UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Sala, G., et al. (2019). Near and Far Transfer in Cognitive Training: A Second-Order Meta-Analysis. *Collabra: Psychology*, 5(1): 18. DOI: https://doi.org/10.1525/collabra.203

ORIGINAL RESEARCH REPORT

Near and Far Transfer in Cognitive Training: A Second-Order Meta-Analysis

Giovanni Sala[*], N. Deniz Aksayli[†], K. Semir Tatlidil[‡], Tomoko Tatsumi[§], Yasuyuki Gondo[*] and Fernand Gobet[‖]

Theory building in science requires replication and integration of findings regarding a particular research question. Second-order meta-analysis (i.e., a meta-analysis of meta-analyses) offers a powerful tool for achieving this aim, and we use this technique to illuminate the controversial field of cognitive training. Recent replication attempts and large meta-analytic investigations have shown that the benefits of cognitive-training programs hardly go beyond the trained task and similar tasks. However, it is yet to be established whether the effects differ across cognitive-training programs and populations (children, adults, and older adults). We addressed this issue by using second-order meta-analysis. In Models 1 (k = 99) and 2 (k = 119), we investigated the impact of working-memory training on near-transfer (i.e., memory) and far-transfer (e.g., reasoning, speed, and language) measures, respectively, and whether it is mediated by the type of population. Model 3 (k = 233) extended Model 2 by adding six meta-analyses assessing the far-transfer effects of other cognitive-training programs (video-games, music, chess, and exergames). Model 1 showed that working-memory training does induce near transfer, and that the size of this effect is moderated by the type of population. By contrast, Models 2 and 3 highlighted that far-transfer effects are small or null. Crucially, when placebo effects and publication bias were controlled for, the overall effect size and true variance equaled zero. That is, no impact on far-transfer measures was observed regardless of the type of population and cognitive-training program. The lack of generalization of skills acquired by training is thus an invariant of human cognition.

Keywords: cognitive intervention; cognitive training; meta-analysis; second-order meta-analysis; transfer

Theory building in science requires findings regarding a particular research question being consistently replicated across independent labs. However, the recent reproducibility crisis (Open Science Collaboration, 2015) has shown that successful replications are rarer than hoped for in psychology and other related fields (e.g., neuroscience). This state of affairs has led to a significant loss of confidence in psychological research among public opinion and policymakers.

This issue has its roots in the low statistical power of most experiments in the behavioral sciences and the higher probability of significant findings to be reported than non-significant ones. The importance of statistically significant (i.e., $p < .050$) results for scientists is well-known. Reporting novel positive findings of the effectiveness of a treatment or previously undiscovered relationships between variables increases the probability of a paper to be accepted in high-impact journals. The same can be said for accessing public and private funding. In brief, enticing positive findings fuel scientists' career.

Crucially, the probability to find a statistically significant effect is a function of statistical power (1–β), which is, in turn, a function of sample size and magnitude of the true (i.e., real and unbiased) effect size. For instance, investigations including only a few dozens of participants are not very likely to find a significant result, unless the true effect size is quite large (e.g., $d > 0.80$ and $r > .50$). However, the probability of a true effect size to be large is relatively small (Button et al., 2013). Thus, small-sample studies reporting significant effects have a good chance to be false positives occurring by accident or stemming from flexibility in data analysis in the deliberate attempt to lower p-values (Simmons, Nelson, & Simonsohn, 2011). Consequently, most replication attempts are bound to find effect sizes smaller than the significant ones estimated by previous underpowered studies (Ioannidis, 2008).

* Osaka University, Graduate School of Human Sciences, Suita, JP

† University of Nottingham, School of Computer Science, Nottingham, UK

‡ Brown University, Department of Cognitive, Linguistic and Psychological Sciences, Providence: RI, US

§ Kobe University, Graduate School of Intercultural Studies, Kobe, JP

‖ University of Liverpool, Department of Psychological Sciences, Liverpool, UK

Corresponding author: Giovanni Sala (sala@hus.osaka-u.ac.jp)

In order to overcome the problem of low statistical power and reliably estimate the size of effects, researchers have extensively employed meta-analysis. As argued by Schmidt and Oh (2013, 2016), integrating the findings in a particular area via meta-analysis is the most effective way for evaluating whether the existing replication studies corroborate or refute the original findings. In fact, one of the major advantages of meta-analysis is reducing sampling error by merging effect sizes from different sample sizes. That allows researchers to produce more precise measures of an effect than the single primary study. Furthermore, the asymmetry in the distribution of the effect sizes due to the systematic suppression of non-significant effects and *p*-hacking can be detected and corrected by publication bias analysis. Finally, meta-analysis provides measures of true (i.e., not due to random error) between-study heterogeneity necessary to assess the degree of consistency between studies in the same field. The role of covariates in accounting for true heterogeneity is evaluated via meta-regression. In brief, meta-analysis offers the best way to build reliable cumulative knowledge, because it makes it possible, among other things, to summarize a large number of studies, identify and correct publication biases, and measure the impact of study methodological features on the effect sizes that are observed.

There are cases where it would be advantageous for theory development to be able to combine not only primary studies, but also whole meta-analyses addressing a similar question. This would make it possible to examine whether different conclusions apply to different domains, countries, ages, populations, etc., or whether the same conclusion applies to all the individual meta-analyses. The technique of second-order meta-analysis (or meta-meta-analysis; Schmidt & Oh, 2013) has been developed with this goal in mind: to integrate the results of first-order meta-analyses.

Schmidt and Oh's (2013) Second-Order Meta-Analysis

Meta-meta-analytical techniques have a long history in the behavioral sciences. Examples include reviews of meta-analyses (e.g., Hattie, 2009) and meta-analyses where the unit of analysis is the overall effect sizes of other meta-analyses (e.g., Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011). However, none of these techniques has been considered fully satisfactory, especially in estimating the amount of between-meta-analysis true variance (Cooper & Koenka, 2012). Schmidt and Oh's (2013) second-order meta-analysis has been developed to address this issue.

Schmidt and Oh's (2013) second-order meta-analysis is defined as "a meta-analysis of a number of statistically independent and methodologically comparable first order meta-analyses examining ostensibly the same relationship in different contexts" (Schmidt & Oh, 2013, p. 204). As seen, conventional (i.e., first-order) meta-analysis increases the reliability of effect-size estimation by reducing sampling error. However, sampling error can never be ruled out entirely because the number of included samples is always less than infinite. Hunter and Schmidt (2004) define this residual sampling error as *second-order sampling error*.

Second-order meta-analysis aims to estimate to what extent second-order sampling error accounts for the difference across overall meta-analytic means in a set of first-order meta-analyses regarding a particular topic. First, first-order meta-analytic means ($\bar{g}_i$) are used to calculate a weighted grand mean ($\bar{\bar{g}}$). Then, the proportion of the between-meta-analysis variance explained by second-order sampling error is calculated and used to produce more accurate estimates in first-order meta-analyses. If second-order sampling error accounts only for a portion of the variance (i.e., true variance is not null; $\sigma^2 > 0$), then one must assume different mechanisms for at least some of the results obtained in the individual meta-analyses. In this case, corrected first-order meta-analytic means are closer, but not necessarily identical, to the grand mean than the uncorrected means. By contrast, if second-order sampling error explains all the observed true variance, then one is entitled to conclude that the same mechanism is likely to occur in the populations studied in the different meta-analyses. Consequently, all first-order meta-analytic means are corrected into the same grand mean because no true variance is observed across first-order meta-analytic means. This is an important conclusion, because it means that the differences between meta-analyses are just apparent and thus a broad but parsimonious conclusion can be reached that summarizes a large number of individual studies. Of course, which one of the two conclusions is correct is an empirical question that can be settled only by carrying out the relevant second-order meta-analysis. Somewhat surprisingly, Schmidt and Oh's (2013) second-order meta-analysis has been neglected in psychology and, more generally, the behavioral sciences in spite of these obvious advantages.

This paper aims to use second-order meta-analysis to address a major theoretical and practical question facing psychology and education: the effectiveness of cognitive training in inducing transfer effects. Do methods claimed to improve overall cognition and lead to educational benefits (e.g., working memory training, music training, or chess playing) really work? In particular, is it possible to improve general cognitive abilities (i.e., far transfer) as opposed to cognitive abilities linked to specific tasks or sets of tasks that share similarities (i.e., near transfer)? In this article, we show that second order meta-analysis provides a surprisingly clear answer to this question across a wide range of cognitive-training methods.

Transfer and Cognitive Training

Transfer of skills is the generalization of skills acquired by training across different domains. Transfer is a central issue in cognitive psychology because it is a manifestation of how humans acquire and process information. It is customary to distinguish between near and far transfer (Barnett & Ceci, 2002): while the former refers to the generalization of skills across similar domains, the latter indicates the transfer of skills across domains that are not, or very weakly, related to each other. Thus, the distinction between near and far transfer relies on the overlap between the source and target domains. In other words, the definition of the type of transfer is directly related to

the extent to which the domains share common features. The more the shared features, the nearer the transfer. Importantly, such features include both perceptual and conceptual information (Singley & Anderson, 1989).

According to the common elements theory (Thorndike & Woodworth, 1901), the likelihood of transfer to take place is directly related to the degree to which the source domain and the target domain share common features. That means that while near transfer is predicted to occur often, far transfer is supposed to be rare. Substantial research into learning, skill acquisition, and expertise has corroborated the theory (Detterman, 1993; Donovan, Bransford, & Pellegrino, 1999; Ritchie, Bates, & Deary, 2015; Sala & Gobet, 2017a). Relying on common sense seems to lead to the same conclusion. For instance, it is reasonable that learning analytic geometry facilitates the acquisition of knowledge in calculus because there is some overlap between the two fields. Conversely, there is no clear reason why learning Latin sentence structure should be of any use for learning calculus (or vice versa).

As seen, it is unanimously acknowledged that near transfer is much more common than far transfer. Nonetheless, far transfer is undoubtedly a much more interesting phenomenon for researchers, policymakers, and practitioners. To begin with, most (if not all) the theories and cognitive architectures of memory and skill acquisition make predictions – implicitly or explicitly – about the possible occurrence of far transfer (e.g., Chase & Simon, 1973; Gobet, 2016; Gobet & Simon, 1996; Singley & Anderson, 1989; Taatgen, 2013). The presence or absence of far transfer is thus a valuable litmus test for theories of human cognition. Furthermore, knowing whether and under what conditions far transfer occurs would represent a breakthrough in education and training in general. Skill acquisition is a costly endeavor and acquiring expertise in more than one specific field is a rare achievement (Ericsson & Charness, 1994; Gobet, 2016). Knowing how to generalize skills acquired in a particular domain across many different domains would help trainees to develop a broad set of skills in many areas more efficiently. Thus, understanding the mechanism of transfer is a major challenge in cognitive science with profound theoretical and societal implications.

Researchers are yet to reach an agreement about the actual possibility of obtaining far transfer of skills. Some authors have suggested, directly or indirectly, that the lack of far transfer is a fundamental characteristic in human cognition (e.g., Chase & Ericsson, 1982; Detterman, 1993; Sala & Gobet, 2017a; Simons et al., 2016). According to them, domain-specific skills acquired by training exert an impact on the relevant domain but hardly generalize to other domains. Moreover, even transfer of skills from one particular field of expertise to one of its sub-domains appears to lead to significant decrease in performance (e.g., Bilalić, McLeod, & Gobet, 2009; Rikers, Schmidt, & Boshuizen, 2002). This line of research is not inconsistent with the fact that some people manage to excel in more than one domain. However, it does not offer an explanation based on transfer. Rather, people with superior cognitive ability are more likely to excel in several domains because they acquire knowledge and process information better and faster than the general population (e.g., Burgoyne et al., 2016; Campitelli & Gobet, 2011; Chassy & Gobet, 2010; Detterman, 2014; Schmidt, 2017).

Other scholars are more optimistic and have suggested that it is possible to elicit far transfer. To date, the most influential and systematic attempt to obtain far transfer of skills is represented by cognitive training (for a review, see Strobach & Karbach, 2016). The cognitive-training program of research assumes that general cognitive ability or, at least, some core cognitive mechanisms (e.g., working memory, inhibition, and processing speed) can be enhanced by engaging in cognitively demanding exercises. While some of these activities, such as working-memory-training and brain-training programs, have been purposely designed to boost cognitive function, other training programs implement mentally challenging activities such as music, video games, and chess (for reviews, see Sala & Gobet, 2017a; Simons et al., 2016; Strobach & Karbach, 2016).

The basic idea underlying cognitive-training programs is that the enhancement of domain-general cognitive mechanisms is a by-product of training in domain-specific activities (Taatgen, 2016). Consistent with the research on skill acquisition and expertise, engaging in cognitive-training programs has been found to improve participants' performance on the trained task and related tasks (e.g., Simons et al., 2016). However, these activities are also believed to foster overall cognitive function or, at least, some domain-general cognitive skills (e.g., memory and processing speed). Once improved, enhanced domain-general cognitive skills are supposed to boost professional and academic domain-specific capabilities that depend on them. Neural plasticity is believed to be the mediator of this process (Karbach & Schubert, 2013).

