38, p = .006). This difference remained large (and significant) even after controlling for all the other moderators coded (see Table 8), suggesting that the difference is unlikely due to factors other than publication status.

Second, in keeping with publication bias, we found a strong positive relationship between effect size and standard error, implying that smaller studies, which are often more prone to publication bias, tended to report larger effect sizes. Again, the relationship remained strong even after controlling for all other coded moderators (see Table 8). Also consistent with publication bias, the relationship between effect size and standard error was observed only in published studies and not in unpublished studies (see Table 6).

Third, a Bayesian analysis that quantifies the likelihood of publication bias under a range of meta-analytic models and assumptions (i.e., robust Bayesian meta-analysis, RoBMA) revealed strong support for the presence of publication bias (BF_{10} = 64.56, suggesting that our data is 64.56 times more likely under the assumption that publication bias was present). Again, the evidence for publication bias was much greater in published than unpublished studies (BF_{10} = 575.84 and BF_{10} = 0.97, respectively). Together, all three methods produced evidence consistent with the presence of publication bias.

We used four different methods to estimate the effect of creativity training programs while correcting for publication bias. The findings of all four methods converged to suggest that the average effect size is considerably lower than our original estimate, likely between 0.29-0.32 *SD*s (PEESE: g = 0.32, p < .001; trim-and-fill: g = 0.30, p < .001; RoBMA: g = 0.29, BF_{10} =11.70; Top10%: g = 0.30, p = .007; see Table 7), a substantial departure (i.e., a 39% to 45% reduction) from our original, uncorrected, estimate.

Despite this reduction, these adjusted effects remained statistically greater than zero, distinguishing our findings from other areas such as ego-depletion (Carter et al., 2015) and growth mindset (Macnamara & Burgoyne, 2023), where adjusted effects were statistically non-significant. Furthermore, our adjusted effects are large when benchmarked against the typical developmental change in creativity over time (Said-Metwaly et al., 2020) and the typical effect size observed in educational interventions (Kraft, 2020). However, our estimates might still be on the higher side, as factors other than publication bias could contribute to overestimation. For example, as we demonstrated earlier, focusing solely on the highest quality studies further reduces the average effect size.

Moderator Analysis

Our study also revealed a large amount of variation across studies (I^2 = 87.37%, τ = 0.56). We conducted a moderator analysis to explore which sample, training, and study characteristics were potentially moderating effect sizes. As mentioned earlier, both *Standard Error* (i.e., study precision) and *Publication Status* (published vs. unpublished) were strongly associated with effect size, likely due to publication bias. *Training Content* also moderated effect size —training programs that included cognitive components, either alone or in combination with non-cognitive components, were associated with larger effect sizes than non-cognitive training programs (see Table 8). This pattern of results is consistent with Scott et al.'s (2004) observation that effective creativity training tended to be based on a cognitive framework. Also, similar to Scott et al.'s (2004) meta-analysis, we observed a positive bivariate relationship between *Total Training Time* and effect size. However, this relationship was weak and only significant when ignoring the contributions of the other moderators.

Many of our findings do not align with those of past meta-analyses of creativity training. For example, unlike Scott et al. (2004), we did not find larger effect sizes for divergent thinking and creative problem solving than for creative product. We also did not observe that studies using a pretest were associated with smaller effect sizes than studies using a posttest-only design (B = -0.07, p = .10), a difference that was very large in Scott et al. (2004) meta-analysis (pre-post: 0.54 *SD*s, posttest only: 1.01 *SD*s). We also did not observe that creativity training programs were more beneficial for older participants, as suggested in Ma's (2006) meta-analysis — in contrast, the pattern in our meta-regression appears to be in the opposite direction (although this was not significant). We observed little difference between the effect sizes associated with the three core dimensions of divergent thinking (i.e., fluency, flexibility, and originality), which is somewhat at odds with the findings of Rose and Lin (1984) and Scott et al. (2004) that creativity training has a stronger impact on originality than on fluency and flexibility.

These discrepancies between our findings and those of previous meta-analyses may be due to our larger sample of studies (which include more recent and more unpublished studies), the use of modern meta-analytic methods that take into account the nested nature of the effect sizes, as well as the differences in inclusion criteria, such as our exclusion of studies with no control group.

Some findings from our moderator analysis also depart from the training literature more generally. For example, publication year, a significant moderator in many fields — where older studies tend to report larger effect sizes (e.g., Ioannidis, 2008; Protzko & Schooler, 2017) — was not associated with effect size in our meta-regression models. As shown in Figure 4, this may be due to the field being slow to adopt more rigorous research practices which tend to reduce effect sizes.

Several study quality indicators, such as randomization at the individual level, use of an active control group, and use of a pretest, were also not significant moderators in our analysis, which might suggest that study quality does not influence effect size—countering the common observation that higher quality studies are associated with smaller effect sizes (e.g., Cheung & Slavin, 2016; Cuijpers et al., 2010). However, study quality is multifaceted, often best represented by a combination of features rather than isolated ones (Higgins et al., 2011). When considering only the highest quality evidence in our sample (studies meeting three quality indicators), the average effect (g = 0.28, p < .001; see Table 10) was notably lower than our original estimate (g = 0.53, p < .001). This substantial reduction is consistent with higher study quality being associated with smaller effect sizes and suggests that researchers aiming for a precise evaluation of the training effect should prioritize high-quality studies; otherwise, their results will likely be inflated. Nevertheless, readers should bear in mind that this analysis was based on a very small sample (19 studies), and there is no clear consensus about the criteria for determining research quality (Slavin, 2008; Valentine, 2019).

The effects of other moderators also deviated from what is expected in the training literature. For example, the origin of outcome measures (researcher-made vs. independent), which were found to be significant moderators in a large-scale meta-analysis of educational interventions (645 studies, Cheung & Slavin, 2016), was not significantly associated with effect size. We suspect that the reason why many of our moderators were not associated with training effectiveness may be partly due to some effects of publication bias not being captured in our analyses — publication bias, like other types of selection biases, can substantially bias estimates of associations (e.g., Munafò et al., 2018). Another potential reason is the high proportion of low-quality studies. In our sample, only 19 studies (out of 169) met at least 3 of the 4 quality indicators (no study met all four indicators). The additional variability introduced by low-quality studies (due to their higher vulnerability to confounding factors) may also have masked the effect of some moderators in our analysis. Given these influences, readers should avoid concluding that our non-significant moderators do not have an important impact on the effectiveness of creativity training. For instance, although age could be an important moderator, publication bias may mean that smaller effect sizes from certain age groups remained unpublished, thereby obscuring the true relationship between age and training effectiveness. A similar situation could arise if study quality tends to be lower in studies focusing on certain age groups.

Methodological Limitations of Studies in the Field

Another goal was to survey the methodological rigor of creativity training literature. A number of studies have explored this aspect of the field (e.g., Mansfield et al., 1978; Valgeirsdottir & Onarheim, 2017), but often in a qualitative fashion, with small samples covering short periods of time — features that prevent, for example, exploring changes in methodological rigor over time. We were particularly interested in quantifying the prevalence of features limiting the internal validity of the creativity training literature, the advancement of underlying theories, and the relevance to practitioners.

Internal Validity. We identified a number of methodological limitations that can pose important threats to the internal validity of creativity training studies.

Type of Control Group. Although all studies included have a control group, only a minority (35%, see Table 2) had an active control group (i.e., a control group that receives an alternative non-creativity training rather than no training). The absence of an active control group makes it difficult to control for the impact of motivation and expectation for improvement on the outcome measures (Boot et al., 2013). Given the significant influence of these two factors on creativity (e.g., Amabile, 1996; Carmeli & Schaubroeck, 2007; Tierney & Farmer, 2004), the absence of an active control group can severely undermine the credibility of the causal claims made in these studies.

Randomization and Pretest. Also problematic is how randomization was achieved. Close to half of the studies in our sample (44%, see Table 2) randomized clusters of participants (e.g., classrooms) — rather than individuals — into the experimental and control conditions, despite analyzing the data at the individual level. This practice is known to underestimate (often substantially) the uncertainty associated with an effect size. Moreover, this also causes the treatment effect to be confounded with differences in the characteristics between clusters (e.g., the intervention effect being confounded with a teacher effect when the clusters are led by different teachers) when the number of clusters is small — like most studies in our sample. Furthermore, 36% of studies in our sample did not conduct a pretest to examine pre-existing differences between the control and experimental groups (see Table 2). The use of a pretest is particularly important for studies that randomize clusters of participants to ensure that the observed effect is not due to pre-existing between-clusters differences (What Works Clearinghouse, 2020).

