THEORETICAL REVIEW

Working memory training in typically developing children: A multilevel meta-analysis

Giovanni Sala [1] · Fernand Gobet [2]

Abstract

Working memory (WM) training in typically developing (TD) children aims to enhance not only performance in memory tasks but also other domain-general cognitive skills, such as fluid intelligence. These benefits are then believed to positively affect academic achievement. Despite the numerous studies carried out, researchers still disagree over the real benefits of WM training. With this meta-analysis ($m = 41$, $k = 393$, $N = 2,375$), we intended to resolve the discrepancies by focusing on the potential sources of within-study and between-study true heterogeneity. Small to medium effects were observed in memory tasks (i.e., near transfer). The size of these effects was proportional to the similarity between the training task and the outcome measure. By contrast, far-transfer measures of cognitive ability (e.g., intelligence) and academic achievement (mathematics and language ability) were essentially unaffected by the training programs, especially when the studies implemented active controls ($\bar{g} = 0.001$, $SE = 0.055$, $p = .982$, $\tau^2 = 0.000$). Crucially, all the models exhibited a null or low amount of true heterogeneity, which was wholly explained by the type of controls (nonactive vs. active) and by statistical artifacts, in contrast to the claim that this field has produced mixed results. Since the empirical evidence shows the absence of both generalized effects and true heterogeneity, we conclude that there is no reason to keep investing resources in WM training research with TD children.

Keywords Academic achievement · Cognitive enhancement · Cognitive training · Meta-analysis · Transfer · Working memory training

It is widely acknowledged that general cognitive ability is a major predictor of academic achievement and job performance (Detterman, 2014; Gobet, 2016; Schmidt, 2017; Wai, Brown, & Chabris, 2018). Finding a way to enhance people's general cognitive ability would thus have a huge societal impact. That is why the idea that engaging in cognitive-training programs can boost one's domain-general cognitive skills has been evaluated in numerous experimental trials over the last two decades (for reviews, see Sala, Aksayli, Tatlidil, Tatsumi, et al., 2019b; Simons et al., 2016). The most influential of such programs has been working memory (WM) training.

✉ Giovanni Sala
sala.giovanni@fujita-hu.ac.jp

1   Fujita Health University, Toyoake, Japan

2   London School of Economics and Political Science, London, UK

WM is the ability to *store* and *manipulate* the information needed to perform complex cognitive tasks (Baddeley, 1992, 2000). The concept of WM thus goes beyond that of short-term memory (STM): Whereas the latter focuses on how much information can be passively stored in one's cognitive system, the former involves an active manipulation of the information, as well (Cowan, 2017; Daneman & Carpenter, 1980).

The importance of WM in cognitive development is well-known. WM capacity—that is, the maximum amount of information that WM can store and manipulate—steadily increases throughout infancy and childhood up to adolescence (Cowan, 2016; Gathercole, Pickering, Ambridge, & Wearing, 2004), due to both maturation and an increase in knowledge (Cowan, 2016; Jones, Gobet, & Pine, 2007). WM capacity is positively correlated with essential cognitive functions such as fluid intelligence and attentional processes (Engle, 2018; Kane, Hambrick, & Conway, 2005; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). WM capacity is also a significant predictor of academic achievement (Peng et al., 2018). Furthermore, low WM capacity is comorbid with

learning disabilities such as dyslexia and attention-deficit hyperactivity disorder (ADHD; Westerberg, Hirvikoski, Forssberg, & Klingberg, 2004). It is thus reasonable to believe that if WM skills could be improved by training, the benefits would spread across many other cognitive and real-life skills.

Three mechanisms, which are not necessarily mutually exclusive, have been hypothesized to explain why WM training might induce generalized cognitive benefits. First, WM and fluid intelligence may share a common capacity constraint (Halford, Cowan, & Andrews, 2007); that is, performance on fluid intelligence tasks is constrained by the amount of information that can be handled by WM. If WM capacity were augmented, then one's fluid intelligence would be expected to improve (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). In turn, individuals with boosted fluid intelligence are expected to improve their real-life skills, such as academic achievement and job performance, of which general intelligence is a major predictor. The second explanation focuses on the role played by attentional processes in both WM and fluid intelligence tasks (Engle, 2018; Gray, Chabris, & Braver, 2003). Cognitively demanding activities such as WM training may foster people's attentional control, which is, once again, a predictor of other cognitive skills and of academic achievement (for a detailed review, see Strobach & Karbach, 2016). Finally, Taatgen (2013, 2016) has claimed that enhancement in domain-general cognitive skills may be a by-product of the acquisition of domain-specific skills. That is, training in a given task (e.g., the *n*-back task) may enable individuals to acquire not only domain-specific skills (i.e., how to correctly perform the trained task) but also elements of more abstract production rules. These elements are assumed to be small enough not to encompass any domain-specific content and, therefore, can be transferred across different cognitive tasks.

Typically developing (TD) children engaging in WM training represent an ideal group on which to test these hypothesized mechanisms, for several reasons. Most obviously, the population of TD children is larger than the population of children with learning disabilities, who suffer from different disorders (e.g., ADHD, dyslexia, and language impairment). Moreover, the distribution of WM skills in TD children encompasses a larger range (which reduces the biases related to range restriction), and it is more homogeneous across studies. The latter features make studies involving TD children easier to meta-analyze than studies including patients with different learning disabilities. The results concerning TD children are thus more generalizable than those obtained from more specific populations. Also, unlike studies examining adult populations, studies involving TD children often include transfer measures of both cognitive skills (e.g., WM capacity and fluid intelligence) and academic achievement (e.g., mathematics and language skills). This feature allows us to directly test the hypothesis that WM training induces near-transfer and far-transfer effects that generalize into benefits in important

real-life skills. Finally, and probably most importantly, TD children represent a population in which cognitive skills are still developing and in which brain plasticity is at its peak. In other words, TD children are the most likely to benefit from cognitive-training interventions. Therefore, a null result in this group would cast serious doubts on the possibility to obtain generalized effects in other populations, as well (e.g., healthy adults).