Inconsistent Findings, Statistical Power, and Current Meta-Analytic Evidence

Hundreds of experimental studies have examined the impact of cognitive-training programs on people's ability to perform cognitive and academic tasks. While some authors have reported significant improvements in cognitive ability induced by cognitive-training regimens (e.g., Green & Bavelier, 2003; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Schellenberg, 2004), replication attempts have not always been successful, especially with regard to far-transfer effects. In fact, the large effect sizes reported by early studies have been subsequently found to be either significantly smaller or null (e.g., Redick et al., 2013; Rickard, Bambrick, & Gill, 2012; van Ravenzwaaij, Boekel, Forstmann, Ratcliff, & Wagenmakers, 2014). Such inconsistency in the findings of cognitive-training research mirrors the general pattern of outcomes in the crisis of reproducibility. Cognitive-training interventions usually count no more than 40 to 60 participants in total. For instance, the median of participants per group in our sample is about 20 (see Supplemental material available online). With 20 participants per group and assuming a statistical power of .80, then the true effect size should be about $d = 0.90$ or $d = 0.80$ to be statistically significant

($p < .050$, two-tailed or one-tailed, respectively). Such an effect is often unrealistic, especially if it represents the presumed enhancement of cognitive function. In fact, it is hard to believe that a short-term training regimen can increase, for example, fluid intelligence, working memory capacity, or attentional control by nearly one standard deviation. Assuming a more realistic effect size (e.g., $d = 0.30$), the number of participants per group should be around 175 to have a statistical power of .80. Nearly no intervention in the field has included so many participants. These considerations lead us to think that the majority of the studies reporting statistically significant effects of a cognitive-training program on people's cognitive function are false positives.

The fact that most of the studies in the field of cognitive training are underpowered means that the reported statistically significant effects are inflated. However, this does not necessarily imply that the true effect equals zero. The meta-analytic evidence gathered so far suggests that, overall, cognitive-training programs exert small to medium near-transfer effects and small to null far-transfer effects (e.g., Melby-Lervåg, Redick, & Hulme, 2016; Sala & Gobet, 2017a). However, it is yet to be understood whether some specific cognitive-training regimens can induce transfer effects better than others (e.g., Cohen, Green, & Bavelier, 2008). Also, it is unclear whether the type of population (e.g., children, young adults, and older adults) that undergoes a particular cognitive-training regimen moderates the degree to which transfer occurs (e.g., Karbach, Könen, & Spengler, 2017). In other words, it is yet to be clarified whether there is genuine between-regimen and between-population variability regarding both near- and far-transfer effects. We here employ second-order meta-analysis to address these questions.

The Present Study

To the best of our knowledge, this is the first time Schmidt and Oh's (2013) second-order meta-analysis is used with standardized mean differences. Also, this is the first second-order meta-analytic investigation regarding cognitive training. We here run three main models. In Models 1 and 2, we examine whether the effects of working-memory (hereafter WM) training on performance in cognitive tasks are mediated by the type of population. Model 1 analyzes the effects of WM training on memory tasks (i.e., near-transfer effects). Model 2 focuses on far-transfer tasks (e.g., fluid reasoning, language, and cognitive control). To date, WM training is the most studied and probably most influential cognitive-training program. Moreover, the average quality of WM-training studies is excellent. In fact, the primary studies often include pre-post-test assessments, active control groups, and measures of both near and far transfer. For these reasons, WM training is the most suitable cognitive-training program for testing the extent to which trained skills transfer across different cognitive tasks and whether there are differences between populations. Finally, Model 3 is an extension of Model 2. In Model 3, we included another six meta-analyses of other cognitive-training programs: action video-game and non-action video-game training, music instruction, chess instruction, and exergames (i.e., cognitive-training games combined with physical activities). (Note that extending Model 1 in that way was not possible because primary studies in these additional six meta-analyses rarely collected near-transfer measures.)

General Method

Second-Order Meta-Analytic Procedure and Omnibus Meta-Analysis

Second-order meta-analysis requires the studies in the first-order meta-analysis to be statistically independent (Schmidt & Oh, 2013). In other words, no study (or sample) must be included in more than one first-order meta-analysis. Since we divided the first-order meta-analyses by type of population and type of cognitive-training program, this assumption was met in all the second-order meta-analytic models.

None of the included first-order meta-analyses corrected the effect sizes for measurement error. Thus, we implemented the equations for second-order meta-analysis of bare-bones meta-analyses (Schmidt & Oh, 2013; pp. 207–209). The data related to first-order meta-analysis (overall effect size, standard error, and amount of true heterogeneity) were calculated by the metafor R package (Viechtbauer, 2010). We used random-effect models for all the first-order meta-analyses. The amount of total between-study heterogeneity (τ^2) was calculated with the REML estimator (Veroniki et al., 2016). We ran two different sets of second-order meta-analytic models. The first one included all the uncorrected (naïve) meta-analytic means. The second one included a corrected effect size from each first-order meta-analysis estimated by the publication bias analysis. This whole procedure was carried out twice: we ran a second-order meta-analysis with first-order meta-analyses including (a) all the primary studies and (b) only the comparisons between experimental and active control groups to check for placebo effects. Therefore, four second-order meta-analyses were carried out for each of the three models.

Finally, we also ran an omnibus meta-analysis—that is, a first-order meta-analysis including all the primary studies of all the first-order meta-analyses (Borenstein et al., 2009)—for each of the three models. This additional analysis was performed as a sensitivity analysis to control for the overall effect sizes, publication-bias estimates, and true between-study heterogeneity calculated in the first-order meta-analyses. It is worth noting that omnibus meta-analysis is not a substitute for second-order meta-analysis (Schmidt & Oh, 2013). In fact, while estimating an overall effect size similar to the grand mean in second-order meta-analysis, omnibus meta-analysis does not provide an estimation of second-order sampling error. As seen, this information is necessary to calculate both the corrected first-order meta-analytic means and the amount of between-meta-analysis true variance (i.e., not due to second-order sampling error). Furthermore, second-order meta-analysis allows us to integrate publication-bias corrected first-order meta-analytic estimates in a single model. By contrast, running a single publication-bias analysis in an omnibus meta-analysis is not as accurate

because it assumes that publication bias is the same across sub-groups. This assumption is, in fact, often violated.

Inclusion Criteria for The First-Order Meta-Analyses

As seen, second-order meta-analysis assumes the first-order meta-analyses to be statistically independent, that is, no study can be included in more than one meta-analysis. To meet this assumption, we had to choose only one meta-analysis per training and population. We thus established five inclusion criteria the first-order meta-analyses had to meet:

1. The meta-analysis included studies employing proper cognitive-training programs, that is, mentally challenging training regimens that aimed to train cognitive mechanisms or skills on objective (i.e., not reported or self-reported) behavioral measures of cognitive skills or academic achievement. Also, we did not include any meta-analysis on the effects of cognitive-training programs on neural patterns;
2. The meta-analysis reported all the raw data necessary for reanalysis;
3. The meta-analysis had to be recent (published/carried in 2015 or later) to include most of the relevant studies in the field of cognitive training;
4. The meta-analyses reported both near- and far-transfer effects (for WM training meta-analyses);
5. The meta-analysis did not report a mix of different types of training (e.g., video-game training and WM training).

Among the eligible meta-analyses, we selected those meta-analyses that were more comprehensive (i.e., included more studies), and more technically sound and comparable in terms of calculation of the effect sizes and inclusion criteria of the primary studies. The details of the selected meta-analyses are reported in the introductory sections of Models 1, 2, and 3. The total number of effect sizes, clusters, and participants included were $n = 1{,}555$, $k = 332$, and $N = 21{,}968$, respectively. The details of the search and the list of the meta-analyses that met the inclusion criteria and excluded are reported in Appendix A.[1]

Inclusion Criteria for The Effect Sizes

We established four inclusion criteria to guarantee a minimum standard of design quality in the primary studies:

1. The primary study included at least one control group;
2. The primary study included a pre-test to assess baseline effects;
3. The experimental samples were not self-selected;
4. The transfer effects were measured by a cognitive/academic task. Self-reported measures were excluded.

Some studies and effect sizes from the published first-order meta-analyses were excluded because of these criteria. For all the details, see Appendix B.

Effect Size Calculation

The effect size used in all the meta-analyses was the corrected standardized mean difference, that is, Hedges' *g* (Hedges & Olkin, 1985). The effect size represented the amelioration of the experimental groups over the controls immediately after the end of the training. Due to their dearth, no follow-up effects were included. For most of the meta-analyses, we used the original effect sizes. In two cases (see Model 3), we recalculated all the effect sizes to uniformize the standard across the meta-analyses. The formula for the effect size was:

$$g = \frac{\left(M_{e\_post} - M_{e\_pre}\right) - \left(M_{c\_post} - M_{c\_pre}\right)}{SD_{pooled_{pre}}} \times \left(1 - \frac{3}{\left(4 \times N\right) - 9}\right)$$

that is, the post-pre between-group mean difference standardized by the pooled pre-test standard deviations and corrected for upward bias.

Active Controls

Active control groups are necessary to control for possible placebo effects. For this reason, we also ran models including only experimental groups matched with active controls. According to commonly accepted guidelines (e.g., Boot, Simons, Stothart, & Stutts, 2013; Simons et al., 2016), we considered a control group as "active" only if it consisted of an engaging and cognitively demanding activity (e.g., non-adaptive training, visual-search training, etc.).[2] Alternative tasks with negligible cognitive demand (e.g., watching videos and filling in questionnaires) were labeled as "non-active." Two coders independently judged whether the primary studies implemented such an active control group.

Correction for Statistical Dependence

Primary studies often report more than one measure of cognitive ability. Measures from the same samples are, by definition, statistically dependent. Modeling these effect sizes as statistically independent does not introduce any systematic bias in the estimation of meta-analytic means (Schmidt & Hunter, 2015; Tracz, Elmore, & Pohlmann, 1992). Nevertheless, not correcting for statistical dependency leads to an underestimation of sampling error variances and an overestimation of the amount of between-study true heterogeneity. Conversely, just merging the effects without applying any additional correction overestimates sampling error variances and underestimates the amount of between-study true heterogeneity (Schmidt & Hunter, 2015). Given that a major goal of second-order meta-analysis is to estimate the amount of between-meta-analysis variability explained by second-order sampling error, this bias must be corrected or at least reduced. To address the problem, we use Cheung and Chan's (2014) samplewise-adjusted-individual correction. This technique has been designed to estimate an adjusted variance based on (a) the number of the dependent effect sizes and (b) inter-effect-size correlation. For the details, see Appendix C.

Finally, we performed a parallel set of meta-analytic models using the Robust Variance Estimation method

(RVE; Hedges et al., 2010). The results are reported in Appendix D (Tables S1, S2, and S3). These additional models were implemented to examine whether the estimates of the meta-analytic models were sensitive to the technique employed to model dependent effect sizes. Furthermore, along with a between-cluster variance estimator (τ^2), RVE provides a within-cluster variance estimator (ω^2). This latter estimator thus offers valuable information about the possible differences across within-study effect sizes.

Publication Bias Analysis

Naïve (i.e., uncorrected) meta-analytic means are often less reliable than the publication-bias corrected estimates (Schmidt & Oh, 2016; Stanley, 2017). We therefore ran a set of publication bias analyses for all the first-order meta-analyses and built a parallel set of second-order meta-analyses using publication-bias corrected estimates. It is usually recommendable to employ multiple publication-bias detection techniques to triangulate the most likely true (i.e., unbiased) overall effect size (e.g., Kepes & McDaniel, 2015). First, we used the "precision effect test" (PET) and the "precision-effect estimate with standard error" (PEESE; Stanley & Doucouliagos, 2014). The PET estimator is the intercept of a weighted (by precision) linear regression where the dependent variable is the effect size and the independent variable is its standard error. The PEESE estimator is obtained by replacing the standard error with the standard error squared (i.e., variance) as the independent variable. If PET suggests the presence of a non-zero effect (at the 10% of significant level, i.e., $p < .100$, one-tailed; Stanley, 2017), the PEESE estimator is employed. Second, when both the PET and PEESE produced inaccurate estimates (i.e., very high standard error) or excessively negative estimates,[3] we used the trim-and-fill analysis with all the three estimators (L0, R0, and Q0) described in Duval and Tweedie (2000). Since the trim-and-fill has been found to underestimate the amount of publication bias, especially when the null is true, (Carter, Schönbrodt, Gervais, & Hilgard, 2017; Moreno et al., 2009; Simonsohn, Nelson, & Simmons, 2014), the PET-PEESE estimates were preferred when they did not suffer from the abovementioned flaws (i.e., high standard errors and large negative values). As a general rule, when the PET test did not show evidence of non-zero effect, the estimate that was the closest to zero was picked up for the second-order meta-analysis. (It is worth noting that selection methods such as selection models [e.g., McShane, Böckenholt, & Hansen, 2016; Vevea & Woods, 2005] and *p*-curve [Simonsohn et al., 2014] are not suitable when statistically dependent effects have been merged because the *p*-value distribution significantly differs from the original one.)