Statistical Power and Replication. Coding studies' effect sizes and associated uncertainty also allowed us to evaluate how many of them were statistically underpowered. Statistical power is important. Underpowered studies produce highly variable estimates of effect, and when paired with publication bias, can result in severe overestimation of their effects. This is particularly important in the present case. In our sample, only 7% of the studies have a power of 80% or above to detect an effect size of 0.32 *SD*s, the highest adjusted effect size estimate (see Figure 5 for the proportion of adequately powered studies for a range of plausible effect sizes), raising further concerns about the robustness of the creativity training literature. Limited statistical power can be mitigated by replication. Indeed, studies following the same procedures can be combined to increase power (e.g., Goulet & Cousineau, 2019). Unfortunately, in our sample of 169 studies, only 9 (5%) reported being replication studies.

Pre-Registration. We also considered the number of preregistered studies. An important reason to preregister studies is to prevent selective reporting bias where researchers selectively report results consistent with their research objective, a practice that inflates effect sizes (e.g., Simmons et al., 2011). Such practice is common in studies using multiple outcome measures (e.g., John et al., 2012), and may have occurred in some of the studies we surveyed. For example, a large number of studies (72 out of 169) used a verbal divergent thinking test, a measure that evaluates creative thinking skills on three dimensions (i.e., fluency, flexibility, and originality; Runco & Jaeger, 2012), yet over 40% reported only a subset of these dimensions, or focused on different, less common dimensions entirely (see supplemental material Table S13 for more details). Although there are multiple reasons why researchers may not have reported scores on the three core dimensions, only preregistration could convincingly rule out selective reporting. In our sample, only one study (Ritter et al., 2020) reported preregistering their protocol before data analysis. Having more preregistered studies would alleviate concerns over selective reporting.

Methodological Improvement Over Time. Finally, one might rightfully suggest that the issues highlighted above were more common in old than in new studies, and thus our results do not accurately reflect the quality of the more recent literature. Indeed, studies in our meta-analysis span more than five decades (1970 to 2022), during which there have been many changes in research practices. Despite this, our analysis found virtually no methodological improvement over time (see Figure 4 and supplemental material Table S10), suggesting that the field is slow in adopting more rigorous research practices.

Research Gaps

Our review of the creativity training literature identified several research gaps. These gaps limit conclusions about the generalizability and persistence of the effects of creativity training. Such limitations impede the identification of boundary conditions for creativity training effectiveness (essential for advancing our understanding of the underlying mechanisms) and reduce the literature's usefulness for practitioners.

Predominance of Divergent Thinking Test as the Sole Measure of Creative Performance. Despite the consensus that creative thinking is a multidimensional construct (Brophy, 1998; Hommel, 2012; Runco, 2007), 85% of the studies in our sample (143 out of 169) used a single instrument to evaluate training effectiveness (see Table 3 for more details). Among these studies, 78% (111 out of 143) used a divergent thinking test as their sole measure of creativity. This could be problematic not only because divergent thinking test performance could be easily influenced by factors unrelated to creative thinking ability, e.g., test instructions, scoring methods, and test settings (Acar et al., 2020; Hattie, 1980; Plucker et al., 2011), but also because divergent thinking is just one of many creative processes that make up creative thinking (Brophy, 1998; Hommel, 2012; Runco, 2007). Said-Metwaly et al. (2022) reviewed 70 studies examining the relationship between divergent thinking performance and real-life creative achievement and estimated their correlation to be $r = .17$, a weak association. Using only divergent thinking tests to evaluate training effectiveness limits our knowledge about the impact of creativity training programs (beyond divergent thinking) and their relevance to practitioners.

Under-Representation of Key Populations. As indicated in Table 2, a majority of the studies in this meta-analysis focused on young children (i.e., US Grade 1-8) and University students. Although promoting creative thinking is a prime concern for organizations and educators (PISA, 2022; World Economic Forum, 2018), non-student adults and secondary school students (the target population of the PISA) are currently underrepresented in the field of creativity training. As such, our average estimate of the effect of creativity training may not generalize to these two populations, again limiting the relevance of these findings to practitioners. This concern may be less pronounced for the adult group, considering that a large portion of the included studies targeted university students (who are predominantly adults), and there seems to be no compelling reason why the impact of creativity training would differ between these two adult populations.

Rarity of the Delayed Posttest. Decision makers are likely to care about how long-lasting the effect of training would be. Thus, it is also important to examine the persistence of the effect over time. Unfortunately, in our sample, only a small fraction of studies (14%, 23 out of 169 studies; see Table 2) conducted a delayed posttest to explore the persistence of the training effect over time. Understanding when the effect of creativity training fades out could also provide valuable insights into the mechanisms underlying the training effect.

Implications

Our findings have implications for both researchers and practitioners.

Implications for Researchers Evaluating Creativity Training Programs. Our findings highlight a number of issues with the current state of creativity training literature. Key amongst these issues is the lack of statistical power. Researchers should conduct power analysis and assume realistic effect size estimates. In our samples, studies that conducted power analysis often assumed the true effect size to be medium to strong (0.5 *SD*s or above). This may be partly due to the large effect size estimates reported in previous meta-analyses (ranging from 0.47 to 1.02 *SD*s). Our meta-analysis suggests that these effect size estimates are inflated to a notable degree. As such, researchers should assume much smaller effect sizes for power analysis, such as our adjusted estimates (which ranged from 0.29 to 0.32 *SD*s).

Another critical issue relates to the internal validity of creativity training studies. For example, we observed a large proportion of studies randomizing at the cluster level (e.g., classroom) but analyzing the data at the individual level, as well as studies using passive control groups and not conducting a pretest. In our analysis, no studies met all four of our quality criteria, and only 19 (11%) met three out of four criteria (see Table 10 for more details). Also disappointing was the lack of notable improvement in study quality over time (see Figure 4). Adopting more rigorous practices would increase the credibility of creativity training literature, facilitating the development of theories and effective interventions (Eronen

& Bringmann, 2021). More preregistration and replication studies, which are still not common practices in the field, would further enhance the credibility of this body of work. We anticipate that when a sizable pool of high-quality studies (ideally registered reports) becomes available, issues related to low quality and publication bias will diminish, making the identification of influential moderators more fruitful. Importantly, when planning these more rigorous studies, researchers should anticipate smaller effect sizes and increase their sample size accordingly.

Our survey of creativity training research also revealed gaps in the literature. To increase the usefulness of the evidence to practitioners of creativity training research, future studies should aim to better understand its generalizability (for example, by using multiple measures of creative performance) and to evaluate the persistence of the effects over time (for example, by using multiple assessment points). Also, given the increasing interest in improving creative thinking in secondary school students (i.e., US Grade 9-12; PISA, 2022) and adults (World Economic Forum, 2018), researchers should conduct more studies to assess the impact of creativity training on these two populations, which are currently underrepresented in the field. Addressing these gaps is also likely to facilitate understanding of the mechanisms underlying the effect of creativity training.

Implications for Practitioners and Researchers. Given our results, we believe that practitioners and researchers should be particularly careful when interpreting findings from creativity training literature. As shown, published effects are likely to be inflated. We found substantial evidence consistent with publication bias — a bias that was not considered in previous meta-analyses — along with a high proportion of studies with methodological limitations (e.g., no active control group, low statistical power, and lack of preregistration and replication). These issues are likely to inflate effects in the creativity training literature. Practitioners and researchers should adjust their expectations in light of these concerns. Moreover, considering that most studies surveyed had small sample sizes (median sample size per effect size was 53), most effects are highly imprecise (i.e., surrounded by large confidence intervals). For this reason, when comparing effect sizes, practitioners and researchers should also consider study precision. Ideally, when selecting a training program, practitioners and researchers should attempt to replicate the effect before scaling it up. If not feasible, more weight should be given to large studies with robust designs. Finally, in light of the suggestive evidence of strong publication bias identified in our meta-analysis, researchers and practitioners should be aware that for every study reporting positive results they encounter, there might be a substantial number of studies reporting negative or null findings that remain hidden.

Limitations of the Present Study

Our findings must be interpreted in light of the following limitations. The first relates to the generalizability of our results. To maximize the relevance of our findings to typical populations, we intentionally excluded atypical groups (e.g., gifted students), thus limiting the generalizability of our findings to these populations. Moreover, we only considered performance-based outcome measures. We made this decision due to the small number of non-performance-based outcomes (such as change in attitude) found in past meta-analyses and in our initial screening of the literature, and due to our belief that improving creative performance is researchers' and practitioners' primary concern. As such, our findings may not apply to such measures. Similarly, we excluded studies using team-level outcome measures, again limiting the generalization of our findings to these outcomes. Relatedly, we only included studies reported in English, which again limits the generalizability of our findings. For example, as shown in Table 2, most of our studies were conducted in populations that are often considered Western, Educated, Industrialized, Rich, and Democratic (i.e., WEIRD populations, Henrich et al., 2010).

Second, although we used different methods to identify and correct for publication bias, these methods make different and untestable assumptions, and none of them can quantify the extent of publication bias with certainty. Kvarven et al. (2020), for instance, found that adjusted effect sizes are sometimes much larger than those found in rigorous replication studies. Whether a similar trend would be observed in the field of creativity training remains to be determined.