The meta-analytic evidence

To date, scholars have disagreed about the effectiveness of WM training programs, and several meta-analytic reviews have been carried out to resolve this issue. The most recent and comprehensive ones—including studies on children, adults, and older adults—are Melby-Lervåg, Redick, and Hulme (2016; number of studies: $m = 87$) and Sala, Aksayli, Tatlidil, Tatsumi, et al. (2019b; $m = 119$). Both meta-analyses reached the conclusion that although WM training exerts a medium effect on memory-task performance (near transfer), no other cognitive or academic skills (far transfer) seem to be affected, regardless of the population examined; in particular, no effects have been observed when active controls are implemented, so as to rule out placebo effects (for a comprehensive list of meta-analyses about WM training, see Sala, Aksayli, Tatlidil, Gondo, & Gobet, 2019a).

Two meta-analyses have focused on children, with results similar to those described above. With TD children (ages 3 to 16), Sala and Gobet (2017) found a medium effect ($\overline{g} = 0.46$) with near transfer and a modest effect ($\overline{g} = 0.12$) with far transfer, with the qualification that the better the quality of the design (in terms of use of an active control group), the smaller the effect sizes. With children with learning disabilities, Sala, Aksayli, Tatlidil, Tatsumi, et al. (2019b) reanalyzed a subsample of the studies from Melby-Lervåg et al. (2016) and found effect sizes of $\overline{g} = 0.37$ for near transfer and $\overline{g} = 0.02$ for far transfer. Similar results were obtained with Cogmed, a commercial WM training program that has been subjected to a considerable amount of research, especially with children with learning disabilities (Aksayli, Sala, & Gobet, 2019).

Critique of the meta-analytic evidence

Some researchers have questioned the conclusions of meta-analytic syntheses concerning WM training. According to Pergher et al. (2019), the diversity of features in the training tasks (e.g., single vs. dual tasks) and the transfer tasks (e.g., numerical vs. verbal tasks) may make any meta-analytic synthesis on the topic essentially meaningless. Exact replications of studies have been rare (where there are any), and the moderators (independent variables in a

meta-regression) that should be added in order to account for all the differences across studies are too numerous to avoid power-related issues in meta-regression models. Therefore, it is not possible to reach strong conclusions from research into WM training. In simple words, this is nothing but the well-known apples-and-oranges argument against meta-analysis (Eysenck, 1994).

It is true that meta-analytic syntheses usually include just a few moderators examining only the most macroscopic study features. Nonetheless, meta-analysis also provides the tools to estimate the amount of variability across different findings in a particular field of research. The total variance observed in any dataset is the sum of sampling error variance and true variance. Sampling error variance is just noise, and therefore does not require any further explanation. By contrast, true variance, also referred to as *true heterogeneity*, is supposed to be accounted for by one or more moderating variables (Schmidt, 2010). In a meta-analysis, it is possible to estimate both within-study and between-study true heterogeneity in order to evaluate whether specific moderating variables are affecting the effect sizes at the level of the single study (e.g., different outcome measures) or across studies (e.g., different types of training or populations involved). Simply put, although it is nearly impossible to test every single potential moderator, it is easy to estimate how big the impact of unknown moderators is on the overall results.

Interestingly, several meta-analyses have estimated within- and between-study true heterogeneity in WM training to be null or low, for both near-transfer and far-transfer effects. When it is present at all, true heterogeneity is accounted for by the type of control group used (active or nonactive), by statistical artifacts such as pre–posttest regression to the mean, due to baseline differences between the experimental and control groups, and, to a lesser extent, by a few extreme effect sizes. This is the case with meta-analyses on younger and older adults (Sala, Aksayli, Tatlidil, Gondo, & Gobet, 2019a) and children with learning disabilities (Aksayli et al., 2019; Melby-Lervåg et al., 2016; Sala, Aksayli, Tatlidil, Tatsumi, et al., 2019b). In brief, despite the many design-related differences across WM training studies, consideration of true heterogeneity has indicated that there are no real differences between the effects produced by such diverse training programs.

The present study

The first aim of the present study was to update the previous meta-analytic synthesis about WM training in TD children (Sala & Gobet, 2017), which included studies only until 2016. Because considerable efforts have been devoted to this field of research, it is important to update this study in order to establish whether the same conclusions obtain. The second aim was to test, with a population of TD children, Pergher et al.'s (2019) claim that the broad variety of features of the training and transfer tasks used in WM training research has led to differential outcomes. Specifically, they hypothesized that some features encourage transfer, while others do not. Thus, resolving Pergher et al.'s claim is tantamount to predicting within-study and between-study true heterogeneity. To estimate both within-study and between-study true heterogeneity, we used multilevel modeling, and more especially robust variance estimation with hierarchical weights (Hedges, Tipton, & Johnson, 2010; Tanner-Smith, Tipton, & Polanin, 2016).

More specifically, we here tested the following study features. First, we examined the role played by the abovementioned design qualities (types of controls) and statistical artifacts (baseline differences and extreme effect sizes). As can be seen, these features have been found to be significant moderators in previous meta-analyses. Therefore, it will be worthwhile to test whether these findings can be replicated. Second, we checked whether transfer effects are influenced by the participants' age. Since WM capacity steadily develops throughout childhood, it is advisable to investigate whether WM training is more effective in TD children in a specific age range. Third, we checked whether such training is more effective for specific far-transfer outcome measures. Fourth, we tested whether the size of near-transfer effects is a function of transfer distance (i.e., the similarity between the training task and the outcome measures). Finally, we examined the effectiveness of different training programs. WM training tasks can be classified according to the type of primary manipulation required in order to perform the training tasks (e.g., Redick & Lindsey, 2013). In fact, whereas a number of WM training experiments have employed only one type of training task (e.g., *n*-back; Jaeggi, Buschkuehl, Jonides, & Shah, 2011), other scholars have suggested that including different kinds of WM tasks could maximize the chances to obtain transfer effects (Byrne, Gilbert, Kievit, & Holmes, 2019).

Method

Literature search

A systematic search strategy was employed to find relevant studies (PRISMA statement; Moher, Liberati, Tetzlaff, & Altman, 2009). The following Boolean string was used: ("working memory training" OR "WM training" OR "cognitive training"). We searched through the MEDLINE, PsycINFO, Science Direct, and ProQuest Dissertation & Theses databases to identify all potentially relevant studies. We retrieved 3,080 records. Also, the references in earlier

meta-analytic and narrative reviews (Aksayli et al., 2019; Melby-Lervåg et al., 2016; Sala, Aksayli, Tatlidil, Tatsumi, et al., 2019b; Sala & Gobet, 2017; Simons et al., 2016) were searched through.