Heterogeneity

Within each meta-analysis, the between-study true heterogeneity was assessed by the τ^2 statistic. High amount of true heterogeneity artificially increases the relative weight assigned to small studies, which in presence of publication bias can result in an overestimation of the overall effect size (Stanley & Doucouliagos, 2017). Moreover, high true heterogeneity can bias publication-bias corrected estimates (Schmidt & Hunter, 2015; Stanley, 2017). Therefore, when the amount of true heterogeneity was statistically significant ($p < .100$),[4] we ran Viechtbauer and Cheung's (2010) influential case analysis. The detected influential cases were excluded to reduce the amount of true heterogeneity and enhance the reliability of the corrected effect sizes. The list of the influential studies is reported in Appendix E.

Model 1: Working Memory Training (Near Transfer)

Description of The First-Order Meta-Analyses

We selected four meta-analyses including typically developing (TD) children (Sala & Gobet, 2017b), children with learning disabilities (LD; a subsample of Melby-Lervåg et al., 2016), healthy adults (a subsample of Melby-Lervåg et al., 2016), and healthy older adults and older adults with mild cognitive impairment (Sala, Aksayli, Tatlidil, Gondo, & Gobet, 2018a). No study or effect size reported in the original meta-analyses was excluded. This section examines the impact of WM training on the participants' performance in memory tasks (i.e., near transfer).

Results

TD children

The overall meta-analytic mean was $\bar{g} = 0.45$, $SE = 0.07$, 95% CI [0.31; 0.58], $k = 16$, $p < .001$. The test of heterogeneity was not significant, $Q = 15.63$, $\tau^2 = 0.004$, $p = .407$. PET and PEESE estimators were $\bar{g} = 0.59$, $SE = 0.20$, $p = .010$[5] and $\bar{g} = 0.50$, $SE = 0.11$, $p < .001$, respectively. Since the PET test showed evidence of a real effect ($p < .100$, one-tailed), the PEESE estimator was used in the relevant second-order meta-analysis.

With regard to the type of control group (active vs. non-active), there was 100% inter-rater agreement. The classification of the type of control group was the same as in Sala and Gobet (2017b). When considering only studies implementing an active control group, the overall meta-analytic mean was $\bar{g} = 0.43$, $SE = 0.09$, 95% CI [0.26; 0.61], $k = 11$, $p < .001$. The test of heterogeneity was not significant, $Q = 6.59$, $\tau^2 = 0.000$, $p = .764$. PET and PEESE estimators were $\bar{g} = 0.54$, $SE = 0.26$, $p = .073$ and $\bar{g} = 0.49$, $SE = 0.13$, $p = .004$, respectively. Since the PET test showed evidence of a real effect ($p < .100$, one-tailed), the PEESE estimator was used in the relevant second-order meta-analysis.

LD children

The overall meta-analytic mean was $\bar{g} = 0.55$, $SE = 0.10$, 95% CI [0.35; 0.76], $k = 17$, $p < .001$. The test of heterogeneity was significant, $Q = 54.65$, $\tau^2 = 0.109$, $p < .001$. PET and PEESE estimators were $\bar{g} = 0.25$, $SE = 0.15$, $p = .110$ and $\bar{g} = 0.40$, $SE = 0.09$, $p < .001$, respectively. Since the model showed a significant amount of true heterogeneity, we ran an influential case analysis. The influential case analysis found one influential study (for the details, see R codes and Appendix E). After removing this study, the overall meta-analytic mean was $\bar{g} = 0.46$, $SE = 0.08$, 95% CI [0.30; 0.62], $k = 16$, $p < .001$. The test of heterogeneity was still significant but the amount of true heterogeneity was much smaller, $Q = 34.15$, $\tau^2 = 0.046$, $p = .003$. PET and PEESE

estimators were $\bar{g}$ = 0.25, *SE* = 0.12, *p* = .057, and $\bar{g}$ = 0.37, *SE* = 0.07, *p* < .001, respectively. The PEESE estimator was used in the relevant second-order meta-analysis.

Regarding the type of control group (active vs. non-active), there was 100% inter-rater agreement. The classification of the type of control group was the same as in Melby-Lervåg et al. (2016). When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g}$ = 0.48, *SE* = 0.11, 95% CI [0.27; 0.69], *k* = 12, *p* < .001. The test of heterogeneity was significant, *Q* = 25.52, τ^2 = 0.080, *p* = .008. PET and PEESE estimators were $\bar{g}$ = –0.19, *SE* = 0.20, *p* = .374 and $\bar{g}$ = 0.08, *SE* = 0.11, *p* = .484, respectively. To reduce the amount of true heterogeneity, we ran an influential case analysis. The influential case analysis found two influential studies. After removing the studies, the overall meta-analytic mean was $\bar{g}$ = 0.32, *SE* = 0.06, 95% CI [0.19; 0.44], *k* = 10, *p* < .001. The test of heterogeneity was not significant, *Q* = 8.95, τ^2 = 0.000, *p* = .441. PET and PEESE estimators were $\bar{g}$ = 0.01, *SE* = 0.17, *p* = .943, and $\bar{g}$ = 0.15, *SE* = 0.09, *p* = .124, respectively. With all the three estimators (L0, R0, and Q0), the analysis filled three studies left of the mean. The overall meta-analytic mean was $\bar{g}$ = 0.26, *SE* = 0.06, 95% CI [0.14; 0.38], *k* = 13, *p* < .001. Considering that no included study reported a null or negative effect, the PET estimator was considered unreliable. The PEESE estimator and the trim-and-fill estimator were more realistic. The trim-and-fill estimator ($\bar{g}$ = 0.26, *SE* = 0.06) was preferred because the effect size of the studies with the largest sample size was about *g* = 0.25.

Adults

The overall meta-analytic mean was $\bar{g}$ = 0.20, *SE* = 0.04, 95% CI [0.12; 0.28], *k* = 31, *p* < .001. The test of heterogeneity was not significant, *Q* = 38.85, τ^2 = 0.012, *p* = .129. PET and PEESE estimators were $\bar{g}$ = 0.15, *SE* = 0.05, *p* = .009 and $\bar{g}$ = 0.17, *SE* = 0.04, *p* < .001, respectively. The PEESE estimator was selected for the second-order meta-analysis.

Concerning the type of control group (active vs. non-active), there was 93% inter-rater agreement. The two coders solved any discrepancy by talk. The classification of the type of control group was slightly different from the one employed by Melby-Lervåg et al. (2016). Three studies whose control group was considered as active in the original meta-analysis were labeled as "non-active." When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g}$ = 0.15, *SE* = 0.04, 95% CI [0.06; 0.23], *k* = 20, *p* = .001. The test of heterogeneity was not significant, *Q* = 19.91, τ^2 = 0.007, *p* = .400. PET and PEESE estimators were $\bar{g}$ = 0.15, *SE* = 0.06, *p* = .017 and $\bar{g}$ = 0.14, *SE* = 0.04, *p* = .002, respectively. The PEESE estimator was selected for the second-order meta-analysis.

Older adults

The overall meta-analytic mean was $\bar{g}$ = 0.29, *SE* = 0.04, 95% CI [0.21; 0.38], *k* = 35, *p* < .001. The test of heterogeneity was significant, *Q* = 52.27, τ^2 = 0.021, *p* = .023. PET and PEESE estimators were $\bar{g}$ = 0.07, *SE* = 0.08, *p* = .374 and $\bar{g}$ = 0.18, *SE* = 0.05, *p* < .001, respectively. The influential case analysis found one influential study. After removing this study, the overall meta-analytic mean was $\bar{g}$ = 0.26, *SE* = 0.04, 95% CI [0.18; 0.33], *k* = 34, *p* < .001. The test of heterogeneity was not significant, *Q* = 40.07, τ^2 = 0.009, *p* = .185. PET and PEESE estimators were $\bar{g}$ = 0.05, *SE* = 0.07, *p* = .487, and $\bar{g}$ = 0.16, *SE* = 0.04, *p* < .001, respectively. The PEESE estimator was preferred over the PET estimator because it was more precise (*SE* = 0.04 vs. *SE* = 0.07). Also, given that the average effect size of the studies with the largest sample size was about *g* = 0.15 and only a few studies reported an effect size close to zero, the adjusted effect size suggested by the PET test seemed too small.

Concerning the type of control group (active vs. non-active), there was 100% inter-rater agreement. When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g}$ = 0.23, *SE* = 0.05, 95% CI [0.13; 0.34], *k* = 19, *p* < .001. The test of heterogeneity was not significant, *Q* = 20.58, τ^2 = 0.011, *p* = .301. PET and PEESE estimators were $\bar{g}$ = 0.06, *SE* = 0.15, *p* = .699 and $\bar{g}$ = 0.16, *SE* = 0.07, *p* = .045, respectively. The PEESE estimator was preferred over the PET estimator because it was more precise (*SE* = 0.07 vs. *SE* = 0.15).

Second-order meta-analysis

Tables 1–4 summarize the results of the second-order meta-analysis of near-transfer effects of WM training. The differences between the first-order meta-analytic means ($\bar{g}_i$) were mostly due to true variance (σ^2) in three out of the four second-order meta-analytic models (**Tables 1, 2** and **3**). The adjusted overall effect sizes

Table 1: Second-order meta-analysis with the uncorrected (naïve) overall effect sizes in Model 1.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
TD Children	16	0.45	0.076	0.004							0.41
LD Children	17	0.55	0.180	0.109							0.49
Adults	31	0.20	0.052	0.012							0.23
Older Adults	35	0.29	0.066	0.021							0.29
					0.30	0.00281	0.01116	0.00836	.25	.75	

Note: (1) Number of samples; (2) First-order overall effect size; (3) Variance of the observed *g*s; (4) Amount of true heterogeneity; (5) Second-order grand mean; (6) Second-order sampling error variance; (7) Observed between-first-order-meta-analysis variance; (8) True between-first-order-meta-analysis variance; (9) Proportion of the variance explained by second-order sampling error; (10) Reliability of the first-order overall effect size; (11) Adjusted first-order overall effect size.

Table 2: Second-order meta-analysis with the corrected overall effect sizes in Model 1.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
TD Children	16	0.50	0.188	0.004							0.43
LD Children	16	0.37	0.081	0.046							0.32
Adults	31	0.17	0.046	0.012							0.18
Older Adults	34	0.16	0.053	0.009							0.17
					0.21	0.00249	0.00928	0.00679	.27	.73	

Note: See Note to Table 1 for abbreviations.

Table 3: Second-order meta-analysis with the uncorrected (naïve) overall effect sizes (only active control groups) in Model 1.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
TD Children	11	0.43	0.087	0.000							0.38
LD Children	12	0.48	0.142	0.080							0.41
Adults	20	0.15	0.040	0.007							0.17
Older Adults	19	0.23	0.055	0.011							0.23
					0.24	0.00377	0.01310	0.00933	.29	.71	

Note: See Note to Table 1 for abbreviations.

Table 4: Second-order meta-analysis with the corrected overall effect sizes (only active control groups) in Model 1.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
TD Children	11	0.49	0.180	0.000							0.35
LD Children	13	0.26	0.047	0.000							0.23
Adults	20	0.14	0.030	0.007							0.17
Older Adults	19	0.16	0.103	0.011							0.17
					0.19	0.00337	0.00721	0.00385	.47	.53	

Note: See Note to Table 1 for abbreviation.

(column 11) were thus close to the original estimates (column 2). When publication bias and possible placebo effects were controlled for, second-order sampling error explained nearly half of the observed between-meta-analysis variance (**Table 4**).

Omnibus meta-analysis

The overall meta-analytic mean was $\bar{g}$ = 0.32, *SE* = 0.03, 95% CI [0.26; 0.38], k = 99, p < .001. The test of heterogeneity was significant, Q = 189.08, τ^2 = 0.036, p < .001. PET and PEESE estimators were $\bar{g}$ = 0.14, *SE* = 0.05, p = .003 and $\bar{g}$ = 0.22, *SE* = 0.03, p < .001, respectively. The influential case analysis found three influential studies. After removing these studies, the overall meta-analytic mean was $\bar{g}$ = 0.28, *SE* = 0.03, 95% CI [0.23; 0.33], k = 96, p < .001. The test of heterogeneity was still significant, but the amount of true heterogeneity was lower, Q = 133.38, τ^2 = 0.015, p = .006. PET and PEESE estimators were $\bar{g}$ = 0.11, *SE* = 0.04, p = .005, and $\bar{g}$ = 0.19, *SE* = 0.02, p < .001, respectively. Both the naïve and corrected overall effect sizes ($\bar{g}$ = 0.32 and $\bar{g}$ = 0.19, respectively) were thus very close to the grand means estimated in the second-order meta-analysis ($\bar{\bar{g}}$ = 0.30 and $\bar{\bar{g}}$ = 0.21 for the uncorrected and corrected models, respectively).