Third, creativity training is inherently complex, with outcomes potentially influenced by numerous factors like participants' motivation, personality, trainer's experience, and preparation. Many of these factors, however, could not be included in our moderation analyses due to their subjective nature and limited reporting in the literature. Consequently, influential moderators may have been omitted from our analyses.

Conclusions

In summary, by evaluating five decades of creativity training studies, this meta-analysis found a positive effect size of creativity training. However, we also found converging evidence consistent with substantial publication bias in the field and our original effect size estimate was considerably reduced once correcting for this bias. Furthermore, despite recurring calls for improving the methodological quality of creativity training studies, we identified numerous methodological shortcomings within the current literature and no sign of improvement over time — a situation that limits the credibility and usefulness of this body of work and hinders theoretical development in the field. We argue that researchers evaluating creativity training should, as a priority, improve the methodological quality of their studies. Additionally, practitioners and researchers should be careful when interpreting current findings in the field.

References

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

*Abdul Hamid, K., Yusoff, A. N., Rahman, S., Osman, S. S., Surat, S., M, M. A., & Azmi, N. H. (2020). Changes in divergent thinking strategy due to creativity training: Findings from dynamic causal modelling. *Journal of Physics: Conference Series*, *1497*(1). https://doi.org/10.1088/1742-6596/1497/1/012022

Acar, S., Runco, M. A., & Park, H. (2020). 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*(1), 39–49. https://doi.org/10.1037/aca0000256

*Albano, C. (1987). The effects of an experimental training program on the creative thinking abilities of adults (Doctoral dissertation). Temple University, US.

*Alencar, E., Feldhusen, J. F., & Widlak, F. W. (1976). Creativity training in elementary schools in brazil. *The Journal of Experimental Education*, *44*(4), 23–27. https://doi.org/10.1080/00220973.1976.11011544

*Alexander, K. D. (2007). *Effects of instruction in creative problem solving on cognition, creativity, and satisfaction among ninth grade students in an introduction to world agricultural science and technology course* [Doctoral dissertation, Texas Tech University]. https://ttu-ir.tdl.org/handle/2346/18066

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ (Clinical research ed.)*, *332*(7549), 1080. https://doi.org/10.1136/bmj.332.7549.1080

Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Research in Organizational Behavior*, *10*(1), 123-167.

Amabile, T. M. (1996). *Creativity in context: Update to "The Social Psychology of Creativity."* Westview Press. https://doi.org/10.4324/9780429501234

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The Work Preference Inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology*, *66*(5), 950–967. https://doi.org/10.1037/0022-3514.66.5.950

Anderson, N. and West, M. (1998) Measuring climate for work group innovation:development and validation of the team climate inventory. *Journal of Organizational Behavior*, *19* (3), 235-258. https://doi.org/10.1002/(SICI)1099-1379(199805)19:3<235::AID-JOB837>3.0.CO;2-C

*Ansburg, P. I., & Dominowski, R. I. (2000). Promoting insightful problem solving. *The Journal of Creative Behavior*, *34*(1), 30–60. https://doi.org/10.1002/j.2162-6057.2000.tb01201.x

*Azevedo, I., Fátima Morais, M., & Martins, F. (2017). The Future Problem Solving Program International: An Intervention to Promote Creative Skills in Portuguese Adolescents. *The Journal of Creative Behavior*, *53*(3), 263–273. https://doi.org/https://doi.org/10.1002/jocb.175

*Baer, J. (1988). Long-term effects of creativity training with middle school students. The *Journal of Early Adolescence*, *8*(2), 183–193. https://doi.org/10.1177/0272431688082006

*Baer, J. (1994). Divergent thinking is not a general trait: A multidomain training experiment. *Creativity Research Journal, 7*(1), 35–46. https://doi.org/10.1080/10400419409534507

*Baer, J. (1996). The effects of task-specific divergent-thinking training. *The Journal of Creative Behavior, 30*(3), 183–187. https://doi.org/10.1002/j.2162-6057.1996.tb00767.x

*Bai, H., Duan, H., Kroesbergen, E. H., Leseman, P. P. M., & Hu, W. (2020). The benefits of the learn to think program for preschoolers' creativity: an explorative study. *The Journal of Creative Behavior, 54*(3), 699–711. https://doi.org/10.1002/jocb.404

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions. *Psychological Science in the Public Interest, 21*(2), 55–97. https://doi.org/10.1177/1529100620915848

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: a practical tutorial. *Evidence-based mental health, 22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

*Barak, M. (2013). Impacts of learning inventive problem-solving principles: Students' transition from systematic searching to heuristic problem solving. *Instructional Science, 41*(4), 657-679. https://doi.org/10.1007/s11251-012-9250-5

Bartoš F., & Maier M. (2020). *RoBMA: An R Package for Robust Bayesian Meta-Analyses* (R package version 2.3.1) [computer software]. Retrieved from https://CRAN.R-project.org/package=RoBMA.

*Basadur, M., Graen, G. B., & Green, S. G. (1982). Training in creative problem solving: Effects on ideation and problem finding and solving in an industrial research organization. *Organizational Behavior & Human Performance, 30*(1), 41–70. https://doi.org/10.1016/0030-5073(82)90233-1

Becker, B. J. (2005). The failsafe N or file-drawer number. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 111–126). Chichester, UK: Wiley. https://doi.org/10.1002/0470870168.ch7

Bellanca, J., & Brandt, R. (2010). *21st Century Skills: Rethinking How Students Learn.* Bloomington, IN: Solution Tree.

*Below, M. E. (1985). *Creative reading materials and the development of students' creative thinking skills (Torrance tests, attitude, ideal child checklist)* (Doctoral Dissertation). University of Georgia, US.

*Benedek, M., Fink, A., & Neubauer, A. C. (2006). Enhancement of Ideational Fluency by Means of Computer-Based Training. *Creativity Research Journal, 18*(3), 317–328. https://doi.org/10.1207/s15326934crj1803_7

*Biles, B. R. (1976). *Creative problem-solving training for graduate and professional* students (Publication No. 7629988) [Doctoral dissertation, Kansas State University]. ProQuest Dissertations Publishing.

Birdi, K.S. (2016). Creativity Training. In P. Sparrow, H. Shipton, P. Budhwar, & A. Brown, (Eds.), *Human Resource Management, Innovation and Performance* (pp. 298-312).

Palgrave Macmillan, Houndmills, Basingstoke, Hampshire, UK https://doi.org/10.1057/9781137465191

*Blissett, S. E., & Mcgrath, R. E. (1996). The relationship between creativity and interpersonal problem-solving skills in adults. *The Journal of Creative Behavior, 30*(3), 173–182. https://doi.org/10.1002/j.2162-6057.1996.tb00766.x

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science, 8*(4), 445-454.

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: John Wiley & Sons, Ltd.

*Bott, N., Quintin, E.-M., Saggar, M., Kienitz, E., Royalty, A., Hong, D. W.-C., Ning, L., Chien, Y. -H., Hawthorne, G., & Reiss, A. L. (2014). Creativity training enhances goal-directed attention and information processing. *Thinking Skills and Creativity, 13,* 120–128. https://doi.org/10.1016/j.tsc.2014.03.005

Brophy, D. R. (1998). Understanding, measuring, and enhancing individual creative problem-solving efforts. *Creativity Research Journal, 11*(2), 123–150. https://doi.org/10.1207/s15326934crj1102_4

Bull, K. S., Montgomery, D., & Baloche, L. (1995). Teaching creativity at the college level: A synthesis of curricular components perceived as important by instructors. *Creativity Research Journal, 8*(1), 83–89. https://doi.org/10.1207/s15326934crj0801_7

*Burstiner, I. (1973). Creativity training. *The Journal of Experimental Education, 41*(4), 17–19. https://doi.org/10.1080/00220973.1973.11011419

*Bush, F. E. (1988). *Creative problem-solving and its effect on eighth-grade students* (Publication No. 8905103) [Doctoral dissertation, West Virginia University]. ProQuest Dissertations Publishing.

*Callahan, C. M., & Renzulli, J. S. (1977). The effectiveness of a creativity training program in the language arts. *Gifted Child Quarterly, 21*(4), 538–545. https://doi.org/10.1177/001698627702100412

Carmeli, A., & Schaubroeck, J.M. (2007). The influence of leaders' and other referents' normative expectations on individual involvement in creative work. *Leadership Quarterly, 18,* 35-48. https://doi.org/10.1016/j.leaqua.2006.11.001

Carter, E. C., Kofler, L. M., Forster, D. E., & McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General, 144*(4), 796–815. https://doi.org/10.1037/xge0000083

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods. *Advances in Methods and Practices in Psychological Science,* 115–144. https://doi.org/10.1177/2515245919847196

*Carter, L. K. (1983). *The effects of multimodal creativity training on the creativity of twelfth graders* (Publication No. 8321139) [Doctoral dissertation, Kent State University]. ProQuest Dissertations Publishing.