Inclusion criteria

The studies were included according to the following seven criteria:

1. The study included children (maximum mean age = 16 years old) not diagnosed with any learning disability or clinical condition;
2. The study included a WM training condition;
3. The study included at least one control group not engaged in any adaptive WM-training program;
4. At least one objective cognitive/academic task was administered. Self-reported measures were excluded. Also, when the active control group was trained in activities closely related to one of the outcome measures (e.g., controls involved in a reading course), the relevant effect sizes were excluded (e.g., tests of reading comprehension);
5. The study implemented a pre–posttest design;
6. The participants were not self-selected;
7. The data were sufficient to compute an effect size.

We searched for eligible published and unpublished articles through July 21, 2019. When the necessary data to calculate the effect sizes were not reported in the original publications, we contacted the researchers by e-mail ($n = 3$). We received one positive reply. In total, we found 41 studies, conducted from 2007 to 2019, that met all the inclusion criteria (see Appendix A in the supplemental materials). These studies included 393 effect sizes and a total of 2,375 participants. The previous most comprehensive meta-analysis concerning WM training in TD children had included 25 studies (conducted between 2007 and 2016), 134 effect sizes, and 1,601 participants (Sala & Gobet, 2017). The present meta-analysis, therefore, adds a significant amount of new data. The procedure is described in Fig. 1.

Meta-analytic models

Each effect size was considered either *near-transfer* or *far-transfer*. The near-transfer effect sizes consisted of memory tasks referring to the *Gsm* construct, as defined by the Cattell–Horn–Carroll model (CHC model; McGrew, 2009). Far-transfer effect sizes referred to all the other cognitive measures. The two authors coded each effect size independently and reached 100% agreement.

Moderators

We evaluated four potential moderators for all studies, based on previous meta-analyses, as well as one moderator apiece that applied only to the far- or to near-transfer models:

1. *Baseline difference* (continuous variable): The corrected standardized mean difference (i.e., Hedges's g) between the experimental and control groups at pretest. This moderator was included to assess the amount of true heterogeneity accounted for by regression to the mean.
2. *Control group* (active or nonactive; dichotomous variable): Whether the WM training group was compared to another cognitively demanding activity (e.g., nonadaptive training); no-contact groups and business-as-usual groups were considered "nonactive." Also, in line with Simons et al.'s (2016) criteria, those control groups involved in activities that were not cognitively demanding were labeled as "nonactive." The interrater agreement was 98%; here and elsewhere, the two raters resolved every discrepancy by discussion.
3. *Age* (continuous variable): The mean age of the participants. A few primary studies did not provide the participants' mean age. In these cases, the participants' mean age was extracted from the median (when the range was reported) or the school grade.
4. *Type of training task* (categorical variable): The type of training task used in the study. This moderator included *updating* tasks (n-back tasks and running tasks; Gathercole, Dunning, Holmes, & Norris, 2019); *span* tasks (e.g., reverse digit span task, Corsi task, odd one out, etc.; Shipstead, Hicks, & Engle, 2012a); and a mix of updating and span tasks (labeled as *mixed*). A few training tasks did not fall into any of these categories and were labeled as *others*. Cohen's kappa was $\kappa = 1.00$.
5. *Outcome measure* (categorical variable): This moderator, which was analyzed only in the far-transfer models, included measures of fluid intelligence (*Gf*; McGrew, 2009), processing speed (*Gs*), mathematical ability, and language ability. The authors coded each effect size for moderator variables independently. Cohen's kappa was $\kappa = .98$.
6. *Type of near transfer* (categorical variable): Whether the task was the same as or similar to the WM training tasks (*nearest transfer*)—that is, referred to the same narrow memory skill—or was a different memory task (*less near transfer*)—that is, referred to different skills in the same broad construct (i.e., *Gsm*; McGrew, 2009). This categorization was the same as that proposed by Noack, Lövdén, Schmiedek, and Lindenberger (2009). This moderator was added only in the near-transfer
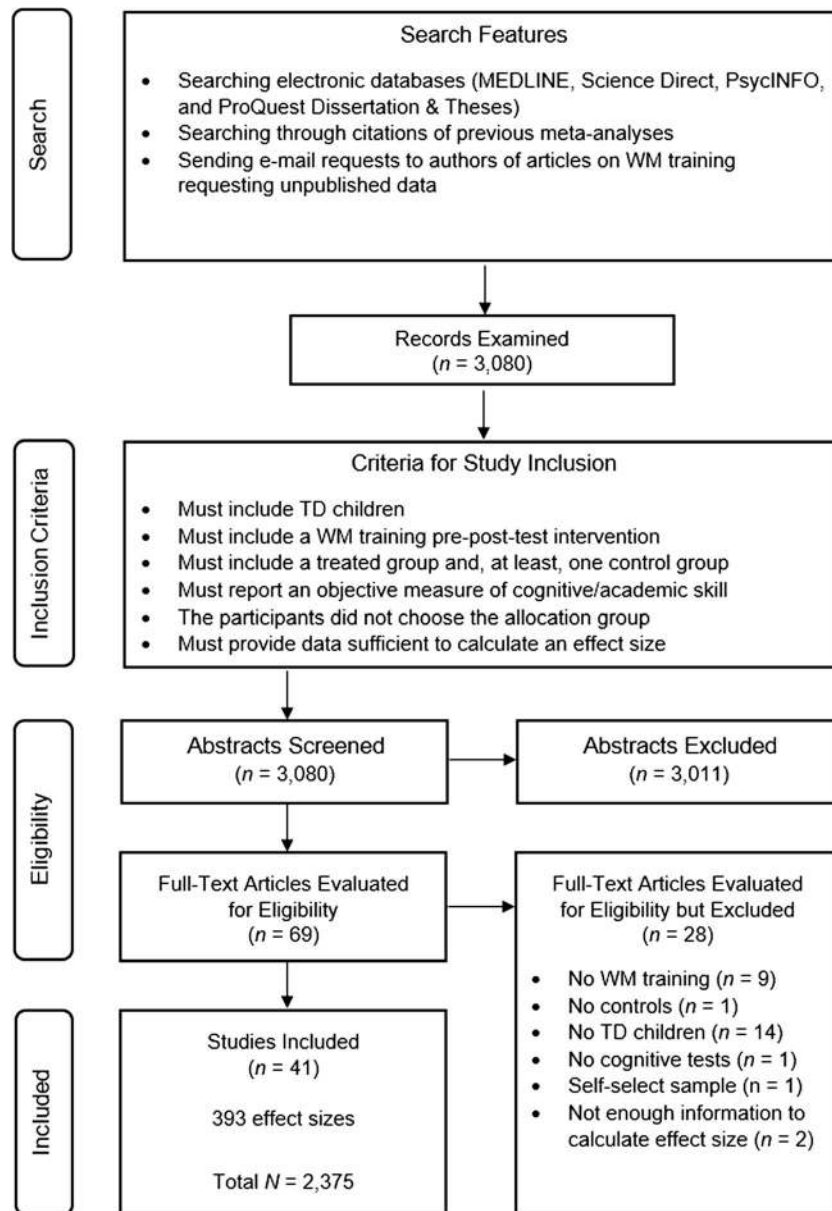
Fig. 1 Flow diagram of the search strategy. TD = typically developing; WM = working memory.

models. The authors coded each effect size for modera-
tor variables independently, and the interrater agree-
ment was 97%.