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g}$ = 0.25, *SE* = 0.03, 95% CI [0.19; 0.32], k = 62, p < .001. The test of heterogeneity was significant, Q = 86.44, τ^2 = 0.016, p = .018. PET and PEESE estimators were $\bar{g}$ = 0.08, *SE* = 0.05, p = .096 and $\bar{g}$ = 0.16, *SE* = 0.03, p < .001, respectively. The influential case analysis found three influential studies. After removing these studies, the overall meta-analytic mean was $\bar{g}$ = 0.25, *SE* = 0.03, 95% CI [0.18; 0.32], k = 59, p < .001. The test of heterogeneity was marginally significant, Q = 73.11, τ^2 = 0.015, p = .087. PET and PEESE estimators were $\bar{g}$ = 0.03, *SE* = 0.08, p = .666,

and $\bar{g}$ = 0.16, *SE* = 0.04, $p < .001$, respectively. Again, the PET estimator appeared to be an underestimation of the true effect, especially because the PET test showed the presence of a non-zero effect when the influential cases were included in the model ($p = .096$). By contrast, the PEESE estimator was more precise and very close to the grand mean in the second-order meta-analysis ($\bar{\bar{g}}$ = 0.19). Finally, in line with the results of the second-order meta-analysis, some true heterogeneity was observed in all the models of the omnibus meta-analysis.

Discussion

The results of Model 1 show that WM training fosters performance in memory tasks in all the reviewed populations. The effect substantially remains even when only comparisons between trained groups and active controls are considered. The effect is, however, quite heterogeneous across populations. In fact, only a portion of the observed between-meta-analysis variance is due to second-order sampling error (25% to 47%).

TD children seem to benefit the most from the training program. Adults and older adults exhibit much smaller effects. Due to the highly asymmetrical distribution of the effect sizes, the effect is less clear in LD children. As seen, the most probable unbiased effect is about $\bar{g}$ = 0.25. This difference probably reflects the different learning pace in the populations. While TD children learn relatively fast, adults and older adults need greater effort to acquire new skills probably because their cognitive system is less flexible and plastic. LD children exhibit better performance than the adult populations but, as expected, are not as good as TD children.

Thus, in line with previous research, near transfer from WM training to memory tasks often (if not always) takes place. Interestingly, the size of this transfer of skills appears to be moderated by the type of population and, arguably, their particular cognitive profiles. It is reasonable that near transfer represents (i.e., exhibits the same pattern of) the general capability of acquiring new skills by practice. In fact, TD children usually learn faster than their peers suffering from some learning disability, adults, and older adults.

On a final note, it is essential to acknowledge that the participants' boosted performance on memory tasks does not necessarily represent evidence of cognitive enhancement. As observed by Shipstead, Redick, and Engle (2012), such an improvement probably denotes the participants' enhanced ability to perform a class of cognitive tasks sharing similar features (e.g., perceptual cues and solving strategies) with the trained task. Either way, regardless of whether it represents cognitive enhancement, the presence of near transfer seems unquestionable.

Model 2: Working Memory Training (Far Transfer)

Description of the First-Order Meta-Analyses

This section examines the effects of WM training on far transfer measures such as tests of fluid reasoning, cognitive control, processing speed, and language. For the details about the four first-order meta-analyses included, see Model 1.

Results

TD children

The overall meta-analytic mean was $\bar{g}$ = 0.13, *SE* = 0.05, 95% CI [0.03; 0.22], $k = 25$, $p = .010$. The test of heterogeneity was not significant, $Q = 21.75$, $\tau^2 = 0.006$, $p = .594$. PET and PEESE estimators were $\bar{g}$ = 0.08, *SE* = 0.10, $p = .441$ and $\bar{g}$ = 0.09, *SE* = 0.06, $p = .145$, respectively. The PET estimator was selected for the second-order meta-analysis.

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g}$ = 0.01, *SE* = 0.07, 95% CI [–0.12; 0.14], $k = 15$, $p = .879$. The test of heterogeneity was not significant, $Q = 7.20$, $\tau^2 = 0.000$, $p = .927$. PET and PEESE estimators were $\bar{g}$ = 0.12, *SE* = 0.15, $p = .440$ and $\bar{g}$ = 0.06, *SE* = 0.08, $p = .495$, respectively. Neither PET nor PEESE provided any evidence of a non-zero effect ($p = .440$ and $p = .495$, respectively). Also, PET estimated a very imprecise overall effect (*SE* = 0.15) suggesting that the detected publication bias was in fact a statistical artifact. We thus ran the trim-and-fill analysis to check for missing studies right of the mean. With the L0 and Q0 estimators, the analysis filled no study. With the R0 estimator, the analysis filled one study right of the mean. The overall meta-analytic mean was $\bar{g}$ = 0.02, *SE* = 0.06, 95% CI [–0.10; 0.15], $k = 16$, $p = .717$. Given that no estimate was significantly different from the null (all $ps \geq .440$), we selected the estimate that was the closest to zero ($\bar{g}$ = 0.01, *SE* = 0.07) as the corrected overall effect size to be included in the second-order meta-analysis.

LD children

The overall meta-analytic mean was $\bar{g}$ = 0.12, *SE* = 0.04, 95% CI [0.03; 0.20], $k = 18$, $p = .006$. The test of heterogeneity was not significant, $Q = 12.34$, $\tau^2 = 0.002$, $p = .779$. PET and PEESE estimators were $\bar{g}$ = 0.06, *SE* = 0.07, $p = .350$ and $\bar{g}$ = 0.10, *SE* = 0.03, $p = .014$, respectively. The PET estimator was selected for the second-order meta-analysis.

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g}$ = 0.08, *SE* = 0.06, 95% CI [–0.04; 0.19], $k = 12$, $p = .202$. The test of heterogeneity was not significant, $Q = 3.50$, $\tau^2 = 0.000$, $p = .982$. PET and PEESE estimators were $\bar{g}$ = 0.02, *SE* = 0.10, $p = .870$ and $\bar{g}$ = 0.03, *SE* = 0.06, $p = .543$, respectively. The PET estimator was selected for the second-order meta-analysis.

Adults

The overall meta-analytic mean was $\bar{g}$ = 0.12, *SE* = 0.03, 95% CI [0.06; 0.18], $k = 44$, $p < .001$. The test of heterogeneity was not significant, $Q = 39.56$, $\tau^2 = 0.003$, $p = .621$. PET and PEESE estimators were $\bar{g}$ = 0.06, *SE* = 0.05, $p = .303$ and $\bar{g}$ = 0.08, *SE* = 0.03, $p = .012$, respectively. The PET estimator was selected for the second-order meta-analysis.

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g}$ = 0.09, *SE* = 0.04, 95% CI [0.01; 0.17], $k = 27$, $p = .032$. The test of heterogeneity was not significant, $Q = 20.08$, $\tau^2 = 0.000$, $p = .788$. PET and PEESE estimators were

$\bar{g}$ = –0.01, *SE* = 0.09, *p* = .949 and $\bar{g}$ = 0.04, *SE* = 0.05, *p* = .473, respectively. The PET estimator was selected for the second-order meta-analysis.

Older adults

The overall meta-analytic mean was $\bar{g}$ = 0.13, *SE* = 0.05, 95% CI [0.03; 0.23], *k* = 32, *p* = .010. The test of heterogeneity was significant, *Q* = 62.39, τ^2 = 0.035, *p* < .001. PET and PEESE estimators were $\bar{g}$ = –0.02, *SE* = 0.06, *p* = .718 and $\bar{g}$ = 0.03, *SE* = 0.05, *p* = .508, respectively. The influential case analysis found one influential study. After removing this study, the overall meta-analytic mean was $\bar{g}$ = 0.10, *SE* = 0.04, 95% CI [0.01; 0.19], *k* = 31, *p* = .029. The test of heterogeneity was still significant, but the amount of true heterogeneity was lower, *Q* = 45.24, τ^2 = 0.017, *p* = .037. PET and PEESE estimators were $\bar{g}$ = –0.04, *SE* = 0.05, *p* = .474, and $\bar{g}$ = 0.01, *SE* = 0.04, *p* = .863, respectively. Since the PET estimator was negative, we selected the PEESE estimator for the second-order meta-analysis.

When considering only studies implementing an active control group, the overall meta-analytic mean was $\bar{g}$ = –0.02, *SE* = 0.03, 95% CI [–0.09; 0.05], *k* = 16, *p* = .574. The test of heterogeneity was not significant, *Q* = 6.72, τ^2 = 0.000, *p* = .965. PET and PEESE estimators were $\bar{g}$ = 0.02, *SE* = 0.03, *p* = .473 and $\bar{g}$ = 0.01, *SE* = 0.02, *p* = .811, respectively. Since the PET test did not find any evidence of a non-zero effect, we selected the PEESE estimator for the second-order meta-analysis because it was the closest to zero.

Second-order meta-analysis

Tables 5–8 summarize the results of the second-order meta-analysis of far-transfer effects of WM training. In all the models, the differences between the first-order meta-analytic means ($\bar{g}_i$) were mostly due solely to second-order sampling error. When ruling out placebo effects and publication bias, the overall effect was nearly null ($\bar{\bar{g}}$ = 0.01; **Table 8**).

Omnibus meta-analysis

The overall meta-analytic mean was $\bar{g}$ = 0.12, *SE* = 0.02, 95% CI [0.08; 0.16], *k* = 119, *p* < .001. The test of heterogeneity was marginally significant, *Q* = 138.41, τ^2 = 0.009, *p* = .097. PET and PEESE estimators were $\bar{g}$ = 0.02, *SE* = 0.03, *p* = .468 and $\bar{g}$ = 0.07, *SE* = 0.02, *p* < .001, respectively. The influential case analysis found nine influential studies. After removing these studies, the overall meta-analytic mean was $\bar{g}$ = 0.12, *SE* = 0.02, 95% CI [0.08; 0.16], *k* = 110, *p* < .001. The test of heterogeneity was not significant, *Q* = 86.58, τ^2 = 0.000, *p* = .944. PET and PEESE estimators were $\bar{g}$ = 0.06, *SE* = 0.04, *p* = .168, and $\bar{g}$ = 0.08, *SE* = 0.02, *p* < .001, respectively.

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g}$ = 0.03, *SE* = 0.02, 95% CI [–0.02; 0.07], *k* = 70, *p* = .195. The test of heterogeneity was not significant, *Q* = 42.17, τ^2 = 0.000, *p* = .996. PET and PEESE estimators were $\bar{g}$ = 0.00, *SE* = 0.03, *p* = .890 and $\bar{g}$ = 0.01, *SE* = 0.02, *p* = .566, respectively.

Discussion

The results provided by the second-order meta-analytic models show that the actual impact of WM training on far-transfer measures is null regardless of the population. The small positive effect sizes disappear when placebo effects and publication bias are controlled for (**Table 8**). That is, the unbiased far-transfer effect exerted by WM training is practically null ($\bar{\bar{g}}$ = 0.01) and consistent (τ^2 = 0). No or low true variance (σ^2) is observed in all the models. Also, it is worth noting that the observed amount of true heterogeneity (τ^2; **Tables 5** and **6**) is entirely accounted for by the type of control group used in the primary studies (**Tables 7** and **8**). Notably, within-study variance (ω^2) was very low or null in all the first-order meta-analytic models (see Table S2 in the Supplemental materials available online). Consequently, comparing effect sizes extracted from different tests of cognitive/academic skills does not add noise to the models. In fact, unbiased far-transfer effects approach zero regardless of the test used to measure them (Melby-Lervåg et al., 2016; Sala & Gobet, 2017b). Finally, the omnibus meta-analysis confirms the findings of the second-order meta-analysis.

Model 3: Adding Other Cognitive-Training Programs

Description of the First-Order Meta-Analyses

This section is an extension of the meta-analytic investigation presented in Model 2. Along with the four first-order meta-analyses of WM training, we used six other first-order meta-analyses of far-transfer effects following cognitive training: action and non-action video-game training in adults, action

Table 5: Second-order meta-analysis with the uncorrected (naïve) overall effect sizes in Model 2.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{\bar{g}}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
TD Children	25	0.13	0.060	0.006							0.12
LD Children	18	0.12	0.032	0.002							0.12
Adults	44	0.12	0.041	0.003							0.12
Older Adults	32	0.13	0.085	0.035							0.12
					0.12	0.00164	0.00004	0	1	0	

Note: See Note to Table 1 for abbreviations.

Table 6: Second-order meta-analysis with the corrected overall effect sizes in Model 2.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
TD Children	25	0.08	0.244	0.006							0.04
LD Children	18	0.06	0.079	0.002							0.04
Adults	44	0.06	0.126	0.003							0.04
Older Adults	31	0.01	0.049	0.017							0.04
					0.04	0.00304	0.00078	0	1	0	

Note: See Note to Table 1 for abbreviations.