*Cassidy, S. E. (2011). *Web-based training for innovation: An examination of training regimens, training environment and the moderating influence of creative personal identity and intrinsic motivation* (Publication No. 3577711) [Doctoral dissertation, The Pennsylvania State University]. ProQuest Dissertations Publishing.

*Chacón-López, H. (2021). Increased creative production in Spanish university students of Education. *Thinking Skills and Creativity, 41*. https://doi.org/10.1016/j.tsc.2021.100908

*Chang, Y. S. (2013). Student technological creativity using online problem-solving activities. *International Journal of Technology and Design Education, 23*(3), 803-816. https://doi.org/10.1007/s10798-012-9217-5

*Cheng, M. C., & Chou, P. I. A Study of the Effect of Creative Thinking Program on Improving Students' Culinary Creativity. *Advances in Education Sciences Volume 10*, 60.

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88–113. https://doi.org/10.1016/j.edurev.2013.01.001

Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283–292. https://doi.org/10.3102/0013189X16656615

*Cheung, C. K., Roskams, T., & Fisher, D. (2006). Enhancement of creativity through a one-semester course in university. *The Journal of Creative Behavior, 40*(1), 1–25. https://doi.org/10.1002/j.2162-6057.2006.tb01264.x

*Chiu, F. C. (2015). Improving your creative potential without awareness: Overinclusive thinking training. *Thinking Skills and Creativity, 15*, 1–12. https://doi.org/10.1016/j.tsc.2014.11.001

*Clapham, M. M. (1997). Ideational skills training: a key element in creativity training programs. *Creativity Research Journal, 10*(1), 33–44. https://doi.org/10.1207/s15326934crj1001_4

*Clapham, M. M., & Schuster, D. H. (1992). Can engineering students be trained to think more creatively? *The Journal of Creative Behavior, 26*(3), 156–162. https://doi.org/10.1002/j.2162-6057.1992.tb01171.x

*Clements, D. H. (1991). Enhancement of creativity in computer environments. American *Educational Research Journal, 28*(1), 173–187. https://doi.org/10.3102/00028312028001173

*Cliatt, M. J. P., Shaw, J. M., & Sherwood, J. M. (1980). Effects of Training on the Divergent-Thinking Abilities of Kindergarten Children. *Child Development, 51*(4), 1061. https://doi.org/10.2307/1129544

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

Collins, M. A., & Amabile, T. M. (1999). *Motivation and creativity.* In R. J. Sternberg (Ed.), Handbook of creativity (pp. 297–312). Cambridge University Press.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

*Cott, A. M. (1991). *On the efficacy of programed instruction on training creative problem-solving strategies.* (Publication No. 9128037) [Doctoral dissertation, Hofstra University]. ProQuest Dissertations Publishing.

*Cropley, D. H., & Cropley, A. J. (2000). Fostering creativity in engineering undergraduates. *High Ability Studies*, *11*(2), 207-219. https://doi.org/10.1080/13598130020001223

Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. The British journal of psychiatry : the journal of mental science, 196(3), 173–178. https://doi.org/10.1192/bjp.bp.109.066001

*Cunningham, J. B., & MacGregor, J. N. (2008). Training insightful problem solving: effects of realistic and puzzle-like contexts. *Creativity Research Journal, 20*(3), 291–296. https://doi.org/10.1080/10400410802278735

*Davidson, M. K. (1981). *The effects of a training program in creativity on creative behavior of head start teachers* (Publication No. 8205768) [Doctoral dissertation, Georgia State University]. ProQuest Dissertations Publishing.

*Davis, G. A., Houtman, S. E., Warren, T. F., Roweton, W. E., Mari, S., & Belcher, T. L. (1972). *A program for training creative thinking: Inner city evaluation* (Rep. No. 224). Madison: Wisconsin Research and Development Center for Cognitive Learning. (ERIC Document Reproduction Service No. ED070809)

Del Re, A. C. & Hoyt, W. T. (2014). *MAd: Meta-Analysis with Mean Differences* (R package version 0.8-1) [computer software]. Retrieved from http://cran.r-project.org/web/packages/MAd

*Dellas, M. (1971). *Effects of creativity training, defensiveness, and intelligence on divergent thinking*. American Educational Research Association Convention in New York, New York. URL: https://eric.ed.gov/?id=ED047327

*Dineen, R., & Niu, W. (2008). The effectiveness of western creative teaching methods in China: An action research project. Psychology of Aesthetics, Creativity, and the Arts, 2(1), 42–52. https://doi.org/10.1037/1931-3896.2.1.42

*Doron, E. (2016). Short term intervention model for enhancing divergent thinking among school aged children. *Creativity Research Journal, 28*(3), 372–378. https://doi.org/10.1080/10400419.2016.1195616

*Doron, E. (2017). Fostering creativity in school aged children through perspective taking and visual media based short term intervention program. *Thinking Skills and Creativity, 23*, 150–160. https://doi.org/10.1016/j.tsc.2016.12.003

*Dow, G. T., & Mayer, R. E. (2004). Teaching Students to Solve Insight Problems: Evidence for Domain Specificity in Creativity Training. *Creativity Research Journal, 16*(4), 389–402. https://doi.org/10.1207/s15326934crj1604_2

Durham Commission. (2019). *Durham Commission on Creativity and Education.* Durham.: Durham University. https://www.dur.ac.uk/creativitycommission/report

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

*Dyson, S. B., Chang, Y. L., Chen, H. C., Hsiung, H. Y., Tseng, C. C., & Chang, J. H. (2016). The effect of tabletop role-playing games on the creative potential and emotional creativity of Taiwanese college students. *Thinking Skills and Creativity*, *19*, 88–96. https://doi.org/10.1016/j.tsc.2015.10.004

*Dziedziewicz, D., Gajda, A., & Karwowski, M. (2014). Developing children's intercultural competence and creativity. *Thinking Skills and Creativity, 13*, 32–42. https://doi.org/10.1016/j.tsc.2014.02.006

*Dziedziewicz, D., Oledzka, D., & Karwowski, M. (2013). Developing 4- to 6-year-old children's figural creativity using a doodle-book program. *Thinking Skills and Creativity, 9*, 85–95. https://doi.org/10.1016/j.tsc.2012.09.004

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology.* Dover.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical research ed.), 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Eisenberger, R., & Shanock, L. (2003). Rewards, Intrinsic Motivation, and Creativity: A Case Study of Conceptual and Methodological Isolation. *Creativity Research Journal, 15*(2-3), 121–130. https://doi.org/10.1207/S15326934CRJ152&3_02

*Engelman, M. A. (1981). The response of older women to a creative problem-solving program. *The Journal of Creative Behavior, 12*(4), 278–278. https://doi.org/10.1080/0380127810060206

Eronen, M. I., & Bringmann, L. F. (2021). The Theory Crisis in Psychology: How to Move Forward. *Perspectives on Psychological Science, 16*(4), 779–788. https://doi.org/10.1177/1745691620970586

European Commission. (2016). *Community Innovation Survey 2016.* https://ec.europa.eu/eurostat/web/microdata/community-innovation-survey

*Fehr, K. K., & Russ, S. W. (2016). Pretend play and creativity in preschool-age children: Associations and brief intervention. *Psychology of Aesthetics, Creativity, and the Arts, 10*(3), 296–308. https://doi.org/10.1037/aca0000054

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science, 7*(6), 555-561. https://doi.org/10.1177/1745691612459059

*Fink, A., Benedek, M., Koschutnig, K., Pirker, E., Berger, E., Meister, S., Neubauer, A. C., Papousek, I., & Weiss, E. M. (2015). *Training of verbal creativity modulates brain activity in regions associated with language- and memory-related demands. Human brain mapping, 36*(10), 4104–4115. https://doi.org/10.1002/hbm.22901

*Fink, A., Koschutnig, K., Benedek, M., Reishofer, G., Ischebeck, A., Weiss, E. M., & Ebner, F. (2012). *Stimulating creativity via the exposure to other people's ideas. Human brain mapping, 33*(11), 2603–2610. https://doi.org/10.1002/hbm.21387

Fischer, B. D., & Afifi, R. (2013). An analysis of creativity and innovation training in organizations. *International Journal of Management and Human Resources, 1*(1), 170-179.

Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. arXiv preprint arXiv:1503.02220.