Effect size calculation

The effect sizes were calculated for each comparison in the
primary studies that met the inclusion criteria. Redundant
comparisons (e.g., rate of correct responses and incorrect re-
sponses) were excluded.

The effect size (Hedges's g) was calculated with the fol-
lowing formula:

$$g = \frac{(M_{e\_post} - M_{e\_pre}) - (M_{c\_post} - M_{c\_pre})}{SD_{pooled_{pre}}}$$

$$\times \left(1 - \frac{3}{(4 \times N) - 9}\right) \qquad (1)$$

where M_{e_post} and M_{e_pre} are the mean performance of the
experimental group at posttest and pretest, respectively,
M_{c_post} and M_{c_pre} are the mean performance of the control
group at posttest and pretest, respectively, SD_{pooled_pre} is the
pooled pretest SDs in the experimental group and the control
group, and N is the total sample size.

The formula used to calculate the sampling error variances was

$$Var_g = \left(\frac{N_e-1}{N_e-3} \times \left(\frac{2 \times (1-r)}{r_{xx}} + \frac{d_e^2}{2} \times \frac{N_e}{N_e-1} \right) \times \frac{1}{N_e} + \frac{N_c-1}{N_c-3} \times \left(\frac{2 \times (1-r)}{r_{xx}} + \frac{d_c^2}{2} \times \frac{N_c}{N_c-1} \right) \times \frac{1}{N_c} \right) \times \left(1 - \frac{3}{(4 \times N)-9} \right)^2 \quad (2)$$

where r_{xx} is the test–retest reliability of the measure, N_e and N_c are the sizes of the experimental group and the control group, d_e and d_c are the within-group standardized mean differences of the experimental group and the control group, and r is the pre–posttest correlations of the experimental group and the control group, respectively (Schmidt & Hunter, 2015, pp. 343–355). The pre–posttest correlations and test–retest coefficients were rarely provided in the primary studies. Therefore, we assumed the reliability coefficient (r_{xx}) to be equal to the pre–posttest correlation (i.e., no treatment-by-subject interaction was postulated; Schmidt & Hunter, 2015, pp. 350–351), and we imposed the pre–posttest correlation to be $r_{xx} = r = .700$. (We replicated the analyses using other correlation values ranging between .500 and .800. No significant differences were observed.)

Some of the studies reported follow-up effects. In these cases, the effect sizes were calculated by replacing the posttest means in Formula 1 with the follow-up means in the two groups.

Modeling approach

Robust variance estimation (RVE) with hierarchical weights was used to perform the intercept and meta-regression models (Hedges et al., 2010; Tanner-Smith & Tipton, 2014; Tanner-Smith et al., 2016). RVE allowed us to model nested effect sizes (i.e., extracted from the same study). Importantly, we used RVE to estimate both within-cluster (ω^2) and between-cluster (τ^2) true heterogeneity—that is, the amount of heterogeneity that was not due to sampling error. The effect sizes extracted from one study were thus grouped into the same cluster. These analyses were performed with the Robumeta R package (Fisher, Tipton, & Zhipeng, 2017).

Sensitivity analysis

A set of additional analyses were run in order to test the robustness of the results. The Metafor R package (Viechtbauer, 2010) was used. We first merged all the statistically dependent effect sizes using Cheung and Chan's (2014; for more details, see Appendix B in the supplemental materials) weighted-sample-wise correction and ran a random-effect model. This analysis was implemented to check whether the results were

sensitive to the way the statistically dependent effect sizes were handled.

Second, we performed Viechtbauer and Cheung's (2010) influential case analysis. This analysis evaluated whether some effect sizes exerted an unusually strong influence on the model's parameters, such as the meta-analytic mean ($\bar{g}$) and amount of between-effect true heterogeneity (τ^2). The RVE models were then rerun without the detected influential effect sizes.

Third, we ran publication bias analyses. We removed those influential effect sizes that increased true heterogeneity in order to rule out heterogeneity-related biases in the publication-bias-corrected estimates (Schmidt & Hunter, 2015). We then merged all the statistically dependent effect sizes and ran a trim-and-fill analysis (Duval & Tweedie, 2000). Trim-and-fill analysis estimates whether some smaller-than-average effects have been systematically suppressed and calculates a corrected overall effect size. We used the *L0* and *R0* estimators described by Duval and Tweedie. Finally, we employed Vevea and Woods's (2005) selection method. This technique estimates the amount of publication bias by assigning to *p*-value ranges different weights. As was suggested by Pustejovsky and Rodgers (2019), the weights employed in the publication bias analysis were not a function of the effect sizes (for more details, see Appendix C in the supplemental materials).

Results

Descriptive statistics

The mean age of the samples included in the present meta-analysis was 8.63 years. The median age was 8.69, the first and third quartiles were 6.00 and 9.85, and the mean age range was 4.27–15.40. The mean baseline difference was 0.037, the median was 0.031, the first and third quartiles were − 0.183 and 0.216, and the range was − 0.912 to 1.274. The descriptive statistics of the categorical/dichotomous moderators are summarized in Tables 1 and 2.

Far transfer

In this section, we examine the effects of WM training on TD children's ability to perform non-memory-related cognitive

Table 1 Numbers of studies and posttest effect sizes, by categorical moderators

Categorical Moderators	No. of Studies	No. of Effect Sizes
Far-Transfer Outcome		
Fluid intelligence	14	27
Processing speed	8	37
Mathematics	19	38
Language	14	44
Type of Near Transfer		
Nearest	20	76
Less near	20	47
Control Group		
Nonactive	22	135
Active	23	134
Type of Training		
Updating	7	26
Span	24	163
Mixed	5	62
Other	4	18

and academic tasks. The tasks did not share any features with the trained tasks.

