Please Don't Stop the Music: A Meta-Analysis of the Benefits of Learning to Play an Instrument on Cognitive and Academic Skills

Rafael Román-Caballero[1], Miguel A. Vadillo[2], Laurel Trainor[3], and Juan Lupiáñez[1]

[1] Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain; Department of Experimental Psychology, University of Granada, Granada, Spain

[2] Department of Basic Psychology, Autonomous University of Madrid, Madrid, Spain

[3] Department of Psychology, Neuroscience, and Behaviour, McMaster University, Hamilton, Canada; McMaster Institute for Music and the Mind, McMaster University, Hamilton, Canada; Rotman Research Institute, Baycrest Hospital, Toronto, Canada

Author Note

Rafael Román-Caballero https://orcid.org/0000-0003-0943-6217

Miguel A. Vadillo https://orcid.org/0000-0001-8421-816X

Laurel Trainor https://orcid.org/0000-0003-3397-2079

Juan Lupiáñez https://orcid.org/0000-0001-6157-9894

All the data and R script for the analyses are fully available in https://osf.io/9y5tp/. The authors declare no conflict of interest.

We would like to thank Prof. Julio Sánchez-Meca for his valuable comments on an earlier version of this article. This work was supported by a predoctoral fellowship from the Spanish Ministry of Education, Culture and Sport to RRC (FPU17/02864), by a research grant from the Autonomous Community of Madrid to MAV (2016-T1/SOC-1395), and by two research project grants from the Spanish Ministry of Economy, Industry, and Competitiveness to JL (PSI2017-84926-P) and to MAV (PSI2017-85159-P).

rrarroca@ugr.es

Abstract

An extensive literature has investigated the impact of musical training on cognition and academic achievement in children and adolescents. However, most of the studies have relied on cross-sectional designs, which makes it impossible to elucidate whether the observed differences are a consequence of the engagement in musical activities. Previous meta-analyses with longitudinal studies have also found inconsistent results, possibly due to their reliance on vague definitions of musical training. In addition, more evidence has appeared in recent years. The current meta-analysis investigates the impact of early programs that involve learning to play musical instruments on cognitive and academic skills, as previous meta-analyses have not focused on this form of musical training. Following a systematic search, 34 independent samples of children and adolescents were included. All the studies had pre-post designs and, at least, one control group. Overall, we found a small but significant benefit ($\bar{g}_{\Delta}$ = 0.24) with short-term programs, regardless of whether they were randomized or not. In addition, a small advantage at baseline was observed in studies with self-selection ($\bar{g}_{pre}$ = 0.28), indicating that participants who had the opportunity to select the activity consistently showed a slightly superior performance prior to the beginning of the intervention. Our findings support a *nature and nurture* approach to the relationship between instrumental training and cognitive performance. Nevertheless, evidence from well-conducted studies is still scarce and more studies are necessary to reach firmer conclusions.

Keywords: musical training, academic achievement, cognitive skills, transfer

Highlights

- Benefits of musical training have been examined across disparate musical activities.
- Instrumental learning is ideal for investigating the causal benefits of musical training.
- Learning to play an instrument has a positive impact on cognitive and academic skills.
- Children who self-select musical training tend to have better performance at baseline.
- Cross-sectional results may reveal both preexisting and caused cognitive advantages.

Please Don't Stop the Music. A Meta-Analysis of the Benefits of Learning to Play an Instrument on Cognitive and Academic Skills

1. Introduction

The literature about the effects of musical training on cognitive and brain function is growing rapidly. Multiple studies have documented that involvement in musical activities enhances auditory and sensorimotor processes (James et al., 2020; Kraus et al., 2014; Slater et al., 2015; for a review, see Herholz & Zatorre, 2012). However, whether musical training impacts general cognitive abilities (e.g., memory or attention) and academic achievement is still debated. Playing an instrument is a complex task involving several perceptual modalities, sensorimotor integration, and higher-order cognitive processes. Moreover, structured instrumental learning is an effortful activity that needs to be maintained across long periods of time; it requires regular and motivated practice, learning of new and progressively more difficult material, and adapting to new contexts. Those characteristics have led some to propose that musical training is an optimal cognitive training activity (Bugos et al., 2007), and extensive evidence has associated musicianship with higher intelligence scores (Bugos & Mazuc, 2014; Schellenberg, 2006; Swaminathan et al., 2017) and advantages in cognitive functions, such as visuospatial abilities (Sluming et al., 2007), processing speed (Bugos & Mazuc, 2014; Jentzsch et al., 2014), executive control (Medina & Barraza, 2019; Jentzsch et al., 2014), attention and vigilance (Kaganovich et al., 2013; Rodrigues et al., 2013; Román-Caballero et al., 2020) and episodic and working memory (Talamini et al., 2017). Also, it might protect against the cognitive decline associated with aging (Román-Caballero et al., 2018). Unfortunately, most of the studies in the field are correlational, which does not allow establishing firm conclusions about the causal role of musical training in those advantages (Schellenberg, 2020).

A plausible alternative explanation for these results is that high-functioning children, with higher musical aptitude, higher socioeconomic status, and/or personality traits associated with cognitive improvements (e.g., openness to experience), are more likely to be interested in music and take music lessons (Corrigall et al., 2013; Swaminathan et al., 2017). Or perhaps individuals with better executive functions are more prepared to resist the temptation to abandon the continued effort that mastering an instrument entails. From this point of view, most of the cognitive and academic advantages observed in correlational studies and interventions without random assignment (where participants and their families chose musical activities) could be due to preexisting differences in children's intelligence, temperament, and environment. In addition, it has been argued that far transfer (i.e., the generalization of training in one domain to skills in a loosely related domain) rarely occurs with most types of cognitive training, because of the small overlap between domain-specific and domain-general skills (Melby-Lervåg et al., 2016; Sala & Gobet, 2019). Extending this logic, it would be unlikely that musical training could enhance general cognitive abilities.

Other theoretical proposals have tried to reconcile both positions, arguing that expert musicians might have preexisting advantages (cognitive, personality, and/or musical aptitudes) that would promote the acquisition of musical skills and motivation to practice, while at the same time this long-term engagement would also result in multiple neural and cognitive changes (e.g., *nature and nurture hypothesis*, Wan & Schlaug, 2010). In this vein, the difference in magnitude between the effects observed in correlational studies (often Cohen's *d* around 0.8–1.0) and in experimental designs with random allocation (≈ 0.2; Corrigall et al., 2013; for a classic example, see Schellenberg, 2004) might be the consequence of musicians' in correlational studies benefitting from both preexisting cognitive advantages *and* musical training itself. Only a small number of experimental studies comply with basic methodological standards, such as randomization, the inclusion of an active control group, and blinding of the assessment, and, in practice, most of them involve short interventions (1–1.5 years long) and relatively small samples (≈ 25 participants per group), with the subsequent lack of statistical power to detect small-to-medium effect sizes[1]. Under those conditions, it is perhaps unsurprising that the results have been inconsistent across studies, with some studies providing evidence of a positive impact of instrumental learning (Frischen et al., 2021; James et al., 2020; Schellenberg, 2004), while other studies have shown null effects (D'Souza & Wiseheart, 2018; Haywood et al., 2015).

This is an ideal context for the application of a meta-analysis, as it allows a quantitative review of the literature and enables drawing firmer conclusions given an increased statistical power. Also, meta-analysis offers numerical estimators of the summary effect and between-studies consistency, which provides the opportunity to assess the relevance of interventions (and not only their statistical significance) and to identify potential moderators. Unfortunately, even at the meta-analytic level, there are inconsistent results concerning the impact of musical training in experimental and quasi-experimental studies (Butzlaff, 2000; Cooper, 2020; Gordon et al., 2015; Hetland, 2000; Sala & Gobet, 2017, 2020; Standley, 2008; Vaughn, 2000). Probably, one of the greatest sources of variability is the vague and inconsistent definition of musical training across meta-analyses, which usually combine highly heterogeneous musical interventions, including instrumental tuition, programs of music education such as Kindermusik, Orff, or Kodály methods, computerized training of musical skills, phonological training with music support, and listening programs, among others.

Arguably, studies examining the effects of formal programs in instrumental training are ideal for investigating the causal role of musical training on cognitive and academic skills. Most correlational studies reporting effects of musical training have compared expert

[1] A power analysis using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, Georg, 2009) for a one-tailed *t*-test and an alpha of .05 indicated that around 310 participants per group would be necessary to achieve an acceptable power of .80 with a Cohen's *d* of 0.20 (small effect), and 51 participants per group for a *d* of 0.50 (medium). Required sample sizes are larger when two-tailed contrast statistics or higher power values are used.

instrumentalists with non-musicians, suggesting that instrumental programs might be advantageous. Formal programs in which the participants learn to play a complex musical instrument and to read music notation are the most similar to the type of training that such expert musicians follow[2]. Additionally, although all types of musical training aim to promote music skills (e.g., rhythm, pitch and timbre discrimination, singing, basic music notation, etc.), learning to play an instrument seems to pose greater cognitive demands than other musical activities, as it requires particularly intensive practice entailing hand dexterity, bimanual coordination, and core cognitive functions such as working memory and attention. For that reason, far transfer might be more probable with instrumental learning. Although some studies have reported cognitive improvements with non-instrumental interventions (for Kindermusik, Orff, Kodály or related methods, see Kaviani et al., 2014; Patscheke et al., 2016; for listening programs, see Bugos, 2010; Hole, 2013), there is evidence of greater benefits with instrumental programs, to such an extent that non-instrumental music programs have even been used as control conditions in some studies (see Bugos, 2010; James et al., 2020). Nevertheless, to the best of our knowledge, the impact of instrumental interventions has not been investigated separately in any previous meta-analysis, nor has it been tested as a moderator.

On the other hand, the most recent and comprehensive meta-analysis (i.e., Sala & Gobet, 2020), which included different musical interventions, found a positive small effect of

---

[2] Structured singing training, such as that received by lyrical singers, is also comparable to the training of expert musicians. Nevertheless, in the literature, is it often difficult to distinguish between formal singing programs (intensive in terms of technique, music theory, and out-of-class practice) and interventions directed at a more diverse population with more informal approaches. It is also the case that singing interventions build on a capacity for singing already present in individuals without training, whereas learning an instrument entails learning completely new skills. Some studies have found smaller effects for vocal training in comparison to instrumental training (Guhn et al., 2020; Kinney, 2008; for a null difference, see Schellenberg, 2004). Although the comparison of formal instrumental and vocal training remains an open question that needs confirmation from studies with experimental designs, the most notable evidence to date is from the study of Guhn et al., who showed advantages for both instrumental and vocal musical training in a remarkably large sample of students ($N \approx 110{,}000$) who chose to take part in music courses or not. This result held even after controlling for several confounding variables (cultural background, SES, sex, and prior academic achievement). Crucially, instrumental learning led to larger differences in comparison to vocal training (*d*s ranging from 0.12 to 0.31), a result that the authors attributed, among other factors, to the complexity involved in learning to play an instrument. They suggested that this complexity might have a particularly positive impact on executive functions and, through them, on other cognitive domains.

Because for much of this literature it is very difficult to determine whether studies used formal or informal vocal training, and considering the evidence from Guhn et al. that cognitive benefits for instrumental training are likely larger than for vocal training, we decided to constrain the scope of our meta-analysis to studies with formal learning of musical instruments.

musical training ($\bar{g}$ = 0.18, p < .001) that was reduced to null when characteristics of design quality (i.e., random allocation and active control) were taken into account ($\bar{g} \approx 0$). However, the difficulty of implementing methodologically rigorous designs adds to the inherent cost of instrumental interventions that require highly specialized material and professionals. This might explain why only a third of the studies included in Sala and Gobet's meta-analysis had instrumental programs (19 out of 54) and why many studies with instrumental programs have not used optimal experimental designs. Indeed, studies involving instrumental training were underrepresented among Sala and Gobet's studies with random assignment and/or with an active control group. More precisely, only 27% of the randomized studies (6 out of 22), 36% of those using an active control group (9 out of 25), and 31% of those using both randomization and an active control group (4 out of 13), had instrumental training. Given that non-instrumental interventions likely have a smaller impact on cognitive and academic performance compared to instrumental ones, the greater representation of the former in Sala and Gobet's study may have led to conclusions mostly related to non-instrumental musical training. Thus, despite all the previous meta-analyses, the overall impact of formal instrumental learning remains uninvestigated. In addition, some outcomes included in the meta-analyses by Sala and Gobet (2017, 2020) were measures of skills trained with active control activities (e.g., linguistic abilities with drama training), and therefore should not be analyzed in a far-transfer meta-analysis (Bigand & Tillmann, 2021). Finally, new studies have appeared since the publication of the most recent meta-analysis (Sala & Gobet, 2020) and we additionally found some studies that have never been included in any previous meta-analysis, including some from unpublished doctoral theses (such as Nering, 2002; Pelletier, 1963).

Considering all the above, it seems crucial to carry out a new comprehensive meta-analysis that separately investigates the impact of instrumental learning programs on cognitive and academic skills. The present work aims to address this issue to shed light on the debate about the causal role of musical training in children and adolescents.

2. Method

2.1. Literature Search

A systematic search strategy was used following the recommendations of PRISMA (Moher et al., 2009). Firstly, we consulted PubMed, ProQuest, Scopus, Web of Science, and ProQuest Dissertation & Theses using the search syntax *"music*" AND ("training" OR "instruction" OR "educati*" OR "practice") AND ("child*" OR "adolescen*")*. Also, references from previous empirical studies, reviews, and meta-analyses on this subject were examined. The latest search was carried out in February 2021, without any time restriction. In total, 8560 potentially relevant results were found, among which 32 met the inclusion criteria described below and were included in our meta-analysis (**Figure 1**). These studies included 34 independent samples, 179 effect sizes, and a total of 5998 participants.

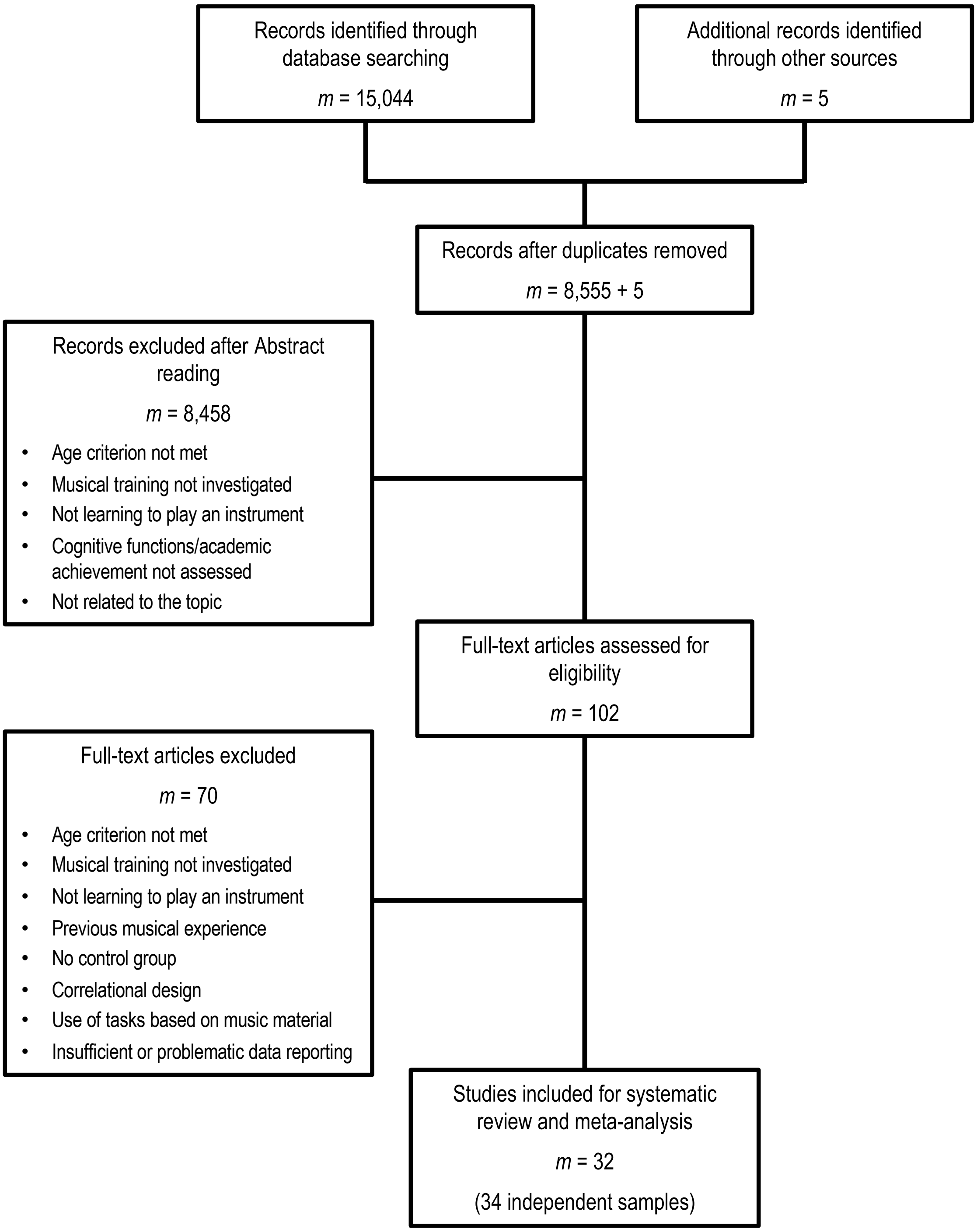

Figure 1. Flowchart of the studies included in the systematic review and meta-analysis.

2.2. Selection Criteria

The studies selected in the review had to meet the following criteria:

1. Published articles or theses that included musical training programs involving at least learning to play an instrument;
2. The design of the studies included pretest and posttest measures, regardless of whether there was a random assignment of children/adolescents to conditions or they themselves (or their parents or their teachers) selected the activity;
3. The studies included a comparison between a music-treated group and, at least, one control group (active or passive);
4. The participants had no previous formal musical training or instrumental learning prior to the program;
5. The studies contained sufficient information to calculate at least one effect size (otherwise, authors were contacted and the studies were included if the information was provided);
6. The studies included at least one non-musical measure of academic and/or cognitive skill (note that near-transfer effects were not included);
7. At the moment of starting the training, participants were between 3 and 16 years old;
8. The participants of the study did not suffer from neurological or psychiatric conditions.

As the included studies used different instruments to assess the outcomes (with different scales) from study to study, we used a standardized estimator of the effect size: Hedges' *g*. There are multiple ways of estimating Hedges' *g* in pre-posttest designs with two groups (see below, *Effect Size*), the most common being the standardized mean difference with posttest measures only, which we will refer to as g_{post}. An alternative index, proposed by Morris (2008), which we will refer to as g_{Δ}, is the standardized mean change difference (i.e., the difference between the two groups in the change of the outcomes between pretest and posttest moments). An advantage of this index over g_{post} is that it controls for preexisting differences at baseline. Collating both types of effect sizes, g_{post} and g_{Δ}, in a single meta-analysis requires making some assumptions. For instance, the variance of the pretest score is assumed to be equal to the variance of the posttest score. Similarly, both groups are assumed to be equivalent in baseline performance. However, it is arguable that these assumptions are not met in most circumstances. Previous research suggests that there are cognitive and personality differences in individuals who choose and continue with musical training as an activity (Corrigall et al., 2013). For that reason, and unlike previous reviews (Sala & Gobet, 2017, 2020; Vaugh, 2000), we constrained our review only to pre-posttest studies.

We included studies with random assigment (*randomized studies*), studies in which the children or their parents or their teachers selected the training group (*self-selection studies*), and studies with other allocation strategies that can not be consider random (such as quasi-randomization; *non-randomized studies*), as all of them can offer valuable information for the debate. On the one hand, randomized studies allow the establishment of more conclusive causal inferences about the effects of training, as randomization of the individuals reduces bias due to preexisting differences in cognitive, academic or musical skills, or other confounds (e.g., personality traits; Corrigall et al., 2013). On the other hand, studies that allowed the participants to choose the program has higher risk of selection effects, which might be observable in the overall difference in the pretest performance. Whereas we based the main inferences about causality, moderating variables and publication bias on randomized and non-randomized studies, the inclusion of self-selection studies was restricted to the assessment of baseline differences and the overall analysis for comparison purpose.

2.3. Statistical Analysis

2.3.1. Effect Size

We used the formula proposed by Morris (2008) for g_{Δ} as an estimator of the effect size in the main analyses,

$$g_{\Delta} = c_{\mathrm{p}} \times \left(\frac{\left(M_{\mathrm{post,\,T}} - M_{\mathrm{pre,\,T}}\right) - \left(M_{\mathrm{post,\,C}} - M_{\mathrm{pre,\,C}}\right)}{SD_{\mathrm{pooled,\,pre}}}\right), \tag{1}$$

where M_{pos} and M_{pre} represent the scores at pretest and posttest, respectively, for the treatment group (T) and the control group (C), and $SD_{\mathrm{pooled,\,pre}}$ is the pooled standard deviation for the pretest scores of both groups. Moreover, c_{p} is a correction factor of the small sample bias, given by

$$c_{\mathrm{p}} = 1 - \frac{3}{4 \times (N_{\mathrm{T}} + N_{\mathrm{C}}) - 9}, \tag{2}$$

where N_{T} and N_{C} are the number of participants in the treatment group and the control group. Positive values of g_{Δ} represent greater benefits in favor of treatment group, and negative values index the contrary. We multiplied by −1 those effects in which it was necessary to keep the mentioned direction. The g values were interpreted according to Cohen's criteria (Cohen, 1992): values close to 0.2, 0.5, and 0.8 or higher are interpreted as small, medium, and large effects, respectively. The variance of g_{Δ} was calculated following the formula by Morris (2008)

$$V_{g_{\Delta}} = 2 \times c_{\mathrm{p}}^2 \times (1 - r) \times \left(\frac{N_{\mathrm{T}} + N_{\mathrm{C}}}{N_{\mathrm{T}} \times N_{\mathrm{C}}}\right) \times \left(\frac{N_{\mathrm{T}} + N_{\mathrm{C}} - 2}{N_{\mathrm{T}} + N_{\mathrm{C}} - 4}\right) \times \left(1 + \frac{g_{\Delta}^2}{2 \times (1 - r) \times \left(\frac{N_{\mathrm{T}} + N_{\mathrm{C}}}{N_{\mathrm{T}} \times N_{\mathrm{C}}}\right)}\right) - g_{\Delta}^2\,, \tag{3}$$

where r is the correlation between pretest and posttest scores. We directly estimated r from raw data when they were available or used the following formula when other reported statistics made it possible:

$$r = \frac{{SD_{\mathrm{pre}}}^2 + {SD_{\mathrm{post}}}^2 - {SD_{\mathrm{Diff}}}^2}{2 \times SD_{\mathrm{pre}} \times SD_{\mathrm{post}}}, \tag{4}$$

$$V_r = \frac{\left(1 - r^2\right)^2}{N - 1}. \tag{5}$$

Using these equations, we could extract 75 correlation coefficients and their respective variances from 14 studies, with a meta-analytic mean r of .71 (**Supplementary Material**). This final value of r is close to .70 that Rosenthal (1991) proposed as a conservative assumption when pre-posttest correlations were not available. Considering that, we conducted our analyses assuming r = .70.

Furthermore, as previous literature pointed to the existence of baseline differences between individuals who chose to take musical training and individuals who did not (Corrigall et al., 2013; Swaminathan et al., 2017), we were also interested in comparing the performance of both groups just at baseline. For that purpose, we calculated the traditional Hedges' *g* only with pretest scores (called here g_{pre})

$$g_{\text{pre}} = c_{\text{p}} \times \left(\frac{M_{\text{pre, T}} - M_{\text{pre, C}}}{SD_{\text{pooled, pre}}}\right), \tag{6}$$

$$V_{g_{\text{pre}}} = c_{\text{p}}^{2} \times \left(\frac{N_{\text{T}} + N_{\text{C}}}{N_{\text{T}} \times N_{\text{C}}} + \frac{g_{\text{pre}}^{2}}{2 \times (N_{\text{T}} + N_{\text{C}})}\right). \tag{7}$$

2.3.2. Meta-Analysis, Heterogeneity and Moderator Analysis

As is often in psychology meta-analyses, most of the included studies contributed with more than one effect size from the same sample, which rendered the outcomes not independent. Most of the conventional meta-analytic procedures, however, assume independence between effect sizes. The robust variance estimation approach (RVE; Hedges et al., 2010) has been developed to deal with correlated structure of outcomes. This method estimates the correlation matrix and sets the weights according to a correlated or a hierarchical estructure. Simulation studies show that RVE is remarkably accurate in estimating the mean effect and the confidence interval, even with a small number of studies (m = 10) and when they include a large number of dependent estimates per study (k = 10; Hedges et al., 2010). We used the *robumeta* package for R (Fisher, Tipton, & Zhipeng, 2017) for implementation of RVE conducted in the main analyses (all the data and R script for the analyses are fully available in the **Supplementary Material**). We chose a correlated dependence model with small-sample corrections (Tipton, 2015).

First, we studied the overall impact of musical training, fitting an overall meta-analytic model with randomized and non-randomized studies, and then for each group of studies separately. For comparison, we repeated the analysis with self-selection studies (combined with the rest of studies and separately). The usual heterogeneity indexes, τ^2 and I^2, were computed. To identify studies with outlying outcomes, we fitted a multilevel model with the *rma.mv*() function of *metafor* (Viechtbauer, 2010) and estimated the Studentized residuals (> 2) and Cook's distance (> 4/*n*). For the analysis of differences at baseline, we fitted separate RVE models for randomized, non-randomized and self-selection studies using the g_{pre} as effect size estimate.

Then, we assessed the influence of the following moderating variables on effect sizes: (1) randomization (randomized vs. non-randomized studies, note that self-selection studies were not included in moderator analyses); (2) type of control group (active vs. passive); (3) whether there was blinding of assessors or the measure was computerized (yes/no); (4) age of the participants at the baseline (in years); (5) duration of the training program (in months); (6) between-groups baseline difference, measured as g_{pre}; (7) low socioeconomic status (SES) of the sample (yes vs. no/not reported); and (8) the type of cognitive or academic outcome (math, literacy, intelligence, processing speed, short-term memory, long-term memory, visuospatial abilities, phonological processing, and executive functions). Regarding the type of control, we conducted the analyses with the effects corresponding to the two comparisons (experimental vs. active control group, and experimental vs. passive control group) in those studies in which both were available in the same study.

2.3.3. Publication Bias

Several lines of evidence indicate that multiple factors of the reporting and the publication procedure can drastically affect the results of a meta-analysis. Studies reporting significant and large effect sizes are more likely to be published or made available than statistically non-significant results or results that contradict an accepted theory (Carter et al., 2019). This phenomenon (called *publication bias*) leads to studies with null or negative estimates being less accessible and underrepresented in meta-analyses. Several methods have been developed to detect publication bias and correct for its adverse consequences over the final effect.

One popular approach is the visual inspection of small-study effects in a funnel plot and the use of the trim-and-fill method to correct the final estimate. The funnel plot is a display of the individual effect sizes on the x-axis against the corresponding standard errors on the y-axis. An asymmetric distribution can be a sign of publication bias, with missing studies in non-significant regions of the plot. The trim-and-fill method (Duval & Tweedie, 2000) detects (and removes) studies causing funnel plot asymmetry and then imputes missing studies to estimate a bias-corrected effect size. Alternatively, the precision-effect test and the precision-effect estimate with standard error procedures (PET and PEESE; Stanley & Doucouliagos, 2014) are based on a meta-regression approach to test for selective reporting and adjust for small-study effects. Both methods use a measure of precision as a covariate in the meta-analytic model (the standard error of the effect size in the case of PET, and sampling variance for PEESE), where the significance of the regression coefficient tests for publication bias, and the intercept of the model is taken as the true underlying effect. Thirdly, selection models (Vevea & Hedges, 1995) assume that the probability of publication depends on the *p* value. In our meta-analysis we use a selection model with a single cut point at $p_{\text{one-tailed}} = .025$, which divides the range of possible *p* values into significant and non-significant values.

The previous methods assume independent effect sizes in their original formulation. A way to account for dependence is to combine the all the effect sizes coming from the same sample generating an average estimate for each study, and conduct the classic methods on these aggregated estimates (Rodgers & Pustejovsky, 2020). In addition, some recent approaches directly handle the issue of dependence. For instance, the logic of PET-PEESE and other regression-based methods can be extended to multilevel models and RVE (Fernández-Castilla et al., 2019; Friese et al., 2017; Rodgers & Pustejovsky, 2020). Mathur and VanderWeele (2020) also proposed a sensitivity analysis that can be fitted with RVE. Assuming that positive results are more likely to be published than null or negative results by an unknown ratio (η, which is > 1 under publication bias), it is posible to estimate how strong this ratio would need to be to make the final effect negligible. Values of 1.5 are frequent in psychology literature, whereas values over 5 are rare (the 95th quantile of the estimated selection ratios, Mathur & VanderWeele, 2020).

Simulation studies show that ignoring dependence results in inflated Type I error (Rodgers & Pustejovsky, 2020). Although the methods that handle correlated effect sizes exhibit better performance, none of them stands as superior in terms of performance. Their performance depends on many parameters, such as the number of studies, heterogeneity, the degree of publication bias, and so on (Carter et al., 2019; Rodgers & Pustejovsky, 2020). A reasonable strategy is to use in combination several of them, and interpret their results taking into account the conditions of the meta-analysis (Carter et al., 2019). In the present meta-analysis we chose four methods to test publication bias and adjust the mean estimate: (i) the trim-and-fill method (with the *L0* and *R0* estimators) and (ii) the selection model, both using aggregates, (iii) the RVE regression-based approaches (RVE PET and RVE PEESE), and (iv) the Mathur and VanderWeele's sensitivity analysis. We used the *MAd* packge in R (Del Re & Hoyt, 2014) to generate within-study aggregates, while we carried out the Vevea and Hedges' selection model (1995) with the *weightr* package (Coburn & Vevea, 2019) and the Mathur and VanderWeele's sensitivity analysis with the *PublicationBias* package (Mathur & VanderWeele, 2020). For the RVE meta-regression test, we chose a modified formula of the sampling variance and, in parallel, a variance-stabilizing transformation for the standardized mean difference to prevent the artifactual dependence between the effect size and its precision estimate (Pustejovsky & Rodgers, 2019; see **Appendix A**).

Regarding the conditions of the present meta-analysis, previous comprehensive meta-analyses of the literature (Sala & Gobet, 2017a, 2020) revealed moderate heterogeneity ($\tau \approx 0.2$), a small sample of studies using instrumental programs ($m \approx 20$), some evidence of publication bias, and a small uncorrected effect ($g \approx 0.20$). Under similar conditions, trim-and-fill, selection model and the RVE meta-regression show acceptable Type I error rates when there is no publication bias (below a nominal level of .1, and RVE meta-regression below .05; Rodgers & Pustejovsky, 2020). When there is selective reporting, the three methods have low power, especially trim-and-fill, although selection model can detect publication bias more often. The limited power of RVE meta-regression was especially sensitive to heterogeneity and the size of the true effect, becoming lower with higher heterogeneity and smaller effects. Regarding the adjustment of the effect, the original PET-PEESE (which assumes independence) performed worse with smaller true effects and higher heterogeneity, consistently underestimating the true effect. Furthermore, its estimate should be interpreted with caution in small meta-analyses (with 20 studies or less; Stanley, 2017). Additionally, we conducted a simulation analysis with the software developed by Carter et al. (2019; http://www.shinyapps.org/apps/metaExplorer/) comparing the performance of the standard versions (not accounting for dependence) of trim-and-fill, PET-PEESE, and selection model under conditions similar to those in previous comprenhensive meta-analyses (Sala & Gobet, 2017a, 2020; for further details, see **Appendix B**). Under the predefined conditions, the selection model achieved the best performance correcting the estimate (in terms of root square mean error and coverage) and trim-and-fill the worst, systematically overestimating the effect. The performance of PET-PEESE fell between both extremes. Finally, the Mathur and VanderWeele's sensitivity analysis seems to be relatively unbiased with values of η below 20 (i.e., a publication probability 20 times higher for positive than null or negative results; Mathur & VanderWeele, 2020).

2.4. Sensitivity Analyses

Only a subset of the studies reported sufficient information to compute the pre-posttest correlation with equations 6 and 7. To confirm that the results of the meta-analysis do not hinge critically on our decision to assume a correlation of .70 for all the studies, we repeated the analyses estimating $V_{g_{\Delta}}$ with $r = .50$ and $r = .60$. In the same vein, we assumed a within-effects correlation of .50 to estimate the sampling variance of the aggregates in the publication bias assessment. We also conducted the analyses with a correlation of .80 and .30. Moreover, we carried out sensitivity analyses following a multilevel Bayesian approach using the *brms* R package (Bürkner, 2017). The results of all the sensitivity analyses were similar to those reported here, showing far transfer with musical training (**Appendices C and D**), the modulating role of several variables on this effect (**Appendix C**), and little evidence of publication bias (**Appendix E**).

3. Results

Thirty-two empirical studies meeting the selection criteria were included in the systematic review, contributing a total of 179 cognitive/academic outcomes from 34 independent samples[3]. As a consequence of our comprehensive search among the gray literature, we identified four theses and a report from a charity foundation (Haywood et al., 2015) that met our inclusion criteria. Moreover, fifteen of the studies have not been included in the most recent meta-analysis by Sala and Gobet (2020), in part because their inclusion

[3] Costa-Giomi (1999) and Costa-Giomi (2004) seem to be reports of the same samples, as well as Roden et al. (2012) and Roden, Grube et al. (2014). We treated each of both pairs as coming from the same sample of participants.

criteria excluded programs in which the participants self-selected the program (although, some self-selection studies were included in their set: Degé et al., 2011; Geoghegan & Mitchelmore, 1996; Habibi et al., 2018; Hogan et al., 2018; Kempert et al., 2016; with contributed with a null mean effect, *g* = 0.03). Ten of the new studies were programs that allowed the selection of the group, two were randomized and three were non-randomized. Additionally, regarding the studies with instrumental traning that the present meta-analysis have in common with the recent one by Sala and Gobet (17 studies), we identified 14 outcomes that had not been previously analyzed that overall showed moderate effects in favor of musical training (mean *g* = 0.40).

Among all the independent samples, eight had random assignment of participants to groups, twelve were non-randomized, and fourteen were self-selection studies. Seven samples had both active and passive control groups, five only active, and 22 only passive. The main characteristics of the studies are summarized in **Table 1**. Regarding sample characteristics, the mean age of the samples included in our meta-analysis was 8 years (*SD* = 2.2; range: 3.9–14.7 years) and the mean duration of the programs was 17 months (*SD* = 16.3; range: 0.75–60 months). A total of 1664 children/adolescents took musical training, whereas 4334 were part of control groups (3670 in passive control groups and 664 in active control groups involving activities such as reading, drama, natural sciences lessons, visual arts, sports, dance, or non-musical computer-based programs).

Table 1

Characteristics of the Studies Included in the Meta-Analysis

Study	$N_{music\ group}$	$N_{control\ group}$	Type of publication	Age at baseline (in years)	Duration (in months)	Type of outcome	Blind assessment	Random assignment	Type of control	Low SES
Costa-Giomi, 1999	43	35	Article	9	36	Intelligence	No	Yes	Passive	No
Costa-Giomi, 2004	45	35	Article	9	36	Literacy, and mathematics	No	Yes	Passive	No
D'Souza & Wiseheart, 2018	24	26 & 25	Article	6–9	0.75	Executive functions, intelligence, literacy, processing speed, and short-term memory	Only computerized measures	Yes (stratified randomization; active control and experimental groups) & No (passive group)	Active and passive	No
Degé et al., 2011	16	18	Article	10	24	Intelligence, and short-term memory	No	No (self-selection)	Passive	Unknown
Fasano et al., 2019	55	58	Article	8–10	3	Executive functions	No	No (selection by teachers)	Passive	No
Fitzpatrick, 2006; Sample 1	78	1535	Article	9	60	Literacy	Yes	No (self-section)	Passive	Yes
Fitzpatrick, 2006; Sample 2	158	1167	Article	9	60	Literacy	Yes	No (self-selection)	Passive	No
Friedman, 1959; Sample 1	76	76	Thesis	10	12	Literacy, and mathematics	No	No (selection by music skills)	Passive	No
Friedman, 1959; Sample 2	51	51	Thesis	11	12	Literacy, and mathematics	No	No (selection by music skills)	Passive	No
Frischen et al., 2021	27	31 & 36	Article	6.6	8.5	Executive functions and short-term memory	Yes	Yes	Active and passive	No
Guo et al., 2018	20	20	Article	6–8	1.5	Executive functions, literacy, processing speed, and short-term memory	Yes	No (quasi-randomization)	Passive	Unknown
Hallberg et al., 2017	26	22	Article	5	1.25	Executive functions, and intelligence	Only computerized measures	Yes	Passive	Yes
Haywood et al., 2015	269	279	Foundation report	11	11	Literacy, and mathematics	Yes	Yes	Active	No
Hennessy et al., 2019	17	17 & 18	Article	6	48	Intelligence	No	No (self-selection)	Active and passive	Yes
James et al., 2020	34	35	Article		48	Executive functions, intelligence, processing speed, short-term memory, and long-term memory	No	Yes (cluster randomization)	Active	Yes
Kinney, 2008; Sample 1	20	85	Article	9	12	Literacy, and mathematics	Yes	No (self-selection)	Passive	Yes
Kinney, 2008; Sample 2	30	97	Article	9	12	Literacy, and mathematics	Yes	No (self-selection)	Passive	No
Legette, 1993	38	47	Thesis	9	8	Literacy, and mathematics	No	No	Passive	Yes
MacCutcheon et al., 2020	26	15	Article	6	12	Short-term memory, and phonological processing	Only computerized measures	No (self-selection)	Active	No
Nan et al., 2018	30	28 & 16	Article	4–5	6	Intelligence, literacy, and phonological processing	Only computerized measures	No (quasi-randomization)	Active and passive	No
Nering, 2002	10	10	Thesis	3.3–7.3	7	Executive functions, intelligence, literacy, mathematics, processing speed, and short-term memory	No	Yes	Passive	No

Table 1 (*continued*)

Study	$N_{music\ group}$	$N_{control\ group}$	Type of publication	Age at baseline (in years)	Duration (in months)	Type of outcome	Blind assessment	Random assignment	Type of control	Low SES
Orsmond & Miller, 1999	21	21	Article	5	4	Intelligence, literacy, and visuospatial abilities	No	No (self-selection)	Passive	No
Pelletier, 1963	55	55	Thesis	8	6.25	Literacy	No	No (quasi-randomization)	Passive	Unknown
Portowitz et al., 2009	45	36	Article	8	24	Intelligence, long-term memory, and visuospatial abilities	No	No	Passive	Yes
Rauscher & Zupan, 2000	34	28	Article	7–9	8	Intelligence, and long-term memory	Yes	No	Passive	No
Rauscher et al., 1997	34	20 & 14	Article	3–4.8	6	Intelligence	Yes	No	Active and passive	Unknown
Roden et al., 2012	25	25 & 23	Article	7.7	18	Long-term memory, and short-term memory	Yes	No	Active and passive	No
Roden, Grube et al., 2014	25	25	Article	7–8	18	Executive functions, and short-term memory	Yes	No	Active	No
Roden, Könen et al., 2014	192	153	Article	7–8	18	Executive functions, and processing speed	No	No	Active	No
Rose et al., 2019	19	19	Article	9	12	Executive functions, intelligence, literacy, processing speed, short-term memory, long-term, and visuospatial abilities	No	No (self-selection)	Passive	No
Said & Abramides, 2020	40	40	Article	10.34	6	Literacy, and mathemathics	No	No (self-selection)	Passive	No
Schellenberg, 2004	30	34 & 36	Article	6	9	Intelligence	Yes	Yes	Active and passive	No
Schellenberg et al., 2015; Sample 1	20	25	Article	8.7	10	Literacy	No	No	Passive	No
Schellenberg et al., 2015; Sample 2	18	21	Article	8.7	10	Literacy	No	No	Passive	No
Slater et al., 2014	23	19	Article	6–9	12	Intelligence, literacy, phonological processing, processing speed, and short-term memory	No	No (quasi-randomization)	Passive	Yes
Tierney et al., 2015	19	21	Article	14.7	36	Phonological processing, processing speed, and short-term memory	No	No (self-selection)	Active	Yes

3.1. Overall Effect

The overall meta-analysis, including both randomized and non-randomized studies, showed a positive and significant average effect of musical training, $\bar{g}_{\Delta}$ = 0.26, 95% CI [0.13, 0.39], p < .001, although heterogeneity was high, τ^2 = 0.15; I^2 = 83.98%. The same result appeared when self-selection studies were also included in the model, $\bar{g}_{\Delta}$ = 0.19, 95% CI [0.10, 0.28], p < .0001; τ^2 = 0.09; I^2 = 76.21%. Interestingly, the pre-posttest difference was similar among the three groups of studies, even numerically smaller for self-selection studies (randomized: $\bar{g}_{\Delta}$ = 0.26, p = .013; non-randomized: $\bar{g}_{\Delta}$ = 0.25, p = .015; self-selection: $\bar{g}_{\Delta}$ = 0.11, p = .023).

Subsequently, we assessed whether the observed heterogeneity could be due to the presence of outliers. Three outliers (Raushcer et al., 1997; Raushcer & Zupan, 2000; Roden, Könen et al., 2014) were detected, as they contributed with implausibly large effect sizes (some of them larger than g_{Δ} = 1). Interestingly, these three studies did not randomly assign participants to groups and had small samples, factors that might have contributed to their outcomes. All subsequent analyses were conducted without these studies. The overall effect (with randomized and non-randomized studies) remained significant but heterogeneity was still substantial, $\bar{g}_{\Delta}$ = 0.26, 95% CI [0.13, 0.39], p < .001; τ^2 = 0.08; I^2 = 70.66%. Again, the final estimates of the three groups of studies did not differ (randomized: $\bar{g}_{\Delta}$ = 0.26, p = .013, see **Figure 2**; non-randomized: $\bar{g}_{\Delta}$ = 0.25, p = .021, see **Figure 3**; self-selection: $\bar{g}_{\Delta}$ = 0.11, p = .023).

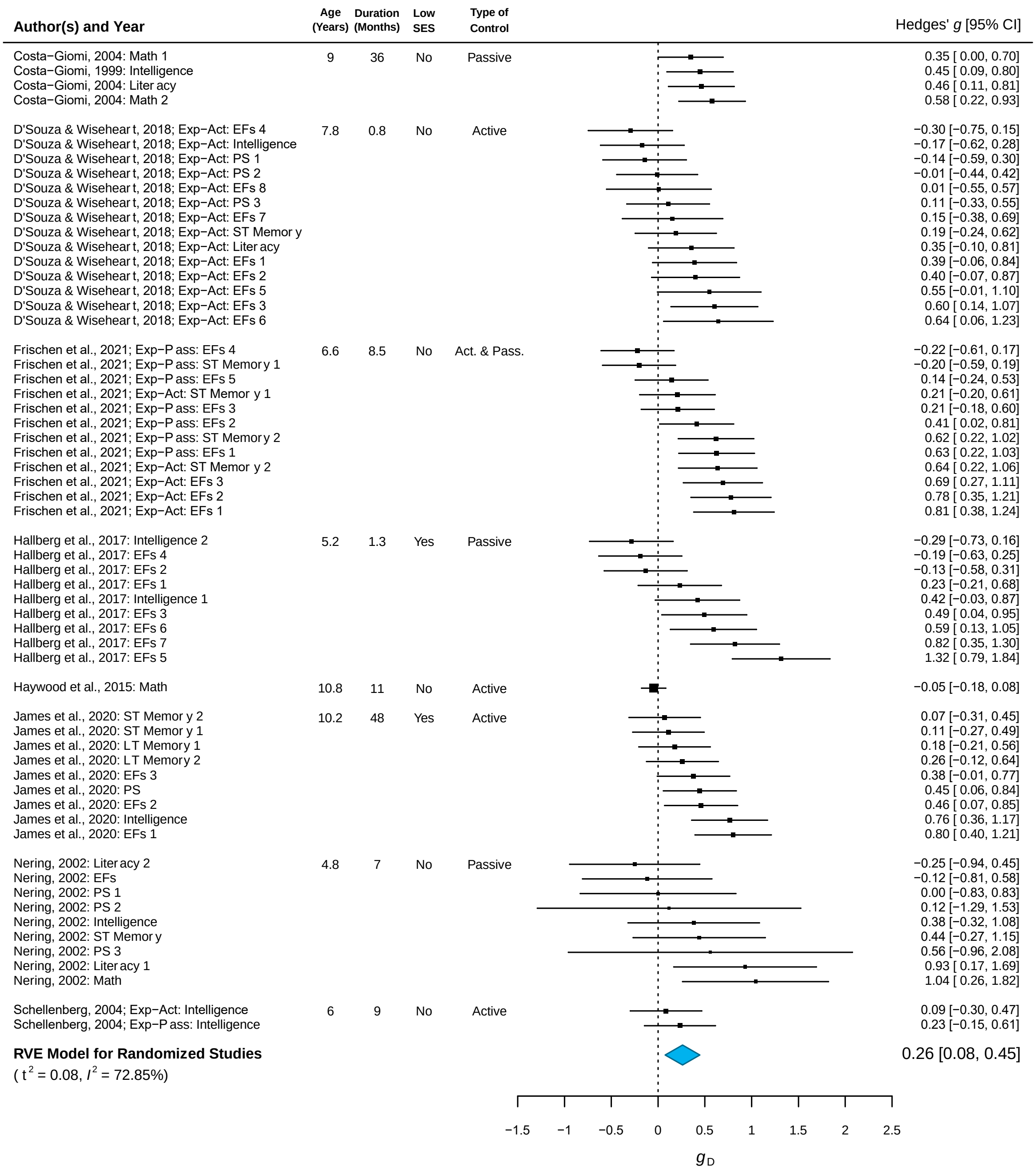

Figure 2. Forest plot of the standardized difference of mean change (pre-posttest difference) in randomized studies. Vertical diagonal of the blue rhombus represents the summary effect size, horizontal diagonal is the confidence interval for that final estimate.

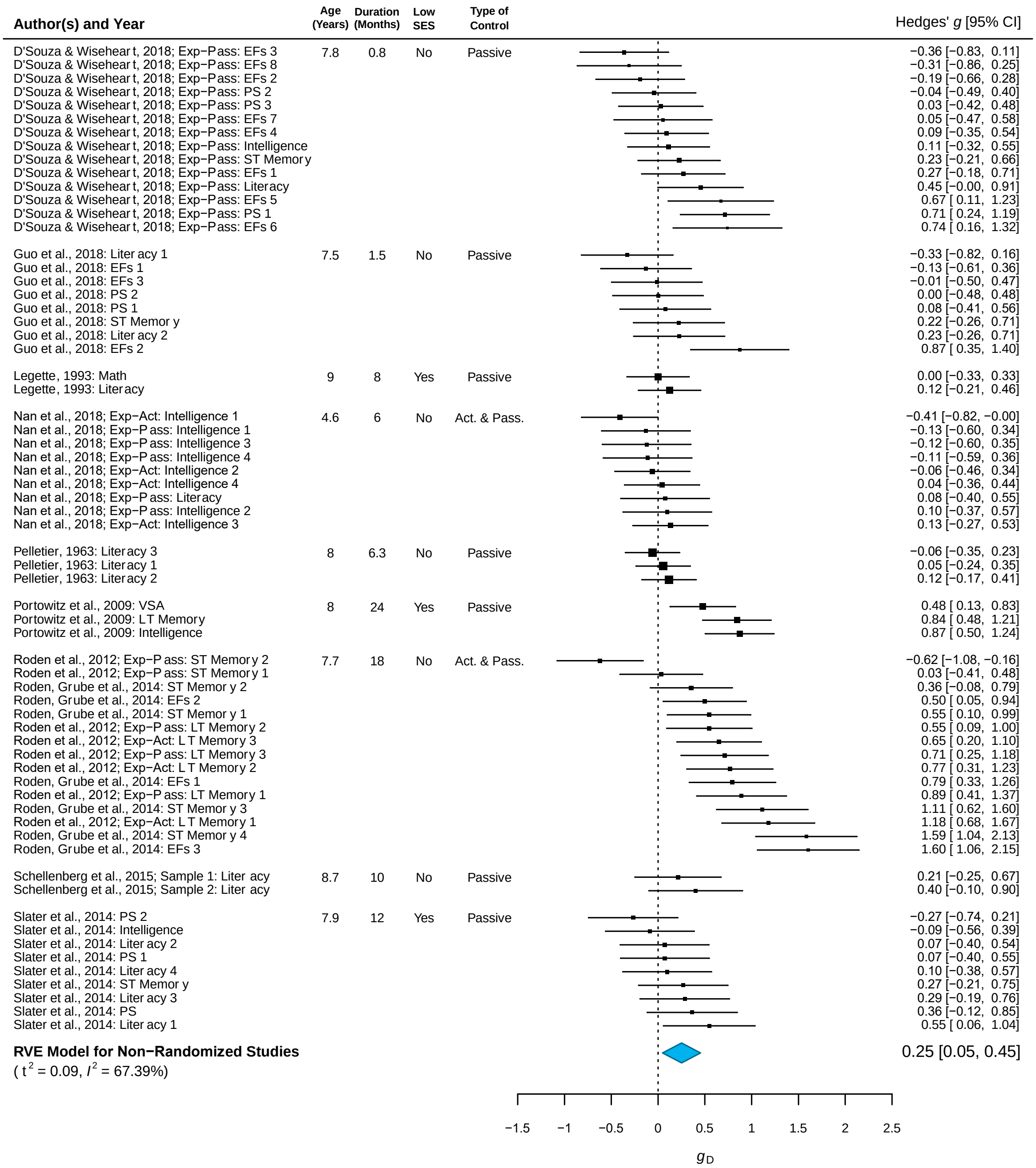

Figure 3. Forest plot of the standardized difference of mean change (pre-posttest difference) in non-randomized studies. Vertical diagonal of the blue rhombus represents the summary effect size, horizontal diagonal is the confidence interval for that final estimate.

3.2. Assessment of Baseline Differences

To test whether there were systematic baseline differences between participants in the treatment and control conditions, we conducted a meta-analysis of g_{pre}. As expected, the mean effect size was non-significant for randomized studies, $\bar{g}_{\text{pre}}$ = 0.00, 95% CI [–0.14, 0.15], p = .957; τ^2 = 0.01; I^2 = 14.20% (see **Appendix F, Figure F.1**), confirming that randomization had been successful in these studies. Also, there was no baseline difference in non-randomized studies without group selection, $\bar{g}_{\text{pre}}$ = 0.03, 95% CI [–0.08, 0.14], p = .510; τ^2 = 0; I^2 = 0% (see **Appendix F, Figure F.2**). On the other hand, there was a positive and significant baseline difference in favor of musical training groups among self-selection studies, $\bar{g}_{\text{pre}}$ = 0.29, 95% CI [0.12, 0.47], p = .003; τ^2 = 0.06; I^2 = 62.96% (see **Appendix F, Figure F.3**), which suggests that children/adolescents who voluntarily selected musical training as an extracurricular activity (over other programs such as sports or drama lessons) showed better initial cognitive and academic skills than their counterparts. A multilevel Bayesian approach using the *brms* R package (Bürkner, 2017) replicated previous results. Whereas there was strong evidence in favor of the lack of difference at baseline in randomized studies, BF_{10} = 0.09, and non-randomized studies, BF_{10} = 0.06; it showed substantial evidence in favor of preexisting differences in self-selection studies, BF_{10} = 7.20.

3.3. Moderator Analyses

Most of the moderators (randomization, active control, blinding of assessors/computerized measures, age of the participants, duration of the program, baseline difference, and low SES) were not significant when they were individually added to the model of the randomized and non-randomized studies (**Table 2**). Overall, the results suggested that several academic or cognitive domains were more sensitive than others to the impact of learning to play an instrument (see **Table 3**). To find out which combination of moderators provided the best fit for the data, we carried out a backward stepwise selection ($\alpha_{exclusion}$ = .10) with all the moderators. The best meta-regressive model did not retain any moderator.

When the influence of moderators was assessed only with randomized studies, no separate variable reached the significance level explaining heterogeneity. However, when more complex structures of moderators were considered, the best meta-regressive model included age, baseline difference, and low SES (remained heterogeneity: τ^2 = 0.08; I^2 = 59.56%). The model suggests that the effect of musical training in randomized studies was smaller in older individuals and individuals with higher performance at baseline, whereas larger effects were found with low SES.

Table 2

Results of the Meta-Regressive Analyses

	Moderator	*F*	*df*	*p*	
Separate models for each moderator					
	Randomization	0.01	1, 13.4	.925	Randomized: $\bar{g}_\Delta$ = 0.26 [0.08, 0.45] Non-randomized: $\bar{g}_\Delta$ = 0.25 [0.05, 0.45]
	Active control	0.00	1, 8.1	.972	Active: $\bar{g}_\Delta$ = 0.28 [−0.06, 0.61] Passive: $\bar{g}_\Delta$ = 0.25 [0.13, 0.37]
	Blinding	0.01	1, 10.5	.916	Blinded: $\bar{g}_\Delta$ = 0.30 [0.03, 0.57] Unblinded: $\bar{g}_\Delta$ = 0.24 [0.08, 0.40]
	Age	0.03	1, 5.4	.875	β = −0.007
	Duration	3.99	1, 2.5	.158	β = 0.009
	Baseline difference	2.43	1, 10.6	.148	β = −0.25
	Low SES	0.70	1, 7.6	.427	Low SES: $\bar{g}_\Delta$ = 0.34 [0.00, 0.68] Middle-high SES: $\bar{g}_\Delta$ = 0.22 [0.06, 0.38]
Separate models for each moderator (randomized studies)					
	Active control	1.4	1, 6	.283	Active: $\bar{g}_\Delta$ = 0.23 [−0.13, 0.58] Passive: $\bar{g}_\Delta$ = 0.32 [0.17, 0.48]
	Blinding	0.41	1, 5.9	.548	Blinded: $\bar{g}_\Delta$ = 0.27 [−0.08, 0.62] Unblinded: $\bar{g}_\Delta$ = 0.27 [−0.00, 0.54]
	Age	0.53	1, 3.7	.509	β = −0.031
	Duration	3.64	1, 2	.198	β = 0.005
	Baseline difference	2.93	1, 4	.162	β = −0.37

Low SES	2.4	1, 1.7	.280	Low SES: $\bar{g}_\Delta$ = 0.38 [0.23, 0.52] Middle-high SES: $\bar{g}_\Delta$ = 0.22 [−0.05, 0.49]
Best meta-regressive model for randomized studies (~ Age + Baseline difference + Low SES)				
Age	19	1, 2	.051	β = −0.06
Baseline difference	10.6	1, 4.6	.026	β = −0.56
Low SES	13.2	1, 1.6	.098	β = 0.28

Table 3

Final Effect of Each Type of Cognitive/Academic Outcome

Type of outcome	$\bar{g}_\Delta$	[95%CI]	*m*	*k*	*p*
Executive functions	**0.41**	**[0.12, 0.70]**	**7**	**41**	**.013**
Intelligence	0.29	[−0.00, 0.58]	9	19	.052
Literacy	**0.21**	**[0.06, 0.35]**	**10**	**18**	**.011**
Mathematics	0.24	[−0.42, 0.90]	4	5	.309
Phonological processing	−0.10	[−0.51, 0.31]	1	2	.999
Processing speed	0.26	[−0.06, 0.57]	5	13	.079
Short-term memory	**0.28**	**[0.15, 0.41]**	**7**	**17**	**.002**
Long-term memory	0.61	[−0.28, 1.5]	3	9	.098
Visuospatial abilities	**0.48**	**[0.13, 0.83]**	**1**	**1**	**.007**

Note. Significant results are depicted in bold; *m* = number of studies, *k* = number of outcomes, *p* = *p* value.

3.4. Publication Bias

Visual inspectioning the funnel plot of the aggregates of randomized and non-randomized studies (self-selection studies were not included in publication bias analyses), there was no clear asymmetry in the distribution of effects (**Figure 4**). Consistent with this, the trim-and-fill and RVE meta-regression showed no evidence of asymmetry of the funnel plot (i.e., publication bias and small-study effects). Trim-and-fill with the *R0* estimator detected one missing study (see white circle in **Figure 4**), but not with the *L0* estimator (no missing studies). On the other hand, the regression coefficients for the standard error and the sampling variance were not significant in the RVE PET/PEESE meta-regressions (see **Table 4**). The likelihood ratio test of the Vevea and Hedges's selection model was not indicative of publication bias either (p = 0.573). Finally, in the Mathur and VanderWeele's sensitivity analysis, no value of η could render the estimate equal to 0 or non-significant. An $\eta > 8.3$ was necessary to diminish the final estimate to $\bar{g}_{\Delta} = 0.10$. Those results suggest that the meta-analytic conclusions are robust regardless of the severity of publication bias. Moreover, the multiple tests of publication bias yielded similar results when they were conducted only with randomized studies, suggesting little evidence of selective reporting or small-study effects (**Table 5**).

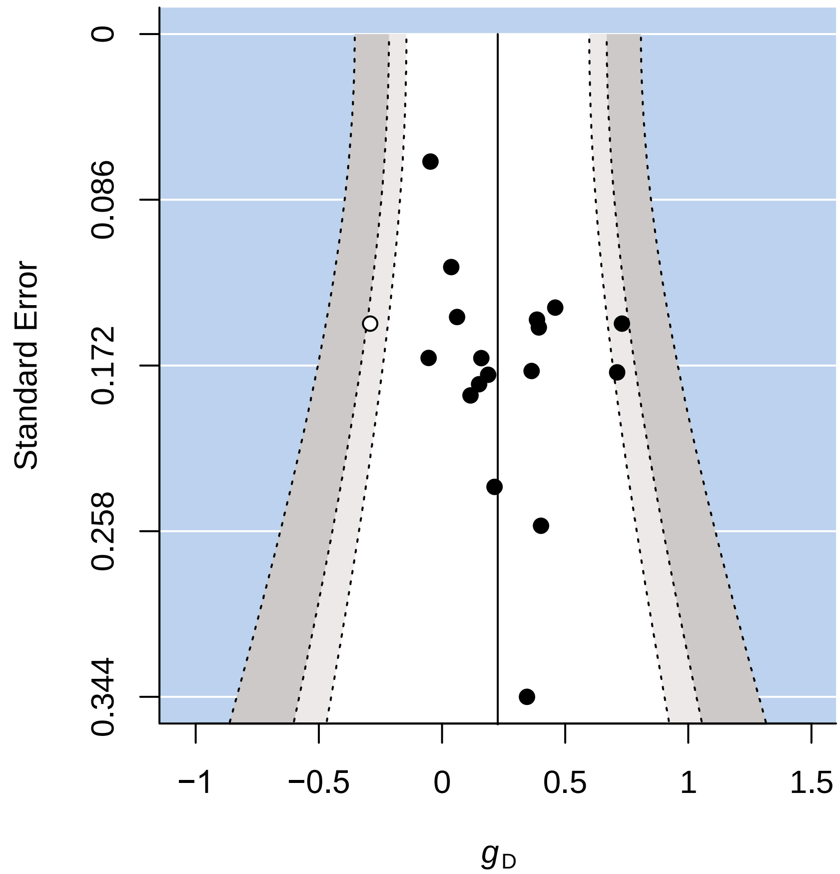

Figure 4. Funnel plot with trim-and-fill of the aggregate effects of randomized and non-randomized studies (black circles). One missing study was imputed with trim-and-fill (white

circle) using the *R0* estimator. The contour of the funnel takes into account the heterogeneity of the trim-and-fill model. The light gray zones show effects between p = .10 and p = .05, and the dark gray zones show effects between p = .05 and p = .01.

Table 4

Tests of Publication Bias for Randomized and Non-Randomized Studies Combined

	Trim and fill	RVE PET	RVE PEESE	Selection model	Sensitivity analysis
Test of publication bias	*L0*: No missing studies *R0*: 1 missing study	Modified SE: β = 0.95, p = .294 Transformation SE: β = 0.93, p = .279	Modified variance: β = 0.84, p = .559 Transformation variance: β = 0.96, p = .700	Likelihood ratio test: $\chi^2(1) = 0.32$, p = .573	$\bar{g}_\Delta = 0$: not possible $\bar{g}_\Delta = 0.10$: η = 8.34
Corrected estimate	*L0*: $\bar{g}_\Delta = 0.26$, $p < .0001$ *R0*: $\bar{g}_\Delta = 0.23$, $p < .001$	Modified SE: $\bar{g}_\Delta = 0.02$, p = .941 Transformation SE: h = 0.01, p = .932	Modified variance: $\bar{g}_\Delta = 0.20$, p = .188 Transformation variance: h = 0.25, p = .020	$\bar{g}_\Delta = 0.30$, p = .004	η = 1.5: $\bar{g}_\Delta = 0.23$, p = .003 η = 5: $\bar{g}_\Delta = 0.12$, p = .003

Table 5

Tests of Publication Bias for Randomized Studies

	Trim and fill	RVE PET	RVE PEESE	Selection model	Sensitivity analysis
Test of publication bias	*L0*: 2 missing studies *R0*: No missing studies	Modified SE: β = 1.19, p = .349 Transformation SE: β = 1.14, p = .352	Modified variance: β = 0.924, p = .659 Transformation variance: β = 0.98, p = .788	Likelihood ratio test: $\chi^2(1) = 0.90$, p = .344	$\bar{g}_\Delta = 0$: not possible $\bar{g}_\Delta = 0.10$: η = 114.24
Corrected estimate	*L0*: $\bar{g}_\Delta = 0.19$, p = .047 *R0*: $\bar{g}_\Delta = 0.26$, p = .002	Modified SE: $\bar{g}_\Delta = -0.03$, p = .911 Transformation SE: $h = -0.01$, p = .937	Modified variance: $\bar{g}_\Delta = 0.19$, p = .322 Transformation variance: h = 0.25, p = .049	$\bar{g}_\Delta = 0.16$, p = .175	η = 1.5: $\bar{g}_\Delta = 0.26$, p = .001 η = 5: $\bar{g}_\Delta = 0.16$, p = .002

Regarding the adjustment of the multiple methods, the corrected effect sightly differed from the uncorrected estimate ($\bar{g}_{\Delta}$ = 0.26) in most of the cases (see **Table 4**). Only regression-based methods yielded non-significant estimates and, among them, only PET returned a negligible corrected estimate ($\bar{g}_{\Delta}$ = 0.02, p = .941). The Mathur and VanderWeele's sensitivity analysis yielded an effect closely identical to the uncorrected one ($\bar{g}_{\Delta}$ = 0.23, p = .003) with the mean value of η in psychology literature (η = 1.5). When the publication probability for positive results was five times the probability for null or negative (the 95th quantile in psychology; Mathur & VanderWeele, 2020), the adjusted effect remained positive and significant ($\bar{g}_{\Delta}$ = 0.12, p = .003). Despite the smaller number of randomized studies (m = 8), the results were similar when the adjustments were aplied in that group of studies (**Table 5**).

In summary, none of the methods detected substantial evidence of publication bias or small-study effects, including those with higher power, such as the Vevas and Hedges' selection model. In addition, the corrected estimate had similar size in most of the cases. Only the RVE PET approach showed a reduction in the effect. However, it is probable that the attenuation with RVE PET was a consequence of its worse performance under the observed conditions of moderate-to-high heterogeneity, small number of studies, and small effect size (Stanley, 2017; also see our performance simulation with Carter et al.'s software, **Appendix B**). Previous simulation studies showed that PET tends to underestimate the true effect under the conditions observed in our meta-analysis (Stanley, 2017). Therefore, taking all the approaches in consideration, the results suggest that the true underlying effect is non-zero.

4. Discussion

The present meta-analysis investigates the causal effects of learning to play an instrument on cognitive and academic skills during the school years. Overall, a small benefit ($\bar{g}_{\Delta}$ = 0.26) was found with relatively short-term programs (with a mean duration of 17 months), regardless of whether or not there was a random assignment of participants to musical training versus control groups, and independently of the type of control group (i.e., active vs. passive). The fact that a positive result was also found in randomized designs taken alone supports the idea of a causal role of musical training in the observed improvements. Complementarily, it is important to note the detection of a bias in baseline performance in favor of music groups across studies in which the participants chose the training group ($\bar{g}_{pre}$ = 0.29). This indicates that participants who self-selected to play an instrument consistently showed better performance prior to the beginning of the intervention, compared to those who decided to enrol in an alternative control activity. As this pretest disparity was small, and most of the studies were underpowered to detect it[4], it is perhaps unsurprising that the authors of those studies usually claimed to have matched groups. However, our meta-analytic evidence reveals that this was not the case. In contrast, and importantly, the pretest differences were null in randomized studies, as one would expect from truly random assignment of participants to groups. Furthermore, there was scarce evidence of publication bias and our conclusions remain valid under almost all the bias-correction methods that we applied (trim-and-fill, selection model, PEESE, and sensitivity analysis), except PET. Simulation studies have found that the last method performed poorly under conditions of moderate-to-high heterogeneity, reduced number of studies, and a putative small true effect as the one explored in the present meta-analysis. Therefore, it reinforces the conclusion that current evidence supports a causal effect of instrumental musical training on cognitive and academic skills in children and adolescents.

Our findings are in line with a *nature and nurture* approach (Wan & Schlaug, 2010). According to this view, preexisting cognitive advantages, such as those we observed at baseline in self-selection studies, would facilitate the learning of musical skills. In addition, engagement

[4] A power analysis using G*Power 3.1 (Faul et al., 2009) for a two-tailed *t*-test and an alpha of .05 indicated that around 188 participants per group would be necessary to achieve an acceptable power of.80 with a Cohen's *d* of 0.29.

in the complex activity of learning to play a musical instrument for a long period of time would lead to neurocognitive adaptations producing further enhancements in general cognitive and academic skills. To our knowledge, only one experimental study has investigated far transfer with instrumental learning using a monozygotic cotwin control design (Nering, 2002). In this study, one of the twins was randomly selected to take a piano training program while the other was assigned to a waitlist group. After 7 months, experimental twins overperformed the control group in intelligence scores. Although the comparison group did not participate in an alternative activity (such as in other experimental studies with positive outcomes; e.g., Frischen et al., 2021), the inclusion of monozygotic twin pairs with a common genotype and an early rearing environment supports that musical training has an impact on extra-musical cognitive skills even when genetic factors and shared environment are controlled.

Under a *nature and nurture* approach, it is not surprising that the differences reported previously between musicians and non-musicians in correlational studies ($\bar{g}$ = 0.8–1; Corrigall et al., 2013) tend to be remarkably larger than the effects of short-term training in children that we observed in our meta-analysis of experimentally controlled studies ($\bar{g} \approx 0.2$). The combination of both initial differences and additional enhancements produced by the involvement in musical training for many years can explain the larger effect observed in correlational studies comparing adult musicians and non-musicians. In a recent study, Mankel and Bidelman (2018) reported similar findings with auditory processing, where listeners with inherently more adept auditory skills but no formal musical training showed better speech encoding than a low-musicality group, whereas formally trained musicians superior musicality and outperformed both groups of non-musicians on speech encoding. Taken together, their results suggest that preexisting factors may play a role in the relationship between musical experience and enhanced auditory functions, at the same time that musical training might provide an additional experience-dependent boost of preexisting differences.

Following this logic, the less controlled the correlational studies, the larger one would expect the observed effect of musical training to be. For instance, Medina and Barraza (2019) observed an extremely large advantage in executive control for professional pianists (d = 1.51) using a visuospatial attentional task (i.e., the *Attentional Networks Test* or ANT; Fan, McCandliss, Sommer, Raz, & Posner, 2002), which correlated with the number of years of musical practice. This exceptionally large effect was likely inflated by the lack of control over several variables potentially also enhancing attention. Indeed, in a similar study with an ANT-like task (i.e., the *Attentional Networks Test for Interactions and Vigilance – executive and arousal components* or ANTI–Vea; Luna, Marino, Roca, & Lupiáñez, 2018), Román-Caballero et al. (2020) found a smaller difference (d = 0.25) when the effect of musical training was measured while controlling for a wide list of sociodemographic and lifestyle confounds. This inflation is still present in observational studies with large samples, such as Guhn et al. (2020; $N \approx$ 110,000), in which reductions around 60% or more were observed in all the measures after controlling for multiple confounders (cultural background, socioeconomic status, sex, and prior academic achievement). Thus, the long history of training in professional musicians (about 12 years in Medina and Barraza, 2019) likely fosters their cognitive capacities, although in a more modest way than reported in uncontrolled cross-sectional studies.

4.1. Influence of Moderators and Individual Differences

4.1.1. Methodological Quality and Other Musical Programs

Unlike previous studies reporting an inverse relationship between design quality and the magnitude of the effects (see Sala & Gobet, 2017a, 2020), we did not find a significant reduction in the size of the outcome of randomized studies compared to non-randomized ones. On the contrary, the benefits for studies with random allocation were numerically greater (randomized: $\overline{g}_{\Delta}$ = 0.26, vs. self-selection: $\overline{g}_{\Delta}$ = 0.11). As noted above, this inconsistency could be a consequence of non-instrumental programs being overrepresented in the studies with higher methodological quality in the meta-analysis by Sala and Gobet (2020; only 31% of those with higher methodological quality involved instrumental training). Previous studies show that the benefits of non-instrumental interventions, such as preschool training of musical skills or active listening, are smaller than those of instrumental training (Bugos, 2010; James et al., 2020). A plausible explanation is that non-instrumental programs are less cognitively demanding and also that the skills they train are more restricted to the music domain compared to instrumental programs.

Indeed, a reanalysis of the data meta-analyzed by Sala and Gobet (2020) supports these impressions. Excluding studies with self-selection of the musical training program (Geoghegan & Mitchelmore, 1996; Hogan et al., 2018; Kempert et al., 2016), those with only posttest designs (five), and those excluded as outliers, we identified 30 non-instrumental studies included in Sala and Gobet's review, which used computerized training of musical skills (4 studies), phonological processing training with music support (1), and Kindermusik, Orff, Kodály or other related methods (25). We compared these non-instrumental studies to the 18 studies with instrumental programs with random or not self-selected assignment included in our meta-analysis. When design quality was not taken into account (i.e, when studies with randomized and non-randomized allocation, as well as active and passive controls, were analyzed), both non-instrumental and instrumental programs showed similar and significant benefits ($\bar{g}_{\Delta,\text{ instrumental}} = 0.26$, $p < .001$, vs. $\bar{g}_{\Delta,\text{ non-instrumental}} = 0.20$, $p = .002$). However, this result changed remarkably when we constrained the analyses to randomized studies, finding that only instrumental programs had a significant effect ($\bar{g}_{\Delta,\text{ instrumental}} = 0.26$, $p = .013$, vs. $\bar{g}_{\Delta,\text{ non-instrumental}} = 0.11$, $p = .197$). Similarly, when the analyses were constrained to studies with active control groups, instrumental programs outperformed non-instrumental interventions ($\bar{g}_{\Delta,\text{ instrumental}} = 0.23$ vs. $\bar{g}_{\Delta,\text{ non-instrumental}} = 0.01$). Therefore, it seems that the null result with high-quality designs reported by Sala and Gobet (2020) was biased by the overrepresentation of non-instrumental interventions (73% of the randomized studies) that, according to our reanalyses, do not seem to produce far transfer benefits. The confound between design quality and the type of musical training in the previous meta-analysis makes it necessary to take their conclusions with caution and limits its generalizability to all musical programs. And again, it reinforces the importance of analyzing instrumental learning programs separately, as in the present meta-analysis.

Altogether, our results support the preferred use of random allocation of participants and pre-posttest designs, rather than only-posttest, to shed light on the debate about the causal role of musical training. However, the duration of studies involving randomized programs tends to be short and many children assigned to the musical training group may not be motivated to learn to play an instrument, both of which might undermine any potential effects of training. While further studies are necessary with randomized pre-posttest designs, active control groups and blind assessment, the evidence from programs where the children select the activity is also interesting on its own, as these studies usually investigate the effects of longer-term interventions and in ecological situations (Habibi et al., 2018; Tervaniemi et al., 2018). In any case, when it comes to instrumental learning, and not musical education in general, conclusions such as "since there is no phenomenon, there is nothing to explain" (Sala & Gobet, 2020, p. 9) or "researchers and policymakers should seriously consider stopping spending resources for this type of research" (Sala & Gobet, 2017b) seem overpessimistic in light of our results, and upcoming investigation will be essential to clarify this debate.

4.1.2. Baseline Differences, Socioeconomic Status, and Age of the Participants

Although randomization and the inclusion of an active control group did not explain between-studies variability in our meta-analysis, other moderators accounted for part of the heterogeneity. In the model of randomized studies, three variables were shown to be influential:

baseline differences between groups, age of the participants, and SES. First, the larger the baseline difference, the smaller the observed effect of musical training. This could be due to children with a lower initial level of performance having a greater window of opportunity, and vice versa. Similar results have been found for general cognitive training (Jaeggi et al., 2011; Whitlock et al., 2012). Conversely, participants who had the chance to choose musical training programs showed better academic and cognitive performance at baseline, but their benefits ($\bar{g}_{\Delta}$ = 0.11) were numerically the smallest compared to those in random ($\bar{g}_{\Delta}$ = 0.26) or other types of non-random allocation ($\bar{g}_{\Delta}$ = 0.25), in which there was no pretest bias. However, when baseline performance is explicitly controlled in the model to the model of self-selection studies, the model predicts a similar pre-postest difference under conditions of no pretest bias (fitted $\bar{g}_{\Delta}$ = 0.19). An alternative explanation for the pretest effect is a regression toward the mean of those samples of children who showed remarkably disparate scores at baseline (higher or lower).

In the same vein, participants with lower SES showed greater improvements compared to those with middle-high SES. Again, this might be the consequence of a large margin for improvement for individuals whose development of cognitive and academic skills is limited by their socioeconomic environments (Diamond, 2012). Therefore, this suggests that, although higher-functioning individuals are more likely to select and maintain musical practice for many years, children with a less favorable background can also benefit from musical training as long as they engage in it for enough time (Fasano et al., 2019; Portowitz et al., 2009; Tierney et al., 2015). If this finding is confirmed by future research, musical training can become an excellent candidate to contribute to reducing cognitive and academic differences due to social disparities.

Finally, the age of the participants at the beginning of the training program seems to modulate the impact of musical training. Our results are consistent with previous cross-sectional studies that show greater neural and cognitive advantages for earlier onsets of the training (Fauvel et al., 2014; Hanna-Pladdy & Gajewski, 2012; Schlaug et al., 1995; Vaquero et al., 2016). This relationship is suggestive of a sensitive period during which instrumental learning is likely to have stronger and more permanent effects on non-musical skills (White-Schwoch et al., 2013), perhaps as a consequence of greater neural plasticity earlier in development, and because those early neurocognitive changes might serve as a scaffold for future training (Vaquero et al., 2016).

4.1.3. Type of Outcome and Other Moderators

Despite the identification of several moderators, heterogeneity remained moderate (I^2 = 59.56%). Part of this variability may be due to artifacts such as differences in measurement error (in relation to the reliability and validity of the tests) or reporting and transcriptional errors (e.g., inaccuracy in coding data, computational errors, errors in reading computer output, or typographical errors). Additionally, although the duration of the programs was known, the participants might have had different levels of engagement and differed in the amount of between-lessons practice. Unfortunately, this information is rarely reported in the studies, so it is hard, if not impossible, to detect this type of biases in most studies (especially, when the outcomes are not outlier values). On the other hand, this heterogeneity may indicate the existence of other unknown variables that can modulate the final effect. In this sense, the type of outcome was a significant moderator when it was individually entered in the overall model, suggesting that the impact of the interventions is not the same for all cognitive and academic domains. Looking at **Table 3,** we observed that some cognitive abilities, such as executive functions ($\bar{g}_{\Delta}$ = 0.41), improved more than others. Unfortunately, the number of observations per type of enhanced cognitive skill was low (only two out of nine were assessed at least in ten studies). Thus, the analysis was overly underpowered and needs to be addressed in future research.

4.2. Transfer in Musical Training

A relevant contribution of our meta-analysis is that the benefits of instrumental learning were observed in cognitive tasks and contexts quite distinct from musical performance. This finding is in accordance with the idea that the involvement in such a stimulating activity, besides improving domain-specific skills, enhances distant functions. Nevertheless, not all cognitive domains and academic skills appear to be equally sensitive to instrumental training (see **Table 3**). For example, executive functions showed the most robust benefits. This is not surprising, as music-making places high demands on the abilities of control and self-regulation, monitoring, planning, and focused and sustained attention, among others.

Some authors have expressed skepticism about far transfer (Thorndike, 1906; Roediger III, 2013; Sala & Gobet, 2019). Specifically, Thorndike (1906) proposed that transfer only occurs when trained and untrained processes share features in common. Under this approach, he concluded that “the most common and surest source of general improvement of a capacity is to train it in many particular connections” (Throndike, 1906, p. 248). Unlike many other cognitive activities involving highly specific contexts and tasks, music-making requires the coordination of several skills and sensory modalities and involves a wide and constantly augmented variety of stimuli, social situations, and types of performance. Therefore, musical training has singular characteristics that made it a plausible cognitive enhancer, even from a skeptical perspective.

One explanation for far transfer is that regular training in a particular basic cognitive process fosters the process itself and, as a consequence, affords advantages to any daily task that also hinges on the same skill. However, this explanation is undoubtedly simplistic, as evoking a “brain as a muscle” metaphor fails to explain why cognitive training programs sometimes fail to extend their benefits to other activities (Gathercole et al., 2019; Roediger III, 2013; Simons et al., 2016; Taatgen, 2013). An alternative proposal conceptualizes transfer as the consequence of acquiring complex cognitive skills that can be applied to untrained tasks with some overlap (Gathercole et al., 2019; Taatgen, 2013). The *cognitive routine framework* (Gathercole et al., 2019) posits that training on unfamiliar or highly demanding tasks, such as learning to play an instrument, leads to the development of new complex cognitive skills. Transfer then occurs when one of these new skills can be applied to a novel activity. In the case of musical training, several studies have reported superior memory scores for adult musicians when they were compared to non-musician counterparts (Franklin et al., 2008; Jakobson et al., 2008; for longitudinal studies, see Portowitz et al., 2009; and Roden et al., 2012), but the evidence suggests that the advantage is largely due to more robust and efficient coding (such as an improved rehearsal mechanism, Franklin et al., 2008; or increased use of semantic information organization strategies, Jakobson et al., 2008). In line with these results, musical training could stimulate the development of singular strategies, such as mental rehearsal or semantic organization, that can be applied in several non-musical tasks. Accordingly, the expansion of cognitive capacities along with the development of new complex skills could explain the broad benefits observed with musical instrumental learning.

Interestingly, musical training and practice may also pose a unique type of ongoing challenge that might facilitate far transfer. No matter what level of technical and artistic mastery a musician achieves, there is always room for improvement. Furthermore, there are always new pieces, interpretations, styles, and genres of music to learn. And different musicians and ensembles to play with, adjust to, and learn from. Thus, improvement through the application of effortful control can remain a rewarding challenge throughout the lifespan. As a representative anecdote, when the virtuoso cellist Pau Casals was asked why he continued to practice four and five hours a day when he was eighty years old, he answered: “Because I think I am making progress”.

4.3. Practical Significance

Our results support that learning to play a musical instrument is an activity with cognitive and academic benefits, although they are fairly small. The overall effect in the present meta-analysis ($\bar{g}_{\Delta} = 0.26$) indicates a probability of 57.3% that a randomly selected person from the musical training group will show higher cognitive and academic performance than a person selected from the control group (only 7.3% above chance level). One pertinent question is the practical significance of this effect, as musical training is an effortful activity that takes many years. In this regard, Hunter and Schmidt (2015) claimed:

> The question for a treatment is really not whether it had an effect but whether the effect is as large as a theory predicts, whether the effect is large enough to be of practical importance, or whether the effect is larger or smaller than some other treatment or some variation of the treatment. (Hunter & Schmidt, 2015, pp. 246–247)

For example, Schellenberg (2004), in one of the first randomized studies with children participants, found that after 36 weeks of intervention the difference between the IQ gain of the keyboard and the passive control groups was only about 2 points. The benefit was even smaller when music participants were compared to children who took drama lessons (a gain difference of one IQ point). Our meta-analysis showed a similar overall increase, corresponding to about 3 IQ points. This contribution is rather small and probably makes very little difference in daily life. For that reason, musical training might be not one of the first-choice interventions if the only purpose is cognitive enhancement. However, additional longitudinal studies have found improvements in other domains, such as emotional development and empathy (Rabinowitch et al., 2013), prosocial skills (Schellenberg et al., 2015), self-esteem (Costa-Giomi, 2004; Rickard et al., 2013), academic self-esteem (Degé et al., 2014; Degé & Schwarzer, 2018), and mood and quality of life (Seinfeld et al., 2013). Thus, the cumulative benefits of music across a broad set of domains likely make it a worthwhile enterprise and, in the long term, there may be synergistic effects with the integration of benefits across domains providing skills to face diverse future challenges. More research is needed to further investigate the impact of musical training in non-cognitive areas, and its interaction with cognition and academic achievement.

5. Conclusions

The present meta-analysis shows that learning to play an instrument during the school years has a modest but significant cognitive and academic impact. Longitudinal evidence suggests both a causal role of musical training and the existence of a self-selection bias, whereby children with favorable backgrounds and higher initial functioning are more likely to choose to learn to play or keep learning an instrument. The contrast of these findings with the null results reported for other types of musical and cognitive programs indicates, once again, the rareness of far transfer. Although the mechanisms for transfer remain unknown, instrumental learning in structured programs would be an optimal framework to investigate them. Finally, given that reliable evidence is still scarce, further studies in this field will be relevant to reach firmer conclusions.

References[5]

Bigand, E., & Tillmann, B. (2021). Near and far transfer: Is music special? *PsyArxiv*. https://doi.org/10.31234/osf.io/gtnza

[5] The references of the studies included in the meta-analysis are in bold and preceded by an asterisk.

Bürkner, P. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bugos, J. A. (2010). The benefits of music instruction on processing speed, verbal fluency, and cognitive control in aging. *Music Education Research International, 4*, 1–9.

Bugos, J. A. (2014). Community music as a cognitive training programme for successful ageing. *International Journal of Community Music, 7*(3), 319–331. https://doi.org/10.1386/ijcm.7.3.319_1

Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging and Mental Health, 11*(4), 464–471. https://doi.org/10.1080/13607860601086504

Bürkner, P. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Butzlaff, R. (2000). Can music be used to teach reading? *Journal of Aesthetic Education, 34*(3/4), 167–178. https://doi.org/10.2307/3333642

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115–144. https://doi.org/10.1177/2515245919847196

Coburn, K. M., & Vevea, J. L. (2019). weightr: Estimating weight-function models for publication bias. R package version 2.0.2. https://cran.r-project.org/web/packages/weightr/

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cooper, P. K. (2020). It's all in your head: A meta-analysis on the effects of music training on cognitive measures in schoolchildren. *International Journal of Music Education, 38*(3), 321–336. https://doi.org/10.1177/0255761419881495

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology, 4*, 222. https://doi.org/10.3389/fpsyg.2013.00222

*** Costa-Giomi, E. (1999). The effects of three years of piano instruction on children's cognitive development. *Journal of Research in Music Education, 47*(3), 198–212. https://doi.org/10.2307/3345779**

*** Costa-Giomi, E. (2004). Effects of three years of piano instruction on children's academic achievement, school performance and self-esteem. *Psychology of Music, 32*(2), 139–152. https://doi.org/10.1177/0305735604041491**

Degé, F., & Schwarzer, G. (2018). The influence of an extended music curriculum at school on academic self-concept in 9-to 11-year-old children. *Musicae Scientiae, 22*(3), 305–321. https://doi.org/10.1177/1029864916688508

*** Degé, F., Wehrum, S., Stark, R., & Schwarzer, G. (2011). The influence of two years of school music training in secondary school on visual and auditory memory. *European Journal of Developmental Psychology, 8*(5), 608–623. https://doi.org/10.1080/17405629.2011.590668**

Degé, F., Wehrum, S., Stark, R., & Schwarzer, G. (2014). Music lessons and academic self-concept in 12-to 14-year-old children. *Musicae Scientiae*, *18*(2), 203–215. https://doi.org/10.1177/1029864914523283

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

Diamond, A. (2012). Activities and programs that improve children's execu- tive functions. *Current Directions in Psychological Science, 21*, 335–341. http://dx.doi.org/10.1177/0963721412453722

*** D'Souza, A. A., & Wiseheart, M. (2018). Cognitive effects of music and dance training in children. *Archives of Scientific Psychology, 6*(1), 178–192. http://dx.doi.org/10.1037/arc0000048**

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

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(3), 340–347. https://doi.org/10.1162/089892902317361886

*** Fasano, M. C., Semeraro, C., Cassibba, R., Kringelbach, M. L., Monacis, L., de Palo, V., ... & Brattico, E. (2019). Short-term orchestral music training modulates hyperactivity and inhibitory control in school-age children: A longitudinal behavioural study. *Frontiers in Psychology, 10*, 750. https://doi.org/10.3389/fpsyg.2019.00750**

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fauvel, B., Groussard, M., Mutlu, J., Arenaza-Urquijo, E. M., Eustache, F., Desgranges, B., & Platel, H. (2014). Musical practice and cognitive aging: Two cross-sectional studies point to phonemic fluency as a potential candidate for a use-dependent adaptation. *Frontiers in Aging Neuroscience, 6*, 227. https://doi.org/10.3389/fnagi.2014.00227

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2021). Detecting selection bias in meta-analyses with multiple outcomes: a simulation study. *The Journal of Experimental Education, 89*(1), 125-144. https://doi.org/10.1080/00220973.2019.1582470

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). robumeta: Robust Variance Meta-Regression. R package version 2.0. https://cran.r-project.org/web/packages/robumeta/

*** Fitzpatrick, K. R. (2006). The effect of instrumental music participation and socioeconomic status on Ohio fourth-, sixth-, and ninth-grade proficiency test performance. *Journal of Research in Music Education, 54*(1), 73–84. https://doi.org/10.1177/002242940605400106**

*** Friedman, B. (1959). *An evaluation of the achievement in reading and arithmetic of pupils in elementary school instrumental music classes* (Doctoral dissertation, New York University, School of Education).**

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

*** Frischen, U., Schwarzer, G., & Degé, F. (2021). Music lessons enhance executive functions in 6-to 7-year-old children. *Learning and Instruction, 74,* 101442. https://doi.org/10.1016/j.learninstruc.2021.101442**

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. https://doi.org/10.1016/j.jml.2018.10.003

Geoghegan, N., & Mitchelmore, M. (1996). Possible Effects of Early Childhood Music on Mathematical Achievement. *Journal for Australian Research in Early Childhood Education, 1*, 57–64.

Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Frontiers in Psychology, 6*, 1777. https://doi.org/10.3389/fpsyg.2015.01777

Guhn, M., Emerson, S. D., & Gouzouasis, P. (2020). A population-level analysis of associations between school music participation and academic achievement. *Journal of Educational Psychology, 112*(2), 308–328. https://doi.org/10.1037/edu0000376

*** Guo, X., Ohsawa, C., Suzuki, A., & Sekiyama, K. (2018). Improved digit span in children after a 6-week intervention of playing a musical instrument: An exploratory randomized controlled trial. *Frontiers in Psychology, 8,* 2303. https://doi.org/10.3389/fpsyg.2017.02303**

Habibi, A., Damasio, A., Ilari, B., Veiga, R., Joshi, A. A., Leahy, R. M., ... & Damasio, H. (2018). Childhood music training induces change in micro and macroscopic brain structure: Results from a longitudinal study. *Cerebral Cortex, 28*(12), 4336–4347. https://doi.org/10.1093/cercor/bhx286

*** Hallberg, K. A., Martin, W. E., & McClure, J. R. (2017). The impact of music instruction on attention in kindergarten children. *Psychomusicology: Music, Mind, and Brain, 27*(2), 113–121. https://doi.org/10.1037/pmu0000177**

Hanna-Pladdy, B., & Gajewski, B. (2012). Recent and past musical activity predicts cognitive aging variability: Direct comparison with general lifestyle activities. *Frontiers in Human Neuroscience*, *6*, 198. https://doi.org/10.3389/fnhum.2012.00198

*** Haywood, S., Griggs, J., Lloyd, C., Morris, S., Kiss, Z., & Skipp, A. (2015). *Creative Futures: Act, Sing, Play*. Evaluation Report and Executive Summary. https://educationendowmentfoundation.org.uk/**

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

*** Hennessy, S. L., Sachs, M. E., Ilari, B. S., & Habibi, A. (2019). Effects of music training on inhibitory control and associated neural networks in school-aged children: A longitudinal study. *Frontiers in Neuroscience, 13*, 1080. https://doi.org/10.3389/fnins.2019.01080**

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, *76*(3), 486–502. https://doi.org/10.1016/j.neuron.2012.10.011

Hetland, L. (2000). Learning to make music enhances spatial reasoning. Journal of Aesthetic Education, 34(3/4), 179–238. https://doi.org/10.2307/3333643

Hogan, J., Cordes, S., Holochwost, S., Ryu, E., Diamond, A., & Winner, E. (2018). Is more time in general music class associated with stronger extra-musical outcomes in kindergarten? *Early Childhood Research Quarterly, 45*, 238–248. https://doi.org/10.1016/j.ecresq.2017.12.004

Hole, K. A. (2013). The impact of an auditory training programme (The Listening Programme®) on the auditory processing and reading skills of mainstream school children (unpublished manuscript). http://etheses.whiterose.ac.uk/id/eprint/5356

Hunter, J. E., & Schmidt, F. L. (2015). *Methods of meta-analysis: Correcting Error and Bias in Research Findings* (3nd ed.). Newbury Park, CA: Sage Publications.

Kempert, S., Götz, R., Blatter, K., Tibken, C., Artelt, C., Schneider, W., & Stanat, P. (2016). Training early literacy related skills: To which degree does a musical training contribute to phonological awareness development? *Frontiers in Psychology, 7*, 1803. https://doi.org/10.3389/fpsyg.2016.01803

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*(25), 10081-10086. https://doi.org/10.1073/pnas.1103228108

*** James, C. E., Zuber, S., Dupuis-Lozeron, E., Abdili, L., Gervaise, D., & Kliegel, M. (2020). Formal string instrument training in a class setting enhances cognitive and sensorimotor development of primary school children. *Frontiers in Neuroscience, 14,* 567. https://doi.org/10.3389/fnins.2020.00567**

Jentzsch, I., Mkrtchian, A., & Kansal, N. (2014). Improved effectiveness of performance monitoring in amateur instrumental musicians. *Neuropsychologia, 52*, 117–124. https://doi.org/10.1016/j.neuropsychologia.2013.09.025

Kaganovich, N., Kim, J., Herring, C., Schumaker, J., MacPherson, M., & Weber-Fox, C. (2013). Musicians show general enhancement of complex sound encoding and better inhibition of irrelevant auditory change in music: An ERP study. *European Journal of Neuroscience, 37*(8), 1295–1307. https://doi.org/10.1111/ejn.12110

Kaviani, H., Mirbaha, H., Pournaseh, M., & Sagan, O. (2014). Can music lessons increase the performance of preschool children in IQ tests? *Cognitive Processing, 15*, 77–84. https://doi.org/10.1007/s10339-013-0574-0

*** Kinney, D. W. (2008). Selected demographic variables, school music participation, and achievement test scores of urban middle school students. *Journal of Research in Music Education, 56*(2), 145–161. https://doi.org/10.1177/0022429408322530**

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., & White-Schwoch, T. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *Journal of Neuroscience, 34*(36), 11913–11918. https://doi.org/10.1523/JNEUROSCI.1881-14.2014

*** Legette, R. M. (1993). The effect of a selected use of music instruction on the self-concept and academic achievement of elementary public school students (Doctoral dissertation, The Florida State University, School of Music).**

Luna, F. G., Marino, J., Roca, J., & Lupiáñez, J. (2018). Executive and arousal vigilance decrement in the context of the attentional networks: The ANTI–Vea task. *Journal of Neuroscience Methods, 306*, 77–87. https://doi.org/10.1016/j.jneumeth.2018.05.011

*** MacCutcheon, D., Füllgrabe, C., Eccles, R., van der Linde, J., Panebianco, C., & Ljung, R. (2019). Investigating the effect of one year of learning to play a musical instrument on speech-in-noise perception and phonological short-term memory in 5-to-7-year-old children. *Frontiers in Psychology, 10*, 2865. https://doi.org/10.3389/fpsyg.2019.02865**

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences, 115*(51), 13129-13134. https://doi.org/10.1073/pnas.1811793115

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 69*(5), 1091–1119. https://doi.org/10.1111/rssc.12440

Medina, D., & Barraza, P. (2019). Efficiency of attentional networks in musicians and non-musicians. *Heliyon, 5*(3), e01315. https://doi.org/10.1016/j.heliyon.2019.e01315

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. *Perspectives on Psychological Science, 11*(4), 512–534. https://doi.org/10.1177/1745691616635612

Moher D, Liberati A, Tetzlaff J, Altman DG, & The PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

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

*** Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., ... & Desimone, R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences, 115*(28), E6630-E6639. https://doi.org/10.1073/pnas.1808412115**

*** Nering, M. E. (2002).** ***The effect of piano and music instruction on intelligence of monozygotic twins*** **(Doctoral dissertation, University of Hawaii).**

*** Orsmond, G. I., & Miller, L. K. (1999). Cognitive, musical and environmental correlates of early music instruction.** ***Psychology of Music, 27*****(1), 18–37. https://doi.org/10.1177/0305735699271003**

Patscheke, H., Degé, F., & Schwarzer, G. (2016). The effects of training in music and phonological skills on phonological awareness in 4-to 6-year-old children of immigrant families. *Frontiers in Psychology*, *7*, 1647. https://doi.org/10.3389/fpsyg.2016.01647

*** Pelletier, H. W. (1963).** ***An investigation of the relation between training in instrumental music and selected aspects of language growth in third grade children*** **(Doctoral dissertation, Arizona State University, College of Education).**

*** Portowitz, A., Lichtenstein, O., Egorova, L., & Brand, E. (2009). Underlying mechanisms linking music education and cognitive modifiability.** ***Research Studies in Music Education, 31*****(2), 107–128. https://doi.org/10.1177/1321103X09344378**

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods, 10*(1), 57-71. https://doi.org/10.1002/jrsm.1332

Rabinowitch, T. C., Cross, I., & Burnard, P. (2013). Long-term musical group interaction has a positive influence on empathy in children. *Psychology of Music, 41*(4), 484–498. https://doi.org/10.1177/0305735612440609

*** Rauscher, F., Shaw, G., Levine, L., Wright, E., Dennis, W., & Newcomb, R. (1997). Music training causes long-term enhancement of preschool children's spatial–temporal reasoning. *Neurological Research, 19*(1), 2–8. https://doi.org/10.1080/01616412.1997.11740765**

*** Rauscher, F. H., & Zupan, M. A. (2000). Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: A field experiment. *Early childhood Research Quarterly, 15*(2), 215–228. https://doi.org/10.1016/S0885-2006(00)00050-8**

Rickard, N. S., Appelman, P., James, R., Murphy, F., Gill, A., & Bambrick, C. (2013). Orchestrating life skills: The effect of increased school-based music classes on children's social competence and self-esteem. *International Journal of Music Education, 31*(3), 292–309. https://doi.org/10.1177/0255761411434824

Rodgers, M. A., & Pustejovsky, J. E. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*. https://doi.org/10.1037/met0000300

Roediger III, H. L. (2013). Applying cognitive psychology to education: Translational educational science. *Psychological Science in the Public Interest, 14*(1), 1–3. https://doi.org/10.1111/medu.12141

*** Roden, I., Kreutz, G., & Bongard, S. (2012). Effects of a school-based instrumental music program on verbal and visual memory in primary school children: A longitudinal study. *Frontiers in Neuroscience, 6*, 572. https://doi.org/10.3389/fpsyg.2012.00572**

*** Roden, I., Grube, D., Bongard, S., & Kreutz, G. (2014). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal**

study. *Psychology of Music, 42*(2), 284–298. https://doi.org/10.1177/0305735612471239

*** Roden, I., Könen, T., Bongard, S., Frankenberg, E., Friedrich, E. K., & Kreutz, G. (2014). Effects of music training on attention, processing speed and cognitive music abilities—Findings from a longitudinal study. *Applied Cognitive Psychology, 28*(4), 545–557. https://doi.org/10.1002/acp.3034**

Rodrigues, A. C., Loureiro, M. A., & Caramelli, P. (2013). Long-term musical training may improve different forms of visual attention ability. *Brain and Cognition*, *82*(3), 229–235. https://doi.org/10.1016/j.bandc.2013.04.009

Román-Caballero, R., Arnedo, M., Trivino, M., & Lupiáñez, J. (2018). Musical practice as an enhancer of cognitive function in healthy aging-A systematic review and meta-analysis. *PloS ONE, 13*(11), e0207957. https://doi.org/10.1371/journal.pone.0207957

Román-Caballero, R., Martín-Arévalo, E., & Lupiáñez, J. (2020). Attentional networks functioning and vigilance in expert musicians and non-musicians. *Psychological Research*. https://doi.org/10.1007/s00426-020-01323-2

*** Rose, D., Jones Bartoli, A., & Heaton, P. (2019). Measuring the impact of musical learning on cognitive, behavioural and socio-emotional wellbeing development in children. *Psychology of Music, 47*(2), 284–303. https://doi.org/10.1177/0305735617744887**

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Beverly Hills, CA: Sage. https://doi.org/10.4135/9781412984997

*** Said, P. M., & Abramides, D. V. M. (2020). Effect of music education on the promotion of school performance in children. *CoDAS, 32*(1). http://doi.org/10.1590/2317-1782/20192018144**

Sala, G., & Gobet, F. (2017a). 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. https://doi.org/10.1016/j.edurev.2016.11.005

Sala, G., & Gobet, F. (2017b). Does far transfer exist? Negative evidence from chess, music, and working memory training. Current directions in psychological science, 26(6), 515–520. https://doi.org/10.1177/0963721417712760

Sala, G., & Gobet, F. (2019). Cognitive training does not enhance general cognition. *Trends in Cognitive Sciences*, *23*(1), 9-20. https://doi.org/10.1016/j.tics.2018.10.004

Sala, G., Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, *48*(8), 1429–1441. https://doi.org/10.3758/s13421-020-01060-2

*** Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15*(8), 511–514. https://doi.org/10.1111/j.0956-7976.2004.00711.x**

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology, 98*(2), 457–68. https://doi.org/10.1037/0022-0663.98.2.457

Schellenberg, E. G. (2020). Correlation = causation? Music training, psychology, and neuroscience. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(4), 475–480. https://doi.org/10.1037/aca0000263

*** Schellenberg, E. G., Corrigall, K. A., Dys, S. P., & Malti, T. (2015). Group music training and children's prosocial skills. *PLoS ONE, 10*(10), e0141449. https://doi.org/10.1371/journal.pone.0141449**

Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia, 33*(8), 1047–1055. https://doi.org/10.1016/0028-3932(95)00045-5

Seinfeld, S., Figueroa, H., Ortiz-Gil, J., & Sanchez-Vives, M. V. (2013). Effects of music learning and piano practice on cognitive function, mood and quality of life in older adults. *Frontiers in Psychology, 4,* 810. https://doi.org/10.3389/fpsyg.2013.00810

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do “brain-training” programs work? *Psychological Science in the Public Interest, 17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Slater, J., Skoe, E., Strait, D. L., O’Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural Brain Research, 291*, 244–252. https://doi.org/10.1016/j.bbr.2015.05.026

*** Slater, J., Strait, D. L., Skoe, E., O'Connell, S., Thompson, E., & Kraus, N. (2014). Longitudinal effects of group music instruction on literacy skills in low-income children. *PLoS ONE, 9*(11), e113383. https://doi.org/10.1371/journal.pone.0113383**

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *Journal of*

Neuroscience, 27(14), 3799–3806. https://doi.org/10.1523/JNEUROSCI.0147-07.2007

Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? *Intelligence, 62,* 119–124. https://doi.org/10.1016/j.intell.2017.03.005

Standley, J. M. (2008). Does music instruction help children learn to read? Evidence of a meta-analysis. *Update: Applications of Research in Music Education, 27*(1), 17–32. https://doi.org/10.1177/8755123308322270

Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science, 8*(5), 581–591. https://doi.org/10.1177/1948550617693062

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

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review, 120*(3), 439. https://doi.org/10.1037/a0033138

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PloS ONE, 12*(10), e0186773. https://doi.org/10.1371/journal. pone.0186773

Tervaniemi, M., Tao, S., & Huotilainen, M. (2018). Promises of music in education? *Frontiers in Education, 3*, 74. https://doi.org/10.3389/feduc.2018.00074

Thorndike, E. L. (1906). *The principles of teaching based on psychology*. New York, NY: A. G. Seiler.

*** Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences, 112*(32), 10062–10067. https://doi.org/10.1073/pnas.1505114112**

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20*(3), 375– 393. https://doi.org/10.1037/met0000011

Vaquero, L., Hartmann, K., Ripollés, P., Rojo, N., Sierpowska, J., François, C., ... & Münte, T. F. (2016). Structural neuroplasticity in expert pianists depends on the age of musical training onset. *Neuroimage*, *126*, 106–119. https://doi.org/10.1016/j.neuroimage.2015.11.008

Vaughn, K. (2000). Music and mathematics: Modest support for the oft-claimed relationship. *Journal of Aesthetic Education, 34*(3/4), 149–166. https://doi.org/ 10.2307/3333641

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika, 60*(3), 419–435. https://doi.org/10.1007/BF02294384

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. http://www.jstatsoft.org/v36/i03/

Wan, C. Y., & Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the life span. *The Neuroscientist, 16*(5), 566–577. https://doi.org/10.1177/1073858410377805

White-Schwoch, T., Carr, K. W., Anderson, S., Strait, D. L., & Kraus, N. (2013). Older adults benefit from music training early in life: Biological evidence for long-term training-driven plasticity. *Journal of Neuroscience, 33*(45), 17667–17674. https://doi.org/10.1523/JNEUROSCI.2560-13.2013

Whitlock, L. A., McLaughlin, A. C., & Allaire, J. C. (2012). Individual differences in response to cognitive training: Using a multi-modal, attentionally demanding game-based intervention for older adults. *Computers in Human Behavior, 28*(4), 1091–1096. https://doi.org/10.1016/j.chb.2012.01.012