*Flaherty, M. A. (1992), The effects of a holistic creativity program on the self-concept and creativity of third graders. *The Journal of Creative Behavior, 26*(3), 165–171. https://doi.org/10.1002/j.2162-6057.1992.tb01173.x

*Franklin, B. S., & Richards, P. N. (1977). Effects on children's divergent thinking abilities of a period of direct teaching for divergent production. *British Journal of Educational Psychology*, *47*(1), 66-70. https://doi.org/10.1111/j.2044-8279.1977.tb03002.x

Friese, M., Frankenbach, J., Job, V., & Loschelder, D. D. (2017). Does self-control training improve self-control? A meta-analysis. *Perspectives on Psychological Science, 12*(6), 1077–1099. https://doi.org/10.1177/1745691617697076

*Furze, C. T., Tyler, J. G., & Reynolds, P. M. (1984). Training creativity of children assessed by the obscure figures test. *Perceptual and Motor Skills, 58*(1), 231–234. https://doi.org/10.2466/pms.1984.58.1.231

*Garaigordobil, M. (2006). *Intervention in Creativity With Children Aged 10 and 11 Years: Impact of a Play Program on Verbal and Graphic-Figural Creativity. Creativity Research Journal*, *18*(3), 329–345. https://doi.org/10.1207/s15326934crj1803_8

*Gendrop, S. C. (1996). Effect of an intervention in synectics on the creative thinking of nurses. *Creativity Research Journal*, *9*(1), 11–19. https://doi.org/10.1207/s15326934crj0901_2

*George, J. C. (1980). *The effects of "talents unlimited" on planning and productive thinking of fifth grade students in Frederick, Maryland* (Publication No. 8107684) [Doctoral dissertation, American University]. ProQuest Dissertations Publishing.

*Gerrard, L. E., Poteat, G. M., & Ironsmith, M. (1996). Promoting children's creativity: Effects of competition, self-esteem, and immunization. *Creativity Research Journal*, *9*(4), 339-346. https://doi.org/10.1207/s15326934crj0904_5

*Gilbert, F. W., Prenshaw, P. J., & Ivy, T. T. (1996). A preliminary assessment of the effectiveness of creativity training in marketing. *Journal of Marketing Education, 18*(3), 46–56. https://doi.org/10.1177/027347539601800306

*Glover, J. A. (1980). A Creativity-Training Workshop: Short-Term, Long-Term, and Transfer Effects. *The Journal of Genetic Psychology*, *136(1*), 3–16. https://doi.org/10.1080/00221325.1980.10534091

Gobet, F., & Sala, G. (2022). Cognitive Training: A Field in Search of a Phenomenon. *Perspectives on Psychological Science.* Advance online publication. https://doi.org/10.1177/17456916221091830

*Golovin, R. W. (1993). *Enhanced creativity as a function of classroom structure and level of originality* (Publication No. 9331616) [Doctoral dissertation, Mississippi State University]. ProQuest Dissertations Publishing.

*Goor, A., & Rapoport, T. (1977). Enhancing creativity in an informal educational framework. *Journal of Educational Psychology, 69*(5), 636–643. https://doi.org/10.1037/0022-0663.69.5.636

Goulet, M. A., & Cousineau, D. (2019). The power of replicated measures to increase statistical power. *Advances in Methods and Practices in Psychological Science, 2*(3), 199–213. https://doi.org/10.1177/2515245919849434

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

Haase, J., Hanel, P. H. P., & Gronau, N. (2023). Creativity enhancement methods for adults: A meta-analysis. Psychology of Aesthetics, Creativity, and the Arts. Advance online publication. https://doi.org/10.1037/aca0000557

*Haley, G. L. (1984). Creative response styles: the effects of socioeconomic status and problem-solving training. *The Journal of Creative Behavior, 18*(1), 25–40. https://doi.org/10.1002/j.2162-6057.1984.tb00987.x

*Hargrove, R. A. (2007). *Creating creativity in the design studio: Assessing the impact of metacognitive skill development on creative abilities* (Publication No. 3329259) [Doctoral dissertation, North Carolina State University]. ProQuest Dissertations Publishing.

Hattie, J. (1980). Should creativity tests be administered under testlike conditions? An empirical study of three alternative conditions. Journal of Educational Psychology, 72(1), 87–98. https://doi.org/10.1037/0022-0663.72.1.87

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hedges, L. V. (Ed.). (2009). Statistical considerations. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 37–46). Russell Sage Foundation.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology, 61*(1), 569–598. https://doi.org/10.1146/annurev.psych.093008.100416

*Hennessey, B. A., & Zbikowski, S. M. (1993). Immunizing children against the negative effects of reward: A further examination of intrinsic motivation training techniques. *Creativity Research Journal, 6*(3), 297–307. https://doi.org/10.1080/10400419309534485

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *The Behavioral and brain sciences*, *33*(2-3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savovic, J., Schulz, K. F., Weeks, L., Sterne, J. A., Cochrane Bias Methods Group, & Cochrane Statistical Methods Group (2011). *The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ (Clinical research ed.)*, 343, d5928. https://doi.org/10.1136/bmj.d5928

*Higuchi, T., Yuizono, T., Miyata, K., Sakurai, K., Kawaji, T. (2016). Creativity Effects of Idea-Marathon System (IMS): Torrance Tests of Creative Thinking (TTCT) Figural Tests for College Students. In: Skulimowski, A., Kacprzyk, J. (eds) *Knowledge, Information and Creativity Support Systems: Recent Trends, Advances and Solutions.*

*Hocking, I., & Vernon, D. (2017). The right tool for the right task: Structured techniques prove less effective on an ill-defined problem finding task. *Thinking Skills and Creativity, 26*, 84–91. https://doi.org/10.1016/j.tsc.2017.08.001

*Hoffmann, J. D., Ivcevic, Z., & Maliakkal, N. (2018). Creative Thinking Strategies for Life: A Course for Professional Adults Using Art. *The Journal of Creative Behavior, 54*(2), 293–310. https://doi.org/10.1002/jocb.366

*Hoffmann, J. D., Ivcevic, Z., & Maliakkal, N. (2020). Emotions, Creativity, and the Arts: Evaluating a Course for Children. *Empirical Studies of the Arts, 39*(2), 123–148. https://doi.org/10.1177/0276237420907864

Hommel, B. (2012). Convergent and divergent operations in cognitive search. In: P. M. Todd, T. T. Hills, & T. W. Robbins (eds.), *Cognitive search: Evolution, algorithms, and the brain* (pp. 221–235). Cambridge, MA: MIT Press.

*Houtz, J. C., & Feldhusen, J. F. (1976). The modification of fourth graders' problem solving abilities. *The Journal of Psychology*, *93*(2), 229-237. https://doi.org/10.1080/00223980.1976.9915817

*Houtz, J. C., & Frankel, A. D. (1992). Effects of incubation and imagery training on creativity. *Creativity Research Journal*, *5*(2), 183–189. https://doi.org/10.1080/10400419209534432

*Hu, W., Wu, B., Jia, X., Yi, X., Duan, C., Meyer, W., & Kaufman, J. C. (2013). Increasing students' scientific creativity: The “learn to think” intervention program. The Journal of *Creative Behavior, 47*(1), 3-21. https://doi.org/10.1002/jocb.20

*Huang, S., Ko, P., Lin, H., Dai, R., & Chen, H. (2021). Creative Thinking Counseling Teaching Program Can Improve the Creativity, Creative Tendency, and Self-Concept of Grade 7 Students: A Quasi-Experimental Study. *The Journal of Creative Behavior, 55*(3), 819–838. https://doi.org/10.1002/jocb.491

*Huber, J., Treffinger, D., Tracy, D., & Rand, D. (1979). Self-instructional use of programmed creativity-training materials with gifted and regular students. *Journal of Educational Psychology, 71*(3), 303–309. https://doi.org/10.1037/0022-0663.71.3.303

*Hueftle, R. K. (1992). *Focusing technique to facilitate creative problem-solving* (Publication No. 9237600) [Doctoral dissertation, University of Denver]. ProQuest Dissertations Publishing.

*Huo, K. (2020). Performance Incentives, Divergent Thinking Training, and Creative Problem Solving. *Journal of Management Accounting Research, 32*(1), 159–176. https://doi.org/10.2308/jmar-52479

IBM Corporation. (2010). *Capitalizing on Complexity: Insights from the Global Chief Exexcutive Officers Study*. https://www.ibm.com/downloads/cas/1VZV5X8J

Inzlicht, M., Gervais, W. M., Berkman, E. T. (2015). *Bias-correction techniques alone cannot determine whether ego depletion is different from zero: Commentary on Carter, Kofler, Forster, & McCullough, 2015*. Retrieved from http://ssrn.com/abstract=2659409

Ioannidis J. P. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*(5), 640–648. https://doi.org/10.1097/EDE.0b013e31818131e7

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling. *Psychological Science, 23*(5), 524–532. https://doi.org/10.1177/0956797611430953

*Jones-Chick, R., Kelloway, E. K., & Birdi, K. (2021). CLEAR IDEAS: Can Idea Implementation Training Enhance the Development of New Ideas Beyond Idea Generation Training? *The Journal of Creative Behavior, 56*(1), 79–91. https://doi.org/10.1002/jocb.517

*Ju Lee, Y., Bain, S. K., & McCallum, R. S. (2007). Improving Creative Problem-Solving in a Sample of Third Culture Kids. *School Psychology International, 28*(4), 449–463. https://doi.org/10.1177/0143034307084135

*Kabanoff, B., & Bottger, P. (1991). Effectiveness of creativity training and its relation to selected personality factors. *Journal of Organizational Behavior, 12*(3), 235–248. https://doi.org/10.1002/job.4030120306

*Karaca, T. & Koray, O (2017). The Effect of Using the Creative Reversal Act in Science Education on Middle School Students' Creativity Levels. *Eurasian Journal of Educational Research, 17*(67), 199–214. https://doi.org/10.14689/ejer.2017.67.12

*Karakelle, S. (2009). Enhancing fluent and flexible thinking through the creative drama process. *Thinking Skills and Creativity, 4*(2), 124–129. https://doi.org/10.1016/j.tsc.2009.05.002

*Kashani-Vahid L., Afrooz, G. A., Shokoohi-Yekta, M., Kharrazi K., & Bonab, B. G. (2015) Creative interpersonal problem solving program: Igniting creativity in elementary students. *2015 Sixth International Conference of Cognitive Science (ICCS)*, 51-56, https://doi.org/10.1109/COGSCI.2015.7426667.