Table 7: Second-order meta-analysis with the uncorrected (naïve) overall effect sizes (only active control groups) in Model 2.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
TD Children	15	0.01	0.064	0.000							0.03
LD Children	12	0.08	0.043	0.000							0.04
Adults	27	0.09	0.049	0.000							0.04
Older Adults	16	–0.02	0.018	0.000							0.02
					0.03	0.00207	0.00250	0.0043	.83	.17	

Note: See Note to Table 1 for abbreviations.

Table 8: Second-order meta-analysis with the corrected overall effect sizes (only active control groups) in Model 2.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
TD Children	15	0.01	0.064	0.000							0.01
LD Children	12	0.02	0.111	0.000							0.01
Adults	27	–0.01	0.217	0.000							0.01
Older Adults	16	0.01	0.009	0.000							0.01
					0.01	0.00184	0.00001	0	1	0	

Note: See Note to Table 1 for abbreviations.

video-game playing in older adults, music training, chess training, and exergame training.

The action video-game training meta-analysis was a subsample of Sala, Tatlidil, and Gobet (2018b). We included only studies examining the effects of training on adult participants' cognitive skills. In line with the specific features of this field of research, the controls playing non-action video games were considered active. All the other comparisons (mostly passive controls) were labeled as non-active. There was 100% inter-rater agreement and the classification was the same as in Sala et al. (2018b).

The non-action video-game training meta-analysis was a subsample of Sala et al. (2018b). We included only those samples consisting of adults. The active control activities consisted of other non-action video games (e.g., The Sims), non-adaptive cognitive training, and brain-training programs (e.g., Lumosity). The inter-rater agreement was 94%. The raters resolved the discrepancies by talk.

We also ran a meta-analysis about the effects of video game training (both action and non-action) on older adults' performance on far-transfer cognitive tasks. The meta-analysis was a subsample of Sala et al. (2018b) as well. Regarding active control groups, we employed the same definition criteria as in the two previous video-game meta-analyses. The inter-rater agreement was 100%. No sample was included in more than one of these three meta-analyses. (We did not include a meta-analysis of video-game training in children because the number of primary studies was too small (n = 6) and the training programs were highly heterogeneous across the primary studies.)

The music-training meta-analysis was a subsample of Sala and Gobet (2017c). We excluded two studies that did not administer a pre-test. All the samples consisted of groups of either children or young adolescents with no diagnosed learning disabilities. Regarding the active control groups,

the classification was the same as in Sala and Gobet (2017c), and the inter-rater agreement was 100%.

To the best of our knowledge, the only meta-analysis of the effects of chess-based interventions on far transfer measures was Sala and Gobet (2016). Like the music-training meta-analysis, the population consisted of children and young adolescents. Eighteen out of 24 studies included in this meta-analysis did not meet our inclusion criteria. Most of these studies and effect sizes were excluded because they were derived from (a) questionnaires rather than cognitive/academic tests, (b) self-selected samples, or (c) because they had no pre-test. To be consistent with the other meta-analyses, all the effect sizes were recalculated. Also, three studies not included in Sala and Gobet (2016) were added (for the details, see Supplemental materials). The final sample consisted of nine studies. Regarding the type of control group (active vs. non-active), there was 100% inter-rater agreement.

The exergame-training meta-analysis was a subsample of Stanmore, Stubbs, Vancampfort, de Bruin, and Firth (2017). The original meta-analysis included 17 randomized controlled trials, most of them analyzing the effect of the intervention on older adults' cognitive function. We excluded two studies that analyzed the effects of the training program in younger populations. We also excluded one study that did not implement any exergame intervention and two studies that did not report enough data to calculate an effect size. All the effect sizes were recalculated. Finally, since this particular type of cognitive-training compares the effects of physical training and cognitive-training games with physical training only, groups who underwent physical activities only were considered as active. Concerning the type of control group (active vs. non-active), there was 100% inter-rater agreement and the classification was the same as in Stanmore et al. (2017).[6]

Results

Action video-game training

The overall meta-analytic mean was $\bar{g} = 0.08$, $SE = 0.05$, 95% CI [–0.01; 0.17], $k = 32$, $p = .094$. The test of heterogeneity was not significant, $Q = 39.76$, $\tau^2 = 0.000$, $p = .134$. PET and PEESE estimators were $\bar{g} = –0.42$, $SE = 0.10$, $p < .001$ and $\bar{g} = –0.11$, $SE = 0.05$, $p = .046$, respectively. Since both the corrected estimates were excessively negative, we ran a trim-and-fill analysis. With the L0 estimator, the analysis filled 11 studies left of the mean. The overall meta-analytic mean was $\bar{g} = –0.01$, $SE = 0.05$, 95% CI [–0.09; 0.08], $k = 43$, $p = .901$. With the R0 estimator, the analysis filled seven studies left of the mean. The overall meta-analytic mean was $\bar{g} = 0.03$, $SE = 0.05$, 95% CI [–0.06; 0.12], $k = 39$, $p = .468$. With the Q0 estimator, the analysis filled 15 studies left of the mean. The overall meta-analytic mean was $\bar{g} = –0.03$, $SE = 0.05$, 95% CI [–0.13; 0.06], $k = 47$, $p = .519$. Since no publication-bias corrected estimate provided any evidence of a non-zero effect, the estimate that was the closest to zero ($\bar{g} = –0.01$, $SE = 0.05$) was selected for the second-order meta-analysis.

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g} = 0.10$, $SE = 0.06$, 95% CI [–0.02; 0.21], $k = 25$, $p = .109$. The test of heterogeneity was not significant, $Q = 31.74$, $\tau^2 = 0.011$, $p = .134$. PET and PEESE estimators were $\bar{g} = –0.43$, $SE = 0.09$, $p < .001$ and $\bar{g} = –0.11$, $SE = 0.05$, $p = .037$, respectively. Since both corrected estimates were excessively negative, we ran a trim-and-fill analysis. With the L0 estimator, the analysis filled nine studies left of the mean. The overall meta-analytic mean was $\bar{g} = –0.01$, $SE = 0.06$, 95% CI [–0.12; 0.10], $k = 34$, $p = .919$. With the R0 estimator, the analysis filled 11 studies left of the mean. The overall meta-analytic mean was $\bar{g} = –0.02$, $SE = 0.06$, 95% CI [–0.13; 0.09], $k = 36$, $p = .674$. With the Q0 estimator, the analysis filled 15 studies left of the mean. The overall meta-analytic mean was $\bar{g} = –0.08$, $SE = 0.07$, 95% CI [–0.21; 0.04], $k = 40$, $p = .200$. Again, the estimate that was the closest to zero ($\bar{g} = –0.01$, $SE = 0.06$) was selected for the second-order meta-analysis.

Non-action video game training

The overall meta-analytic mean was $\bar{g} = 0.15$, $SE = 0.05$, 95% CI [0.04; 0.25], $k = 16$, $p = .006$. The test of heterogeneity was not significant, $Q = 17.16$, $\tau^2 = 0.012$, $p = .309$. PET and PEESE estimators were $\bar{g} = 0.09$, $SE = 0.07$, $p = .191$ and $\bar{g} = 0.12$, $SE = 0.04$, $p = .010$, respectively. The PEESE estimate was selected for the second-order meta-analysis (PET's $p < .100$, one-tailed).

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g} = 0.03$, $SE = 0.06$, 95% CI [–0.09; 0.16], $k = 6$, $p = .592$. The test of heterogeneity was not significant, $Q = 2.72$, $\tau^2 = 0.000$, $p = .744$. PET and PEESE estimators were $\bar{g} = –0.07$, $SE = 0.12$, $p = .592$ and $\bar{g} = 0.00$, $SE = 0.07$, $p = .989$, respectively. The PEESE estimate was selected for the second-order meta-analysis.

Video game training in older adults

The overall meta-analytic mean was $\bar{g} = 0.04$, $SE = 0.06$, 95% CI [–0.07; 0.15], $k = 10$, $p = .493$. The test of heterogeneity was not significant, $Q = 5.93$, $\tau^2 = 0.000$, $p = .747$. PET and PEESE estimators were $\bar{g} = –0.07$, $SE = 0.09$, $p = .445$ and $\bar{g} = 0.00$, $SE = 0.05$, $p = .950$, respectively. The PEESE estimate was selected for the second-order meta-analysis.

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g} = –0.03$, $SE = 0.09$, 95% CI [–0.21; 0.14], $k = 4$, $p = .709$. The test of heterogeneity was not significant, $Q = 0.05$, $\tau^2 = 0.000$, $p = .997$. PET and PEESE estimators were $\bar{g} = –0.05$, $SE = 0.02$, $p = .156$ and $\bar{g} = –0.04$, $SE = 0.01$, $p = .082$, respectively. The estimate that was the closest to zero ($\bar{g} = –0.03$, $SE = 0.09$) was selected for the second-order meta-analysis.

Music

The overall meta-analytic mean was $\bar{g} = 0.19$, $SE = 0.05$, 95% CI [0.10; 0.29], $k = 36$, $p < .001$. The test of heterogeneity was significant, $Q = 93.00$, $\tau^2 = 0.042$, $p < .001$. PET and PEESE estimators were $\bar{g} = –0.15$, $SE = 0.07$, $p = .039$, and $\bar{g} = 0.00$, $SE = 0.05$, $p = .921$, respectively. The influential case analysis found one influential study. After removing the study, the overall meta-analytic mean was $\bar{g} = 0.16$, $SE = 0.04$, 95% CI [0.07; 0.25], $k = 35$, $p < .001$. The test

of heterogeneity was still significant, but the amount of heterogeneity was lower, $Q = 72.64$, $\tau^2 = 0.028$, $p < .001$. PET and PEESE estimators were $\bar{g} = -0.14$, $SE = 0.06$, $p = .039$, and $\bar{g} = 0.00$, $SE = 0.04$, $p = .938$, respectively. Since the PET estimator was negative, the PEESE estimator was selected for the second-order meta-analysis because it was the closest to zero.

When considering only studies implementing an active control group, the overall meta-analytic mean was $\bar{g} = 0.03$, $SE = 0.06$, 95% CI [–0.10; 0.15], $k = 18$, $p = .678$. The test of heterogeneity was marginally significant, $Q = 26.99$, $\tau^2 = 0.023$, $p = .058$. PET and PEESE estimators were $\bar{g} = -0.20$, $SE = 0.08$, $p = .022$ and $\bar{g} = -0.10$, $SE = 0.06$, $p = .122$, respectively. With the L0 estimator, the analysis filled six studies left of the mean. The overall meta-analytic mean was $\bar{g} = -0.09$, $SE = 0.07$, 95% CI [–0.23; 0.05], $k = 24$, $p = .189$. With the R0 estimator, the analysis filled no study. With the Q0 estimator, the analysis filled nine studies left of the mean. The overall meta-analytic mean was $\bar{g} = -0.15$, $SE = 0.07$, 95% CI [–0.28; –0.01], $k = 27$, $p = .037$. The influential case analysis found one influential study. After removing the study, the overall meta-analytic mean was $\bar{g} = -0.02$, $SE = 0.06$, 95% CI [–0.13; 0.09], $k = 17$, $p = .704$. The test of heterogeneity was not significant, $Q = 18.92$, $\tau^2 = 0.012$, $p = .273$. PET and PEESE estimators were $\bar{g} = -0.20$, $SE = 0.07$, $p = .012$, and $\bar{g} = -0.11$, $SE = 0.05$, $p = .051$, respectively. With the L0 estimator, the analysis filled five studies left of the mean. The overall meta-analytic mean was $\bar{g} = -0.10$, $SE = 0.05$, 95% CI [–0.20; 0.01], $k = 22$, $p = .076$. With the R0 estimator, the analysis filled no study. With the Q0 estimator, the analysis filled eight studies left of the mean. The overall meta-analytic mean was $\bar{g} = -0.14$, $SE = 0.06$, 95% CI [–0.26; –0.03], $k = 25$, $p = .012$. All the corrected and uncorrected estimates were either negative or non-significantly different from zero. We thus selected the estimate that was the closest to zero ($\bar{g} = -0.02$, $SE = 0.06$) for the second-order meta-analysis.

Chess

The overall meta-analytic mean was $\bar{g} = 0.13$, $SE = 0.07$, 95% CI [–0.02; 0.27], $k = 9$, $p = .089$. The test of heterogeneity was significant, $Q = 56.58$, $\tau^2 = 0.031$, $p < .001$. PET and PEESE estimators were $\bar{g} = 0.12$, $SE = 0.11$, $p = .300$, and $\bar{g} = 0.13$, $SE = 0.07$, $p = .094$, respectively. No influential case was found. The PET estimate was selected for the second-order meta-analysis.