Immediate posttest The overall effect size of the RVE intercept model was $\overline{g} = 0.092$, $SE = 0.033$, 95% CI [0.021; 0.163], $m = 34$, $k = 146$, $df = 14.8$, $p = .015$, $\omega^2 = 0.000$, $\tau^2 = 0.000$. The random-effect (RE) model (with Cheung & Chan's, 2014, correction) yielded very similar estimates: $\overline{g} = 0.105$, $SE = 0.040$, $p = .013$, $\tau^2 = 0.005$ ($p = .291$). Baseline was a

Table 2 Numbers of studies and follow-up effect sizes, by categorical moderators

Categorical Moderators	No. of Studies	No. of Effect Sizes
Far-Transfer Outcome		
Fluid intelligence	4	6
Processing speed	5	22
Mathematics	8	17
Language	8	21
Type of Near Transfer		
Nearest	7	30
Less near	9	28
Control Group		
Nonactive	8	74
Active	8	50
Type of Training		
Updating	2	10
Span	10	63
Mixed	3	51
Other	0	0

statistically significant moderator ($b = -0.376$, $SE = 0.065$, $p < .001$), whereas age was not ($p = .117$). Regarding the categorical moderators, the control group was the only statistically significant moderator ($p = .030$). No significant differences were found across different outcome measures ($p = 1.000$ in all pairwise comparisons; Holm's correction) or type of training task (all $ps \geq .563$).

Analysis of the control group moderator Since the control group moderator was statistically significant, we performed the sensitivity analysis on the subsamples separately. When nonactive controls were used, the overall effect size was $\overline{g} = 0.139$, $SE = 0.045$, 95% CI [0.034; 0.243], $m = 21$, $k = 75$, $df = 8.2$, $p = .015$, $\omega^2 = 0.000$, $\tau^2 = 0.005$. The RE model yielded very similar results, $\overline{g} = 0.177$, $SE = 0.056$, $p = .005$, $\tau^2 = 0.012$ ($p = .176$). Five influential cases were found. Excluding these effects did not meaningfully affect the results, $\overline{g} = 0.150$, $SE = 0.050$, 95% CI [0.040; 0.261], $m = 20$, $k = 70$, $df = 9.9$, $p = .013$, $\omega^2 = 0.000$, $\tau^2 = 0.000$. The two influential cases inflating heterogeneity were excluded for the following analyses. The trim-and-fill analysis retrieved four missing studies with the *L0* estimator, and the corrected estimate was $\overline{g} = 0.116$, 95% CI [0.020; 0.211]. No missing study was retrieved with the *R0* estimator. Vevea and Woods's (2005) selection model calculated a similar estimate ($\overline{g} = 0.097$).

When active controls were used, the overall effect size was $\overline{g} = 0.032$, $SE = 0.049$, 95% CI [−0.073; 0.138], $m = 18$, $k = 71$, $df = 12.3$, $p = .517$, $\omega^2 = 0.000$, $\tau^2 = 0.000$. The RE model yielded very similar results, $\overline{g} = 0.001$, $SE = 0.055$, $p = .982$, $\tau^2 = 0.000$. One influential case was found. Excluding this effect did not meaningfully affect the results, $\overline{g} = 0.046$, $SE = 0.047$, 95% CI [−0.055; 0.148], $m = 17$, $k = 70$, $df = 12.0$, $p = .339$, $\omega^2 = 0.000$, $\tau^2 = 0.000$. No missing study was retrieved with either the *L0* or *R0* estimator. The selection model estimate was $\overline{g} = -0.002$.

Follow-up The overall effect size of the RVE intercept model was $\overline{g} = 0.006$, $SE = 0.022$, 95% CI [−0.048; 0.059], $m = 13$, $k = 66$, $df = 6.2$, $p = .809$, $\omega^2 = 0.002$, $\tau^2 = 0.000$. The RE model provided very similar estimates: $\overline{g} = 0.014$, $SE = 0.056$, $p = .809$, $\tau^2 = 0.000$. Due to the limited number of studies included in this model, no further analysis was conducted.

Near transfer

In this section, we examine the effects of WM training on TD children's ability to perform memory tasks.

Immediate posttest The RVE model included all the effect sizes related to near-transfer measures. The overall effect size was $\overline{g} = 0.389$, $SE = 0.056$, 95% CI [0.271; 0.507], $m = 29$, $k = 123$, $df = 18.8$, $p < .001$, $\omega^2 = 0.006$, $\tau^2 = 0.059$. The RE model yielded very similar estimates: $\overline{g} = 0.365$, $SE = 0.056$, $p <$

.001, $\tau^2 = 0.036$ ($p = .002$). The meta-regression showed that neither baseline nor age was a significant moderator ($p = .154$ and $p = .914$, respectively). The type of control group and type of training were not significant moderators, either ($p = .845$ and ps ≥ .477, respectively). By contrast, type of near transfer (i.e., nearest vs. less near) was a significant moderator ($p = .005$).

Type of near transfer Since the type of near transfer moderator was statistically significant, we performed the sensitivity analysis on these two subsamples separately. With regard to nearest-transfer effects, the meta-analytic mean was $\overline{g} = 0.468$, $SE = 0.072$, 95% CI [0.310; 0.626], $m = 20$, $k = 76$, $df = 11.9$, $p < .001$, $\omega^2 = 0.011$, $\tau^2 = 0.054$. The RE model yielded very similar results, $\overline{g} = 0.457$, $SE = 0.064$, $p < .001$, $\tau^2 = 0.022$ ($p = .090$). One influential case was found. Excluding this effect did not meaningfully affect the results, $\overline{g} = 0.451$, $SE = 0.071$, 95% CI [0.297; 0.605], $m = 20$, $k = 75$, $df = 11.8$, $p < .001$, $\omega^2 = 0.000$, $\tau^2 = 0.052$. Merging the effects after excluding the influential case lowered the between-study true heterogeneity to a nonsignificant amount ($\tau^2 = 0.015$, $p = .158$). The trim-and-fill analysis retrieved seven missing studies with the $L0$ and $R0$ estimators, and the corrected estimate was $\overline{g} = 0.356$, 95% CI [0.221; 0.492]. The selection model estimate was $\overline{g} = 0.391$.