*Khatena, J. (1970). Training college adults to think creatively with words. *Psychological Reports, 27*(2), 279–281. https://doi.org/10.2466/pr0.1970.27.2.27

*Khatena, J. (1971). A second study training college adults to think creatively with words. *Psychological Reports, 28*(2), 385–386. https://doi.org/10.2466/pr0.1971.28.2.385

*Khatena, J. (1971). Teaching disadvantaged preschool children to think creatively with pictures. *Journal of Educational Psychology, 62*(5), 384–386. https://doi.org/10.1037/h0031634

*Khatena, J. (1973). Creative level and its effects on training college adults to think creatively with words. *Psychological Reports, 32*(1), 336–336. https://doi.org/10.2466/pr0.1973.32.1.336

*Khatena, J., & Dickerson, E. C. (1973). Training sixth grade children to think creatively with words. *Psychological Reports, 32*(3), 841–842. https://doi.org/10.2466/pr0.1973.32.3.841

*Kienitz, E., Quintin, E.-M., Saggar, M., Bott, N. T., Royalty, A., Hong, D. W.-C., Liu, N., Chien, Y.-h., Hawthorne, G., & Reiss, A. L. (2014). Targeted intervention to increase creative capacity and performance: A randomized controlled pilot study. *Thinking Skills and Creativity, 13*, 57–66. https://doi.org/10.1016/j.tsc.2014.03.002

*Kim, H. J., Park, J. H., Yoo, S., & Kim, H. (2016). Fostering creativity in tablet-based interactive classrooms. *Journal of Educational Technology & Society, 19*(3), 207-220.

*Kim, S., Choe, I., & Kaufman, J. C. (2019). The development and evaluation of the effect of creative problem-solving program on young children's creativity and character. *Thinking Skills and Creativity, 33*, 100590. https://doi.org/10.1016/j.tsc.2019.100590

*Kim, S., Chung, K., & Yu, H. (2013). Enhancing digital fluency through a training program for creative problem solving using computer programming. *The Journal of Creative Behavior*, *47*(3), 171–199. https://doi.org/10.1002/jocb.30

*Kirisci, N., Sak, U., & Karabacak, F. (2020). The effectiveness of the selective problem solving model on students' mathematical creativity: A Solomon four-group research. *Thinking Skills and Creativity, 38*. https://doi.org/10.1016/j.tsc.2020.100719

*Kobe, L. M. (2001). *Computer-based creativity training: Training the creative process* (Publication No. 3022642) [Doctoral dissertation, The University of Nebraska]. ProQuest Dissertations Publishing.

*Komarik, E., & Brutenicova, E. (2003). Effect of creativity training on preschool children. *Studia Psychologica*, *45*, 37–42.

*Kovác, T. (1998). Effects of creativity training in young soccer talents. *Studia Psychologica*, *40*(3), 211.

Kraemer, H. C., Noda, A., & O'Hara, R. (2004). Categorical versus dimensional approaches to diagnosis: Methodological challenges. *Journal of Psychiatric Research, 38*, 17-25. https://doi.org/10.1016/S0022- 3956(03)00097-9

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PloS one, 9*(9), e105825. https://doi.org/10.1371/journal.pone.0105825

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*(4), 423–434. https://doi.org/10.1038/s41562-019-0787-z

Lane, D. M., & Dunlap, W. P. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology, 31*(2), 107–112. https://doi.org/10.1111/j.2044-8317.1978.tb00578.x

Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *BMJ (Clinical research ed.)*, *333*(7568), 597–600. https://doi.org/10.1136/bmj.333.7568.597

*Leach, C. E. (2009). An investigation of training in creative problem solving and its relationship to affective and effective idea generation of entrepreneurial learners (Publication No. 350010) [Doctoral dissertation, Nova Southeastern University]. ProQuest Dissertations Publishing.

*Lee, D. W. (2006). *The impact of nurturing creativity in language arts classrooms for improving students' creative products* (Publication No. 3211919) [Doctoral dissertation, Tennessee State University]. ProQuest Dissertations Publishing.

*Liao, Y. H., Chen, Y. L., Chen, H. C., & Chang, Y. L. (2018). Infusing creative pedagogy into an English as a foreign language classroom: Learning performance, creativity, and motivation. *Thinking Skills and Creativity, 29,* 213–223. https://doi.org/10.1016/j.tsc.2018.07.007

*Lin, W. L., & Shih, Y. L. (2016). Designing EEG neurofeedback procedures to enhance open-ended versus closed-ended creative potentials. *Creativity Research Journal, 28*(4), 458–466. https://doi.org/10.1080/10400419.2016.1229979

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*(12), 1181–1209. https://doi.org/10.1037/0003-066x.48.12.1181

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Sage Publications, Inc.

*Liu, H. Y., Wang, I. T., Huang, D. H., Hsu, D. Y., & Han, H. M. (2020). Nurturing and Enhancing Creativity of Nursing Students in Taiwan: A Quasi-Experimental Study. The Journal of Creative Behavior, 54(4), 799-814. https://doi.org/10.1002/jocb.407

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management, 64,* 19-40. https://doi.org/10.1016/j.jom.2018.10.003

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology, 51*(3), 485–504. https://doi.org/10.1002/ejsp.2752

Ma, H. H. (2006). A synthetic analysis of the effectiveness of single components and packages in creativity training programs. *Creativity Research Journal, 18*(4), 435-446. https://doi.org/10.1207/s15326934crj1804_3

Macnamara, B. N., & Burgoyne, A. P. (2023). Do growth mindset interventions impact students' academic achievement? A systematic review and meta-analysis with recommendations for best practices. Psychological Bulletin, 149(3-4), 133–173. https://doi.org/10.1037/bul0000352

*Maddox, E. N. (1987). *The effects of problem-solving strategy and outcome expectancy cues on creative problem-solving performance* (Publication No. 8723136) [Doctoral dissertation, The Florida State University]. ProQuest Dissertations Publishing.

*Mahboub, K. C., Portillo, M. B., Liu, Y., & Chandraratna, S. (2004). Measuring and enhancing creativity. *European Journal of Engineering Education*, *29*(3), 429-436. https://doi.org/10.1080/03043790310001658541

*Maier, M., Bartoš, F., & Wagenmakers, E. J. (2022). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*, 10.1037/met0000405. Advance online publication. https://doi.org/10.1037/met0000405

Maier, M., Bartoš, F., & Wagenmakers, E. J. (2022). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological methods, 10.1037/met0000405*. Advance online publication. https://doi.org/org/10.1037/met0000405

Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J., & Wagenmakers, E. J. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences*, *119*(31), e2200300119. https://doi.org/10.1073/pnas.2200300119

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43*(6), 304-316. https://doi.org/10.3102/0013189X14545513

*Malycha, C. P., & Maier, G. W. (2017). Enhancing creativity on different complexity levels by eliciting mental models. *Psychology of Aesthetics, Creativity, and the Arts, 11*(2), 187–201. https://doi.org/10.1037/aca0000080

*Malycha, C. P., & Maier, G. W. (2017). The Random-Map Technique: Enhancing Mind-Mapping with a Conceptual Combination Technique to Foster Creative Potential. *Creativity Research Journal, 29*(2), 114–124. https://doi.org/10.1080/10400419.2017.1302763

Mansfield, R. S., Busse, T. V., & Krepelka, E. J. (1978). The Effectiveness of Creativity Training. *Review of Educational Research, 48*(4), 517 536. https://doi.org/10.3102/00346543048004517

*Markewitz, D. A. (1982). *The influence of creativity intervention training on the adjustment potential of kindergarten children.* (Publication No. 8400811) [Doctoral dissertation, University of Santo Tomas(The Philippines)]. ProQuest Dissertations Publishing.