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g} = 0.05$, $SE = 0.10$, 95% CI [–0.15; 0.25], $k = 3$, $p = .623$. The test of heterogeneity was not significant, $Q = 0.47$, $\tau^2 = 0.000$, $p = .791$. PET and PEESE estimators were $\bar{g} = -0.05$, $SE = 0.22$, $p = .849$ and $\bar{g} = 0.01$, $SE = 0.10$, $p = .927$, respectively. The PEESE estimate was selected for the second-order meta-analysis because it was the closest to zero.

Exergames

The overall meta-analytic mean was $\bar{g} = 0.15$, $SE = 0.08$, 95% CI [–0.01; 0.32], $k = 11$, $p = .071$. The test of heterogeneity was not significant, $Q = 13.96$, $\tau^2 = 0.021$, $p = .175$. PET and PEESE estimators were $\bar{g} = -0.03$, $SE = 0.07$, $p = .659$ and $\bar{g} = 0.03$, $SE = 0.05$, $p = .554$, respectively. The PEESE estimate was selected for the second-order meta-analysis because it was the closest to zero.

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g} = 0.08$, $SE = 0.05$, 95% CI [–0.02; 0.17], $k = 8$, $p = .131$. The test of heterogeneity was not significant, $Q = 10.85$, $\tau^2 = 0.000$, $p = .145$. PET and PEESE estimators were $\bar{g} = -0.02$, $SE = 0.09$, $p = .835$ and $\bar{g} = 0.03$, $SE = 0.06$, $p = .620$ respectively. The PET estimate was selected for the second-order meta-analysis.

Second-order meta-analysis

Tables 9–12 summarize the results of the second-order meta-analysis of far-transfer effects for Model 3. Like in Model 2, publication-bias corrected estimates in studies

Table 9: Second-order meta-analysis with the uncorrected (naïve) overall effect sizes in Model 3.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{\bar{g}}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
WM (TD Children)	25	0.13	0.060	0.006							0.12
WM (LD Children)	18	0.12	0.032	0.002							0.12
WM (Adults)	44	0.12	0.041	0.003							0.12
WM (Older Adults)	32	0.13	0.085	0.035							0.12
Action VG (Adults)	32	0.08	0.073	0.000							0.12
Non-Action VG (Adults)	16	0.15	0.047	0.012							0.12
VG (Older Adults)	10	0.04	0.033	0.000							0.12
Music (TD Children)	36	0.19	0.087	0.042							0.12
Chess (TD Children)	9	0.13	0.049	0.031							0.12
Exergames (Older Adults)	11	0.15	0.079	0.021							0.12
					0.12	0.00235	0.00129	0	1	0	

Note: See Note to Table 1 for abbreviations.

Table 10: Second-order meta-analysis with the corrected overall effect sizes in Model 3.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
WM (TD Children)	25	0.08	0.244	0.006							0.04
WM (LD Children)	18	0.06	0.079	0.002							0.04
WM (Adults)	44	0.06	0.126	0.003							0.04
WM (Older Adults)	31	0.01	0.049	0.017							0.04
Action VG (Adults)	43	–0.01	0.087	0.000							0.04
Non-Action VG (Adults)	16	0.12	0.025	0.012							0.04
VG (Older Adults)	10	0.00	0.028	0.000							0.04
Music (TD Children)	35	0.00	0.067	0.028							0.04
Chess (TD Children)	9	0.12	0.099	0.031							0.04
Exergames (Older Adults)	11	0.03	0.028	0.021							0.04
					0.04	0.00263	0.00210	0	1	0	

Note: See Note to Table 1 for abbreviations.

Table 11: Second-order meta-analysis with the uncorrected (naïve) overall effect sizes (only active control groups) in Model 3.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
WM (TD Children)	15	0.01	0.064	0.000							0.04
WM (LD Children)	12	0.08	0.043	0.000							0.04
WM (Adults)	27	0.09	0.049	0.000							0.04
WM (Older Adults)	16	–0.02	0.018	0.000							0.04
Action VG (Adults)	25	0.10	0.089	0.011							0.04
Non-Action VG (Adults)	6	0.03	0.024	0.000							0.04
VG (Older Adults)	4	–0.03	0.033	0.000							0.04
Music (TD Children)	18	0.03	0.071	0.023							0.04
Chess (TD Children)	3	0.05	0.031	0.000							0.04
Exergames (Older Adults)	8	0.08	0.020	0.000							0.04
					0.04	0.00300	0.00214	0	1	0	

Note: See Note to Table 1 for abbreviations.

implementing an active control group are all around zero. We estimated the unbiased overall effect size to be $\bar{\bar{g}} = 0.00$. Finally, second-order sampling error accounted for the observed between-meta-analysis variance (i.e., $\sigma^2 = 0$) in all the models.

Omnibus meta-analysis

The overall meta-analytic mean was $\bar{g} = 0.13$, $SE = 0.02$, 95% CI [0.10; 0.16], $k = 233$, $p < .001$. The test of heterogeneity was significant, $Q = 368.69$, $\tau^2 = 0.015$, $p < .001$. PET and PEESE estimators were $\bar{g} = 0.03$, $SE = 0.02$, $p = .137$, and $\bar{g} = 0.07$, $SE = 0.01$, $p < .001$, respectively. The influential case analysis found twelve influential studies. After removing these studies, the overall meta-analytic mean was $\bar{g} = 0.12$, $SE = 0.02$, 95% CI [0.09; 0.15], $k = 221$, $p < .001$. The test of heterogeneity was not significant, $Q = 235.91$, $\tau^2 = 0.006$, $p = .220$. PET and PEESE estimators were $\bar{g} = 0.00$, $SE = 0.03$, $p = .872$, and $\bar{g} = 0.06$, $SE = 0.02$, $p < .001$, respectively.

When considering only studies implementing an active control group the overall meta-analytic mean was $\bar{g} = 0.03$, $SE = 0.02$, 95% CI [0.00; 0.07], $k = 134$, $p = .077$. The test of heterogeneity was not significant, $Q = 121.14$, $\tau^2 = 0.003$, $p = .761$. PET and PEESE estimators were $\bar{g} = –0.07$, $SE = 0.03$, $p = .007$ and $\bar{g} = –0.03$, $SE = 0.02$, $p = .133$, respectively.

Discussion

Like in Model 2, when placebo effects and publication bias are controlled for, the actual impact of cognitive-training programs on far-transfer measures is null regardless of the training regimen employed or population examined.

Table 12: Second-order meta-analysis with the corrected overall effect sizes (only active control groups) in Model 3.

Population	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	k_i	$\bar{g}_i$	$S^2_{g_i}$	τ^2	$\bar{\bar{g}}$	σ^2_e	$\sigma^2_{\bar{g}}$	σ^2	Pro_{var}	R_{xx}	$Adj.\bar{g}_i$
WM (TD Children)	15	0.01	0.064	0.000							0.00
WM (LD Children)	12	0.02	0.111	0.000							0.00
WM (Adults)	27	–0.01	0.217	0.000							0.00
WM (Older Adults)	16	0.01	0.009	0.000							0.00
Action VG (Adults)	34	–0.01	0.107	0.011							0.00
Non-Action VG (Adults)	6	0.00	0.033	0.000							0.00
VG (Older Adults)	4	–0.03	0.033	0.000							0.00
Music (TD Children)	17	–0.02	0.055	0.012							0.00
Chess (TD Children)	3	0.01	0.032	0.000							0.00
Exergames (Older Adults)	8	–0.02	0.072	0.000							0.00
					0.00	0.00303	0.00014	0	1	0	

Note: See Note to Table 1 for abbreviations.

In all the models, the differences across the first-order meta-analytic means ($\bar{g}_i$) are accounted for by second-order sampling error ($\sigma^2 = 0$). Also, all the corrected first-order meta-analytic means are associated with null or low amount of true heterogeneity (none of them significant). Again, within-study variance (ω^2) was low or null in the omnibus meta-analyses and most of the other cases (see Tables S2 and S3 in the Supplemental materials available online). This outcome confirms that far transfer is null regardless of the measure employed to assess it. The small or null overall effect sizes and near-zero amount of true heterogeneity in the omnibus meta-analysis corroborate the findings of the second-order meta-analytic models.

General Discussion

Cognitive training is currently one of the most studied and controversial topics in the behavioral sciences. As in many other areas of the social and behavioral sciences, the initial promising findings have been challenged by more recent replications. The broad meta-analytic investigation reported in this paper (n = 1,555, k = 332, N = 21,968) has evaluated, via first-order meta-analysis, the impact of a variety of cognitive-training programs on different populations' cognitive and academic skills. Critically, second-order meta-analyses were carried out to assess whether the differences across first-order meta-analytic means were due to true variance or second-order sampling error.

The results are highly consistent: near transfer frequently occurs and, interestingly, seems to be moderated by the type of population; by contrast, far transfer is very modest at best. Moreover, once publication bias and placebo effects are ruled out, far-transfer effects are null regardless of the type of far-transfer measure, type of cognitive training program, and population. This latter conclusion can be summarized by three equations:

$$Adj.\bar{g}_i = \bar{g}_i - \left(PB_i + PE_i\right) = 0$$
$$\tau^2_{Adj.\bar{g}_i} = 0$$
$$\sigma^2_{Adj.\bar{g}} = 0$$

where $Adj.\bar{g}_i$ is the adjusted overall effect size in the ith first-order meta-analysis, $\bar{g}_i$ is the naïve overall effect size, PB_i is publication bias in the ith first-order meta-analysis, PE_i is placebo effects in the ith first-order meta-analysis, $\tau^2_{Adj.\bar{g}_i}$ is the amount of true heterogeneity in the ith first-order meta-analysis with adjusted overall effect sizes, and $\sigma^2_{Adj.\bar{g}}$ is the true variance between first-order adjusted overall effect sizes.

Beyond first- and second-order meta-analytic evidence on cognitive training, the observed lack of generalized cognitive benefits is consistent with a well-established corpus of findings in other disciplines. For example, although education is positively associated with scores on cognitive tests, its impact on general intelligence or domain-general cognitive skills appears to be modest (Detterman, 2016; Finn et al., 2014; Mosing, Madison, Pedersen, & Ullén, 2016; Ritchie et al., 2015), yet relatively consistent (Ritchie & Tucker-Drob, 2018). If even years of mentally challenging activities in school exert only a small effect on people's overall cognitive ability, it is hard to see how a few weeks (or months) of cognitive training can lead to more appreciable benefits. Also, as seen earlier, research into learning and the psychology of expertise has repeatedly shown that far transfer is rare because skill acquisition relies on domain-specific perceptual and conceptual information (Ericsson & Charness, 1994; Gobet, 2016). Furthermore, other non-cognitive-training-based interventions too have failed to induce appreciable generalized effects (e.g., Berggren, Nilsson, Brehmer, Schmiedek, & Lövdén, 2018; Sisk, Burgoyne, Sun, Butler, & Macnamara, 2018). Put together, the convergent insights from different fields of scientific research about far transfer represent a successful example of triangulation

(Campbell & Fiske, 1959; Munafò & Smith, 2018), and lead us towards the conclusion that while human cognition is malleable to training, the benefits are, to a large extent, domain-specific.

Moreover, domain-specific benefits such as the near-transfer effects observed in Model 1 do not necessarily imply that the participants' memory-related cognitive skills have improved. Certainly, this is a possible explanation: the observed near-transfer effects might represent true cognitive enhancement (e.g., increased WM capacity). However, WM training may simply make participants more able to perform a certain typology of tasks. Shipstead, et al. (2012) have observed that even those tasks that are not usually part of training regimens (e.g., complex span tasks) share some amount of overlap with the trained tasks (e.g., simple-span tasks). Thus, people undergoing WM training may just acquire the ability to perform near-transfer tasks slightly better than controls. This interpretation is in line with the absence of far-transfer effects. In fact, WM capacity is an important predictor of job and academic performance and is highly correlated with fluid reasoning. Also, deficits in WM capacity are comorbid with several learning disabilities (e.g., Swanson, 2006). Enhanced WM capacity is expected to make information processing more efficient, which, in turn, should lead to broad benefits in other domains of cognition. Thus, if WM training program enhanced the participants' WM capacity, improvements in other cognitive and academic tasks should occur. However, this does not seem to be the case. Thus, in line with Shipstead et al. (2012), our opinion is that near-transfer effects do not represent true cognitive enhancement. That being said, the topic deserves further investigation.

Overall, the implications are profound. From the theoretical point of view, those theories of human cognition predicting minimal or no far transfer of skills are corroborated by our findings (e.g., chunking-based theories; for a review, see Gobet, 2016). Conversely, those theories predicting the generalization of skills acquired by training across multiple domains are refuted (e.g., Bavelier, Green, Pouget, & Schrater, 2012; Jaeggi et al., 2008; Tierney, Krizman, & Kraus, 2015). Regarding practical implications, the obvious conclusion is that, to date, professional and educational curricula should focus on domain-specific knowledge rather than general and allegedly transferable skills.