The less-near-transfer overall effect size was $\overline{g} = 0.261$, $SE = 0.092$, 95% CI [0.060; 0.462], $m = 20$, $k = 47$, $df = 12.0$, $p = .015$, $\omega^2 = 0.000$, $\tau^2 = 0.051$. The RE model yielded similar results, $\overline{g} = 0.292$, $SE = 0.070$, $p < .001$, $\tau^2 = 0.030$ ($p = .086$). One influential case was found. Excluding these effects did not meaningfully affect the results, $\overline{g} = 0.284$, $SE = 0.089$, 95% CI [0.090; 0.477], $m = 20$, $k = 46$, $df = 12.2$, $p = .008$, $\omega^2 = 0.000$, $\tau^2 = 0.039$. Excluding the influential effect and merging the statistically dependent effects lowered the between-study true heterogeneity to a nonsignificant amount ($\tau^2 = 0.010$, $p = .234$). No missing study was retrieved with either the $L0$ or $R0$ estimator. Finally, the selection model estimated some publication bias ($\overline{g} = 0.196$).

Follow-up The overall effect size of the RVE intercept model was $\overline{g} = 0.239$, $SE = 0.103$, 95% CI [– 0.012; 0.489], $m = 12$, $k = 58$, $df = 6.1$, $p = .059$, $\omega^2 = 0.000$, $\tau^2 = 0.045$. The results with the RE model were $\overline{g} = 0.276$, $SE = 0.084$, $p = .007$, $\tau^2 = 0.031$ ($p = .080$). Due to the limited number of studies included in this model, no further analysis was conducted.

Discussion

In this article we have analyzed the impact of WM training on TD children's cognitive skills and academic achievement. The findings were clear: whereas WM training fosters performance on memory tasks, small (with nonactive controls) to

null (with active controls) far-transfer effects are observed. Therefore, the impact of training on far-transfer measures does not go beyond placebo effects. The follow-up overall effects are consistent with this pattern of results. These results are also in line with Sala and Gobet (2017; a reanalysis with RVE of the data used in that study yielded similar results; for the details, see the supplemental materials) and, more broadly, with the conclusions of previous meta-analytic syntheses concerning WM training in the general population (Aksayli et al., 2019; Melby-Lervåg et al., 2016; Sala, Aksayli, Tatlidil, Tatsumi, et al., 2019b). The findings are summarized in Table 3.

The examination of true heterogeneity revealed that the meta-analytic models exhibit high internal consistency. No appreciable within-study true heterogeneity was observed ($\omega^2 \approx 0.000$ in all the models). This result supports the validity of Noack et al.'s (2009) taxonomy of transfer distance, which was used here. If near-transfer tasks had incorrectly been classified as far-transfer tasks (or vice versa), some within-study true heterogeneity would have been present. In addition, this result suggests that the memory tests (near transfer) used in the primary studies are correlated with each other and can be averaged by study to get more precise measures. Analogously, as we reported in the meta-regression analysis, there is no significant variability across diverse far-transfer measures. The important implication is that WM training fails to induce far transfer in every type of outcome measure (e.g., fluid intelligence, mathematics, etc.).

The models report some between-study true heterogeneity ($\tau^2 > 0.000$). Regarding far transfer, this heterogeneity is very low and is accounted for by the type of control group, baseline differences, and a few influential cases. The near-transfer models show slightly higher between-study true heterogeneity, which is partly explained by the type of near transfer (nearest vs. less near). The remaining true heterogeneity almost completely disappears when the statistically dependent (i.e., belonging to the same study) effects are averaged into more precise measures of memory skills. This corroborates the idea that most of the observed between-study heterogeneity is a statistical artifact related to measurement error in memory tasks. Otherwise, between-study true heterogeneity would occur even after averaging the effect sizes within the same study.

Finally, no significant amount of true heterogeneity appears to be accounted for by either the participants' mean age or the type of training task. The various training programs seem equally (in)effective in eliciting transfer effects. This outcome is in line with the findings of Melby-Lervåg et al. (2016) and corroborates the idea that transfer is a function of distance between the training task and the target task, rather than the features of the training program per se (e.g., Byrne et al., 2019; Pergher et al., 2019). Analogously, since age exerts no appreciable impact on the amount of transfer, we can conclude that

Table 3 Overall effects in the two meta-analyses, sorted by significant moderators

Model (1)	$\bar{g}$ (RVE) (2)	Adj. $\bar{g}$ (range) (3)	Heterogeneity (4)	Residual Heterogeneity (5)	RE τ^2 (6)
Present Meta-Analysis					
Far	0.092	–	$\omega^2 = 0.000$, $\tau^2 = 0.000$	$\omega^2 = 0.000$, $\tau^2 = 0.000$	$\tau^2 = 0.005$ (n.s.)
Nonactive	0.139	0.097–0.150	$\omega^2 = 0.000$, $\tau^2 = 0.005$	$\omega^2 = 0.000$, $\tau^2 = 0.000$	$\tau^2 = 0.000$ (n.s.)
Active	0.032	− 0.002–0.001	$\omega^2 = 0.000$, $\tau^2 = 0.000$	$\omega^2 = 0.000$, $\tau^2 = 0.000$	$\tau^2 = 0.000$ (n.s.)
Near	0.389	–	$\omega^2 = 0.006$, $\tau^2 = 0.059$	$\omega^2 = 0.001$, $\tau^2 = 0.058$	$\tau^2 = 0.036$
Nearest	0.468	0.356–0.391	$\omega^2 = 0.011$, $\tau^2 = 0.054$	$\omega^2 = 0.000$, $\tau^2 = 0.052$	$\tau^2 = 0.015$ (n.s.)
Less near	0.261	0.196–0.292	$\omega^2 = 0.000$, $\tau^2 = 0.051$	$\omega^2 = 0.000$, $\tau^2 = 0.039$	$\tau^2 = 0.010$ (n.s.)

(1) The meta-analytic model. (2) The overall robust variance estimation (RVE) effect size. (3) The range of the publication bias-adjusted estimates. (4) The amount of true heterogeneity of the model. (5) The true heterogeneity after excluding influential cases and running a meta-regression. (6) The random-effect between-study true heterogeneity after merging the statistically dependent effect sizes.

the stage of WM development in TD children does not play any role in making training programs more (or less) effective. That being said, it is worth noting that most of the primary studies investigated the effects of WM training in preschool and primary school TD children (see the Descriptive Statistics section). Only a fraction of the primary studies included adolescent samples, which makes our findings somewhat less generalizable to typically middle/high school students (e.g., 12–16 years of age).

Overall, Pergher et al.'s (2019) claim that the outcomes of WM training might be mediated by specific characteristics of the training and transfer tasks is not supported by our analyses: The estimated true heterogeneity, when present at all, was explained by a few moderators (distance of transfer and type of control group) and statistical artifacts (baseline differences and a few extreme effects). Therefore, searching for other potential moderators (e.g., duration of the intervention) seems pointless, and could even be perceived as a questionable research practice (i.e., capitalizing on sampling error; Schmidt & Hunter, 2015). In other words, even though, just as in pretty much any field of research in the behavioral sciences, there are a number of design-related differences across the primary studies (as was correctly observed by Pergher and colleagues), almost none of these differences exert any influence on the ability of WM training to induce near- or far-transfer effects. In fact, without quantitative evidence for within- and between-study true heterogeneity, appealing to generic differences across studies risks ending up being just a smokescreen behind which anybody can question the conclusions of meta-analytic syntheses and justify the need to carry out further research (Schmidt, 2017; Schmidt & Hunter, 2015).

Moreover, it is unlikely that WM training exerts positive far-transfer effects on subgroups of individuals (e.g., underachievers at baseline assessment; Jaeggi et al., 2011). Assuming so would necessarily lead to implausible conclusions. Since the meta-analytic far-transfer mean is null when placebo effects are ruled out, postulating nonartifactual

between-individual differences would imply that, whereas WM training enhances cognitive/academic skills in some children (positive effect), other individuals have their skills damaged by the training (negative effect). However, there is no theoretical reason nor any empirical evidence to believe that WM training exerts a detrimental effect on one's cognition. Instead, the reported between-study and between-individual differences are simply statistical fluctuations (e.g., sampling error and regression to the mean).

Therefore, given the circumstances, it is possible to apply Occam's razor (Schmidt, 2010), and conclude that WM training does not produce any generalized (far-transfer) effect in TD children. Furthermore, because the same pattern of results has been found in adults, older adults, and children with learning disabilities (Aksayli et al., 2019; Melby-Lervåg et al., 2016; Sala, Aksayli, Tatlidil, Tatsumi, et al., 2019b), the most parsimonious and plausible conclusion is that WM training does not lead to far transfer. Thus, on the basis of the available scientific evidence, the rational decision should be to redirect research efforts and resources to other means of fostering cognitive and academic skills, most likely using domain-specific methods (Gobet, 2016; Gobet & Simon, 1996).

Practical and theoretical implications

The practical implications of our results are the most obvious ones to highlight. Given the absence of appreciable far-transfer effects, especially in those studies implementing active controls, WM training should not be recommended as an educational tool. Although there seems to be no reason to believe that WM training negatively affects children's cognitive skills or academic achievement, implementing such programs would represent a waste of financial and time resources.

Given that positive effects were observed in our meta-analyses with respect to near transfer, one might nonetheless wonder whether WM training is worth the effort. In our opinion, it is not. First, nearest-transfer effects do not constitute

robust evidence for cognitive enhancement. Rather, they are clearly a measure of children's boosted ability to perform the training task or one of its variants. This fact reflects the well-known psychometric principle according to which cognitive tests are not reliable proxies for the cognitive constructs of interest if the participant has the opportunity to carry out the task multiple times. Second, less-near-transfer effects are not evidence of improved domain-general memory skills either. As was noted by Shipstead, Redick, and Engle (2012b), even though some less-near-transfer memory tasks (e.g., odd-one-out task) are not part of the training programs, they still share some overlap with some training tasks (e.g., simple-span tasks). Simply put, individuals engaging in WM training do not expand their WM capacity. Rather, they most likely acquire the ability to perform some memory tasks somewhat better than controls, which explains the small effect sizes reported in less-near-transfer measures, and the absence of far transfer.

Two main theoretical implications stem from our findings. First, on the behavioral level, we observe that the amount of transfer is a function of the similarity between the training task and the outcome task. This pattern of results has been replicated in many different domains and appears to be a constant in human cognition (for a review, see Sala & Gobet, 2019). Second, and most important, our findings support recent empirical evidence showing that WM and fluid intelligence do not share the same neural mechanisms, as was previously hypothesized (e.g., Halford et al., 2007; Jaeggi et al., 2008; Strobach & Karbach, 2016; Taatgen, 2013, 2016). Brain-imaging data suggest that WM performance is associated with increased network segregation, whereas the opposite pattern occurs when participants are asked to solve fluid intelligence tasks (Lebedev, Nilsson, & Lövdén, 2018). In the same vein, Burgoyne, Hambrick, and Altman (2019) have recently failed to find any evidence of a causal link between WM capacity and fluid intelligence. In fact, this study shows that the correlation between performance in WM tasks and fluid intelligence tasks is not a function of the capacity demands of the items of fluid intelligence tasks. This finding is in direct contradiction to the predictions of the common-capacity-constraint hypothesis. Thus, WM and fluid intelligence do not appear isomorphic, or even causally related, which would explain why WM training fails to induce any far-transfer effect, despite the well-known correlation between measures of WM capacity, fluid intelligence, and academic achievement.

Pessimism about the possibility to stimulate cognitive enhancement through WM training has thus been upheld by a robust corpus of evidence that goes beyond our meta-analytic results. Such convergent findings at different levels of empirical evidence (experimental, correlational, and neural) provide a successful example of triangulation that does not leave much room for further debate (Campbell & Fiske, 1959; Munafò & Smith, 2018). Indeed, it is our conviction that the data collected so far should lead researchers involved in WM training to entirely reconsider the theoretical bases of the field, or even to dismiss this branch of research.

Conclusions

In this meta-analysis we examined the impact of WM training on TD children's performance on cognitive and academic tasks, using a multilevel approach. The results significantly extend and corroborate the conclusions reached in a previous meta-analysis (Sala & Gobet, 2017): First, training programs exert an appreciable effect on memory task performance. The size of this effect is a function of the similarity between the training task and the outcome task. By contrast, small to null effects are found on far-transfer measures (i.e., fluid intelligence, attention, language, and mathematics). The magnitude of these effects equals zero in studies implementing active controls, suggesting that the small benefits reported in some studies have been the product of placebo effects. Finally, the meta-analytic models exhibit a low to null amount of true heterogeneity that is entirely explained by transfer distance, type of control group, baseline between-group differences, and a few extreme effect sizes. The lack of residual true heterogeneity means that there is no variance left to explain and implies that systematically comparing the features of training tasks and far-transfer outcome measures in order to identify successful WM training regimens, as was suggested by Pergher et al. (2019), is bound to fail.

Data availability statement The data supporting the findings of this study are openly available at the Open Science Foundation, at doi:10.17605/OSF.IO/BW8PG.

References

Aksayli, N. D., Sala, G., & Gobet, F. (2019). The cognitive and academic benefits of Cogmed: A meta-analysis. *Educational Research Review*, *29*, 229–243. doi:https://doi.org/10.1016/j.edurev.2019.04.003

Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559. doi:https://doi.org/10.1126/science.1736359

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. doi:https://doi.org/10.1016/S1364-6613(00)01538-2

Burgoyne, A. P., Hambrick, D. Z., & Altman, E. M. (2019). Is working memory capacity a causal factor in fluid intelligence? *Psychonomic Bulletin & Review*, *26*, 1333–1339. doi:https://doi.org/10.3758/s13423-019-01606-9

Byrne, E. M., Gilbert, R. A., Kievit, R., & Holmes, J., (2019, April 16). Evidence for separate backward recall and *n*-back working memory factors: A large-scale latent variable analysis. doi:10.31234/osf.io/bkja7

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

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

Cowan, N. (2016). Working memory maturation: Can we get at the essence of cognitive growth? *Perspective on Psychological Science*, *11*, 239–264. doi:https://doi.org/10.1177/1745691615621279

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*, 1158–1170. doi:https://doi.org/10.3758/s13423-016-1191-6

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466. doi:https://doi.org/10.1016/S0022-5371(80)90312-6

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

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

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*, 190–193. doi:https://doi.org/10.1177/1745691617720478

Eysenck, H. J. (1994). Systematic reviews: Meta-analysis and its problems. *BMJ*, *309*, 789. doi:https://doi.org/10.1136/bmj.309.6957.789

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). Package "robumeta." Retrieved from https://cran.r-project.org/web/packages/robumeta/robumeta.pdf

Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. (2019). Working memory training involves learning new skills. *Journal of Memory and Language*, *105*, 19–42. doi:https://doi.org/10.1016/j.jml.2018.10.003

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*, 177–190. doi:https://doi.org/10.1037/0012-1649.40.2.177

Gobet, F. (2016). Understanding expertise: A multi-disciplinary approach. London, UK: Palgrave/Macmillan.

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

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*, 316–322. doi:https://doi.org/10.1038/nn1014

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, *11*, 236–242. doi:https://doi.org/10.1016/j.tics.2007.04.001

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

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*, *105*, 6829–6833. doi:https://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, *108*, 10081–10086. doi:https://doi.org/10.1073/pnas.1103228108

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science*, *10*, 853–873. doi:https://doi.org/10.1111/j.1467-7687.2007.00638.x

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 66–71. doi:https://doi.org/10.1037/0033-2909.131.1.66

Lebedev, A. V., Nilsson, J., & Lövdén, M. (2018). Working memory and reasoning benefit from different modes of large-scale brain dynamics in healthy older adults. *Journal of Cognitive Neuroscience*, *30*, 1033–1046. doi:https://doi.org/10.1162/jocn_a_01260

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10. doi:https://doi.org/10.1016/j.intell.2008.08.004

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

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264–269. doi:https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Munafò, M. R., & Smith, G. D. (2018). Robust research needs many lines of evidence. *Nature*, *553*, 399–401. doi:https://doi.org/10.1038/d41586-018-01023-3

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. *Restorative Neurology and Neuroscience*, *27*, 435–453. doi:https://doi.org/10.3233/RNN-2009-0496

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., . . . Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*, 48–76. doi:https://doi.org/10.1037/bul0000124

Pergher, V., Shalchy, M. A., Pahor, A., Van Hulle, M. M, Jaeggi, S. M., & Seitz, A. R. (2019). Divergent research methods limit understanding of working memory training. *Journal of Cognitive Enhancement*. Advance online publication. doi:https://doi.org/10.1007/s41465-019-00134-7

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods*, *10*, 57–71. doi:https://doi.org/10.1002/jrsm.1332

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and *n*-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *20*, 1102–1113. doi:https://doi.org/10.3758/s13423-013-0453-9

Sala, G., Aksayli, N. D., Tatlidil, K. S., Gondo, Y., & Gobet, F. (2019a). Working memory training does not enhance older adults' cognitive skills: A comprehensive meta-analysis. *Intelligence,* *77*, 101386. doi:https://doi.org/10.1016/j.intell.2019.101386.

Sala, G., Aksayli, N. D., Tatlidil, K. S., Tatsumi, T., Gondo, Y., & Gobet, F. (2019b). Near and far transfer in cognitive training: A second-order meta-analysis. *Collabra: Psychology*, *5*, 18. doi:https://doi.org/10.1525/collabra.203

Sala, G., & Gobet, F. (2017). 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. (2019). Cognitive training does not enhance general cognition. *Trends in Cognitive Sciences*, *23*, 9–20. doi:https://doi.org/10.1016/j.tics.2018.10.004

Schmidt, F. L. (2010). Detecting and correcting the lies that data tell. *Perspectives on Psychological Science*, *5*, 233–242. doi:https://doi.org/10.1177/1745691610369339

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.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012a). Cogmed working memory training: Does the evidence support the claims? *Journal of Applied Research in Memory and Cognition*, *1*, 185–193. doi:https://doi.org/10.1016/j.jarmac.2012.06.003

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012b). 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

Strobach, T., & Karbach, J. (Eds.). (2016). Cognitive training: An overview of features and applications. New York, NY: Springer.

Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*, 261–288. doi:https://doi.org/10.1016/S0160-2896(01)00100-3

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). Cham, Switzerland: Springer.

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

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*, 85–112. doi:https://doi.org/10.1007/s40865-016-0026-5

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

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

Wai, J., Brown, M. I., & Chabris, C. F. (2018). Using standardized test scores to include general cognitive ability in education research and policy. *Journal of Intelligence*, *6*, 37. doi:https://doi.org/10.3390/jintelligence6030037

Westerberg, H., Hirvikoski, T., Forssberg, H., & Klingberg, T. (2004). Visuo-spatial working memory span: A sensitive measure of cognitive deficits in children with ADHD. *Child Neuropsychology*, *10*, 155–161. doi:https://doi.org/10.1080/09297040490911014