*Mathisen, G. E., & Bronnick, K. S. (2009). Creative self-efficacy: An intervention study. *International Journal of Educational Research, 48*(1), 21–29. https://doi.org/10.1016/j.ijer.2009.02.009

*May, J., Redding, E., Whatley, S., Łucznik, K., Clements, L., Weber, R., ... & Reed, S. (2020). Enhancing creativity by training metacognitive skills in mental imagery. *Thinking Skills and Creativity, 38*. https://doi.org/10.1016/j.tsc.2020.100739

*McCormack, A. J. (1974). Training Creative Thinking in General Education Science. *Journal of College Science Teaching, 4*(1), 10-15. https://www.jstor.org/stable/42965425

*McNair, M. (2008). *Creativity enhancement of information systems personnel: A study of the effects of creativity training and individual creative support system software* (Publication No. 3317640) [Doctoral dissertation, TUI University]. ProQuest Dissertations Publishing.

*Meador, K. S. (1995). The Effect of Synectics Training on Gifted and Nongifted Kindergarten Students. *Journal for the Education of the Gifted, 18*(1), 55–73. https://doi.org/10.1177/016235329401800105

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220–232. https://doi.org/10.1037/h0048850

*Meichenbaum, D. (1975). Enhancing creativity by modifying what subjects say to themselves. *American Educational Research Journal, 12*(2), 129–145. https://doi.org/10.2307/1162416

*Merrill, J. (2014). Increasing Creativity in Design Education: Measuring the E/Affect of Cognitive Exercises on Student Creativity (Publication No. 3627756) [Doctoral dissertation, Kansas State University]. ProQuestion Dissertation Publishing.

*Missett, T. C. (2012). *The Development of Critical and Creative Thinking Skills for 21st Century Learning* (Publication No. ED545497) [Doctoral dissertation, University of Virginia]. ProQuest LLC.

Mohr, D. C., Spring, B., Freedland, K. E., Beckner, V., Arean, P., Hollon, S. D., Ockene, J., & Kaplan, R. (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychotherapy and psychosomatics, 78*(5), 275–284. https://doi.org/10.1159/000228248

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M., & Davey Smith, G. (2018). Collider scope: when selection bias can substantially influence observed associations. *International Journal of Epidemiology, 47*(1), 226-235. https://doi.org/10.1093/ije/dyx206

*Nelson, A., & Lalemi, B. (1991). The role of imagery training on Tohono O'odham children's creativity scores. *Journal of American Indian Education, 9*(3), 24-32. https://www.jstor.org/stable/24397982

Nielsen, K., & Miraglia, M. (2017). What works for whom in which circumstances? On the need to move beyond the 'what works?' question in organizational intervention research. *Human Relations, 70*(1), 40–62. https://doi.org/10.1177/0018726716670226

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual review of psychology, 73*, 719-748. https://doi.org/10.1146/annurev-psych-020821-114157

*O'Connor, P. J., Gardiner, E., & Watson, C. (2016). Learning to relax versus learning to ideate: Relaxation-focused creativity training benefits introverts more than extraverts. *Thinking Skills and Creativity, 21*, 97–108. https://doi.org/10.1016/j.tsc.2016.05.008

*Pannells, T. C. (2010). *The effects of training preservice teacher in creative problem solving and classroom management* (Publication No. ED522568) [Doctoral dissertation, The University of Oklahoma]. ProQuest LLC.

*Parikh, C., Maddulety, K., & Meadows, C. (2020). Improving creative ability of base of pyramid (BOP) students in India. *Thinking Skills and Creativity, 36*. https://doi.org/10.1016/j.tsc.2020.100652

*Parnes, S. J., & Noller, R. B. (1972). Applied creativity: The creative studies project: II. Results of the two-year program. *The Journal of Creative Behavior, 6*(3), 164–186. https://doi.org/10.1002/j.2162-6057.1972.tb00927.x

*Piers, E. V., & Morgan, F. T. (1973). Effects of free association training on children's ideational fluency. *Journal of Personality, 41*(1), 42–49. https://doi.org/10.1111/j.1467-6494.1973.tb00658.x

PISA (2022). *PISA 2022 Assessment and Analytical Framework*. https://www.oecd.org/pisa/publications/pisa-2021-assessment-and-analytical-framework.htm

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

Protzko, J., & Schooler, J. W. (2017). Decline effects: Types, mechanisms, and personal reflections. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 85–107). Wiley Blackwell. https://doi.org/10.1002/9781119095910.ch6

Pustejovsky, J. (2016). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections* (R package version 0.2.1.9000) [computer software]. Retrieved from https://github.com/jepusto/clubSandwich

Reed, W. R. (2015). A Monte Carlo analysis of alternative meta-analysis estimators in the presence of publication bias. *Economics, 9*, 1–40. doi: 10.7910/DVN/OI8XSG

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*(2), 144–152. https://doi.org/10.1037/aca0000227

*Richard, V., Lebeau, J. C., Becker, F., Boiangin, N., & Tenenbaum, G. (2018). Developing cognitive and motor creativity in children through an exercise program using nonlinear pedagogy principles. *Creativity Research Journal, 30*(4), 391-401. https://doi.org/10.1080/10400419.2018.1530913

*Ritter, S. M., Gu, X., Crijns, M., & Biekens, P. (2020). Fostering students' creative thinking skills by means of a one-year creativity training program. *PLOS ONE, 15*(3), e0229773. https://doi.org/10.1371/journal.pone.0229773

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods, 26*(2), 141–160. https://doi.org/10.1037/met0000300

Rose, L. H., & Lin, H.-T. (1984). A meta-analysis of long-term creativity training programs. *The Journal of Creative Behavior, 18*(1), 11–22. https://doi.org/10.1002/j.2162-6057.1984.tb00985.x

*Rostafinski, T. J. (1981). *Brief creativity training: A controlled comparison of self-instructional and associational procedures* (Publication No. 8201455) [Doctoral dissertation, Ohio University]. ProQuest Dissertations Publishing.

Runco, M. A. (2007). *Creativity: Theories and themes: Research, development, and practice.* Elsevier Academic Press.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92-96. https://doi.org/10.1080/10400419.2012.650092

*Saggar, M., Quintin, E. M., Bott, N. T., Kienitz, E., Chien, Y. H., Hong, D. W., Liu, N., Royalty, A., Hawthorne, G., & Reiss, A. L. (2017). Changes in Brain Activation Associated with Spontaneous Improvization and Figural Creativity After Design-Thinking-Based Training: A Longitudinal fMRI Study. *Cerebral Cortex, 27*(7), 3542–3552. https://doi.org/10.1093/cercor/bhw171

*Şahin, F. (2014). The effectiveness of mentoring strategy for developing the creative potential of the gifted and non-gifted students. *Thinking Skills and Creativity, 14*, 47–55. https://doi.org/10.1016/j.tsc.2014.07.002

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., Van den Noortgate, W., & Barbot, B. (2020). Does the Fourth-Grade Slump in Creativity Actually Exist? A Meta-analysis of the Development of Divergent Thinking in School-Age Children and Adolescents. *Educational Psychology Review, 33*(1), 275–298. https://doi.org/10.1007/s10648-020-09547-9

Said-Metwaly, S., Taylor, C. L., Camarda, A., & Barbot, B. (2022). Divergent thinking and creative achievement—How strong is the link? An updated meta-analysis. *Psychology* *of Aesthetics, Creativity, and the Arts*. Advance online publication. https://doi.org/10.1037/aca0000507

*Sannomiya, M., & Yamaguchi, Y. (2016). Creativity training in causal inference using the idea post-exposure paradigm: Effects on idea generation in junior high school students. *Thinking Skills and Creativity, 22*, 152–158. https://doi.org/10.1016/j.tsc.2016.09.006

*Sannomiya, M., Mashimo, T., & Yamaguchi, Y. (2021). Creativity training for multifaceted inferences of reason behind others' behaviors. *Thinking Skills and Creativity, 39*. https://doi.org/10.1016/j.tsc.2020.100757

*Santos, S., Jiménez, S., Sampaio, J., & Leite, N. (2017). Effects of the Skills4Genius sports-based training program in creative behavior. *PLOS ONE, 12*(2), e0172520. https://doi.org/10.1371/journal.pone.0172520

Sawyer, R. K. (2006). *Explaining creativity: The science of human innovation.* New York Oxford University Press.

Scott, G., Leritz, L. E., & Mumford, M. D. (2004). The effectiveness of creativity training: A quantitative review. *Creativity Research Journal, 16*(4), 361-388. https://doi.org/10.1080/10400410409534549

*Shadinger, T. V. (2004). *Optimization of creativity: The roles of processes and components* (Publication No. 3128181) [Doctoral dissertation, The University of Alabama]. ProQuest Dissertations Publishing.

*Shaffer-Zamari, M. (1983). *Effects of an instructional program and teacher's verbal praise on the creative performance of fourth grade children* (Publication No. 8323145) [Doctoral dissertation, Hofstra University]. ProQuest Dissertations Publishing.

*Sherief, N. M. S. (1978). *The effects of creativity training, classroom atmosphere and cognitive style on the creative thinking abilities of egyptian elementary school children* (Publication No. 7914969) [Doctoral dissertation, Purdue University]. ProQuest Dissertations Publishing.