Limitations

As noted in the General Method section, none of the first-order meta-analyses included effects that were corrected for measurement error. However, we think that the practical consequences of this flaw are negligible, especially in Models 2 and 3. In fact, pretty much all the uncorrected and corrected far-transfer overall effects were close or equal to zero. Thus, applying such a correction would leave all these estimates virtually unaltered. Interestingly, correcting for measurement error would increase the effect sizes' sampling error variances and, consequently, reduce the total amount of between-study true heterogeneity (when any). Thus, the actual heterogeneity is probably even smaller than the one observed in some of the models.

A second limitation concerns the choice of the most appropriate corrected estimate to be included in the second-order meta-analytic models. As seen, our criterion is based on the PET: if the PET-corrected estimate is not significantly ($p < .100$, one-tailed) positive, we select the estimate that is the closest to zero. In our opinion, this criterion is sensible and reflects the rationale of PET (Stanley, 2017). Furthermore, the fact that none of the PET-corrected estimates in Models 2 and 3 were both positive and statistically ($p < .100$, one-tailed) different from zero represents substantial evidence in favor of our hypothesis (i.e., no far transfer regardless of population and training regimen). However, when the PET appeared to overcorrect (especially in Model 1), we preferred the PEESE or trim-and-fill estimates. That being said, preferring a corrected estimate over another one always implies a certain degree of arbitrariness that is impossible to rule out completely, especially because the true mechanism introducing the bias (e.g., selective reporting and *p*-hacking) is unknown. In any case, given the high degree of consistency observed across the findings, we expect the overall results to essentially remain the same regardless of the publication-bias corrected estimate employed. Moreover, the impact of publication bias on the overall results appears to be somewhat limited. In all the three Models (**Tables 1–12**), the publication-bias corrected grand means do not differ from the uncorrected estimates for more than 0.10 standardized mean differences (0.05 in the models including only active comparisons). Thus, using other publication-bias detection techniques would produce negligible differences.

Third, the selection of some of the first-order meta-analyses is, to a certain degree, arbitrary too. Specifically, along with Melby-Lervåg et al. (2016) and Sala and Gobet (2017b), several other recent meta-analyses examining the effects of WM training have been carried out (e.g., Au et al., 2015; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017). These meta-analyses are substantially in line with our conclusions. Soveri et al. (2017) report no far-transfer effects in healthy adults. Au et al. (2015) claim that *n*-back training has a small overall positive impact ($\bar{g} = 0.24$) on fluid intelligence. However, this effect disappears in studies using active control groups (Dougherty, Hamovits, & Tidwell, 2016). A similar consideration applies to action video-game training. Along with Sala et al. (2018b), another meta-analysis examining the effects of action video-game training has been carried out recently (Bediou et al., 2018). While reporting a small to medium overall effect size ($\bar{g} = 0.34$), this meta-analysis shows a highly asymmetrical distribution of the effect sizes which suggests that the uncorrected overall effect is an overestimation. This conclusion is upheld by the results of the two publication bias analyses (trim-and-fill and PET-PEESE) included in the original article. Thus, we think that our findings are robust regardless of the particular meta-analytic study selected for our second-order meta-analyses.

Finally, another issue concerns the limited total number of primary studies included in some of the first-order

meta-analyses (e.g., video-games in older adults, chess, and exergames). Small number of effect sizes provides less accurate estimates (large standard errors) and limits the power of publication bias analysis. This problem cannot be overcome until new experiments have been carried out. Nonetheless, the results provided by the omnibus meta-analyses, which do not suffer from low statistical power, confirm the results of the first-order and second-order meta-analyses.

Conclusions and Recommendation for Future Research

This study aimed to examine, using second-order meta-analysis, to what extent cognitive-training programs induce near-transfer and far-transfer effects. Near transfer occurs in all the examined populations but, interestingly, young populations seem to benefit from the treatment (i.e., WM training) more than adult populations. It is also worth remembering that the observed near transfer probably reflects an improvement in the ability to perform memory tasks rather than enhanced cognitive function (Shipstead et al., 2012). On the other hand, there is no evidence of far transfer regardless of the population and type of cognitive-training program. These findings are consistent with substantial research into education, skill acquisition, and expert performance.

Despite being a trivial caveat, it is worth mentioning that, strictly speaking, our findings apply only to the populations and training programs examined. We cannot exclude that alternative cognitive-training programs may provide appreciable cognitive benefits in some special populations (e.g., cancer survivors). That being said, we think that the lack of far transfer is an invariant in human cognition, at least with regard to the general population, regardless of age and population types (e.g., LD children).

Further research is required to extend the present second-order meta-analytic investigation in order to test our hypothesis. First, the investigation should include a meta-analysis of brain-training programs. To our knowledge, no eligible first-order meta-analysis on the topic has been published so far. For example, meta-analytic investigations about brain-training programs usually include other cognitive-training programs such as WM training or action video-game training (e.g., Mewborn, Lindbergh, & Miller, 2017). Thus, adding such a meta-analysis would violate the assumption of statistical independence between studies. It is worth noting that, in line with our conclusions, Simons et al. (2016) have concluded that no convincing evidence of far transfer has been provided by brain-training experimental studies so far (see also Rebok et al., 2014). Nonetheless, a meta-analysis examining both near- and far-transfer effects of brain-training programs would provide valuable additional information to test this claim further.

Second, novel cognitive-training programs have been designed in recent years (e.g., Daugherty et al., 2018), and new studies analyzing the effect of old cognitive-training programs on different populations are being currently (or have just been) carried out (e.g., action video-game training for dyslexic children; Franceschini et al., 2017). Once the number of experimental studies is sufficient to run additional first-order meta-analyses, it will be possible to carry out more extensive second-order meta-analytic models. The same applies to those studies about the effects of cognitive-training programs on those populations that are included in our analyses. New primary studies will contribute to the updating of the first- and second-order meta-analyses.

Finally, the current paper has shown that second-order meta-analysis is a powerful tool for settling debates in the behavioral sciences. In our case, it has unambiguously showed that, when publication bias and placebo effects are controlled for, presumed far-transfer effects of cognitive training vanish, between-meta-analyses heterogeneity dissolves, and true variance between first-order adjusted overall effect sizes disappears. Once the debates and controversies about the effects of cognitive training are reined by second-order meta-analysis, everything is distilled to a parsimonious answer and a single number: zero.

Data Accessibility Statement

All the raw data and analysis scripts can be found on this paper's project page on OSF (https://osf.io/qk2vu/).

Notes

[1] It must be observed that there are no established guidelines (e.g., PRISMA) for selecting first-order meta-analyses in second-order meta-analysis. We thus used those criteria that we thought most sensible and suitable in this particular case.

[2] The only exception is the meta-analysis of exergame training. In this field, the active controls usually consist of participants involved in physical activities. For more details, see Model 3.

[3] We postulated that the true transfer effect of a training program could not be negative.

[4] The alpha was set at 0.100 (rather than 0.050) to make estimates more conservative (i.e., potential influential cases more likely to be excluded) and reliable (i.e., the less heterogeneity, the more trustworthy corrected estimates).

[5] Unless specified otherwise, the reported *p*-values are two-tailed.

[6] It is worth noting that the results we report here are sometimes different from the ones presented in the original meta-analyses. This happened for two main reasons, in addition to the exclusion of some studies. First, in the original publications, the authors used different methods to model nested effect sizes. For example, in some cases the authors (a) applied no correction with either merged or unmerged effects (Melby-Lervåg et al., 2016; Sala & Gobet, 2016; Stanmore et al., 2017), (b) applied corrections based on Cheung and Chan's (2004, 2008) method (Sala & Gobet, 2017b, c; Sala et al., 2018), or (c) used robust variance estimation (Sala et al., 2018). Second, the original meta-analyses employed different publication-bias detection techniques (e.g., *p*-curve, selection models, and trim-and-fill). Despite these differences, in most of the cases both uncorrected and corrected

estimates are very close to the ones reported in the original meta-analyses. The only significant exception is Stanmore et al. (2017). This meta-analysis, however, shows a highly asymmetrical distribution of the effect sizes suggesting the presence of severe publication bias.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Meta-Analyses Search. DOI: https://doi.org/10.1525/collabra.203.s1
- **Appendix B.** List of Included/Excluded Studies. DOI: https://doi.org/10.1525/collabra.203.s2
- **Appendix C.** Cheung and Chan's (2014) Correction. DOI: https://doi.org/10.1525/collabra.203.s3
- **Appendix D.** Robust Variance Estimation Analyses. DOI: https://doi.org/10.1525/collabra.203.s4
- **Appendix E.** List of Influential Studies. DOI: https://doi.org/10.1525/collabra.203.s5

Acknowledgements

The authors gratefully thank Thomas Redick for providing useful information about some studies included in Melby-Lervåg et al. (2016). We also thank Frank L. Schmidt for providing some clarification on the use of publication-bias-corrected estimates in second-order meta-analysis.

Funding Information

GS receives funding from the Japan Society for the Promotion of Science [17F17313].

Competing Interests

The authors have no competing interests to declare.

Author Contributions

- Contributed to conception and design: GS, FG
- Contributed to acquisition of data: GS, NDA, KST, TT
- Contributed to analysis and interpretation of data: GS
- Drafted and/or revised the article: GS, NDA, KST, TT, YG, FG
- Approved the submitted version for publication: GS, NDA, KST, TT, YG, FG

References

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & **Jaeggi, S. M.** (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*, 366–377. DOI: https://doi.org/10.3758/s13423-014-0699-x

Barnett, S. M., & **Ceci, S. J.** (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637. DOI: https://doi.org/10.1037//0033-2909.128.4.612

Bavelier, D., Green, C. S., Pouget, A., & **Schrater, P.** (2012). Brain plasticity through the life span: learning to learn and action video games. *Annual Review of Neuroscience*, *35*, 391–416. DOI: https://doi.org/10.1146/annurev-neuro-060909-152832

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & **Bavelier, D.** (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, *144*, 77–110. DOI: https://doi.org/10.1037/bul0000130

Berggren, R., Nilsson, J., Brehmer, Y., Schmiedek, F., & **Lövdén, M.** (2018). No evidence that foreign language learning in older age improves cognitive ability: A randomized controlled study. DOI: https://doi.org/10.31219/osf.io/8y4ga

Bilalić, M., McLeod, P., & **Gobet, F.** (2009). Specialization effect and its influence on memory and problem solving in expert chess players. *Cognitive Science*, *33*, 1117–1143. DOI: https://doi.org/10.1111/j.1551-6709.2009.01030.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*, 445–454. DOI: https://doi.org/10.1177/1745691613491271

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & **Rothstein, H. R.** (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley. DOI: https://doi.org/10.1002/9780470743386

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & **Hambrick, D. Z.** (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence*, *59*, 72–83. DOI: https://doi.org/10.1016/j.intell.2016.08.002

Campbell, D., & **Fiske, D.** (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105. DOI: https://doi.org/10.1037/h0046016

Campitelli, G., & **Gobet, F.** (2011). Deliberate practice: Necessary but not sufficient. *Current Directions in Psychological Science*, *20*, 280–285. DOI: https://doi.org/10.1177/0963721411421922

Carter, E., Schönbrodt, F., Gervais, W. M., & **Hilgard, J.** (2017). Correcting for bias in psychology: A comparison of meta-analytic methods. Preprint available at: https://psyarxiv.com/9h3nu/. DOI: https://doi.org/10.17605/OSF.IO/9H3NU

Chase, W. G., & **Ericsson, K. A.** (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 1–58). New York: Academic.

Chase, W. G., & **Simon, H. A.** (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81. DOI: https://doi.org/10.1016/0010-0285(73)90004-2

Chassy, P., & **Gobet, F.** (2010). Speed of expertise acquisition depends upon inherited factors. *Talent Development and Excellence*, *2*, 17–27.

Cheung, S. F., & **Chan, D. K.** (2004). Dependent effect sizes in meta-analysis: Incorporating the degree of interdependence. *Journal of Applied Psychology*, *89*, 780–791. DOI: https://doi.org/10.1037/0021-9010.89.5.780

Cheung, S. F., & **Chan, D. K.** (2008). Dependent correlations in meta-analysis: The case of heterogeneous dependence. *Educational and Psychological*

Measurement, *68*, 760–777. DOI: https://doi.org/10.1177/0013164408315263

Cheung, S. F., & **Chan, D. K.** (2014). Meta-analyzing dependent correlations: An SPSS macro and an R script. *Behavioral Research Methods*, *46*, 331–345. DOI: https://doi.org/10.3758/s13428-013-0386-2

Cohen, J. E., Green, C. S., & **Bavelier, D.** (2008). Training visual attention with video games: Not all games are created equal. In H. O'Neil, & R. Perez (Eds.), *Computer games and adult learning* (pp. 205–227). Amsterdam: Elsevier.

