



Near and far transfer: Is music special?

E. Bigand^{1,2}  · B. Tillmann^{3,4}

Accepted: 29 July 2021 / Published online: 30 August 2021
© The Psychonomic Society, Inc. 2021

Abstract

Far transfer rarely occurs, and a recent meta-analysis suggests that music is no exception. The overall effect of musical training on cognition was claimed to be null when considering studies with active control groups or implemented randomization procedures (Sala & Gobet, *Memory & Cognition*, 48: 1429–1441, 2020). Using the authors' data file and program (<https://osf.io/rquye>), we did not confirm the effect of randomization, and we demonstrated that their conclusion is based in part on the failure to differentiate near and far transfer, with near transfer effect sizes being selectively excluded for the musical training group studies, but not for the active control group studies. Reanalyzing their data file resulted in a significant effect size ($g = .234$), and also provided new evidence that far-transfer effects of musical training can challenge near-transfer effects of linguistic training. Music is a recreational activity that may be special in allowing for small but statistically significant far-transfer effects.

Keywords Music cognition · Meta-analyses · Music training · Near and far transfer

Introduction

Recent research investigating neuroplasticity has reignited the debate about near and far transfer, which has a long history in cognitive psychology. The human brain remains plastic throughout life, and this plasticity has important theoretical and social implications for health, wellness, and education. The possibility to compensate for aging and to rehabilitate brain-damaged patients or children with neurodevelopmental disorders by cognitive stimulation opens new perspectives and leads to understandable enthusiasm. However, this enthusiasm must be qualified: while there is no doubt that specific training improves performance directly linked to the trained skill (referred to as *near transfer*), it remains unclear whether training in a given task can improve other skills that are not directly related to the training activities (referred to as *far transfer*). Most cognitive stimulation programs claim to provide far-transfer effects, but there are almost as many studies

confirming this claim as those denying it. The heterogeneity of the training methods and tests used, the differences between participants, the relatively small sample sizes, and the types of implemented control groups undoubtedly contribute to this puzzling scenario. Three questions remain critical: Does far transfer following musical training exist? If yes, is its size of practical interest, and is it actually *caused* by the training program?

Meta-analyses of cognitive training programs

Numerous meta-analyses have led to intense moderation of the initial enthusiasm about far transfer. Neither working memory training, nor brain training, nor computer games, nor other more leisurely training, such as chess games, video games, or exergames, were found to provide far transfer (see Sala & Gobet, 2019a, for a review). A second-order meta-analysis of these meta-analyses reported no impact of training on far-transfer measures, regardless of the type of populations and cognitive training programs (Sala et al., 2019). According to the authors, these analyses provide converging evidence that when the allocation procedure (randomization) and the implementation of control groups (i.e., active groups) were controlled for, the far-transfer effects were almost nil, suggesting that previously reported far-transfer effects came mostly from scientifically poor empirical studies. The failure to observe far transfer suggests that “the lack of generalization of skills

✉ E. Bigand
Emmanuel.Bigand@u-bourgogne.fr

¹ University of Burgundy, LEAD, CNRS-UMR 5020, Dijon, France

² Institut Universitaire de France, Dijon, France

³ Lyon Neuroscience Research Center, CNRS, UMR5292, INSERM, U1028, Auditory Cognition and Psychoacoustics Team, F-69000 Lyon, France

⁴ University Lyon 1, F-69000 Lyon, France

acquired by training is thus an invariant of human cognition” (Sala et al., 2019, abstract). As a consequence, “researchers and policymakers should seriously consider stopping spending resources for this type of research. Rather than searching for a way to improve overall domain-general cognitive ability, the field should focus on clarifying the domain-specific cognitive correlates underpinning expert performance” (Sala & Gobet, 2017a, p. 519).

Meta-analyses are a powerful statistical tool, but have some limitations (Borenstein et al., 2009). Although numerous statistical procedures are available, their outcome can be based on several decisions performed without being double blind in an a posteriori process. Our present paper illustrates this limitation by showing how the authors’ decisions led to underestimate far-transfer effects for musical training (Sala & Gobet, 2020). Using open science resources of their publication (<https://osf.io/rquye>), we revisited their analysis by using their program and data pool, and reached a rather different conclusion.

Music and far transfer

Music is an interesting domain in which to investigate far transfer. It is a joyful activity, which is easily accessible over the lifespan, from young children to the elderly, and for patients with brain deficiency. It can be practiced alone or in a group. Music is also a demanding activity that requires numerous cognitive resources (Patel, 2011). It stimulates brain regions beyond the auditory cortex, including the frontal cortex and the motor cortex, emotional and reward systems, as well as socio-affective brain networks. Because the engaged neural network is rather large, music is a good candidate for far-transfer training, which could have “transformational power” over the brain (Patel, 2018). Correlational studies have reported brain differences associated with musical training (Herholz & Zatorre, 2011), and a recent meta-analysis confirmed that musically trained individuals show better memory performance than untrained individuals (Talamini et al., 2017). Correlational studies provide a necessary but insufficient demonstration for a *causal* relationship. The repetition of a demanding task over months and even years could plausibly shape the brain, but an alternative explanation is that only the smarter individuals manage to pursue this training. Put differently, music does not make people smarter, but smarter individuals are more likely to start to learn and to succeed to play music (Schellenberg, 2020). Experimental studies implementing musical training in a longitudinal approach have led to disparate findings (Sala & Gobet, 2017b). Once again, meta-analyses seem promising to further assess whether musical training might be inducing far-transfer effects.

To the best of our knowledge, seven published meta-analyses have addressed this issue to date. The first three have

included rather small sets of studies and effect sizes (Gordon et al., 2015; Hetland & Winner, 2001; Vaughn, 2000). A more elaborate meta-analysis was performed by Sala and Gobet (2017b) that included 38 studies investigating 3- to 16-year old children, leading to the inclusion of 118 effect sizes and 3,085 participants. This same pool of studies (minus two studies) was reanalyzed in a second-order meta-analysis (Sala et al., 2019). The majority of these studies were then combined with more recent studies in Sala and Gobet (2019b), leading to 43 studies, 204 effect sizes, and 3,780 participants. Finally, a selection of these studies was combined with 11 new studies in a multilevel meta-analysis, including 54 studies, 254 effect sizes, and 6,984 participants, which are the focus of our present paper (Sala & Gobet, 2020; referred to as S&G2020 hereafter). Another recent meta-analysis (Cooper, 2020) was performed with 21 studies and 100 effect sizes (many had also been included in Sala & Gobet, 2017b, 2020). In contrast to Sala and Gobet (2017b, 2019b, 2020), Cooper’s (2020) meta-analysis reported a moderate overall effect of musical training for both active and non-active control group studies ($g = .28$), but this effect failed to be significant for studies performed in a laboratory setting instead of classroom or community-center setting.

Sala and Gobet’s (2020) meta-analyses of music training programs

The present paper focuses on S&G2020’s multilevel meta-analysis approach. Several moderators relative to randomization, type of control, baseline differences, age, duration of training, and type of outcome measures were included. These outcome measures were organized into four categories: non-verbal ability (fluid reasoning, mathematical and spatial skills), verbal ability (vocabulary and reading skills, phonological processing), memory (short-term/working-memory tasks) and speed (processing speed and inhibition tasks). Three modeling approaches were used, notably the robust variance estimation (RVE), a random effect model (RE), and Bayesian analysis. Only far-transfer tests following musical training run with typically developing 3- to 16-year-old children were considered.

As a main outcome, S&G2020 reported an overall impact of music training programs on cognitive and academic outcomes ($g = 0.184$, $p < .001$) that dropped close to zero when either only the active control group studies ($g = 0.056$; $p = .350$) or the randomized non-active control group studies ($g = 0.064$, $p = .381$) were considered separately (see Table 1 for details). S&G2020 concluded that when confounding factors, such as type of controls or the lack of random assignment of participants to the groups, were neutralized, the overall effect of music training was null. Neither age, duration of training, and outcome

Table 1 Summary presentation of the analyses presented in Sala and Gobet (2020, p. 1435f)

	Main analyses	Sensitivity analysis Step 1 (3 studies removed)	
	<i>m</i> = 54; <i>k</i> = 254	<i>m</i> = 51; <i>k</i> = 235	
RVE	g = 0.184; SE = 0.041; p < .001 df = 38.36; 95% CI [0.101; 0.268] τ^2 = 0.041; I ² = 43.16%	g = 0.166; SE = 0.041; p < .001 df = 34.9; 95% CI [0.083; 0.249] τ^2 = 0.036; I ² = 40.62%	
RE	g = 0.176; SE = 0.037; p < .001 τ^2 = 0.033	g = 0.149; SE = 0.035; p < .001 τ^2 = 0.024	
	Non-active control group studies	Non-active control group studies	Step 2 (5 influential cases removed)
	<i>m</i> = 41; <i>k</i> = 144	<i>m</i> = 40; <i>k</i> = 139	Non-active control group studies
RVE	g = 0.228; SE = 0.045; p < .001 df = 30.1; 95% CI [0.137; 0.320] τ^2 = 0.042; I ² = 43.11%	g = 0.226; SE = 0.045; p < .001 df = 29.2; 95% CI [0.133; 0.319] τ^2 = 0.041; I ² = 42.96%	<i>m</i> = 39; <i>k</i> = 134 g = 0.181; SE = 0.042; p < .001 df = 21.9; 95% CI [0.093; 0.268] τ^2 = 0.018; I ² = 24.92%
RE	g = 0.201; SE = 0.041; p < .001 τ^2 = 0.023	g = 0.200; SE = 0.041; p < .001 τ^2 = 0.024	g = 0.161; SE = 0.037; p < .001 τ^2 = 0.013
Moderator	<i>p</i> = .487 ¹	<i>p</i> = .480 ¹	<i>p</i> = .042 ²
Randomization			<i>Non-randomized studies</i> g = 0.246; SE = 0.049; p < .001 95% CI [0.140; 0.352] <i>Randomized studies</i> g = 0.064; SE = 0.065; p = .381 95% CI [-0.116; 0.244]
	Active control group studies	Active control group studies	
	<i>m</i> = 23; <i>k</i> = 110	<i>m</i> = 20; <i>k</i> = 96	
RVE	g = 0.056; SE = 0.058; p = 0.350 df = 12.6; 95% CI [-0.069; 0.182] τ^2 = 0.025; I ² = 23.10%	g = -0.021; SE = 0.032; p = .558 df = 4.2; 95% CI [-0.109; 0.068] τ^2 = 0; I ² = 0	
RE	g = 0.090; SE = 0.060; p = .136 τ^2 = 0.032	g = -0.010; SE = 0.035; p = .787 τ^2 = 0	

Note: *m* refers to the number of studies, *k* refers to the number of effect sizes. We used S&G2020’s R script and data file, available at <https://osf.io/rquye>. Supplementary Table 1 lists the studies excluded at the various analysis steps

¹ We ran this moderator analysis (not reported in the S&G2020 article) with the authors’ program. It included the three moderators Randomization, Baseline, and Age

² The authors’ program showed (lines 737-739) that the authors ran three separate analyses, one for each moderator: (1) randomization, (2) baseline, (3) age. When we ran one moderator analysis with the three moderators, similar to the other moderator analyses performed before by the authors (lines 579 and 665), the influence of the moderator Randomization was at *p* = .08

measures were found to have a significant contribution. Accordingly, “researchers’ optimism about the benefits of music training is empirically unjustified and stems from misinterpretation of the empirical data and, possibly, confirmation bias” (Sala & Gobet, 2020, page 1429). This finding was considered to be consistent with their previous conclusions, summarized as “Music is over” (Sala & Gobet, 2017b) and “Elvis has left the building” (Sala & Gobet, 2019b). For the authors, “the obvious practical implication is that music training should not be used as a tool for cognitive enhancement” (Sala & Gobet, 2019b, page 991) and “Educators and policymakers should be aware that music training provides no benefits on non-music-related cognitive or academic skills” (ibid.).

We here propose to reconsider their conducted meta-analysis and its conclusions in three steps. We focus first on the potential influence of randomization. S&G2020 found a significant effect of this factor, but Sala and Gobet (2019b) and Sala and Gobet (2017b, after their sensitivity analysis) did not. We then demonstrate for S&G2020 that the active control group studies instilled an unfair comparison, notably by including near-transfer effects in the control group studies, but only far-transfer effects in the musical training studies. Finally, using S&G2020’s data file and R program (<https://osf.io/rquye>), we ran a set of meta-analyses that removed both concerns aiming to have a more appropriate estimation of the effect of music training (here based on the studies included in their data file).

Revisiting Sala and Gobet's (2020) meta-analyses on musical training

Randomization

One of the two main conclusions of S&G2020 was that the observed effect of musical training vanishes when only randomized studies were considered. Randomization was not a significant moderator in their previous meta-analysis (Sala & Gobet, 2019b), which involved 204 of the 254 effect sizes of S&G2020. Randomization was also not a significant moderator in the main analysis of S&G2020 ($p = .518$,¹ based on all effect sizes). In contrast to Sala and Gobet (2019b), S&G2020 ran a two-step sensitivity analysis. After the first step of the sensitivity analysis of S&G2020, randomization was not a significant moderator ($p = .693$), but type of control was (as in the main analysis). This led S&G2020 to perform the subgroup analyses: For the non-active control group studies, the effect size remained significant ($g = 0.226$; $p < .001$, see Table 1, middle), and no moderator analysis was reported at that point. When running a moderator analysis with their program, it revealed that randomization was not a significant moderator ($p = .480$). At that stage, the second step of S&G2020's sensitivity analysis intervened: S&G2020 ran an influential case analysis and wrote "Five effect sizes were found to be significantly inflating the true heterogeneity" (p. 1435). Removing these values resulted in the reduced g of 0.181 (Table 1, right), and a moderator analysis was run. Although not explicitly stated in the article, this moderator analysis was slightly different from that performed for the other analyses. Instead of running one moderator analysis with all moderators (here randomization, baseline, age), the program (see lines 737–739 of their program) reveals that S&G2020 ran here three separate moderator analyses, each one with one moderator. With this change in model, randomization was a significant moderator, $p = .042$. However, when we ran one moderator analysis with the three moderators (similar to the other moderator analyses performed by S&G2020, see lines 579 and 665 of their program), the contribution of randomization did not reach significance ($p = .08$). When performing the analysis with their program, we also observed that the influential case analysis suggests nine influential effect sizes for this data set, and not five as stated by S&G2020 (p. 1435). When all nine influential cases were removed, the effect size for the non-active control group studies remained significant (g

$= 0.203$; $p < .0001$; see Table 2, left) and, most importantly, heterogeneity dropped to 0 (RVE: I^2 and τ^2), indicating that these additional four influential cases were actually increasing heterogeneity. Once again randomization was not a significant moderator ($p = .194$ or $p = .158$ when running one or three moderator analyses, respectively).

In sum, these findings suggest that randomization is not a robust moderator, and was obtained by S&G2020 via a two-step sensitivity analysis, with the second step applying the influential case analysis only to the non-active control group studies, by removing only five influential cases (out of nine) and changing the implementation of the moderator analysis. In contrast to S&G2020, Sala and Gobet (2019b) ran a simpler, one-step process: an influential case analysis was run on *all* studies of the main analysis and did not find evidence for a significant influence of the moderator randomization. We applied this one-step process to the present data file of S&G2020 (i.e., the full data set as used in S&G2020's main analysis). The influential case study revealed 16 influential effect sizes (see Online Supplementary Table 1 for details of studies excluded for all analyses reported in this article, available at <https://osf.io/w5kx9/>). Without them, the effect size was significant ($p < .0001$; see Table 2, right), and the moderators randomization and type of control did not reach significance ($p = .476$ and $p = .064$). For comparison purposes with S&G2020, we nevertheless ran the separate analysis for non-active control group studies, and the effect size remained significant ($g = 0.202$; $p < .0001$; see Table 2, right), the heterogeneity was 0 (I^2 and τ^2) and randomization was not a significant moderator ($p = .185$ or $p = .157$ when running one or three moderator analyses, respectively). (See Online Supplementary Material for Bayesian analyses.) In agreement with the previous findings of Sala and Gobet (2019b), our reanalysis here provides converging evidence that randomization is not a significant moderator.

Near versus far transfer in control versus experimental training programs

The second main conclusion of S&G2020 is that the effect of musical training is null for active control group studies. In the following, we demonstrate that this conclusion is based on their failure to differentiate far transfer from near transfer. When active control groups perform sport, computer, or video game activities, the various pre- and post-test tasks measure far-transfer effects, as for the experimental musical training groups. However, when the control group follows drama lessons and is evaluated, just like the musical training group, on linguistic performance, this raises the question of the relevance of the active control group and its equidistance to the tests. Given that drama stimulates different facets of linguistic abilities, drama training is closer to the linguistic target tasks

¹ The authors ran this moderator analysis, but the article did not provide this p -value; we recalculated it with the open source program provided by S&G2020.

Table 2 Alternative sensitivity analyses of Sala and Gobet's data set, see main text for details

	Sensitivity analysis Step 2	Influential case analysis on all studies (main analysis data)
RVE		$m = 51; k = 238$ $g = 0.171$; SE = 0.038; $p < .0001$ df = 33.40; 95% CI [0.095; 0.248] $\tau^2 = 0.015$; $I^2 = 14.41\%$
RE		$g = 0.152$; SE = 0.032; $p < .0001$ $\tau^2 = .018$
	Non-active control group studies Step 2 (all 9 influential cases removed) $m = 38; k = 130$	Non-active control group studies $m = 39; k = 134$
RVE	$g = 0.203$; SE = 0.036; $p < .0001$ df = 19.77; 95% CI [0.128; 0.278] $\tau^2 = 0$; $I^2 = 0\%$	$g = 0.202$; SE = 0.035; $p < .0001$ df = 20.18; 95% CI [0.129; 0.276] $\tau^2 = 0$; $I^2 = 0\%$
RE	$g = 0.175$; SE = 0.034; $p < .0001$ $\tau^2 = .001$	$g = 0.170$; SE = 0.032; $p < .0001$ $\tau^2 = 0.0002$
Moderator Randomization	$p = .194^1$	$p = .185^2$

Note: m refers to the number of studies, k refers to the number of effect sizes. Note that for the influential case analysis on all studies (Table 2, right), the moderators randomization and type of control did not reach significance ($p = .48$ and $p = .06$, respectively), but we reported the non-active control group studies data separately for comparison with Table 1 and Table 2. left. See Online Supplementary Material for Bayesian analyses. We used S&G2020's R script and data file, available at <https://osf.io/rquye>. Online Supplementary Table 1 lists the studies excluded at the various analysis steps

¹ All three moderators (i.e., randomization, baseline, age) were entered in one moderator analysis. When three separate analyses were ran, then the moderator randomization was at $p = .158$

² All three moderators (i.e., randomization, baseline, age) were entered in one moderator analysis. When three separate analyses were ran, then the moderator randomization was at $p = .157$

than is musical training, and we know from a previous meta-analysis that a group with classroom drama training (i.e., enacting text) outperforms an active control group (passive reading) on different verbal skills, such as writing, story understanding and recall, as well as oral understanding (Hetland & Winner 2001). The concern raised here becomes even more important when the active control group is directly trained on linguistic tasks that are similar to the target tasks used in pre- and post-tests. For instance, when the control group receives phonological training and is tested on phonological awareness, or when the control group is trained on reading and evaluated for reading, the active control group is tested for near transfer, while the musical training group is tested for far transfer (with these same language tests). This results in a somewhat biased or unfair comparison, notably with effect sizes having a different meaning here than do effect sizes in an equidistant control training implementation. In such an unequal comparison, an effect size close to zero does not mean that musical training does not create transfer effects. It indicates that musical training creates far-transfer effects that are not stronger than the near-transfer effects instilled by the given control trainings. S&G2020 were well aware of differences between far transfer and near transfer, and they correctly removed all effect sizes associated with musical tests (e.g., pitch, rhythm) and even environmental sound discrimination tests. However, they did not

apply the same caution to the active control group studies. This unequal treatment leads to underestimation of the effect of musical training in the analyses.²

Aiming to assess the extent of this underestimation, we removed from S&G2020's data file the 21 effect sizes that are related to the most unbalanced comparisons. As the boundary between “near” and “far” transfer might be a matter of debate, we excluded only effect sizes that tap into highly similar constructs in training and test in the control group: 18 effect sizes were related to the comparison with an active control group receiving phonological training and being tested on phonological processing, two effect sizes were related to the comparison with an active control group receiving a reading intervention and being tested on reading, and one effect size was related to the comparison with an active control group receiving visual art lessons and being tested on visual form analysis (see Online Supplementary Table 1 for details). All other effect sizes, including those linked to drama or dance, remained included. Using the program of S&G2020,

² One might argue that phonological processing may benefit from musical training because musical activities train listening skills. However, phonological tests tap into processes that go beyond sound discrimination and that have no equivalent in the musical training program, such as for example rhyme detection, phoneme-grapheme correspondence or phoneme deletion, which requires participants to remove one phoneme from a nonword to create a word. As a consequence, phonological training is necessarily closer to the phonological tasks than is musical training.

we observed comparable (or even slightly increased) effect sizes for this reduced data set (RVE: $g = 0.208$; $p < .0001$; RE: $g = 0.195$; $p < .0001$; see Table 3, left, for details), with similar heterogeneity to the main analysis of the authors (see Table 1, left). As the authors did in the main analysis, we ran a moderator analysis with type of control, randomization, baseline, and age as moderators, but neither randomization nor type of control was significant ($p = .610$ and $p = .193$, respectively) (see Online Supplementary Material for complementary analyses).

To further investigate whether *near*-transfer effects induced by linguistic training (and art lessons for one measure, i.e., the set of active control training studies described above) was significantly stronger than *far*-transfer effects induced by musical training, a subgroup analysis was performed on the 21 removed effect sizes (corresponding to eight studies). We observed an effect size of $g = -0.126$ ($SE = .071$; 95% CI [-0.350; 0.099]; $df = 3.01$; $p = .17$; $I^2 = 0$; $\tau^2 = 0$) with the RVE model, and $g = -0.117$ ($SE = .107$; $p = .275$; $\tau^2 = 0$) for the RE model. This absence of difference was further supported by Bayesian analyses; the Bayes factor ($BF_g = 0.447$) provided some evidence that g was more likely to be null than non-null (i.e., H_0 almost 2.24 times more likely to be true than H_1) (see Online Supplementary Material for details).³ This finding provides new evidence suggesting that *far*-transfer effects induced by musical training could even compete with *near*-transfer effects induced by linguistic training (and art lessons for one measure).

Meta-analysis without post-test-only studies

A further concern with S&G2020's sensitivity analysis was to include studies that did not report pretest measures of the targeted tests. In these cases, the program of S&G2020 assumes that experimental and control groups did not differ at pre-test (coded as a difference of 0 in the baseline moderator), which is certainly unlikely, in particular in developmental psychology. In another meta-analysis run with a similar set of studies, the authors had excluded post-test-only studies (Sala et al., 2019), and this was also done by Gordon et al. (2015). We thus applied this rationale to the present data set (i.e., removing all effect sizes without pre-test measures from

³ One of the included studies (Herrera et al., 2011) used a phonological outcome measure and compared experimental and control groups that both received phonological training (i.e., near transfer) with the experimental group performing this training in music. In response to the argumentation that this study should not have been included here, we reran the analysis by excluding the eight effect sizes of this study. The results of this additional analysis confirmed the previous finding, notably an effect size of $g = -0.126$ ($SE = .076$; 95% CI [-0.385; 0.131]; $df = 2.68$; $p = .20$; $I^2 = 0$; $\tau^2 = 0$) with the RVE model, and $g = -0.118$ ($SE = .110$; $p = .287$; $\tau^2 = 0$) for the RE model. This absence of difference was further supported by Bayesian analyses; the Bayes factor ($BF_g = 0.383$) provided evidence that g was more likely to be null than non-null (see Online Supplementary Material for details).

the data file of S&G2020), while still focusing on studies testing for far-transfer effects. This analysis confirmed a significant overall effect size (RVE: $g = 0.243$; $p < .0001$ RE: $g = .226$; $p < .0001$; see Table 3, middle), and neither randomization nor type of control was a significant moderator ($p = .676$ and $p = .181$, respectively) (see Online Supplementary Material for complementary analyses).

We then applied an influential case analysis on the effect sizes of this data set, following the procedure of Sala and Gobet (2019b). This influential case analysis led us to remove seven effect sizes (i.e., six were positive), which reduced heterogeneity (see Table 3, right, for details).⁴ The overall effect size was significant (RVE: $g = 0.234$; $p < .0001$; RE: $g = 0.213$; $p < .0001$), and again, the moderators type of control and randomization were not significant ($p = .163$ and $p = .319$, respectively) (see Online Supplementary Material for complementary analyses).

Our finding is consistent with two recent meta-analyses reporting a significant effect size of musical training, albeit slightly stronger, $g = .26$ in Román-Caballero et al. (2021) and $g = .28$ in Cooper (2020). These small differences in effect sizes might be explained by several minor decisions made by the different authors about potential selection or merging of effect sizes for a given study. For example, Cooper (2020) included six positive effect sizes for Bilhartz et al. (1999; notably $d = .37$, $d = .56$, $d = .68$, $d = .70$, $d = .75$, $d = .78$), while S&G2020 included only one effect size ($d = .19$). Similarly, for the study of Costa Giomi (2004), three positive effect sizes were included by Roman-Caballero et al. (2021, notably $d = .34$, $d = .40$, and $d = .53$), while S&G2020 included only one ($d = .209$). On the contrary, S&G2020 included all 26 effect sizes of Rickard et al. (2012), who failed to find an effect of music training, while Cooper (2020) included only ten. These observations suggest that S&G2020 favored an approach with more conservatism, which is further supported by other changes between Sala and Gobet (2017b) and S&G2020 (i.e., from a set of 13 effect sizes, ten effect sizes decreased and two positive effect sizes were excluded in S&G2020). All of these points might contribute to underestimating the potential effect size of music training in S&G2020, and when revisiting their data file, we inherited this tendency.

This said, it might be argued that effect sizes ranging from .234 to .28 remain small according to Hattie's (2008) barometer of influence. In his book, Hattie (2008) analyzed more than 800 meta-analyses and reported that the median value of intervention effect sizes in education is 0.40. One simplistic way to understand this contribution would be to recommend that all effects below this value should be ignored, as 50% of

⁴ Note that in comparison to the data set of S&G2020 used in their main analyses, the data set used here was overall reduced by as many effect sizes in randomized (26) as in non-randomized (27) studies.

Table 3 Reanalysis of the data set of Sala and Gobet (2020) with a new approach (see main text for details)

	Removal of near-transfer active control group studies	Additional removal of studies without pre-tests	Additional removal of studies without pre-tests, combined with influential case analyses (7 effect sizes)
	$m = 51; k = 233$	$m = 42; k = 197$	$m = 41; k = 190$
RVE	$g = 0.208$; SE = 0.043; $p < .0001$ df = 36.99; 95% CI [0.121; 0.295] $\tau^2 = 0.044$; $I^2 = 45.81\%$	$g = 0.243$; SE = 0.052; $p < .0001$ df = 34.41; 95% CI [0.136.; 0.345] $\tau^2 = 0.059$; $I^2 = 36.42\%$	$g = 0.234$; SE = 0.046; $p < .0001$ df = 28.84; 95% CI [0.141; 0.327] $\tau^2 = 0.017$; $I^2 = 13.03\%$
RE	$g = 0.195$; SE = 0.039; $p < .0001$ $\tau^2 = 0.033$	$g = 0.226$; SE = 0.045; $p < .0001$ $\tau^2 = 0.037$	$g = 0.213$; SE = 0.041; $p < .0001$ $\tau^2 = 0$

Note: m refers to the number of studies, k refers to the number of effect sizes. Note that for these three analyses, the moderators randomization and type of control did not reach significance (see main text for details). For comparison purpose with S&G's analyses, we present separate analyses for the non-active and active control group studies respectively in the Online Supplementary Material. We used S&G2020's R script and data file, available at <https://osf.io/rquye>. Online Supplementary Table 1 lists the studies excluded at the various analysis steps

all interventions obtained at least such an effect. Along this line, the effects of music training on cognitive abilities could be considered as too small to be of any practical use, a view consistent with S&G2020's conclusion. However, according to Hattie (2008), this effect size of .40 "is not a magic number that should become like a $p < .05$ cut-off point" (page 17). "Effect lower than $d = .40$ can be regarded as a need of more consideration, although it is not as simple as saying that all effect below $d = .40$ are not worth having" (page 16). "There are many examples that show small effects may be important" (page 9), and Hattie insists on the fact that the value of an effect size also depends on the cost of its implementation. For instance, the effect of homework, which is typically $d = .29$ according to Hattie (page 234), is of interest because of its low cost of implementation. A similar situation occurs with music training, which is a recreational activity with low cost and with effect sizes likely to be in the same range as visual/audio-visual learning ($d = .22$, page 229) or programmed instruction ($d = .24$, page 231) and larger than effect sizes of extra curricula activities ($d = .17$, page 159), sport ($d = .10$), and numerous teaching approaches explicitly designed to improve achievement, such as "mentoring" ($d = .15$, page 188), aptitude treatment interaction ($d = .19$, page 194), problem-based learning ($d = .15$, page 211), web-based learning ($d = .18$, page 227), or home school programs ($d = .16$, page 234). A more recent publication, which involved 1,200 meta-analyses (Hattie, 2015), even reported an effect size of .37 for music-based programs, which thus placed music at rank 94 among the 195 variables that influence school achievement.

Conclusion

Over the last 5 years, Sala and Gobet have published several meta-analyses providing converging evidence that cognitive training does not enhance general cognition (see Sala & Gobet

2019a, for a review). Their finding about music training fits well with this claim (Sala & Gobet, 2017b, 2019b, 2020) and has led them to conclude that "researchers' optimism about the benefits of music training is empirically unjustified and stems from misinterpretation of the empirical data and, possibly, confirmation bias" (Sala & Gobet, 2020, page 1429).

Thanks to the resources made available by S&G2020 within the Open Science Framework related to their article, we revisited their meta-analysis of 2020. We provided some evidence that their findings are based on decisions that led them to underestimate potential far-transfer effects created by music training. Our findings show the importance of testing experimental and control groups for far transfer. Without such a fair comparison, the effect of music training is underestimated. As all meta-analyses of Sala and Gobet on musical training included studies with an unbalanced far- versus near-transfer comparison, which is unfavorable to musical training, their conclusions need to be re-evaluated. This issue also applies to Sala et al.'s (2019) second-order meta-analysis that combined their meta-analysis on music training (Sala & Gobet, 2017b) with meta-analyses on other cognitive training types. Revisiting this second-order meta-analysis is now needed to further evaluate whether music might actually be a special case allowing for far transfer.

Our findings converge with those of other recent meta-analyses towards a consistent conclusion for the first two questions we raised in our introductory section: yes, there is a significant effect of musical training on general cognition, and it can be considered of practical interest, even if its size remains small according to Hattie (2008)'s barometer of influence. It would not seem reasonable to expect that a couple of hours of musical training per week could place music among the most efficient educational interventions (i.e., those with effect sizes d superior to .40). The fact that this recreational activity of low implementation cost succeeds to create significant far-transfer effects to general cognition is of

interest for cognitive psychology, and it may have practical implications for educational science. The third question raises the need to specify whether music may *cause* these benefits. Revisiting S&G2020's data file here supported a general effect of music training, which is not significantly modulated by randomization or type of control. A causal interpretation thus cannot be rejected, even though this point needs further new studies that should rigorously control for randomization, fair active control group comparisons, the inclusion of both pre- and post-test measurements, as well as the measurement of IQ at baseline. For now, all findings together lead us to conclude that music is not over and Elvis is still on stage.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13421-021-01226-6>.

Acknowledgements We thank G. Sala and F. Gobet for our e-mail exchanges related to their meta-analyses and the information provided about their work. The team *Auditory Cognition and Psychoacoustics* of the Lyon Neuroscience Research Center is part of the framework LabEx CeLyA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060).

Authors' contributions The authors contributed equally to this article.

Funding Auditory Cognition and Psychoacoustics Team is part of the framework LabEx CeLyA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060).

Data availability The data set and R program used in the present article are those of Sala and Gobet (2020), who made them openly available at the Open Science Framework (OSF) at <https://osf.io/rquye/>. The Online Supplementary Table 1 reproduces the data file of Sala and Gobet (2020) with additional information about the excluded studies in each of the analyses reported in Tables 1, 2 and 3. It is openly available at the OSF at <https://osf.io/w5kx9/>.

Declarations

Conflicts of interest/Competing interests N/A

Ethics approval N/A

Consent to participate N/A

Consent for publication N/A

Code availability We used the program (R script) of Sala & Gobet (2020; <https://osf.io/rquye/>).

References

- Bilhartz, T. D., Bruhn, R. A., & Olson, J. E. (1999). The effect of early music training on child cognitive development. *Journal of Applied Developmental Psychology, 20*, 615–636.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons.
- Cooper, P. K. (2020). It's all in your head: A meta-analysis on the effects of music training on cognitive measures in school children. *International Journal of Music Education, 38*:3, 321–336 <https://doi.org/10.1177/0255761419881495>.
- 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>
- 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>
- Hattie, J. A. C. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- Hattie, J. (2015). The Applicability of Visible Learning to Higher Education. *Scholarship of Teaching and Learning in Psychology, 1* (1), 79–91. <https://doi.org/10.1037/stl0000021>
- Herholz, S. & Zatorre, R. (2011). 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., & Winner, E. (2001). The arts and academic achievement: What the evidence shows. *Arts Education Policy Review, 102*, 3e6. <https://doi.org/10.1080/10632910109600008>.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2*, 142. <https://doi.org/10.3389/fpsyg.2011.00142>
- Patel, A.D. (2018) Music as a Transformative Technology of the Mind: An Update (pp. 113–126), In H. Honing (Ed.), *The origin of musicality*, MIT Press: Cambridge MA.
- Rickard, N. S., Bambrick, C. J., & Gill, A. (2012). Absence of widespread psychosocial and cognitive effects of school-based music instruction in 10–13-year-old students. *International Journal of Music Education, 30*, 57–78. <https://doi.org/10.1177/025576141431399>
- Román-Caballero, R., Vadillo, M. A., Trainor, L., & Lupiáñez, J. (2021). Please Don't Stop the Music: A Meta-Analysis of the Benefits of Learning to Play an Instrument on Cognitive and Academic Skills. <https://doi.org/10.31234/osf.io/4bm8v>
- Sala, G., & Gobet, F., (2017a). 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. (2017b). 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. (2019a). Cognitive training does not enhance general cognition. *Trends in Cognitive Sciences, 23*, 9–20. <https://doi.org/10.1016/j.tics.2018.10.004>
- Sala, G., & Gobet, F. (2019b). Elvis has left the building: Correlational but not causal relationship between music skill and cognitive ability. *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (pp. 987–993)*, Montreal, Canada.
- Sala, G., N. D. Aksayli, K. S. Tatlidil, T. Tatsumi, Y. Gondo & F. Gobet (2019) Near and Far Transfer in Cognitive Training: A Second-Order Meta-Analysis. *Collabra: Psychology, 5*(1): 18. <https://doi.org/10.1525/collabra.203>
- Sala, G. & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition, 48*:1429–1441; <https://doi.org/10.3758/s13421-020-01060-2>
- Schellenberg, E. G. (2020). Music Training, Individual Differences, and Plasticity (pp. 415–442), In M.S.C. Thomas, D. Mareschal & I. Dumontheil, *Educational Neuroscience, Frontiers of Developmental Science*, Routledge Taylor & Francis Group: New York and London.
- Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLoS One 12*: e0186773. <https://doi.org/10.1371/journal.pone.0186773>

- Vaughn, K. (2000). Music and mathematics: Modest support for the oft-claimed relationship. *Journal of Aesthetic Education*, 34, 149e166. <https://doi.org/10.2307/3333641>.
- Herrera, L., Lorenzo, O., Defior, S., Fernandez-Smith, G., & Costa-Giomi, E. (2011). Effects of phonological and musical training on the reading readiness of native- and foreign-Spanish-speaking

children. *Psychology of Music*, 39, 68–81. <https://doi.org/10.1177/0305735610361995>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.