*Sikka, A. (1991). *The effect of creativity training methods on the creative thinking of fourth, fifth, and sixth-grade minority students* (Publication No. 9312052) [Doctoral dissertation, The Mississippi State University]. ProQuest Dissertations Publishing.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *p*-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. *Perspectives on Psychological Science, 9*(6), 666–681. https://doi.org/10.1177/1745691614553988

Sio, U. N., & Lortie-Forgues, H. (2023, December 7). Meta-Analysis of Creativity Training. Retrieved from osf.io/8z3k5

Slavin, R. E. (2008). Perspectives on Evidence-Based Research in Education—What Works? Issues in Synthesizing Educational Program Evaluations. *Educational Researcher, 37*(1), 5–14. https://doi.org/10.3102/0013189X08314117

*Smogorzewska, J. (2014). Developing children's language creativity through telling stories– An experimental study. *Thinking Skills and Creativity, 13*, 20-31. https://doi.org/10.1016/j.tsc.2014.02.005

Stanley T. D. (2008). Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *Oxford Bulletin of Economics and Statistics, 70*, 103–127.

Stanley T. D., Doucouliagos C. H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5,* 60–78.

Stanley, T. D., & Doucouliagos, H. (2007). Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. *Economics Series*, *11*, 2007.

Stanley, T. D., Jarrell, S. B., & Doucouliagos, H. (2010). Could it be better to discard 90% of the data? A statistical paradox. *The American Statistician, 64*(1), 70-77. https://doi.org/10.1198/tast.2009.08205

*Staso, W. H. (1977). *The effects of simulation games and creativity training on children's divergent thinking* (Publication No. 7729102) [Doctoral dissertation, The University of Texas at Austin]. ProQuest Dissertations Publishing.

Sternberg, R. J., & O'Hara, L. A. (1999). Creativity and intelligence. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 251–272). Cambridge University Press.

Sterne, J. A., Egger, M., & Smith, G. D. (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ (Clinical research ed.), 323*(7304), 101–105. https://doi.org/10.1136/bmj.323.7304.101

Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. Journal of clinical epidemiology, 53(11), 1119–1129. https://doi.org/10.1016/s0895-4356(00)00242-0

*Stevenson, C. E., Kleibeuker, S. W., de Dreu, C. K. W., & Crone, E. A. (2014). Training creative cognition: adolescence as a flexible period for improving creativity. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00827

*Sun, J., Chen, Q., Zhang, Q., Li, Y., Li, H., Wei, D., Yang, W., & Qiu, J. (2016). Training your brain to be more creative: brain functional and structural changes induced by divergent thinking training. *Human Brain Mapping, 37*(10), 3375–3387. https://doi.org/10.1002/hbm.23246

*Sun, M., Wang, M., & Wegerif, R. (2020). Effects of divergent thinking training on students' scientific creativity: The impact of individual creative potential and domain knowledge. *Thinking Skills and Creativity, 37*. https://doi.org/10.1016/j.tsc.2020.100682

*Suwantra, P. (1994). *Effects of the creativity training program on preschoolers* (Publication No. 9521343) [Doctoral dissertation, Illinois State University]. ProQuest Dissertations Publishing.

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*, 13–30. https://doi.org/10.1002/jrsm.1091

Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009). Transfer of management training from alternative perspectives. *Journal of Applied Psychology*, *94*(1), 104. https://doi.org/10.1037/a0013006

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: its causes and consequences. *Journal of clinical epidemiology*, *53*(2), 207–216. https://doi.org/10.1016/s0895-4356(99)00161-4

Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal, 45*(6), 1137–1148. https://doi.org/10.2307/3069429

Tierney, P., & Farmer, S. M. (2004). The Pygmalion Process and Employee Creativity. *Journal of Management*, *30*(3), 413–432. https://doi.org/org/10.1016/j.jm.2002.12.001

*Tisone, J. M. (1985). *Enhancement of creative abilities in middle school students using computer technologies compared to traditional approaches* (Publication No. 8526262) [Doctoral dissertation, The Ohio State University]. ProQuest Dissertations Publishing.

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking-Norms-Technical Manual Research Edition-Verbal Tests, Forms A and B-Figural Tests, Forms A and B.* Princeton, NJ: Personnel Press.

Tsai, K. C. (2013). A review of the effectiveness of creative training on adult learners. *Journal of Social Science Studies*, *1*(1), 17-30. https://doi.org/10.5296/jsss.v1i1.4329

*Tull, C. M. (2014). *An Experiment in Facilitating Creative Thinking in Second Degree Accelerated Nursing Students.* Doctoral dissertation. Texas State University.

*Tweet, C. A. C. (1980). *The effects of the implementation of creativity training in the elementary school social studies curriculum* (Publication No. 8017751) [Doctoral dissertation, Montana State University]. ProQuest Dissertations Publishing.

UK Innovation Survey (2019). *UK Innovation Survey 2019: Statistical Annex.* Department for Business, Energy & Industrial Strategy. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/903591/UK_Innovation_Survey_2019_Statistical_Annex.xlsx

*Ulger, K. (2016). The creative training in the visual arts education. *Thinking Skills and Creativity, 19*, 73–87. https://doi.org/10.1016/j.tsc.2015.10.007

UN General Assembly (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. UN General Assembly. https://www.refworld.org/docid/57b6e3e44.html

UNESCO Institute for Statistics (2017). *Summary Report of the 2015 UIS Innovation.* The UNESCO Institute for Statistics http://uis.unesco.org/sites/default/files/documents/ip37-summary-report-of-the-2015-uis-innovation-data-collection-2017-en.pdf

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and Enhancing Spatial Thinking: Links to Achievement in Science, Technology, Engineering, and Mathematics? *Current Directions in Psychological Science, 22*(5), 367–373. https://doi.org/10.1177/0963721413484756

Valentine, J. C. (2019). Incorporating judgments about study quality into research syntheses. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 129–140). Russell Sage Foundation.

Valgeirsdottir, D., & Onarheim, B. (2017). Studying creativity training programs: A methodological analysis. *Creativity and Innovation Management, 26*, 430-439. https://doi.org/10.1111/caim.12245 van Aert, R., Wicherts, J. M., & van Assen, M. (2019). Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PloS one*, *14*(4), e0215052. https://doi.org/10.1371/journal.pone.0215052

*Vernon, D., & Hocking, I. (2014). Thinking hats and good men: Structured techniques in a problem construction task. *Thinking Skills and Creativity, 14,* 41–46. https://doi.org/10.1016/j.tsc.2014.07.001

*Vernon, D., & Hocking, I. (2016). Beyond belief: Structured techniques prove more effective than a placebo intervention in a problem construction task. *Thinking Skills and Creativity, 19,* 153–159. https://doi.org/10.1016/j.tsc.2015.10.009

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03

*Walinga, J., Cunningham, J. B., & MacGregor, J. N. (2011). Training insight problem solving through focus on barriers and assumptions. *The Journal of Creative Behavior, 45*(1), 47-58. https://doi.org/10.1002/j.2162-6057.2011.tb01084.x

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. N. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empiricially," all must have prizes." *Psychological bulletin*, *122*(3), 203. https://doi.org/10.1037/0033-2909.122.3.203

*Waterstreet, M. A. (1977). *The effects of amount and spacing of creativity training sessions on immediate and enduring gains in the creative production of third-grade children* (Doctoral Dissertation). University of Georgia.

West, M. A. (2002). Sparkling fountains or stagnant ponds: An integrative model of creativity and innovation implementation in work groups. *Applied psychology, 51*(3), 355-387. https://doi.org/10.1111/1464-0597.00951

*Westberg, K. L. (1996). The effects of teaching students how to invent. *The Journal of Creative Behavior, 30*(4), 249–267. https://doi.org/10.1002/j.2162-6057.1996.tb00772.x

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook (Version 4.1)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/handbooks

Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods, 6*(4), 413–429. https://doi.org/10.1037/1082-989X.6.4.413

World Economic Forum. (2018). *The future of jobs report 2018*. https://www.weforum.org/reports/the-future-of-jobs-report-2018

Yasin, R. M., & Yunus, N. S. R. (2014). A meta-analysis study on the effectiveness of creativity approaches in technology and engineering education. *Asian Social Science, 10*(3), 242-252. https://doi.org/10.5539/ass.v10n3p242

*Zahra, P., Yusooff, F., & Hasim, M. S. (2013). Effectiveness of training creativity on preschool students. *Procedia-Social and Behavioral Sciences, 102,* 643-647. https://doi.org/10.1016/j.sbspro.2013.10.782

*Zhu, H. (2015). *The effectiveness of a creativity course on developing Chinese design students' creative thinking* (Publication No. 3720939) [Doctoral dissertation, University of the Pacific]. ProQuest Dissertations Publishing.

*Zivkovic, Z., Nikolic, S. T., Doroslovacki, R., Lalic, B., Stankovic, J., & Zivkovic, T. (2015). Fostering creativity by a specially designed Doris tool. *Thinking Skills and Creativity*, *17*, 132–148. https://doi.org/10.1016/j.tsc.2015.06.004