Cooper, H., & **Koenka, A. C.** (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist*, *67*, 446–462. DOI: https://doi.org/10.1037/a0027119

Daugherty, A. M., Zwilling, C., Paul, E. J., Sherepa, N., Allen, C., Kramer, A. F., Barbey, A. K., et al. (2018). Multi-modal fitness and cognitive training to enhance fluid intelligence. *Intelligence*, *66*, 32–44. DOI: https://doi.org/10.1016/j.intell.2017.11.001

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. K. Detterman, & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 1–24). Westport, CT, US: Ablex Publishing.

Detterman, D. K. (2014). Introduction to the intelligence special issue on the development of expertise: is ability necessary? *Intelligence*, *45*, 1–5. DOI: https://doi.org/10.1016/j.intell.2014.02.004

Detterman, D. K. (2016). Education and intelligence: Pity the poor teacher because student characteristics are more significant than teachers or schools. *The Spanish Journal of Psychology*, *19*, E93. DOI: https://doi.org/10.1017/sjp.2016.88

Donovan, M. S., Bransford, J. D., & **Pellegrino, J. W.** (1999). *How people learn: Bridging research and practice*. Washington, DC: National Academies Press.

Dougherty, M. R., Hamovits, T., & **Tidwell, J. W.** (2016). Revaluating the effect of n-back training on transfer through the Bayesian lens: Support for the null. *Psychonomic Bulletin & Review*, *23*, 306–316. DOI: https://doi.org/10.3758/s13423-015-0865-9

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*, 276–284. DOI: https://doi.org/10.1111/j.0006-341X.2000.00455.x

Ericsson, K. A., & **Charness, N.** (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*, 725–747. DOI: https://doi.org/10.1037/0003-066X.49.8.725

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., Gabrieli, J. D. E., et al. (2014). Cognitive skills, student achievements tests, and schools. *Psychological Science*, *25*, 736–744. DOI: https://doi.org/10.1177/0956797613516008

Franceschini, S., Trevisan, P., Ronconi, L., Bertoni, S., Colmar, S., Double, K., Gori, S., et al. (2017). Action video games improve reading abilities and visual-to-auditory attentional shifting in English-speaking children with dyslexia. *Scientific Reports*, *7*: 5863. DOI: https://doi.org/10.1038/s41598-017-05826-8

Gobet, F. (2016). *Understanding expertise: A multi-disciplinary approach*. London: Palgrave/Macmillan. DOI: https://doi.org/10.1007/978-1-137-57196-0

Gobet, F., & **Simon, H. A.** (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*, 1–40. DOI: https://doi.org/10.1006/cogp.1996.0011

Green, C. S., & **Bavelier, D.** (2003). Action video game modifies visual selective attention. *Nature*, *423*, 534–537. DOI: https://doi.org/10.1038/nature01647

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Hedges, L. V., & **Olkin, I.** (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

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

Hunter, J. E., & **Schmidt, F. L.** (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage. DOI: https://doi.org/10.4135/9781412985031

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & **Perrig, W. J.** (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 6829–6833. DOI: https://doi.org/10.1073/pnas.0801268105

Karbach, J., Könen, T., & **Spengler, M.** (2017). Who benefits the most? Individual differences in the transfer of executive control training across the lifespan. *Journal of Cognitive Enhancement*, *1*, 394–405. DOI: https://doi.org/10.1007/s41465-017-0054-z

Karbach, J., & **Schubert, T.** (2013). Training-induced cognitive and neural plasticity. *Frontiers in Human Neuroscience*, *9*, 6. DOI: https://doi.org/10.3389/fnhum.2013.00048

Kepes, S., & **McDaniel, M. A.** (2015). The validity of conscientiousness is overestimated in the prediction of job performance. *PLoS ONE*, *10*, e0141468. DOI: https://doi.org/10.1371/journal.pone.0141468

McShane, B. B., Böckenholt, U., & **Hansen, K. T.** (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*, 730–749. DOI: https://doi.org/10.1177/1745691616662243

Melby-Lervåg, M., Redick, T. S., & **Hulme, C.** (2016). Working memory training does not improve performance on measures of intelligence or other measures of far-transfer: Evidence from a meta-analytic review. *Perspective on Psychological Science*, *11*, 512–534. DOI: https://doi.org/10.1177/1745691616635612

Mewborn, C. M., Lindbergh, C. A., & **Miller, L. S.** (2017). Cognitive interventions for cognitively healthy, mildly impaired, and mixed samples of older adults:

A systematic review and meta-analysis of randomized-controlled trials. *Neuropsychology Review*, *27*, 403–439. DOI: https://doi.org/10.1007/s11065-017-9350-8

Moreno, S. G., Sutton, A. J., Ades, A. E., Stanley, T. D., Abrams, K. R., Peters, J. L., & **Cooper, N. J.** (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC medical research methodology*, *9*, 2. DOI: https://doi.org/10.1186/1471-2288-9-2

Mosing, M. A., Madison, G., Pedersen, N. L., & **Ullén, F.** (2016). Investigating cognitive transfer within the framework of music practice: Genetic pleiotropy rather than causality. *Developmental Science*, *19*, 504–512. DOI: https://doi.org/10.1111/desc.12306

Munafò, M. R., & **Smith, G. D.** (2018). Replication is not enough. *Nature*, *553*, 399–401. DOI: https://doi.org/10.1038/d41586-018-01023-3

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*. DOI: https://doi.org/10.1126/science.aac4716

Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H.-Y., King, J. W., Willis, L. S., et al. (2014). Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. *Journal of the American Geriatrics Society*, *62*, 16–24. DOI: https://doi.org/10.1111/jgs.12607

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., Engle, R. W., et al. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*, 359–379. DOI: https://doi.org/10.1037/a0029082

Rickard, N. S., Bambrick, C. J., & **Gill, A.** (2012). Absence of widespread psychosocial and cognitive effects of school-based music instruction in 10–13-year-old students. *International Journal of Music Education*, *30*, 57–78. DOI: https://doi.org/10.1177/0255761411431399

Rikers, R. M. J. P., Schmidt, H. G., & **Boshuizen, H. P. A.** (2002). On the constraints of encapsulated knowledge: Clinical case representations by medical experts and subexperts. *Cognition and Instruction*, *20*, 27–45. DOI: https://doi.org/10.1207/S1532690XCI2001_2

Ritchie, S. J., Bates, T. C., & **Deary, I. J.** (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology*, *51*, 573–582. DOI: https://doi.org/10.1037/a0038981

Ritchie, S. J., & **Tucker-Drob, E. M.** (2018). How much does education intelligence? A meta-analysis. *Psychological Science*. Advance online publication. DOI: https://doi.org/10.1177/0956797618774253

Sala, G., Aksayli, N. D., Tatlidil, K. S., Gondo, Y., & **Gobet, F.** (2018a, August 1st). Working memory training does not enhance older adults' cognitive skills: A meta-analysis. https://psyarxiv.com/5frzb/

Sala, G., & **Gobet, F.** (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis. *Educational Research Review*, *18*, 46–57. DOI: https://doi.org/10.1016/j.edurev.2016.02.002

Sala, G., & **Gobet, F.** (2017a). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Direction in Psychological Science*, *26*, 515–520. DOI: https://doi.org/10.1177/0963721417712760

Sala, G., & **Gobet, F.** (2017b). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology*, *53*, 671–685. DOI: https://doi.org/10.1037/dev0000265

Sala, G., & **Gobet, F.** (2017c). When the music's over. Does music skill transfer to children's and young adolescents' cognitive and academic skills? A meta-analysis. *Educational Research Review*, *20*, 55–67. DOI: https://doi.org/10.1016/j.edurev.2016.11.005

Sala, G., Tatlidil, K. S., & **Gobet, F.** (2018b). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin*, *144*, 111–139. DOI: https://doi.org/10.1037/bul0000139

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*, 511–514. DOI: https://doi.org/10.1111/j.0956-7976.2004.00711.x

Schmidt, F. L. (2017). Beyond questionable research methods: The role of omitted relevant research in the credibility of research. *Archives of Scientific Psychology*, *5*, 32–41. DOI: https://doi.org/10.1037/arc0000033

Schmidt, F. L., & **Hunter, J. E.** (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Newbury Park, CA: Sage. DOI: https://doi.org/10.4135/9781483398105

Schmidt, F. L., & **Oh, I.-S.** (2013). Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Processes*, *121*, 204–218. DOI: https://doi.org/10.1016/j.obhdp.2013.03.002

Schmidt, F. L., & **Oh, I.-S.** (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology*, *4*, 32–37. DOI: https://doi.org/10.1037/arc0000029

Shipstead, Z., Redick, T. S., & **Engle, R. W.** (2012). Is working memory training effective? *Psychological Bulletin*, *138*, 628–654. DOI: https://doi.org/10.1037/a0027473

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & **Stine-Morrow, E. A. L.** (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*, 103–186. DOI: https://doi.org/10.1177/1529100616661983

Simonsohn, U., Nelson, L. D., & **Simmons, J. P.** (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. DOI: 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*, 666–681. DOI: https://doi.org/10.1177/1745691614553988

Singley, M. K., & **Anderson, J. R.** (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & **Macnamara, B. N.** (2018). To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. *Psychological Science*, *29*, 549–571. DOI: https://doi.org/10.1177/0956797617739704

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & **Laine, M.** (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*, *24*, 1077–1096. DOI: https://doi.org/10.3758/s13423-016-1217-0

Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science*, *8*, 581–591. DOI: https://doi.org/10.1177/1948550617693062

Stanley, T. D., & **Doucouliagos, H.** (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*, 60–78. DOI: https://doi.org/10.1002/jrsm.1095

Stanley, T. D., & **Doucouliagos, H.** (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research Synthesis Methods*, *8*, 19–42. DOI: https://doi.org/10.1002/jrsm.1211

Stanmore, E., Stubbs, B., Vancampfort, D., de Bruin, E. D., & **Firth, J.** (2017). The effect of active video games on cognitive functioning in clinical and non-clinical populations: A meta-analysis of randomized controlled trials. *Neuroscience and Biobehavioral Reviews*, *78*, 34–43. DOI: https://doi.org/10.1016/j.neubiorev.2017.04.011

Strobach, T., & **Karbach, J.** (2016). *Cognitive training: An overview of features and applications*. New York: Springer. DOI: https://doi.org/10.1007/978-3-319-42662-4

Swanson, H. L. (2006). Working memory and reading disabilities: Both phonological and executive processing deficits are important. In T. P. Alloway, & S. E. Gathercole (Eds.), *Working memory and neurodevelopmental disorders* (pp. 59–88). Hove, England: Psychology Press.

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, *120*, 439–471. DOI: https://doi.org/10.1037/a0033138

Taatgen, N. A. (2016). Theoretical models of training and transfer effects. In T. Strobach, & J. Karbach (Eds.), *Cognitive training: An overview of features and applications* (pp. 19–29). New York: Springer. DOI: https://doi.org/10.1007/978-3-319-42662-4_3

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & **Schmid, R. F.** (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, *81*, 4–28. DOI: https://doi.org/10.3102/0034654310393361

Thorndike, E. L., & **Woodworth, R. S.** (1901). The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review*, *8*, 247–261. DOI: https://doi.org/10.1037/h0074898

Tierney, A. T., Krizman, J., & **Kraus, N.** (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 10062–10067. DOI: https://doi.org/10.1073/pnas.1505114112

Tracz, S. M., Elmore, P. B., & **Pohlmann, J. T.** (1992). Correlational meta-analysis: Independent and nonindependent cases. *Educational and Psychological Measurement*, *52*, 879–888. DOI: https://doi.org/10.1177/0013164492052004007

van Ravenzwaaij, D., Boekel, W., Forstmann, B. U., Ratcliff, R., & **Wagenmakers, E. J.** (2014). Action video games do not improve the speed of information processing in simple perceptual tasks. *Journal of Experimental Psychology: General*, *143*, 1794–1805. DOI: https://doi.org/10.1037/a0036923

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Salanti, G., et al. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *7*, 55–79. DOI: https://doi.org/10.1002/jrsm.1164

Vevea, J. L., & **Woods, C. M.** (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*, 428–443. DOI: https://doi.org/10.1037/1082-989X.10.4.428

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48. Retrieved from: http://brieger.esalq.usp.br/CRAN/web/packages/metafor/vignettes/metafor.pdf. DOI: https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & **Cheung, M. W. L.** (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*, 112–125. DOI: https://doi.org/10.1002/jrsm.11

Peer review comments
The author(s) of this paper chose the Open Review option, and Streamlined Review option, and all new peer review comments (but not ported comments from the prior review) are available at: http://doi.org/10.1525/collabra.203.pr

How to cite this article: Sala, G., Aksayli, N. D., Tatlidil, K. S., Tatsumi, T., Gondo, Y., & Gobet, F. (2019). Near and Far Transfer in Cognitive Training: A Second-Order Meta-Analysis. *Collabra: Psychology*, 5(1): 18. DOI: https://doi.org/10.1525/collabra.203

Senior Editor: Rolf Zwaan

Editor: Peter Verkoeijen

Submitted: 14 October 2018 **Accepted:** 25 March 2019 **Published:** 26 April 2019

UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Collabra: Psychology is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS