Article:
Aryanto, C.B., Aisyah, A.R.K., Timmers, R. et al. (2026) How is musical expertise related to executive functions? A systematic review and meta-analysis. Psychonomic Bulletin and Review, 33 (3). 64. ISSN: 1069-9384

https://doi.org/10.3758/s13423-025-02769-4

Word count: ~9616

How Is Musical Expertise Related to Executive Functions? A Systematic Review and Meta-Analysis

Christ B. Aryanto[1, 2], Aireen R. K. Aisyah[2], Renee Timmers[3], Emma Blakey[1], Claudia C. von Bastian[1]

[1]School of Psychology, University of Sheffield, UK

[2]Faculty of Psychology, Atma Jaya Catholic University of Indonesia, Indonesia

[3]School of Languages, Arts and Societies, University of Sheffield, UK

Author Note

Christ B. Aryanto https://orcid.org/0000-0001-6347-2505

Aireen R. K. Aisyah

Renee Timmers https://orcid.org/0000-0002-1981-0834

Emma Blakey https://orcid.org/0000-0003-3685-3649

Claudia C. von Bastian https://orcid.org/0000-0002-0667-2460

This study has been preregistered on Open Science Framework (https://osf.io/evj5d). The data and analysis scripts are available on the Open Science Framework (https://osf.io/8h6tc/).

Parts of these findings were presented as a poster at Psychonomics 62nd Annual Meeting, virtual conference, 2021, and European Society for Cognitive Psychology, Lille, France, 2022, and as a spoken presentation at Learning and Plasticity Meeting, Finland, 2022. This study has been published as a preprint on the PsyArxiv (https://doi.org/10.31234/osf.io/6r4d5).

This research was supported by a PhD scholarship to Christ B. Aryanto from the Indonesia Endowment Fund for Education, Ministry of Finance, Republic of Indonesia (Lembaga Pengelola Dana Pendidikan, Kementerian Keuangan, Republik Indonesia) with the scholarship ID number 20200422341285 and is part of his doctoral thesis. We are grateful to all the authors who kindly responded to our queries and who provided additional data for the meta-analysis. We further thank Dilara Steenken for her assistance in the literature search.

Correspondence concerning this article should be addressed to Christ B. Aryanto, Faculty of Psychology, Atma Jaya Catholic University of Indonesia, Jl. Jenderal Sudirman No.51, Jakarta, 12930, Indonesia, or Claudia C. von Bastian, School of Psychology, Interdisciplinary Centre of the Social Sciences (ICOSS), 219 Portobello, Sheffield, S1 4DP, United Kingdom. E-mails: christ.billy@atmajaya.ac.id, c.c.vonbastian@sheffield.ac.uk.

Disclaimer: This manuscript has been peer-reviewed and accepted for publication in Psychonomic Bulletin & Review (3rd November 2025). This version serves as the author accepted manuscript, representing the final manuscript version after peer review and acceptance, but before copyediting and typesetting by the publisher. This is not the final copy of record and may contain minor differences in formatting, layout, or style compared to the authoritative document published later by the journal. When citing this work, please use the final published article citation once it is available.

How Is Musical Expertise Related to Executive Functions? A Systematic Review and Meta-Analysis

Abstract

A person's lifetime involvement with musical expertise has been hypothesized to be positively associated with executive functions, including inhibition, shifting, and working memory. However, results of past research have been inconclusive. This preregistered systematic review and three-level meta-analysis of 47 studies encompassing 235 effect sizes from 4651 healthy adult participants investigated how strongly musical expertise relates to each of the three executive functions. The results showed significant medium associations between musical expertise and executive functions ($g = 0.43$), with small to medium associations between musical expertise and each of the functions: inhibition ($g = 0.31$), shifting ($g = 0.22$), and working memory ($g = 0.49$). Risk of bias moderated the relationship between musical expertise and inhibition, and the paradigm used to assess executive functions moderated the association between musical expertise and working memory. The results suggest that individuals who engage in musical activities have higher levels of executive function, with a particularly critical role of working memory in musical expertise. The literature review further identified several methodological issues, including predominant reliance on dichotomizing continuous variables and the use of small samples yielding low statistical power. The review offers methodological recommendations and directions for further investigating the relation between working memory and musical expertise.

Keywords: Musical expertise, executive functions, meta-analysis

Abstrak (Indonesian Language)

Keterlibatan seseorang sepanjang hidup dalam keahlian bermusik telah dihipotesiskan berhubungan positif dengan fungsi eksekutif, yaitu inhibisi, peralihan (*shifting*), dan memori kerja. Namun, hasil penelitian sebelumnya menunjukkan temuan yang tidak konsisten. Kajian sistematis yang telah diregistrasi sebelumnya dan meta-analisis tiga tingkat pada 47 studi yang mencakup 235 ukuran efek dari 4651 partisipan dewasa sehat ini menyelidiki sejauh mana keahlian bermusik berkaitan dengan masing-masing dari tiga fungsi eksekutif tersebut. Hasil penelitian menunjukkan adanya hubungan signifikan dengan ukuran sedang antara keahlian bermusik dan fungsi eksekutif (g = 0,43), serta hubungan lemah hingga sedang antara keahlian bermusik dan masing-masing fungsi eksekutif: inhibisi (g = 0,31), peralihan (g = 0,22), dan memori kerja (g = 0,49). Risiko bias memoderasi hubungan antara keahlian bermusik dan inhibisi, sedangkan paradigma yang digunakan untuk mengukur fungsi eksekutif memoderasi hubungan antara keahlian bermusik dan memori kerja. Temuan ini menunjukkan bahwa individu yang terlibat dalam aktivitas musikal memiliki tingkat fungsi eksekutif yang lebih tinggi, dengan peran memori kerja yang sangat penting dalam keahlian musik. Kajian literatur juga mengidentifikasi beberapa isu metodologis, termasuk banyaknya penelitian sebelumnya yang melakukan dikotomisasi variabel kontinu dan penggunaan sampel kecil yang menghasilkan daya statistik yang rendah. Kajian ini menawarkan rekomendasi metodologis serta arahan untuk penelitian lebih lanjut mengenai hubungan antara memori kerja dan keahlian bermusik.

Kata kunci: Keahlian bermusik, fungsi eksekutif, meta-analisis

How Is Musical Expertise Related to Executive Functions? A Systematic Review and Meta-Analysis

Executive function is an umbrella term for higher-order cognitive skills underpinning goal-directed behavior (Friedman & Miyake, 2017; Miyake & Friedman, 2012). Executive functions are essential when people make decisions, plan, and solve problems in everyday life (Ferguson et al., 2021). Higher levels of executive functions have been shown to be associated with academic attainment, positive levels of physical and mental health, marital satisfaction, and less social problems (Diamond, 2013).

Past research has established three factors of executive functions: inhibition, shifting, and working memory updating (Miyake et al., 2000). *Inhibition* is a deliberate, intended, and controlled action to override dominant but incorrect responses. *Shifting*, also referred to as cognitive flexibility, is the ability to switch from one task to another, activating the currently relevant task and inhibiting the currently irrelevant task. Finally, *updating* refers to replacing no longer relevant information in working memory with new information. These three executive functions have shown to be correlated yet distinct, leading to the development of the unity/diversity framework of executive functions (for a review, see Miyake & Friedman, 2012). The commonality (or *unity*) between the three executive functions suggests that a common underlying ability exists which presumably serves maintenance of task goals and goal-related information. The separability (or *diversity*) of executive functions suggests additional individual differences in more specific functions. Previous studies have shown that, once unity is accounted for, diversity only emerges for shifting and updating, which both demonstrate specific individual differences. In contrast, individual differences in inhibition can be fully explained by their commonalities with individual differences in shifting and updating.

Playing music and musical training have been shown to relate to cognitive function in general (Schellenberg & Weiss, 2013; Schellenberg & Lima, 2024) and executive functions more specifically (Okada & Slevc, 2020). Here, we use the term *musical expertise* to refer to people's lifetime engagement with *actively making* music, including regularly playing an instrument and singing with or without formal training, musical sophistication (e.g., measured using Ollen Musical Sophistication index; Ollen, 2006), and musical ability (e.g., measured using Melodic Discrimination Testing; Harrison et al., 2017), as opposed to passively or receptively listening to music. Associations between musical expertise and executive functions gave rise to the idea that musical training could be a viable intervention to improve executive functions (Bugos, 2019; Bugos et al., 2022; Frischen et al., 2019; for reviews, see Degé & Frischen, 2022; Rodriguez-Gomez & Talero-Gutiérrez, 2022). However, like many interventions aimed at boosting cognition (see von Bastian et al., 2022), the effectiveness of musical training interventions is highly contentious (Sala & Gobet, 2017b, 2020; for a review, see Schellenberg & Lima, 2024). Before even considering musical training as means for interventions, however, it is imperative to first to better understand whether and how musical expertise is related to executive functions.

Musical expertise is thought to be related to cognitive performance because it is a complex activity that involves processes akin to executive functions. For example, musicians need to maintain and process the musical information while playing their musical instruments or singing, and, at the same time, also adjust for the tempo, volume, and timbre (Okada & Slevc, 2020). Furthermore, musicians must coordinate and shift their attention to sensory inputs, such as a different sound, hold their impulse to play the music at the right tempo, and anticipate the music to play the right notes (Okada & Slevc, 2020).

However, while intuitively plausible, the empirical evidence for this relationship between musical expertise and executive functions is inconsistent. Studies measuring single executive functions separately found that, relative to non-musicians, musicians performed better in tasks assessing inhibition (i.e., showed smaller effects of cognitive conflict; e.g., Moussard et al., 2016), shifting (e.g., Moradzadeh et al., 2015; Moser, 2003), and working memory (e.g., Franklin et al., 2008; Gagnon & Nicoladis, 2021; Grassi et al., 2017). Yet, studies that measured all three factors of executive functions in the same sample of participants often found that only one or two of these factors were related to musical expertise. Specifically, some studies found that musical expertise correlated with inhibition and working memory but not with shifting (Okada & Slevc, 2018; Porflitt & Rosas-Díaz, 2019). Another study found that musical expertise was related only to working memory, but not inhibition and shifting (Slevc et al., 2016). Although these studies did not explicitly model the unity and diversity of executive functions, these patterns of varying correlations may suggest that musical expertise more likely taps specific aspects of executive function rather than commonality.

The present review systematically synthesizes these different findings to examine the relation between musical expertise and inhibition, shifting, and working memory updating across the body of the existing literature. By meta-analytically investigating the relationship with musical expertise both across all executive functions indicators and for each executive function separately, we gauged the evidence for general and specific associations. Moreover, we tested the impact of at least four critical methodological differences between studies that may explain the inconsistent results of past empirical studies.

First, a major difference between studies is how musical expertise is defined and assessed. Musical expertise can encompass concepts such as music training, musical ability, and musical sophistication. Music training is often assessed by self-reports of years of musical training or playing an instrument (e.g., Gray & Gow, 2020; Lu & Greenwald, 2016), and sometimes simply by asking whether participants regard themselves as musicians. Much of past research used only such self-reports (Criscuolo et al., 2019; D'Souza et al., 2018; Okada & Slevc, 2018). In contrast, musical ability refers to music perception (e.g., der Nederlanden et al., 2020) and music production (e.g., Okada, 2018), and it is typically assessed through performance-based measures. Finally, musical sophistication pertains to a person's perception of their own music-making habits and musical skills (Müllensiefen et al., 2014), and is often assessed through a mix of self-report and performance-based measures. Accordingly, some past studies combined self-report questionnaires with performance-based measures such as melodic and rhythm test scores (Hansen et al., 2013; Slevc et al., 2016; Talamini et al., 2016).

Assessing musical expertise solely by self-report measures can be problematic. While self-report measures have the advantage that they allow for assessing people's lifetime engagement with musical activity and how they perceive their own musical ability, these self-reported data are prone to measurement noise that can be systematically related to individual differences in other traits. For example, participants may give inaccurate responses because they have forgotten for how long they have played music, under- or overestimate their capacity to play music, or exhibit social desirability bias. Performance-based measures can complement the results of the self-report measures by providing a more objective assessment of musical skill and sophistication. This variability in definitions and measurement approaches may have contributed to inconsistent past findings. Therefore, the present work considers how musical expertise is operationalized in different studies to better understand its relation to executive functions.

Second, the criteria for classifying musicians and non-musicians vary between studies. For example, some studies classed participants as musicians if they reported more than eight years of music practice experience (D'Souza et al., 2018), whereas other studies set the threshold to only five years (Criscuolo et al., 2019). Yet other studies based their classification on a cut-off score for a musical questionnaire (Porflitt & Rosas-Díaz, 2019). In most cases, these criteria are often not further justified or explained. Yet, relations between musical expertise and executive functions may only emerge with a certain amount of musical expertise experience.

Third, a related but more fundamental issue concerns whether considering musical expertise as a continuous variable in correlational designs (i.e., years of music experience in a sample varying in their engagement with music) or as a categorical variable in between-groups designs (i.e., group comparisons of people classified as musicians and non-musicians) may lead to different associations with executive functions. For example, correlational studies found musical expertise and inhibition were unrelated (Okada & Slevc, 2018; Slevc et al., 2016), whereas group-comparison studies found increased inhibition in musicians compared to non-musicians (Criscuolo et al., 2019; Porflitt & Rosas-Díaz, 2019). Even *within* individual studies, the results varied depending on the design used. Porflitt and Rosas (2020) reported the results of both an analysis of covariance (ANCOVA), testing differences in cognitive performance between musicians and non-musicians, and correlations between musical sophistication and cognitive performance. While musicians performed significantly better in a shifting task than non-musicians, no significant correlations were found between musical sophistication as a continuous variable and shifting. Hence, based on the type of analysis, opposite conclusions could be drawn. However, artificially dichotomizing people into groups is analytically problematic and often yields misleading results (MacCallum et al., 2002). For example, the frequent practice of using median splits neglects that people around the median are more similar to each other than to the other members of their artificial groups. Indeed, median splits add random error that can lead to false-positive as well as false-negative results (McClelland et al., 2015). Using extreme groups by selecting only the top and bottom portion of the distribution (e.g., comparing people who never make music to professional musicians) is similarly problematic, as it increases the risk of inflated effect sizes, reduces measurement reliability, and may be based on erroneous assumptions about linearity and group membership (Preacher et al., 2005). Furthermore, people in these extreme groups will likely differ on several other, correlated dimensions (e.g., socioeconomic background, education, openness to new experiences, etc.), rendering it difficult to disentangle differences due to musical expertise from variations in these confounded variables (Unsworth et al., 2015).

Fourth, the concern related to extreme groups is even more aggravated as the impact of possible confounding – or moderating – variables is yet unclear. For example, the relation between musical expertise and executive functions decreased once demographic characteristics such as age, gender, education, and socioeconomic status were taken into account in some studies (Correia et al., 2023; Okely et al., 2022; Vincenzi et al., 2022) but not in others (Arndt et al., 2023; Criscuolo et al., 2019; D'Souza et al., 2018; Okada & Slevc, 2018; Slevc et al., 2016). The present review and meta-analysis will take into account these methodological concerns and examine the moderating effects of variation in the operationalization, assessment, and analytical treatment of musical expertise, and demographic characteristics (age, gender, socioeconomic status, and education).

Prior Reviews

Most prior reviews primarily focused on the effects of music training intervention on executive functions in children (Degé & Frischen, 2022; Rodriguez-Gomez & Talero-Gutiérrez, 2022, Román-Caballero et al., 2022), or cognitive abilities in general (Sala & Gobet, 2017a, 2017b, 2020). Three meta-analyses investigated the relationship between musical expertise and executive functions in adults. Hernández et al. (2020) identified twelve articles (k = 60) and found a large difference between musicians' and non-musicians' executive functions, d = 0.71, 95% CI [0.57, 0.85]. However, they included only group-comparison studies and considered executive functions only as a single variable, without distinguishing between inhibition, shifting, and working memory.

In another meta-analysis, Román-Caballero et al. (2018) included nine studies comparing inhibition, shifting, and working memory in adults older than 59 years who were classified either as musicians and non-musicians'. These group-comparison studies showed significant effects for inhibition, g = 1.77, 95% CI [0.60, 2.93], shifting, g = 0.57, 95% CI [0.00, 1.14], and verbal working memory, g = 0.88, 95% CI [0.03, 1.72]. However, these meta-analytic effects averaged only small numbers of effect sizes (k = 3 for inhibition and shifting; k = 6 for verbal working memory). Thus, any single additional study could strongly affect the findings.

Finally, Talamini et al. (2017) focused their meta-analysis on memory differences between groups of musicians (i.e., people who had a high level of formal music training) and non-musicians. On average, musicians showed better working memory performance than non-musicians, g = 0.56, 95% CI [0.33, 0.80]. However, again, this meta-analysis was based on a relatively moderate sample of only k = 19. Furthermore, it remains unclear how much formal training participants had, due to a lack of reporting in the original studies that consisted exclusively of group comparisons.

Present Meta-Analysis

The few existing previous meta-analyses focusing on adults suggest that musical expertise and executive functions are positively related. However, these findings are based on only small numbers of studies comparing categorical groups of musicians to non-musicians, thereby neglecting correlational studies. Therefore, it remains unclear how musical expertise and executive functions are related in adulthood when considering the degree of experience in musical expertise and when distinguishing between the three factors of executive functions. The present preregistered systematic review and meta-analysis fills these gaps by focusing on healthy, adult participants and include both studies comparing groups of musicians and non-musicians and studies reporting correlations between musical expertise and executive functions to address the following research questions:

1. How strongly are musical expertise and the three factors of executive functions (inhibition, shifting, and updating/working memory) related?
2. To what extent is the relationship between musical expertise and executive functions affected by moderators such as study type, risk of bias score, music measures, task paradigms, age, gender, socioeconomic status, and education?

Method

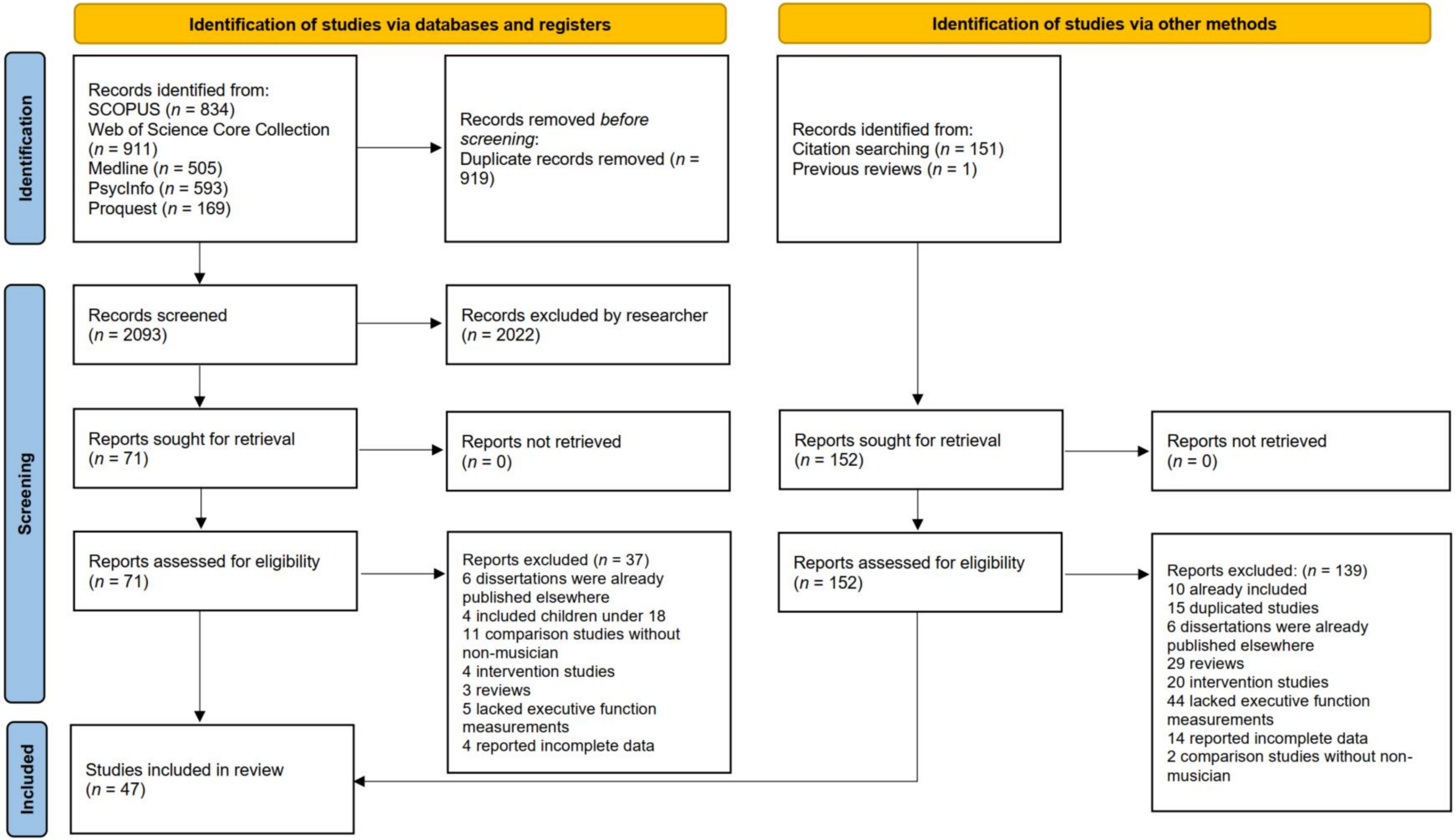
This meta-analysis followed the PRISMA statement 2020 (Page et al., 2021). The study protocol was preregistered on the Open Science Framework on January 19, 2021 (https://osf.io/evj5d).

Search Strategy

The flow diagram in Figure 1 summarizes the literature search. The search terms used in the study were the following: music* AND cogni* AND (“executive function*” OR “working memory” OR “cognitive control” OR “cognitive flexibility” OR “attention* control” OR “executive control” OR inhibit* OR shifting OR switching OR updating). A systematic search was conducted using Scopus, Web of Science Core Collection, Medline, and PsycInfo to identify peer-reviewed articles, and ProQuest Dissertations and Theses to identify grey literature. The databases were searched on February 8, 2021, and the search was conducted again on March 17, 2021, to add new reports, which resulted in 2,093 references after removing duplicates. An initial screening based on titles and abstracts was done, which resulted in excluding 2,022 references that did not meet the inclusion and exclusion criteria. The literature search and initial screening were completed by the first author (CBA). The remaining 71 references were assessed by CBA for eligibility for inclusion by obtaining and inspecting full texts. In addition, forward citation chasing based on eligible papers was done on April 25, 2021, which resulted in 152 additional references being inspected for eligibility.

Figure 1

PRISMA Flowchart of Systematic Search and Study Selection



Inclusion Criteria

We included published journal articles and unpublished undergraduate, master, and PhD theses written in English that reported quantitative data of any associations (i.e., zero-order correlations or comparisons of groups with high versus low levels of musical expertise) between the measures of at least one executive function (i.e., inhibition, shifting, updating/working memory) and musical expertise (e.g., musicality, musicianship level, years of music training, duration of music playing, music perception skills). Furthermore, we included only samples of healthy adults (at least 18 years old). Forty-seven of the 223 full texts assessed met these criteria.

Coding

The information from all articles was extracted by the first author (CBA) and the second author (ARKA), who independently used a pre-defined coding protocol to extract the following information: (1) General study information (title, author, publication year, country in which the research was done), (2) participants' characteristics, (3) type of study (correlational or group comparison), and (4) risk of bias. The complete coding protocol can be accessed from the Open Science Framework (https://osf.io/8h6tc/). Effect sizes were coded as Hedges' *g* to account for small sample bias. Hedges' *g* was computed from means and standard deviations and, where these were not available, converted from correlation coefficients or other effect size approximations (e.g., *r, t*; Harrer et al., 2021). Only the data from the full sample was coded for studies reporting effects for both a full and a subsample.

Risk of Bias Assessment

The risk of bias was assessed for all included studies using eight criteria adapted from the Risk of Bias Assessment Tool for Nonrandomized Studies (RoBANS; Kim et al., 2013): (1) Selection of participants, (2) confounding variables, (3) measures of exposure (cognitive assessment), (4) measures of exposure (music assessment), (5) blinding of participants, (6) blinding of experimenters, (7) incomplete outcome data, and (8) selective outcome reporting. Selection of participants referred to biases of selecting participants (e.g., musicians and non-musicians were not selected from comparable populations). Confounding variables referred to bias arising from the inadequate confirmation and consideration of confounding musical expertise with other variables such as participants socio-economic background. Measures of exposure referred to performance biases caused by inadequate measures for assessing cognitive and musical expertise variables (e.g., data were obtained through self-reported methods instead of performance-based tasks). Blinding of participants and experiments referred to their awareness to the hypotheses of the study (e.g., were participants recruited explicitly for a study investigating benefits of musical expertise). Incomplete outcome data referred to how missing data were handled and reported. Selective outcome reporting referred to omitting primary outcomes in the result section despite the study protocol.

Each study was independently rated by CBA and ARKA. Studies received -1 point for a low risk of bias, +1 point for a high risk of bias, and 0 points for an unclear risk of bias. Hence, high sum scores across all eight criteria reflect stronger risk of bias. Disagreements among coders were resolved through discussion and through consultation of a third reviewer (CvB).

Interrater Reliability

The interrater reliability was assessed for coding of the study type (i.e., comparison or correlation) and risk of bias. Interrater reliability was acceptable for study type, $\kappa = 0.67$. For the risk of bias variables, κ ranged from 0.04 (music assessment), to 0.23 (selection of participants), 0.66 (incomplete outcome data) to 1 (selective outcome). Interrater reliability could not be computed for cognitive assessment and blinding of experimenters because the two raters were in perfect agreement, rating all studies as having low bias and unclear bias, respectively. Interrater reliability could also not be computed for the two criteria confounding variables and blinding of participants, because one of the raters rated all studies the same. Even though disagreements were resolved, the poor interrater reliability, especially for music assessment and selection of participants, demands caution when interpreting the analysis outcomes of influences of risk of bias.

Meta-Analytic Procedure

Primary Analysis

A three-level meta-analytic approach was used to account for statistical non-independence between effect sizes (e.g., multiple tasks assessing the same outcome, multiple groups of musicians being compared to one control group of non-musicians) and between-study heterogeneity of effect sizes, providing estimates of the variance between (Level 3) and within (Level 2) studies. The restricted maximum-likelihood estimator from the *metafor* package was used (Version 3.4-0, Viechtbauer, 2010). Forest plots and funnel plots were created using code provided by Fernández-Castilla et al. (2020).

Moderator Analysis

We considered study type, the sum of the risk of bias score, music measure, task paradigm, age, gender distribution, socioeconomic status, and years of education, as potential moderators. We preregistered to analyse age and gender as potential moderators to explore whether basic demographic variables impact any relations found, without specific theoretical hypotheses. Each task paradigm was considered a separate category if at least 10 cases were available; otherwise, the paradigm was categorized as ‘others’. The gender distribution was counted based on the proportion of females to males.

Frequentist Inference

The homogeneity of effect sizes was evaluated with Q tests and complemented by examining I^2 and τ^2. Cochran’s Q is the difference between sampling error and between-study heterogeneity. I^2 estimates the variance of the effect sizes not caused by sampling error, and

τ^2 reflects the variance of the true effect size, both between and within studies. We conducted meta-regression omnibus *Q* tests for moderator analyses, complemented by subgroup comparisons examining overlap in confidence intervals. To reduce the risk of underpowered moderator analyses, we ran moderator analyses only for those categorical moderators where at least 10 cases per moderator level were available for analysis (Deeks et al., 2019), as per our pre-registered analysis plan.

Bayesian Inference

To evaluate the strength of evidence available in the included body of literature, we computed Bayes factors (BF) for each study, the average effect size per outcome, and the inclusion of moderators. We assumed medium effect sizes for the prior Cauchy distributions ($r = 0.5$ for average effect sizes and categorical moderators; $r = 0.3$ for continuous moderators). Furthermore, we conducted sensitivity analyses assuming small effect sizes ($r = 0.2$ for the average effect size and categorical moderators; $r = 0.1$ for continuous moderators) and large effect sizes ($r = 0.8$ for the average effect size and categorical moderators; $r = 0.5$ for continuous moderators). For the prior distribution of the homogeneity parameter τ, we assumed an inversed Gamma distribution ($\alpha = 1.23$, $\beta = 0.16$ and boundaries ranging from 0.01 to ∞). This prior was based on the heterogeneity observed in mean-difference effect sizes reported in *Psychological Bulletin* between 1990 and 2013 (Van Erp et al., 2017). All Bayesian analysis were computed using *brms* package (version 2.19.0; Bürkner, 2017, 2018)). BFs for each study were computed using *BayesFactor* package (version 0.9.12-4.4; Morey & Rouder, 2015) and the BF against a null region of $-\infty$ and 0 and marginal likelihood between the alternative model and the null model for the three-level meta-analysis and moderator analysis using *bayestestR* package (version 0.13.0; Makowski et al., 2019). BFs are reported using Wetzels and Wagenmakers's (2012) categorical verbal labels as guidelines to facilitate interpretation (see Table 1).

Table 1

Categorical Verbal Labels for Guiding Interpretation of Bayes Factors

Bayes Factors		Strength of evidence
BF_{10}	BF_{01}	
>100	<1/100	Decisive
30 to 100	1/100 to 1/30	Very strong
10 to 30	1/30 to 1/10	Strong
3 to 10	1/10 to 1/3	Substantial
1 to 3	1/3 to 1	Ambiguous
1	1	No evidence

Note. BF_{10} = evidence in favor of the alternative hypothesis, BF_{01} = evidence in favor of the null hypothesis.

Publication Bias

We assessed publication bias with funnel plots and by conducting Egger's and Begg’s and Mazumdar's tests. Egger's test aims to analyze the relationship between the standardized effect estimates and the standard error to determine the possibility of publication bias using linear regression (Egger et al., 1997). Begg’s and Mazumdar’s test determines whether there is a statistically significant correlation between the effect estimates' ranks and the variances of their values (Begg & Mazumdar, 1994). Both tests were conducted using the *metafor* package (Version 3.4-0, Viechtbauer, 2010).

Results

Table 2 lists the methodological details and effect sizes of the included studies. Of the 47 studies included in the meta-analysis, 33 reported between-group comparisons (k = 124) and 14 correlational associations (k = 111), from a total of 4651 participants (M age = 33.39 years, SD = 20.01 years). The correlational studies included in this systematic review investigated musicians only, combined the data of both musicians and non-musicians, or investigated students and then asked for their musical experience to ensure the degree varied from non-music players to highly experienced music players, which were sufficient to include a spread of musical expertise, as expected for correlational analysis. Most of the studies were conducted in North America (*n* = 29), followed by Europe (*n* = 9) and Asia (*n* = 6). The remaining studies were conducted in South America (*n* = 2) and Australia (*n* = 1). As the inclusion criteria were any healthy adults above 18 years old, the studies included 30 studies with young adults (e.g., D'Souza et al., 2018; Hou et al., 2014; Okada & Slevc, 2018; Slevc et al., 2016), 10 studies with older adults (e.g., Grassi et al., 2017; Gray & Gow, 2020; Hanna-Pladdy & Gajewski, 2012; Hanna-Pladdy & MacKay, 2011), and 7 studies did not specify participants' age (e.g., Anaya, 2013; Kempe et al., 2012; Porflitt & Rosas-Díaz, 2019; Strait et al., 2010).

Table 2

Sample, Study Type, Demographics, Pooled Effect Size, and Strength of Evidence for Studies Included in the Review

						g [95% CI]			BF_{10} [Sensitivity]		
Author	Study Type	*N*	Age	Sex	Education	Inhibition	Shifting	Working memory	Inhibition	Shifting	Working memory
Amer et al. (2013)	Comparison	42	60.00		18.56	0.41 [-0.22, 1.04]		0.87 [0.22, 1.52]	1.24 [1.44, 1.00]		11.31 [7.78, 11.60]
Anaya (2013)	Correlation	48		56.25	16.20	0.55 [-0.03, 1.13]		0.49 [-0.09, 1.07]	1/4.54 [1/2.32, 1/7.14]		1/3.85 [1/2.00, 1/5.88]
Bialystok & DePape (2009)	Comparison	71	24.23			0.88 [0.36, 1.40]	0.34 [-0.15, 0.83]	0.31 [-0.18, 0.80]	69.59 [41.76, 74.40]	1.18 [1.47, 1/1.10]	1.04 [1.34, 1/1.26]
Bianco et al. (2017)	Comparison	36	29.10	13.89	16.00	0.64 [-0.07, 1.35]			2.34 [2.21, 2.10]		
Blumenthal (2013)	Comparison	54	23.90			0.20 [-0.35, 0.75]		0.18 [-0.37, 0.73]	1/1.59 [1/1.09, 1/2.17]		1/1.67 [1/1.14, 1/2.32]
Caldwell (2015)	Correlation	12	20.80					1.09 [-0.14, 2.32]			1/1.82 [1/1.22, 1/2.50]
Chang-Arana & Luck (2018)	Comparison	29	27.42			-0.21 [-0.97, 0.55]			1/2.94 [1/1.78, 1/4.35]		
Clayton et al. (2016)	Comparison	34	21.48			0.00 [-0.67, 0.67]	-0.19 [-0.86, 0.48]	0.37 [-0.32, 1.06]	1/2.44 [1/1.51, 1/3.33]	1/3.33 [1/1.89, 1/4.77]	1.00 [1.22, 1/1.26]
Criscuolo et al. (2019)	Comparison	101	29.15	54.70	17.79	3.73 [3.08, 4.38]		2.33 [1.84, 2.82]	>100 [>100, >100]		>100 [>100, >100]
D'Souza et al. (2018), comp. 1	Comparison	81	22.05			0.33 [-0.11, 0.78]		0.92 [0.45, 1.39]	1.33 [1.66, 1.01]		>100 [>100, >100]
D'Souza et al. (2018), comp. 2	Comparison	72	22.00			0.16 [-0.31, 0.63]		0.30 [-0.17, 0.78]	1/1.82 [1/1.18, 1/2.56]		1.06 [1.38, 1/1.25]
der Nederlanden et al. (2020)	Correlation	60	20.94	45.00		-0.18 [-0.69, 0.33]	0.12 [-0.39, 0.63]	0.01 [-0.50, 0.52]	1/5.55 [1/2.63, 1/8.33]	1/5.55 [1/2.56, 1/8.33]	1/5.26 [1/2.50, 1/8.33]

Franklin et al. (2008)	Comparison	20	21.60	51.52	3.60			0.95 [0.00, 1.89]			3.23 [2.59, 3.23]
Gagnon & Nicoladis (2021)	Comparison	190	19.60					0.21 [-0.08, 0.51]			1.03 [1.55, 1.39]
Grassi et al. (2017)	Comparison	40	72.47	30.00	12.62			0.81 [0.09, 1.54]			7.35 [5.46, 7.27]
Gray & Gow (2020)	Correlation	30	69.20	50.00	15.90	0.36 [-0.37, 1.09]	0.40 [-0.33, 1.13]	0.98 [0.22, 1.74]	4.35 [1/2.22, 1/6.67]	1.4.54 [1/2.27, 1/6.67]	1/2.63 [1.51, 1/4.00]
Hanna-Pladdy & Gajewski (2012)	Comparison	70	68.63			-0.054 [-0.52, 0.42]	0.03 [-0.44, 0.50]	0.23 [-0.24, 0.70]	1/3.57 [1/1.96, 1/5.26]	1/2.86 [1/1.64, 1/4.17]	1/1.35 [1.05, 1/1.85]
Hanna-Pladdy & MacKay (2011)	Comparison	70	70.00		17.07		0.58 [0.05, 1.11]	0.43 [-0.08. 0.95]		4.27 [3.85, 3.77]	1.82 [2.01, 1.48]
Hansen et al. (2012)	Correlation	60	21.10	43.33	13.27			0.08 [-0.43, 0.59]			1/5.00 [1/2.44, 1/7.69]
Hou et al. (2014), comp. 1	Comparison	44	20.77	54.70		-0.41 [-1.02, 0.20]	-0.05 [-0.65, 0.56]	0.20 [-0.41, 0.81]	1/5.00 [1/2.63, 1/7.69]	1/2.86 [1/1.69, 1/4.17]	1/1.56 [1/1.10, 1/2.13]
Hou et al. (2014), comp. 2	Comparison	44	20.37	53.84		0.50 [-0.11, 0.20]	0.35 [-0.26, 0.95]	0.89 [0.25, 1.53]	1.82 [1.92, 1.53]	1.01 [1.25, 1/1.27]	14.70 [9.75, 15.32]
Jain & Nataraja (2019)	Comparison	51	33.28	0.00				0.56 [0.00, 1.13]			3.00 [2.85, 2.61]
Kempe et al. (2012)	Correlation	114		50.00		0.00 [-0.37, 0.37]		0.20 [-0.17, 0.57]	1/7.14 [1/3.12, 1/11.11]		1/6.25 [1/2.94, 1/10.00]
Lee et al. (2007)	Comparison	40	22.00					-0.08 [-0.70, 0.55]			1/2.94 [1/1.75, 1/4.17]
Lu & Greenwald (2016)	Correlation	60	22.62	90.00	15.62			1.32 [0.74, 1.91]			1/3.57 [1/1.85, 1/5.55]
Mansens et al. (2018)	Comparison	1101	74.35	26.25				0.24 [0.10, 0.38]			56.56 [76.96, 39.77]
Meyer et al. (2020)	Comparison	72	20.14			0.43 [-0.10, 0.96]	1.00 [0.45, 1.55]	0.75 [0.21, 1.28]	1.78 [1.97, 1.44]	>100 [84.84, >100]	14.00 [10.06, 13.72]
Moradzadeh et al. (2015), comp. 1	Comparison	81	22.05				0.29 [-0.14, 0.73]			1.08 [1.42, 1/1.25]	

Moradzadeh et al. (2015), comp. 2	Comparison	72	22.00				0.48 [0.01, 0.96]			3.19 [3.20, 2.67]	
Moser (2003)	Comparison	276					1.65 [1.38, 1.92]			>100 [>100, >100]	
Moussard et al. (2016)	Comparison	34	69.55	52.94	16.45	3.71 [2.59, 4.83]			>100 [>100, >100]		
Okada (2018)	Correlation	165	19.86	60.00				0.24 [-0.07, 0.55]			1/7.69 [1/3.70, 1/12.50]
Okada & Slevc (2018)	Correlation	150	19.26			0.28 [-0.04; 0.60]	-0.20 [-0.51, 0.11]	0.55 [0.22, 0.89]	1/7.14 [1/3.12, 1/11.11]	1/8.33 [1/3.70, 1/14.28]	1/6.25 [1/2.86, 1/10.00]
Parbery-Clark et al. (2011)	Comparison	37	54.50					0.82 [0.14, 1.49]			>100 [62.34, >100]
Porflitt & Rosas (2020)	Correlation	70	30.50	32.90		0.82 [0.34, 1.31]	0.36 [-0.11, 0.83]	1.06 [0.56, 1.57]	1/4.00 [1/2.00, 1/6.25]	1/4.76 [1/2.32, 1/7.69]	1/3.85 [1/1.92, 1/6.25]
Porflitt & Rosas-Díaz (2019)	Comparison	141				0.54 [0.15, 0.92]	0.32 [-0.05, 0.69]	0.88 [0.49, 1.27]	13.53 [11.43, 11.88]	1.61 [2.02, 1.22]	>100 [>100, >100]
Posedel et al. (2011)	Correlation	45	18.80	40.00				0.48 [-0.12, 1.08]			1/3.85 [1/1.96, 1/5.88]
Ramachandra et al. (2012)	Comparison	60	19.45					0.71 [0.20, 1.23]			12.11 [8.96, 11.68]
Schroeder et al. (2016)	Comparison	107	22.55	75.78		0.28 [-0.11, 0.67]			1.16 [1.55, 1/1.18]		
Slater et al. (2017)	Comparison	60	23.50	0.00		1.12 [0.55, 1.69]			>100 [>100, >100]		
Slevc et al. (2016)	Correlation	93	20.84			-0.07 [-0.48, 0.34]	-0.22 [-0.63, 0.19]	1.19 [0.74, 1.64]	1/6.67 [1/2.94, 1/10.00]	1/6.67 [1/3.12, 1/11.11]	1/4.17 [1/2.00, 1/6.25]
Strait et al. (2010)	Comparison	33						0.24 [0.93, 0.45]			1/3.45 [1/1.96, 1/5.00]
Strong & Mast (2019)	Comparison	58	73.13	49.69	16.10		0.55 [-0.08, 1.18]	0.17 [0.46, 0.80]		2.14 [2.14, 1.84]	1/1.69 [1/1.17, 1/2.32]
Strong & Midden (2020)	Comparison	46	73.30	52.91		0.23 [-0.40, 0.86]		0.29 [0.34, 0.92]		1/1.45 [1/1.05, 1.92]	1/1.23 [1.06, 1/1.61]

Suárez et al. (2016)	Comparison	54	22.59	80.00	15.09			0.62 [0.07, 1.17]			4.69 [4.06, 4.25]
Talamini et al. (2016)	Correlation	36	22.67	61.11	15.80			0.51 [-0.16, 1.18]			1/3.45 [1/1.82, 1/5.26]
Vasuki et al. (2016)	Comparison	40		74.75		0.15 [-0.48, 0.78]		1.16 [0.49, 1.83]	1/1.82 [1/1.22, 1/2.44]		65.04 [35.20, 76.23]
Vuvan et al. (2020)	Correlation	242	20.64	68.18				0.79 [0.53, 1.05]			1/7.14 [1/3.22, 1/11.11]
Weiss et al. (2014)	Comparison	75	23.35	64.18				0.21 [-0.25, 0.68]			1/1.41 [1.03, 1/1.96]
Zuk et al. (2014)	Comparison	30		40.00		0.35 [-0.38, 1.08]	0.56 [-0.18, 1.30]	1.19 [0.41, 1.97]	1/1.09 [1.14, 1/1.39]	1.63 [1/1.64], 1.40]	22.79 [13.12, 26.25]

Note. Age (years), sex (percentage of female participants relative to male participants), and education (years) scores are given in means. Bayes factors from sensitivity analyses for small-effect priors (*d* = 0.20) and large-effect priors (*d* = 0.80) are provided in angular brackets. BF_{10} = Bayes factor in favor of the alternative hypothesis; comp = comparison.

Three-Level Meta-Analysis

Influential Cases

We conducted outlier analyses to identify influential cases both across all executive function factors and for each factor of executive functions (see Supplementary Material for statistics and plots). Following Viechtbauer and Cheung's (2010) recommendation, for each study, we inspected the confidence interval not overlapping with the average confidence interval of the pooled effect and with a fit difference (DFFITS) above 0.4 *SD*s, Cook's distance above 0.15, and covariance ratio below 1. We flagged studies that fit those criteria, which resulted in removing the same two influential cases for executive functions and inhibition, and one influential case each for shifting and working memory. Here, we report the results without influential cases. Appendix A summarizes the results including those influential cases.

Musical Expertise and Executive Functions

First, we ran an omnibus three-level meta-analytic model across all three executive functions, with executive function factor (inhibition, shifting, and working memory) as moderator, thereby testing whether the association between musical expertise and executive functions depended on the specific executive functions factor assessed. We found a significant association supported by decisive evidence, $g = 0.43$, 95% CI [0.34, 0.53], $p < .001$, $BF_{10} > 100$. The heterogeneity was significant, $Q(df = 231) = 805.34$, $p < .001$, with an estimated true between-study variance of $\tau^2_{level\ 3} = 0.05$, and an estimated true within-study variance of $\tau^2_{level\ 2} = 0.11$. The proportion of variability from true heterogeneity relative to the sampling error was small for between-study variance, $I^2_{level\ 3} = 22.09\%$, and larger for within-study variance, $I^2_{level\ 2} = 51.41\%$. The average association was significantly moderated by executive functions factor, $F(2, 229) = 4.88$, $p = .008$. Although the Bayesian evidence for the presence of this moderation effect was only ambiguous, $BF_{10} = 1.39$, these results tentatively indicate potential differences in the strength of associations between musical expertise and single executive functions factors. Indeed, the association between musical expertise and working memory was significantly stronger (p = .015) than for the other executive functions.

Next, we ran separate meta-analyses for each executive functions factor despite the ambiguous evidence of the executive functions as the moderator because, theoretically, it was in line with the unity framework of executive functions model (Friedman & Miyake, 2017; Miyake et al., 2000; Miyake & Friedman, 2012). Table 3 summarizes the results. The three-level meta-analytic model for inhibition revealed a small but significant effect size supported by decisive evidence, g = 0.31, 95% CI [.16, .46], p < 0.001, BF_{10} > 100. Figure 2 shows a forest plot of pooled effect sizes for the association between musical expertise and inhibition. There was significant heterogeneity in the effect sizes, $Q(df = 60) = 182.09$, p < .001, with an estimated true between-study variance $\tau^2_{level\ 3}$ = 0.06, and an estimated true within-study variance of $\tau^2_{level\ 2}$ = 0.08. The proportion of variability from true heterogeneity relative to the sampling error was substantial for between-study variance, $I^2_{level\ 3}$= 27.63%, but bigger for within-study variance, $I^2_{level\ 2}$ = 40.51%.

Table 3

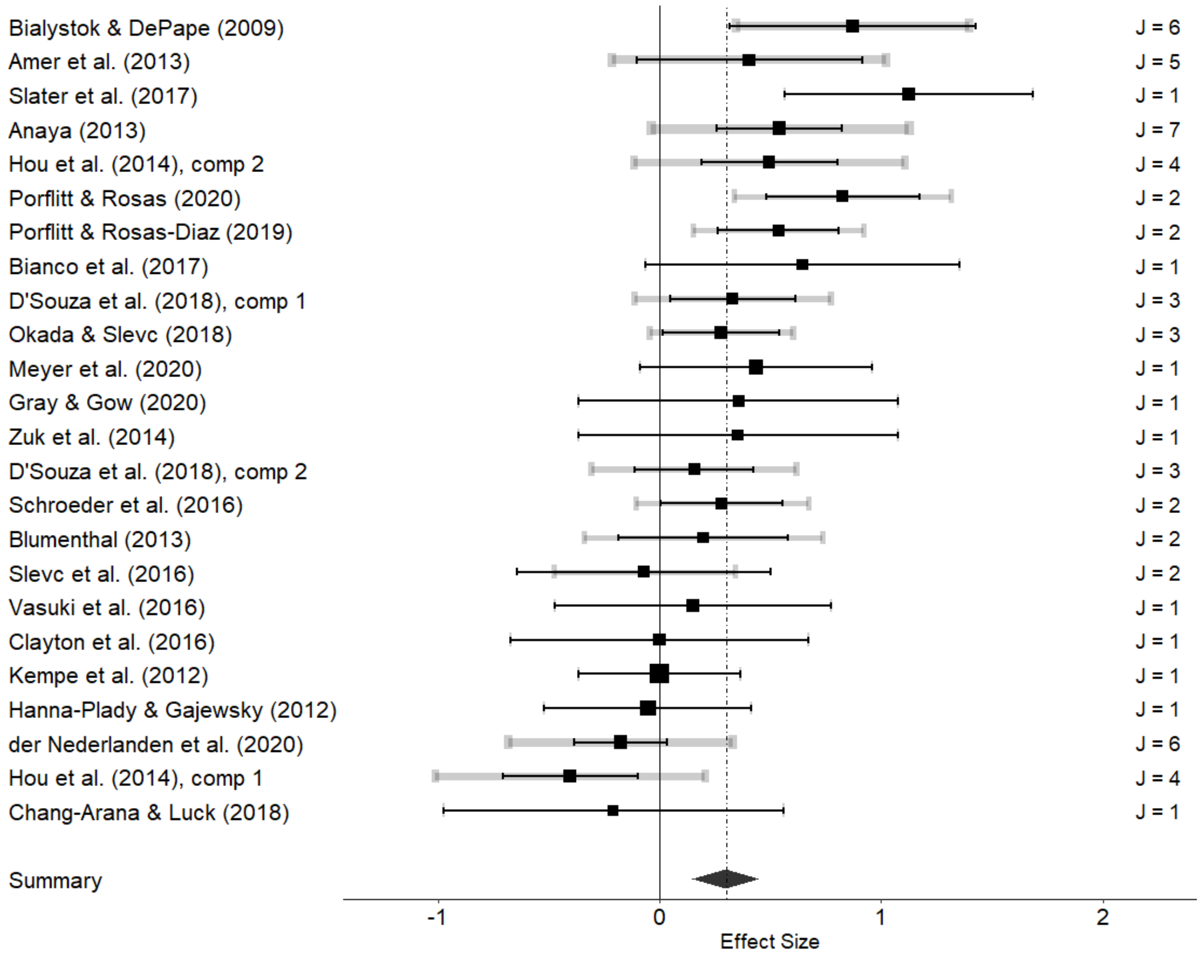
Average Effect Sizes for the Relation between Musical Expertise and Executive Functions

						I^2		τ^2	
Factors	*n*	*k*	*g*	95% CI	BF_{10} [Sensitivity]	Level 2	Level 3	Level 2	Level 3
Inhibition	15/7	39/22	0.31	[0.16 - 0.46]	>100 [>100, >100]	40.51%	27.63%	0.06	0.08
Shifting	12/5	31/13	0.22	[0.07 – 0.38]	>100 [73.99, >100]	0.77%	48.80%	<0.01	0.06
Working memory	24/14	50/76	0.49	[0.37 – 0.61]	>100 [>100, >100]	55.08%	20.98%	0.12	0.05

Note. All effect sizes were significant at $p < .001$. Bayes factors from sensitivity analyses for small-effect priors ($d = 0.20$) and large-effect priors ($d = 0.80$) are provided in angular brackets. Level 2 refers to the within-study level, level 3 to the between-study level. n = Number of studies (comparisons/correlations); k = Number of effect sizes (comparisons/correlations); CI = Confidence interval; BF_{10} = Bayes factor in favor of the alternative hypothesis.

Figure 2

Multilevel Meta-Analysis of Musical Expertise and Inhibition

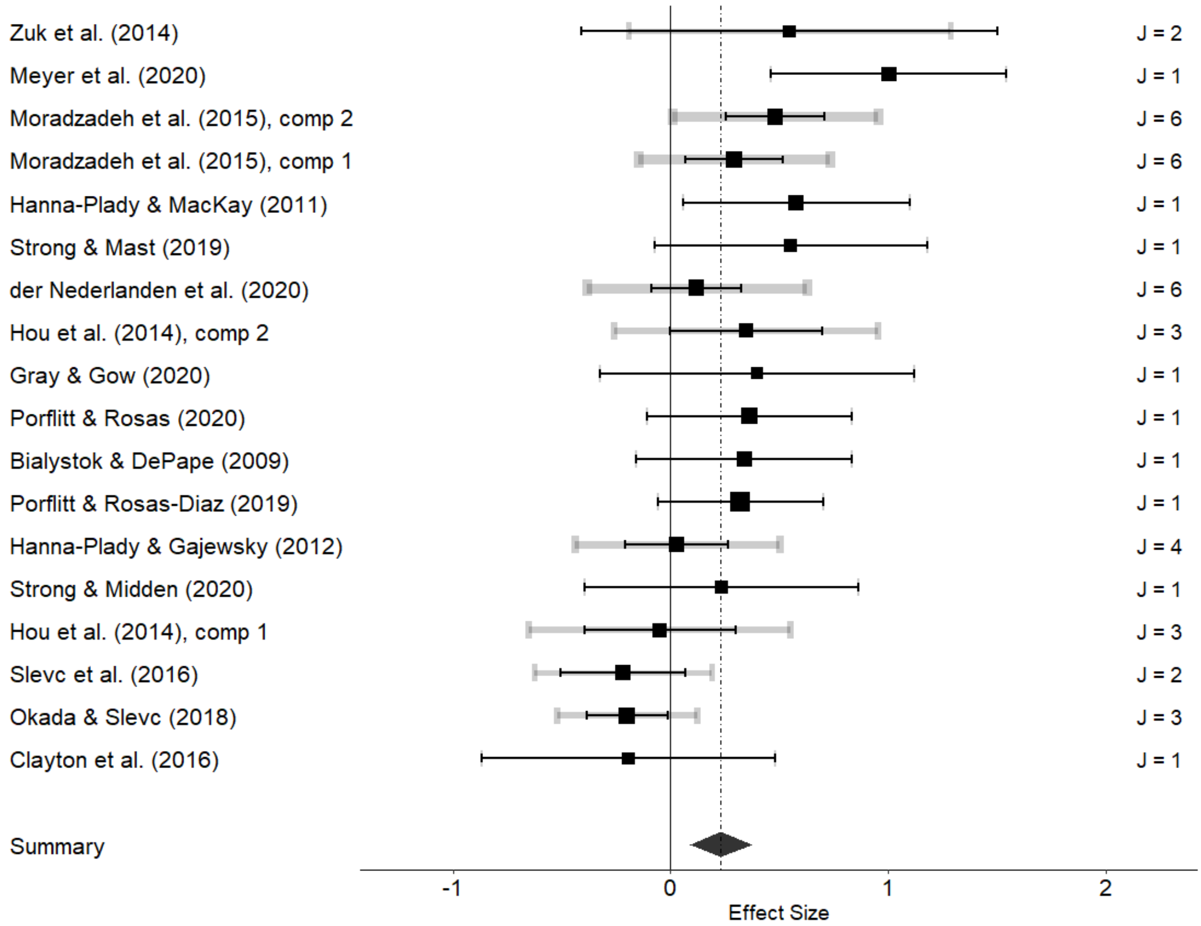


Note. Effect sizes are weighted averages and can include multiple effect sizes *J*. Black error bars represent 95% confidence intervals of the weighted averages. Gray error bars reflect the sampling variance of individual observed effect sizes of each study. The thickness of the gray error bars is proportional to the number of effect sizes within individual studies.

Musical expertise was also significantly related to the shifting factor (see Figure 3), with a small effect size supported by decisive Bayesian evidence, $g = 0.22$, 95% CI [0.07, 0.38], $p = .004$, $BF_{10} > 100$. The heterogeneity was significant, $Q(df = 43) = 82.07$, $p < .001$, $\tau^2_{level\ 3} = 0.06$, $\tau^2_{level\ 2} < 0.01$, $I^2_{level\ 3} = 48.80\%$, $I^2_{level\ 2} = 0.77\%$.

Figure 3

Multilevel Meta-Analysis of Musical Expertise and Shifting

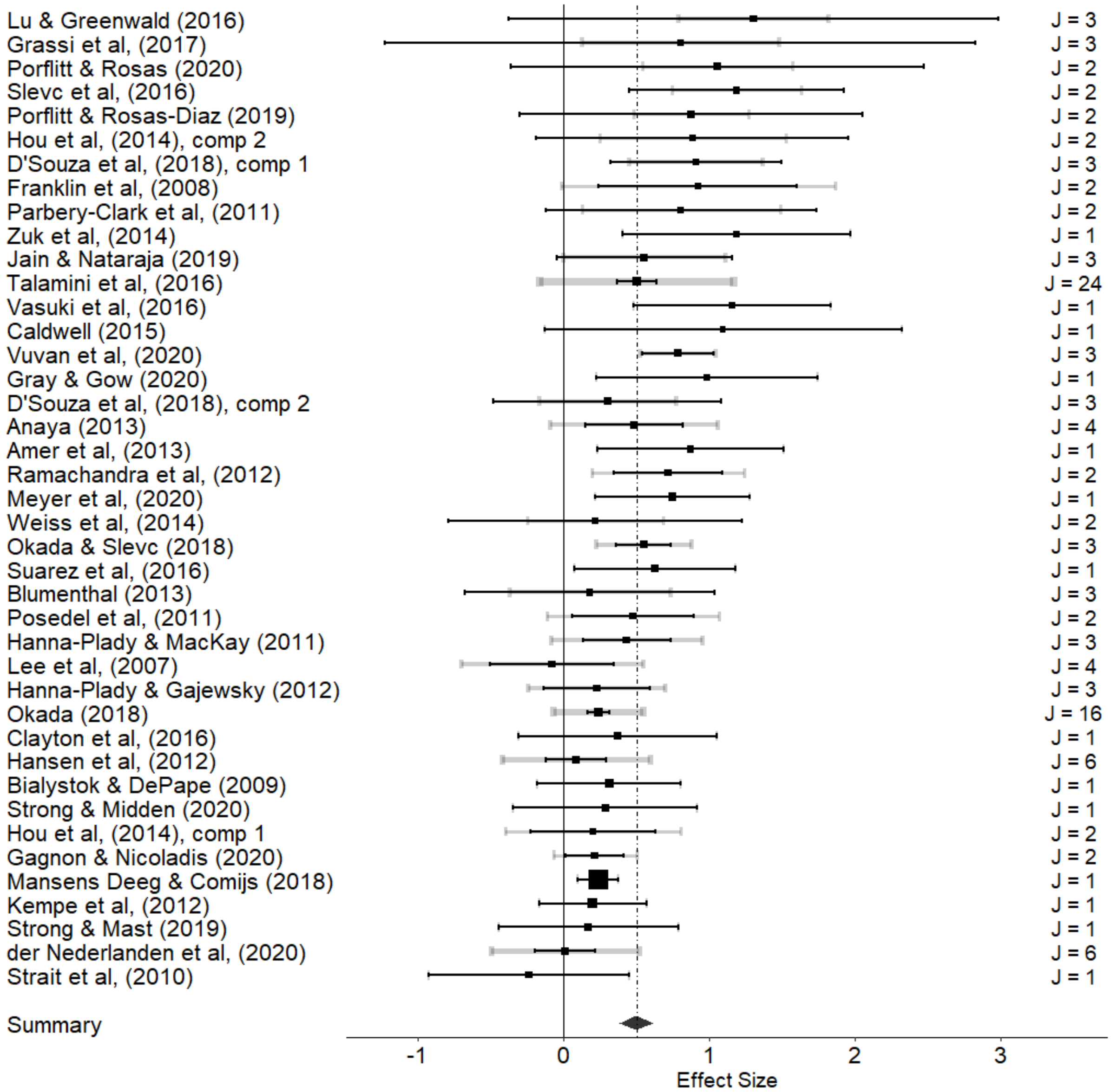


Note. Effect sizes are weighted averages and can include multiple effect sizes *J*. Black error bars represent 95% confidence intervals of the weighted averages. Gray error bars reflect the sampling variance of individual observed effect sizes of each study. The thickness of the gray error bars is proportional to the number of effect sizes within individual studies.

Finally, we also found a significant medium association between musical expertise and working memory (see Figure 4), supported by decisive evidence, $g = 0.49$, 95% CI [0.38, 0.61], $p < .001$, BF > 100. Like for inhibition and shifting, the heterogeneity between effect sizes was significant for working memory, $Q(df = 125) = 426.82$, $p < .001$, $\tau^2_{level\ 3} = 0.05$, $\tau^2_{level\ 2} = 0.12$, $I^2_{level\ 3} = 21.16\%$, $I^2_{level\ 2} = 54.49\%$.

Figure 4

Multilevel Meta-Analysis of Musical Expertise and Working Memory



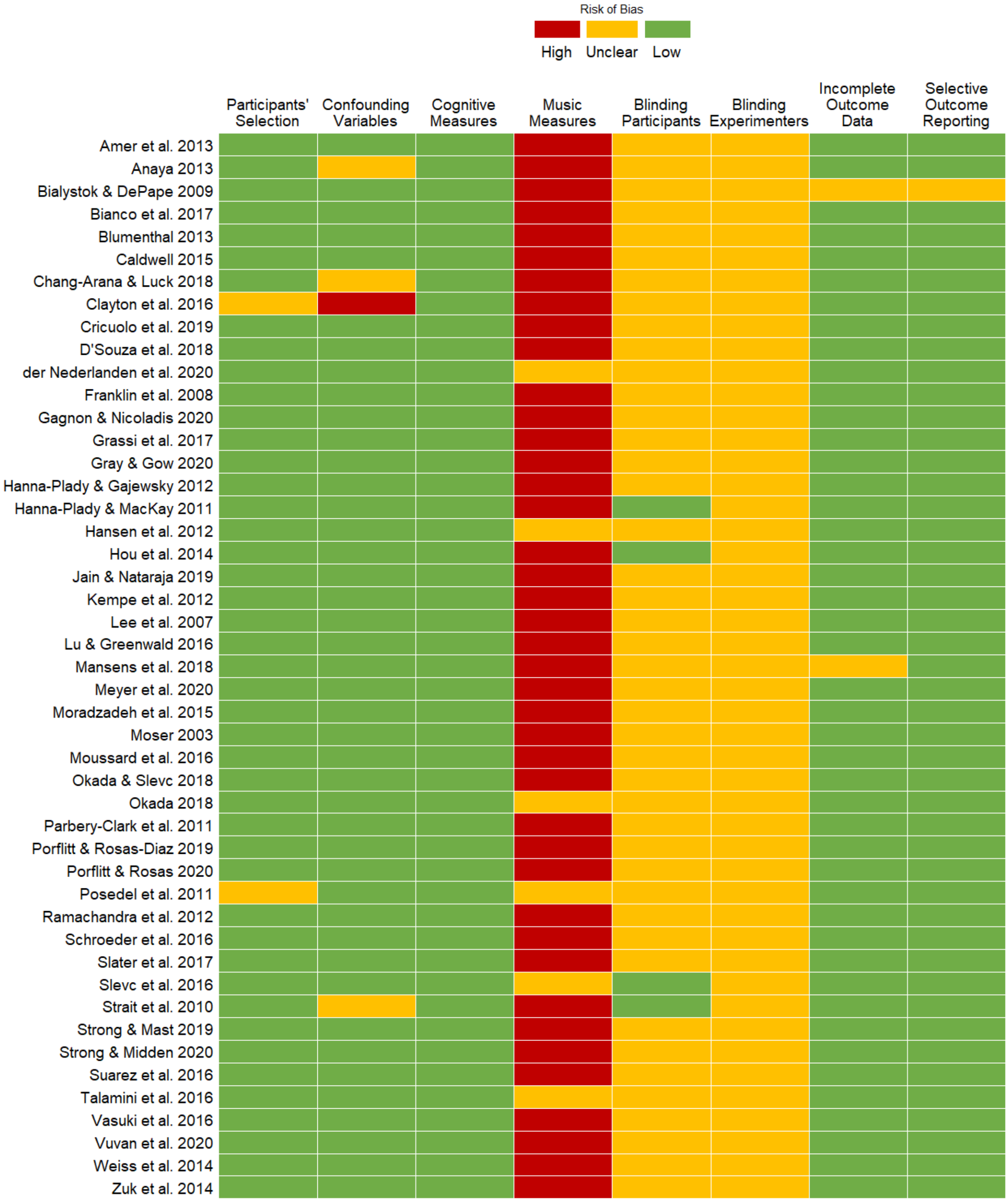
Note. Effect sizes are weighted averages and can include multiple effect sizes *J*. Black error bars represent 95% confidence intervals of the weighted averages. Gray error bars reflect the sampling variance of individual observed effect sizes of each study. The thickness of the gray error bars is proportional to the number of effect sizes within individual studies.

Risk of Bias Assessment

Figure 5 displays the results of the risk of bias ratings. The highest risk of bias was found for music assessment, with 41 studies (87.2%) being rated as having a high risk of bias because these studies relied only on self-report musical expertise measures, for example by asking participants' years of music training. For the remaining studies, the risk of bias for music assessment was rated as unclear because self-report and objective measures of musical expertise (e.g., both a self-report questionnaire and melodic testing) were combined into a composite score measuring musical ability. The risk of bias from blinding of both participants and experimenters (i.e., whether they were informed about the hypotheses of the studies prior to data collection) was rated to be largely unclear, because most studies did not explicitly describe their blinding procedures. The risk of bias was considered low for all other criteria (i.e., selection of participants, confounding variables, cognitive measures, incomplete outcome, and selective outcome).

Figure 5

Risk of Bias Ratings



Moderator Analyses

Table 4 lists the results of the moderator analyses in regard to study type (correlational vs. comparison studies), study risk of bias, measure (self-report questionnaire or performance-based task) to assess musical expertise (music measures), cognitive task used to assess executive functions (paradigm), the modalities (i.e, auditory, verbal, or visual) of the cognitive task (modality), and participant characteristics (age, gender distribution, and education). Different to our pre-registered plans, no moderator analyses could be conducted for socioeconomic status, because many studies did not report any relevant data, and the remaining studies greatly varied in the scales used, rendering them incomparable. For example, studies used the Hollingshead Four-Factor Index (Criscuolo et al., 2019), MacArthur Scale of Subjective Social Status (Okada, 2018; Okada & Slevc, 2018), and mother's education (Blumenthal, 2013; Moradzadeh et al., 2015). Nonetheless, none of the previous studies showed differences between musicians' and non-musicians socioeconomic status (Blumenthal, 2013; Criscuolo et al., 2019; D'Souza & Wiseheart, 2018; Moradzadeh et al., 2015) or did not find an effect on the relationship between musical expertise and executive functions (Okada, 2018; Okada & Slevc, 2018; Slevc et al., 2016; Suárez et al., 2016). In addition, no moderator analysis was run for years of education on the association between musical expertise and shifting, because less than 10 studies reported the relevant data. We also refrained from testing paradigm as a moderator across all executive functions, as the paradigm used is inherently confounded with the executive functions factor assessed (e.g., a Stroop task assesses always only inhibition but not shifting or working memory).

Only modality significantly affected the association between musical expertise and executive functions across all three factors, $F(2, 218) = 3.27$, $p = .040$, with significantly larger effect sizes for tasks with auditory stimuli than verbal and visual stimuli, $\beta = .43$, $p < .001$. However, the evidence for modality as a moderator was ambiguous, $BF_{10} = 1/2.76$.

None of the other moderators significantly affected the association between musical expertise and executive functions across all three factors, with these null associations being supported by at least substantial evidence. The association between musical expertise and inhibition specifically was moderated only by risk of bias, $F(1, 59) = 16.80$, $p < .001$, $BF_{10} = 6.65$, with larger effect sizes of studies with higher risk of bias scores (see Figure 6). For shifting, we found a significant moderation effect of study type, $F(1, 42) = 5.50$, $p = .024$, with smaller effect sizes for correlational than between-group comparison studies, $\beta = -.32$, $p = .023$; however, the evidence for this moderation effect was only ambiguous, $BF_{10} = 2.38$. Moreover, we found a significant moderation effect of music measure on shifting, $F(2, 41) = 3.31$, $p = .046$, yet only supported by ambiguous evidence, $BF_{10} = 1.15$. Furthermore, we found that a significant moderating effect of gender distribution, $F(1, 19) = 6.46$, $p = .020$, suggesting that the more female participants in the study the smaller the effect size. However, this effect was not supported by the Bayesian evidence, which indeed instead favored the null hypothesis, $BF_{10} = 1/4.09$. Finally, the association between musical expertise and working memory was also significantly moderated by music measure, $F(2, 123) = 3.76$, $p = .026$, with however ambiguous the evidence for the presence of this effect, $BF_{10} = 2.80$. Furthermore, the paradigm used also emerged as a significant moderator, $F(2, 123) = 3.76$, $p = .026$, but, again, the evidence for this effect was only ambiguous, $BF_{10} = 1/1.28$. Heterogeneity in effect sizes was still significant for all tested moderators, $Q(\leq 124) = < 435.54$, all ps $< .001$.

Table 4

Effects of Moderators on the Relationship between Musical Expertise and Executive Functions

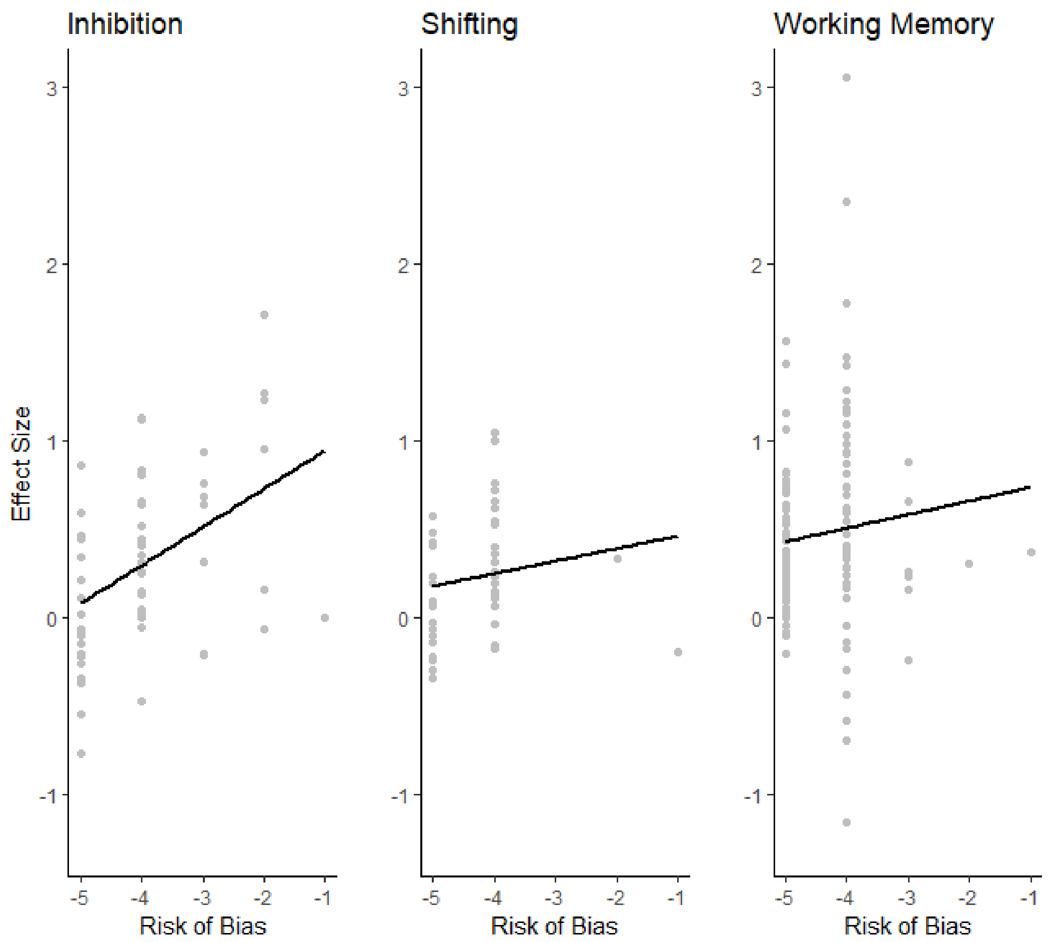
Moderator	*k*	*F*	DF	*p*	BF_{10} [sensitivity]
Executive Functions					
Study type	232	0.01	1, 230	.905	1/6.46 [1/2.94, 1/10.15]
Risk of bias	232	1.18	1, 230	.278	1/3.07 [1/1.49, 1/5.05]
Music measure	232	0.40	2, 229	.673	1/22.73 [1/5.29, 1/55.55]
Modality	221	3.27	2, 218	.040	1/2.76 [1.44, 1/6.53]
Age	206	0.02	1, 204	.897	<1/100 [1/47.32, <1/100]
Gender distribution	146	<0.01	1, 144	.946	<1/100 [1/40. <1/100]
Education	74	0.63	1, 72	.431	1/9.43 [1/3.43, 1/15.12]
Inhibition					
Study type	61	0.46	1, 59	.499	1/3.41 [1/1.85, 1/5.16]
Risk of bias	61	16.80	1, 59	<.001	6.65 [6.56, 4.99]
Music measure	61	1.12	2, 58	.333	1/3.73 [1/1.64, 1/7.19]
Paradigm	61	1.82	1, 59	.183	1/2.58 [1/1.41, 1/3.85]
Modality	61	1.92	2, 58	.155	1/4.38 [1/2.06, 1/7.94]
Age	49	0.03	1, 47	.860	1/59.13 [1/20.07, 1/99.21]
Gender distribution	34	3.92	1, 32	.056	1/9.59 [1/3.16, 1/16.14]
Education	15	0.22	1, 13	.649	1/3.23 [1/1.66, 1/5.03]
Shifting					
Study type	44	5.50	1, 42	.024	2.38 [2.75, 1.79]
Risk of bias	44	0.31	1, 42	.583	1/3.42 [1/1.67, 1/5.47]
Music measure	44	3.31	2, 41	.046	1.15 [2.07, 1/1.62]
Paradigm	44	0.97	2, 41	.387	1/8.55 [1/2.78, 1/18.22]
Modality	44	1.18	2, 41	.317	1/7.19 [1/3.42, 1/12.19]
Age	41	0.70	1, 39	.410	1/69.73 [1/22.49, <1/100]
Gender distribution	21	6.46	1, 19	.020	1/4.09 [1/1.41, 1/6.80]
Working Memory					
Study type	126	0.19	1, 124	.659	1/5.55 [1/2.66, 1/8.36]
Risk of bias	126	0.07	1, 124	.796	1/4.48 [1/1.98, 1/7.00]
Music measure	126	4.72	2, 123	.011	2.80 [6.19, 1.34]
Paradigm	126	3.76	2, 123	.026	1/1.28 [1.40, 1/2.40]
Modality	115	1.91	2, 112	.153	1/5.02 [1/1.93, 1/10.10]

Age	116	<0.01	1, 114	.988	<1/100 [1/39.68, <1/100]
Gender distribution	91	1.01	1, 89	.316	1/56.94 [1/19.38, 1/93.72]
Education	52	0.22	1, 50	.643	1/9.91 [1/3.67, 1/16.30]

Note. Age and education are given in years. Risk of bias is the total sum score of risk of bias scores. DF = degrees of freedom; BF_{10} = Bayes factor in favor of the alternative hypothesis.

Figure 6

Risk of Bias as Moderator of the Association between Musical Expertise and Executive Functions



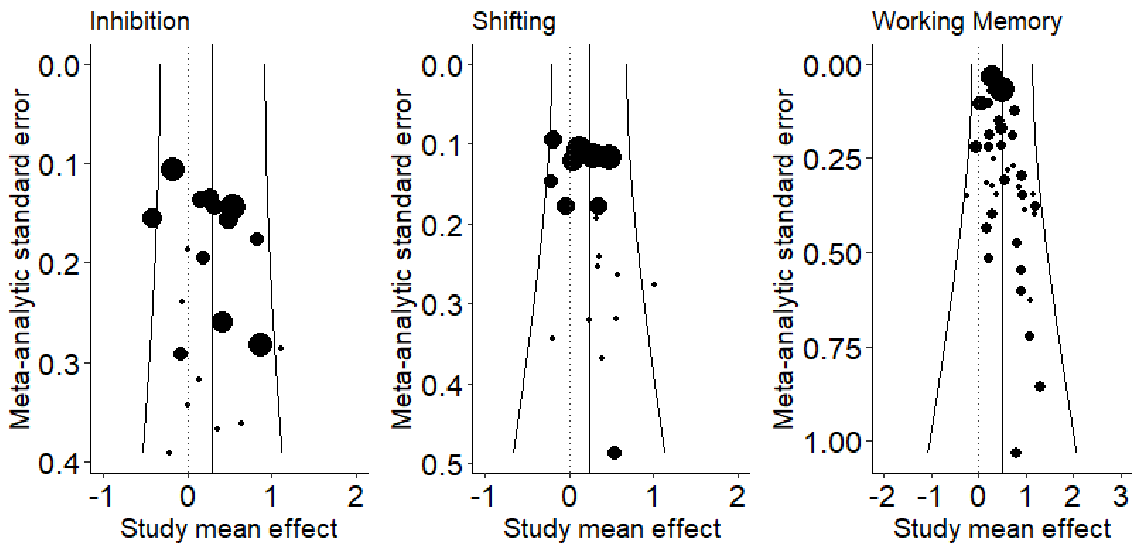
Note. Scatterplots relating pooled effect sizes of the association between musical expertise and each factor of executive functions with the summed risk of bias scores for each study.

Publication Bias

Funnel plots relating standard errors and effect sizes estimates for each study were inspected for each factor of executive functions (see Figure 7). Asymmetries were not apparent for inhibition. However, five working memory studies (Caldwell, 2015; Franklin et al., 2008; Grassi et al., 2017; Strong & Midden, 2020; Zuk et al., 2014) and one shifting study (Zuk et al., 2014) had high standard errors ($SE \geq 0.4$), potentially due to small study sample sizes. To test the funnel plot asymmetry quantitatively, we conducted Egger's test and Begg's and Mazumdar's tests with a multilevel approach. Egger's test regressions were not significant for inhibition, $p = .394$, $BF_{10} = 0.89$, but were significant for shifting ($p = .006$, $BF_{10} = 1.59$) and working memory ($p = .001$, $BF_{10} = 6.28$). Begg and Mazumdar's correlations between the ranks of effect sizes and their variance were non-significant for inhibition, Kendall's $\tau = .04$, $p = .623$, and shifting, Kendall's $\tau = .21$, $p = .050$ but for working memory, Kendall's $\tau = .27$, $p < .001$. Taken together, this indicates a potential bias arising from studies with small sample sizes on shifting and inhibition.

Figure 7

Funnel Plots of Effect Sizes of the Association between Musical Expertise and Executive Functions



Note. Study funnel plots where the size of the dots is proportional to the number of effect sizes included in the studies. The dotted vertical line represents the reference line at zero. The solid vertical line represents the overall effect size. The tilted lines represent the 95% confidence interval.

Discussion

This preregistered systematic review and meta-analysis examined the relationship between musical expertise and executive functions in adults. The present three-level meta-analysis synthesizing 47 between-group comparisons and correlational studies, and considering Bayesian inference, is the most comprehensive meta-analysis to date. We found that musical expertise was significantly related to executive functions overall ($g = 0.43$) and all three factors of executive functions, with the significantly largest average association for working memory ($g = 0.49$), followed by inhibition ($g = 0.31$) and shifting ($g = 0.22$). Evidence for these associations was decisive for executive functions overall and across all three factors ($BF_{10} > 100$). The moderator analyses we conducted revealed mostly non-significant (age and education) or ambiguous effects (gender distribution, study type, music measure, and paradigm). The only exception was that a larger risk of bias increased the relationship between musical expertise and inhibition, which was supported by substantial evidence ($BF_{10} > 6$).

Hence, overall, the present findings are in line with previous meta-analyses (Hernández et al., 2020; Román-Caballero et al., 2022), showing associations between musical expertise and executive functions. However, those meta-analyses only included studies comparing extreme groups (i.e., musicians and non-musicians), while the current study included both comparison and correlational studies increasing the robustness of the results of the present meta-analysis. Moreover, going beyond previous meta-analyses, we systematically distinguished between inhibition, shifting, and working memory, enabling us to determine that the association is particularly strong for working memory. The large effect sizes found in previous meta-analyses can therefore likely be attributed to associations with working memory.

Unity and Diversity of Executive Functions in Musical Expertise?

Although none of the studies directly modeled executive functions in the unity/diversity framework (Friedman & Miyake, 2017; Miyake et al., 2000; Miyake & Friedman, 2012), our analyses provide some pointers to the theoretical roots of the associations found. Musical expertise was associated with all three executive functions, with only ambiguous evidence for a moderator effect of the specific executive function assessed, which suggests that musical expertise may primarily tap the unity between executive functions. Hence, theoretically, associations observed may be due to a functional overlap in the maintenance of task goals and goal-related information. Previous studies have argued that musicians need to maintain the goals of many activities while making music, such as reading music sheets, pressing the right keys on the instruments, listening to oneself and other players, and adjusting for pitch, tempo, and volume (Frischen et al., 2019; Okada & Slevc, 2020).

Although the evidence for the moderation effect was ambiguous, the relatively stronger association with working memory suggests that musical expertise may also reflect working memory-specific aspects of executive functions. Hence, musical expertise and executive functions may functionally overlap in demanding the short-term storage, processing, and updating of information. Consistent with this notion, previous research showed that musicians update information in their working memory to sight-read new music (Okada & Slevc, 2020). For example, pianists who were asked to sight-read five pieces of music found that professional pianists can keep more notes in working memory before playing it compared to amateur pianists (Furneaux & Land, 1999). Similarly, Meinz and

Hambrick's (2010) found a significant positive association between sight-reading performance and working memory capacity.

Another, not mutually exclusive, possible reason for the relationship between musical expertise and working memory may be overlapping third cognitive processes. For example, musicians have been shown to have better auditory processing skills (Wang, 2022) such as the ability to discriminate between two auditory stimuli. Recent findings have shown a positive association between auditory discrimination ability and working memory capacity (Tsukahara et al., 2020). This suggests that the relationship with musical expertise may not be a direct relationship but could be mediated by auditory discrimination ability, a hypothesis warranting further investigation (see Aryanto, 2024).

While working memory exhibited the highest effect size, inhibition and shifting were also significantly related to musical expertise. These executive functions factors are vital for musicians to coordinate their performance with others, adapt to changes in the musical structure, and maintain focus throughout a piece. Inhibition is essential for controlling a musician's performance, ensuring that music is played with the correct rhythm and tempo (Okada & Slevc, 2020). Study on percussionists showed that rhythmic expertise mediated the relationship between drum playing and inhibition (Slater et al., 2018). Shifting, on the other hand, is crucial for switching between different auditory streams where musicians must continuously adjust their contributions to the overall performance and transition between various musical parts when performing in an ensemble (Okada & Slevc, 2020). Indeed, previous studies reported quicker reaction times in musicians relative to non-musicians when they switched between different rules within a task using verbal stimuli (Hao et al., 2023) and also within musician groups as musical expertise develops when using music stimuli (Slama et al., 2017).

Impact of Methodological and Publication Biases on the Association between Musical Expertise and Executive Functions

Analyses of the risk of bias and publication bias in our meta-analysis highlighted several methodological issues in existing research. Notably, increased risk of bias was significantly associated with larger effect sizes of the association between musical expertise and inhibition, suggesting that biases may contribute to inflating meta-analytic effect sizes. The highest risk of bias was identified for the assessment of musical expertise. Specifically, most studies included in the current meta-analysis assessed musical expertise with self-report questionnaires, thereby relying exclusively on participants' memory and perception of their own musical expertise ability. Moreover, these self-reports are often used as sole reference for dichotomizing participants into groups of musicians and non-musicians. To aggravate these issues further, the group-comparison studies reviewed here often also applied different criteria for musicianship. One study, for example, considered musicians as people with formal music training before the age of 10 and at least nine years of music training (Franklin et al., 2008), whereas another study classified people as musicians if they had formal private music lessons for at least a year (Strong & Midden, 2020). This illustrates the arbitrariness and variability of the criteria used for classifying people as musicians or non-musicians, rendering between-group comparisons less robust and hard to interpret.

While we found no unambiguous evidence for a moderating effect of study type, only coarse categorization based on questionnaires likely increases measurement noise and, thus, may have contributed to the heterogeneity between studies. To further clarify the relation between musical expertise and executive functions, or other cognitive abilities, objective measures can assess musical expertise more precisely and on a continuous spectrum. Self-report measures were significantly more strongly related to shifting and working memory than objective measures and musical sophistication; however, all of these effects were supported by ambiguous evidence only. Objective measures such as the beat alignment (Harrison & Müllensiefen, 2018) or the melodic discrimination test (Harrison et al., 2017) have been shown to be significantly related to years of music training (Okada, 2018), duration of music rehearsal (Mosing et al., 2014), duration of music lessons (Swaminathan et al., 2021), and musical sophistication (Correia et al., 2022). Therefore, although the moderating effect of music measures showed ambiguous evidence, complementing the assessment of musical expertise with objective measures in addition to self-report questionnaires can enhance our understanding of the relationship between musical expertise and executive functions.

The analysis of publication bias suggested a potential small-sample bias on the relationship between musical expertise, shifting, and working memory. Further exploration of the studies with high standard error showed that all of those studies were comparison studies with less than 30 participants in each group. Even for a true medium effect, the theoretical power for this sample size is only about 50%. Low statistical power can not only lead to false-negative findings but also false-positives (Button et al., 2013) and inflated effect sizes (Halsey et al., 2015), further aggravating the methodological and statistical concerns associated with group comparisons derived from artificial dichotomization of musical expertise measures (MacCallum et al., 2002; McClelland et al., 2015; Preacher et al., 2005). Thus, future studies should instead consider operationalizing musical expertise as a continuous variable and follow sample size recommendations for assessing robust and stable correlations (e.g., Schönbrodt & Perugini, 2013).

Methodological Recommendations for Future Research

Taken together, future research investigating the relationship between musical expertise and cognitive abilities should consider correlational study designs with sufficient numbers of participants to ensure adequate statistical power and stable estimates of any associations. Furthermore, future research may benefit from complementing self-report assessments of the experience of musical expertise by objective, performance-based measures of musical expertise. Preregistering research and reporting the results transparently would further decrease the risk of biases.

Limitations

One limitation pertaining to the coding of risk of bias was the low interrater reliability for music assessment and selection of participants. The low interrater reliability for music assessment was due to initial differences in interpreting the criterion as to whether it related to the assessment of musical expertise (self-report vs. objective instruments) or to the study design (correlational vs. group-comparison designs). After discussion, the raters agreed to interpret the criterion based on the assessment of musical expertise rather than the study design, yielding perfect agreement. The low interrater reliability of coding the risk of bias arising from selection of participants was due to disagreement for only 3 out of 47 studies. The overall reliability of a measurement or assessment is influenced not only by the quantity of disagreements, but also by the proportion of agreements in comparison to the total number of cases. When the number of raters is relatively small (i.e., two raters), Cohen's Kappa can be sensitive to minor shifts in agreement. Critically, all initial disagreements were resolved by discussion prior to any analysis.

Another limitation of our risk of bias analysis is the difficulty in uniformly assessing the quality of participant recruitment and the definition of musicianship across studies. While we adapted the RoBANS risk of bias criteria (Kim et al., 2013) and evaluated factors such as the presence of music perception tests, a more robust approach would be to focus on the specificity and transparency of recruitment criteria. Future meta-analyses and systematic reviews on musical expertise should consider developing a more granular risk of bias criterion that specifically evaluates whether studies provide sufficient detail on participant inclusion (e.g., years of formal music training, current music making status, music conservatory or university students) to ensure the validity of their musician and non-musician groups. This would help standardize the definition of musical expertise and improve the comparability of findings across different studies.

A further limitation is that we were unable to include socioeconomic status as a moderator. Socioeconomic status correlates with executive functions, especially in children (Blakey et al., 2020; Cuartas et al., 2022) and it is plausible that being from a more advantaged background affords greater access and opportunities to play music. However, the studies included in the review varied strongly in how they measured socioeconomic status, both in terms of their scales and scores but also in how they may influence both cognition and music playing. For example, parental education may shape cultural beliefs and practices whereas household income would influence the ability to partake in musical activities because of the costs associated with doing so. Finally, some measures of socioeconomic status are highly country specific. This makes drawing comparisons between studies, and furthermore, any conclusions related to the influence of socioeconomic status difficult. To address this limitation, future research could include measures of socioeconomic status that can be compared across contexts, such as income-to-needs-ratio or the MacArthur Scale of Subjective Social Status (Adler et al., 2000).

Conclusions

This pre-registered systematic review and meta-analysis confirmed that musical expertise is positively associated with performance in tasks testing executive functions. This association was strongest for working memory, followed by inhibition and then shifting, and supported by decisive Bayesian evidence. The moderator analyses suggested that these associations may be affected by risks of methodological biases (for inhibition) and small participant samples (for shifting and working memory). Therefore, methodological recommendations include using objective assessments of musical expertise in larger, correlational samples. Furthermore, this review suggests exciting avenues for further research to explore the mechanisms underpinning the relation between musical expertise and executive functions, in particular working memory.

Declarations

Funding This research was supported by a PhD scholarship to Christ B. Aryanto from the Indonesia Endowment Fund for Education, Ministry of Finance, Republic of Indonesia (Lembaga Pengelola Dana Pendidikan, Kementerian Keuangan, Republik Indonesia) with the scholarship ID number 20200422341285 and is part of his doctoral thesis. We are grateful to all the authors who kindly responded to our queries and who provided additional data for the meta-analysis. We further thank Dilara Steenken for her assistance in the literature search.

Conflicts of interest/Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and materials The data are available on the Open Science Framework (https://osf.io/8h6tc/) and this study was preregistered (copy of redacted preregistration available at https://osf.io/8h6tc/?view_only=1e92639423da4c9ab8df0ea3a4f1702f).

Code availability The analysis scripts are available on the Open Science Framework (https://osf.io/evj5d).

Authors' contributions Christ Billy Aryanto (conceptualisation, methodology, software, formal analysis, investigation, data curation, writing – original draft, project administration, funding acquisition)

Aireen Rhammy Kinara Aisyah (investigation, data curation)

Emma Blakey (conceptualisation, methodology, writing – review and editing, supervision)

Renee Timmers (conceptualisation, writing – review and editing, supervision)

Claudia C. von Bastian (conceptualisation, methodology, formal analysis, writing – review and editing, supervision)

References

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

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health Psychology*, *19*(6), 586–592. https://doi.org/10.1037/0278-6133.19.6.586

*Amer, T., Kalender, B., Hasher, L., Trehub, S. E., & Wong, Y. (2013). Do Older Professional Musicians Have Cognitive Advantages? *PLoS ONE*, *8*(8), e71630. https://doi.org/10.1371/journal.pone.0071630

*Anaya, E. M. (2013). *Musicians' enhanced cognitive and perceptual skills: Modality-specific or modality-general?* Indiana University.

Arndt, C., Schlemmer, K., & Van Der Meer, E. (2023). The Relationship of Musical Expertise, Working Memory, and Intelligence. *Music Perception: An Interdisciplinary Journal*, *40*(4), 334–346. https://doi.org/10.1525/mp.2023.40.4.334

Aryanto, C. B. (2024). *The Structure and Mechanism of the Relation between Music Making, Executive Functions, and Sensory Discrimination* (Doctoral dissertation, University of Sheffield).

Begg, C. B., & Mazumdar, M. (1994). Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*, *50*(4), 1088. https://doi.org/10.2307/2533446

*Bialystok, E., & DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 565–574. https://doi.org/10.1037/a0012735

*Bianco, V., Berchicci, M., Perri, R. L., Quinzi, F., & Di Russo, F. (2017). Exercise-related cognitive effects on sensory-motor control in athletes and drummers compared to non-athletes and other musicians. *Neuroscience*, *360*, 39–47. https://doi.org/10.1016/j.neuroscience.2017.07.059

Blakey, E., Matthews, D., Cragg, L., Buck, J., Cameron, D., Higgins, B., Pepper, L., Ridley, E., Sullivan, E., & Carroll, D. J. (2020). The Role of Executive Functions in Socioeconomic Attainment Gaps: Results From a Randomized Controlled Trial. *Child Development*, *91*(5), 1594–1614. https://doi.org/10.1111/cdev.13358

*Blumenthal, G. (2013). *An examination of musical experience effects on working memory and interference control*. Library and Archives Canada = Bibliothèque et Archives Canada.

Bugos, J. A. (2019). The Effects of Bimanual Coordination in Music Interventions on Executive Functions in Aging Adults. *Frontiers in Integrative Neuroscience*, *13*, 68. https://doi.org/10.3389/fnint.2019.00068

Bugos, J. A., DeMarie, D., Stokes, C., & P. Power, L. (2022). Multimodal music training enhances executive functions in children: Results of a randomized controlled trial. *Annals of the New York Academy of Sciences*, *1516*(1), 95–105. https://doi.org/10.1111/nyas.14857

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

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

*Caldwell, K. A. (2015). *Speech in noise ability, outer hair cell function, and working memory for trained flute players*.

*Chang-Arana, Á. M., & Luck, G. (2018). *Inhibitory Control in Transposing Musicians, Non-Transposing Musicians, and Non- Musicians*. 4.

*Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd, G. (2016). Executive Function, Visual Attention and the Cocktail Party Problem in Musicians and Non-Musicians. *PLOS ONE*, *11*(7), e0157638. https://doi.org/10.1371/journal.pone.0157638

Correia, A. I., Vincenzi, M., Vanzella, P., Pinheiro, A. P., Lima, C. F., & Schellenberg, E. G. (2022). Can musical ability be tested online? *Behavior Research Methods*, *54*(2), 955–969. https://doi.org/10.3758/s13428-021-01641-2

Correia, A. I., Vincenzi, M., Vanzella, P., Pinheiro, A. P., Schellenberg, E. G., & Lima, C. F. (2023). Individual differences in musical ability among adults with no music training. *Quarterly Journal of Experimental Psychology*, *76*(7), 1585–1598. https://doi.org/10.1177/17470218221128557

*Criscuolo, A., Bonetti, L., Särkämö, T., Kliuchko, M., & Brattico, E. (2019). On the Association Between Musical Training, Intelligence and Executive Functions in Adulthood. *Frontiers in Psychology*, *10*, 1704. https://doi.org/10.3389/fpsyg.2019.01704

Cuartas, J., Hanno, E., Lesaux, N. K., & Jones, S. M. (2022). Executive function, self-regulation skills, behaviors, and socioeconomic status in early childhood. *PLOS ONE*, *17*(11), e0277013. https://doi.org/10.1371/journal.pone.0277013

Deeks, J. J., Higgins, J. P., Altman, D. G., & Cochrane Statistical Methods Group. (2019). Analysing data and undertaking meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions*, 241–284.

Degé, F., & Frischen, U. (2022). The impact of music training on executive functions in childhood—A systematic review. *Zeitschrift Für Erziehungswissenschaft*, *25*(3), 579–602. https://doi.org/10.1007/s11618-022-01102-2

*der Nederlanden, C. M. V. B., Zaragoza, C., Rubio-Garcia, A., Clarkson, E., & Snyder, J. S. (2020). Change detection in complex auditory scenes is predicted by auditory memory, pitch perception, and years of musical training. *Psychological Research*, *84*(3), 585–601. https://doi.org/10.1007/s00426-018-1072-x

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

*D'Souza, A. A., Moradzadeh, L., & Wiseheart, M. (2018). Musical training, bilingualism, and executive function: Working memory and inhibitory control. *Cognitive Research: Principles and Implications*, *3*(1), 11. https://doi.org/10.1186/s41235-018-0095-6

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

Ferguson, H. J., Brunsdon, V. E. A., & Bradford, E. E. F. (2021). The developmental trajectories of executive function from adolescence to old age. *Scientific Reports*, *11*(1), 1382. https://doi.org/10.1038/s41598-020-80866-1

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). Visual representations of meta-analyses of multiple outcomes: Extensions to forest plots, funnel plots, and caterpillar plots. *Methodology*, *16*(4), 299–315. https://doi.org/10.5964/meth.4013

*Franklin, M. S., Sledge Moore, K., Yip, C.-Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*(3), 353–365. https://doi.org/10.1177/0305735607086044

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Frischen, U., Schwarzer, G., & Degé, F. (2019). Comparing the Effects of Rhythm-Based Music Training and Pitch-Based Music Training on Executive Functions in Preschoolers. *Frontiers in Integrative Neuroscience*, *13*, 41. https://doi.org/10.3389/fnint.2019.00041

*Gagnon, R., & Nicoladis, E. (2021). Musicians show greater cross-modal integration, intermodal integration, and specialization in working memory than non-musicians. *Psychology of Music*, *49*(4), 718–734. https://doi.org/10.1177/0305735619896088

*Grassi, M., Meneghetti, C., Toffalini, E., & Borella, E. (2017). Auditory and cognitive performance in elderly musicians and nonmusicians. *PLOS ONE*, *12*(11), e0187881. https://doi.org/10.1371/journal.pone.0187881

*Gray, R., & Gow, A. J. (2020). How is musical activity associated with cognitive ability in later life? *Aging, Neuropsychology, and Cognition*, *27*(4), 617–635. https://doi.org/10.1080/13825585.2019.1660300

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, *12*(3), 179–185. https://doi.org/10.1038/nmeth.3288

*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*. https://doi.org/10.3389/fnhum.2012.00198

*Hanna-Pladdy, B., & MacKay, A. (2011). The relation between instrumental musical activity and cognitive aging. *Neuropsychology*, *25*(3), 378–386. https://doi.org/10.1037/a0021895

*Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*, *41*(6), 779–793. https://doi.org/10.1177/0305735612452186

Hao, J., Zhong, Y., Pang, Y., Jing, Y., Liu, Y., Li, H., Li, J., & Zheng, M. (2023). The relationship between music training and cognitive flexibility: An ERP study. *Frontiers in Psychology*, *14*, 1276752. https://doi.org/10.3389/fpsyg.2023.1276752

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing Meta-Analysis With R: A Hands-On Guide* (1st ed.). Chapman & Hall/CRC Press. https://www.routledge.com/Doing-Meta-Analysis-with-R-A-Hands-On-Guide/Harrer-Cuijpers-Furukawa-Ebert/p/book/9780367610074

Harrison, P. M. C., Collins, T., & Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, *7*(1), 3618. https://doi.org/10.1038/s41598-017-03586-z

Harrison, P. M. C., & Müllensiefen, D. (2018). Development and Validation of the Computerised Adaptive Beat Alignment Test (CA-BAT). *Scientific Reports*, *8*(1), 12395. https://doi.org/10.1038/s41598-018-30318-8

Hernández, M., Molina-Delgado, M., Smith-Castro, V., & Rodríguez-Villagra, O. A. (2020). Funciones ejecutivas entre músicos y no músicos: Un metaanálisis. *Interdisciplinaria Revista de Psicología y Ciencias Afines*, *37*(2), 39–55. https://doi.org/10.16888/interd.2020.37.2.3

*Hou, J., Chen, C., Wang, Y., Liu, Y., He, Q., Li, J., & Dong, Q. (2014). Superior pitch identification ability is associated with better executive functions. *Psychomusicology: Music, Mind, and Brain*, *24*(2), 136–146. https://doi.org/10.1037/a0036963

*Jain, S., & Nataraja, N. P. (2019). The Effect of Fatigue on Working Memory and Auditory Perceptual Abilities in Trained Musicians. *American Journal of Audiology*, *28*(2S), 483–494. https://doi.org/10.1044/2019_AJA-IND50-18-0102

*Kempe, V., Thoresen, J. C., Kirk, N. W., Schaeffler, F., & Brooks, P. J. (2012). Individual Differences in the Discrimination of Novel Speech Sounds: Effects of Sex, Temporal Processing, Musical and Cognitive Abilities. *PLoS ONE*, *7*(11), e48623. https://doi.org/10.1371/journal.pone.0048623

Kim, S. Y., Park, J. E., Lee, Y. J., Seo, H.-J., Sheen, S.-S., Hahn, S., Jang, B.-H., & Son, H.-J. (2013). Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *Journal of Clinical Epidemiology*, *66*(4), 408–414. https://doi.org/10.1016/j.jclinepi.2012.09.016

*Lee, Y., Lu, M., & Ko, H. (2007). Effects of skill training on working memory capacity. *Learning and Instruction*, *17*(3), 336–344. https://doi.org/10.1016/j.learninstruc.2007.02.010

*Lu, C.-I., & Greenwald, M. (2016). Reading and working memory in adults with or without formal musical training: Musical and lexical tone. *Psychology of Music*, *44*(3), 369–387. https://doi.org/10.1177/0305735614568881

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40. https://doi.org/10.1037/1082-989X.7.1.19

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

*Mansens, D., Deeg, D. J. H., & Comijs, H. C. (2018). The association between singing and/or playing a musical instrument and cognitive functions in older adults. *Aging & Mental Health*, *22*(8), 970–977. https://doi.org/10.1080/13607863.2017.1328481

McClelland, G. H., Lynch, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, Type II errors, and false–positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology*, *25*(4), 679–689. https://doi.org/10.1016/j.jcps.2015.05.006

*Meyer, J., Oguz, P. G., & Moore, K. S. (2020). Superior fluid cognition in trained musicians. *Psychology of Music*, *48*(3), 434–447. https://doi.org/10.1177/0305735618808089

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

*Moradzadeh, L., Blumenthal, G., & Wiseheart, M. (2015). Musical Training, Bilingualism, and Executive Function: A Closer Look at Task Switching and Dual-Task Performance. *Cognitive Science*, *39*(5), 992–1020. https://doi.org/10.1111/cogs.12183

*Moser, S.R. (2003). *Beyond the mozart effect: Age-related cognitive functioning in instrumental music participants* [Doctoral Dissertation, The University of Southern Mississippi]. ProQuest Dissertations Publishing.

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice Does Not Make Perfect: No Causal Effect of Music Practice on Music Ability. *Psychological Science*, *25*(9), 1795–1803. https://doi.org/10.1177/0956797614541990

*Moussard, A., Bermudez, P., Alain, C., Tays, W., & Moreno, S. (2016). Life-long music practice and executive control in older adults: An event-related potential study. *Brain Research*, *1642*, 146–153. https://doi.org/10.1016/j.brainres.2016.03.028

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

*Okada, B. M. (2018). *Towards a comprehensive model of musical ability*. University of Maryland, College Park.

*Okada, B. M., & Slevc, L. R. (2018). Individual differences in musical training and executive functions: A latent variable approach. *Memory & Cognition*, *46*(7), 1076–1092. https://doi.org/10.3758/s13421-018-0822-8

Okada, B. M., & Slevc, L. R. (2020). Music Training: Contributions to Executive Function. In J. M. Novick, M. F. Bunting, M. R. Dougherty, & R. W. Engle, *Cognitive and Working Memory Training: Perspectives From Psychology, Neuroscience, and Human Development* (pp. 487–507). Oxford University Press. https://doi.org/10.1093/oso/9780199974467.003.0011

Okely, J. A., Overy, K., & Deary, I. J. (2022). Experience of Playing a Musical Instrument and Lifetime Change in General Cognitive Ability: Evidence From the Lothian Birth Cohort 1936. *Psychological Science*, *33*(9), 1495–1508. https://doi.org/10.1177/09567976221092726

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings*.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

*Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical Experience and the Aging Auditory System: Implications for Cognitive Abilities and Hearing Speech in Noise. *PLoS ONE*, *6*(5), e18082. https://doi.org/10.1371/journal.pone.0018082

*Porflitt, F. I., & Rosas-Díaz, R. R. (2019). Behind the scene: Cognitive benefits of playing a musical instrument. Executive functions, processing speed, fluid intelligence and divided attention / Detrás de la escena: beneficios cognitivos de tocar un instrumento musical. Funciones ejecutivas, velocidad de procesamiento, inteligencia fluida y atención dividida. *Estudios de Psicología*, *40*(2), 464–490. https://doi.org/10.1080/02109395.2019.1601474

*Porflitt, F., & Rosas, R. (2020). Core music elements: Rhythmic, melodic and harmonic musicians show differences in cognitive performance (*Elementos básicos de la música: músicos rítmicos, melódicos y armónicos muestran diferencias de desempeño cognitivo*). *Studies in Psychology*, *41*(3), 532–562. https://doi.org/10.1080/02109395.2020.1795493

*Posedel, J., Emery, L., Souza, B., & Fountain, C. (2012). Pitch perception, working memory, and second-language phonological production. *Psychology of Music*, *40*(4), 508–517. https://doi.org/10.1177/0305735611415145

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the Extreme Groups Approach: A Critical Reexamination and New Recommendations. *Psychological Methods*, *10*(2), 178–192. https://doi.org/10.1037/1082-989X.10.2.178

*Ramachandra, V., Meighan, C., & Gradzki, J. (2012). The impact of musical training on the phonological memory and the central executive: a brief report. *North American Journal of Psychology*, *14*(3).

Rodriguez-Gomez, D. A., & Talero-Gutiérrez, C. (2022). Effects of music training in executive function performance in children: A systematic review. *Frontiers in Psychology*, *13*, 968144. https://doi.org/10.3389/fpsyg.2022.968144

Román-Caballero, R., Arnedo, M., Triviño, 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., Vadillo, M. A., Trainor, L. J., & Lupiáñez, J. (2022). Please don't stop the music: A meta-analysis of the cognitive and academic benefits of instrumental musical training in childhood and adolescence. *Educational Research Review*, *35*, 100436. https://doi.org/10.1016/j.edurev.2022.100436

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. (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., & Lima, C. F. (2024). Music Training and Nonmusical Abilities. *Annual Review of Psychology*, *75*(17), 17.1-17.42.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

*Schroeder, S. R., Marian, V., Shook, A., & Bartolotti, J. (2016). Bilingualism and Musicianship Enhance Cognitive Control. *Neural Plasticity*, *2016*, 1–11. https://doi.org/10.1155/2016/4058620

Slama, H., Rebillon, E., & Kolinsky, R. (2017). Expertise and cognitive flexibility: A Musician's Tale. *Journal of Cultural Cognitive Science*, *1*(2), 119–127. https://doi.org/10.1007/s41809-017-0011-5

Slater, J., Ashley, R., Tierney, A., & Kraus, N. (2018). Got Rhythm? Better Inhibitory Control Is Linked with More Consistent Drumming and Enhanced Neural Tracking of the Musical Beat in Adult Percussionists and Nonpercussionists. *Journal of Cognitive Neuroscience*, *30*(1), 14–24. https://doi.org/10.1162/jocn_a_01189

*Slater, J., Azem, A., Nicol, T., Swedenborg, B., & Kraus, N. (2017). Variations on the theme of musical expertise: Cognitive and sensory processing in percussionists, vocalists and non-musicians. *European Journal of Neuroscience*, *45*(7), 952–963. https://doi.org/10.1111/ejn.13535

*Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, *152*, 199–211. https://doi.org/10.1016/j.cognition.2016.03.017

*Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*(1–2), 22–29. https://doi.org/10.1016/j.heares.2009.12.021

*Strong, J. V., & Mast, B. T. (2019). The cognitive functioning of older adult instrumental musicians and non-musicians. *Aging, Neuropsychology, and Cognition*, *26*(3), 367–386. https://doi.org/10.1080/13825585.2018.1448356

*Strong, J. V., & Midden, A. (2020). Cognitive differences between older adult instrumental musicians: Benefits of continuing to play. *Psychology of Music*, *48*(1), 67–83. https://doi.org/10.1177/0305735618785020

*Suárez, L., Elangovan, S., & Au, A. (2016). Cross-sectional study on the relationship between music training and working memory in adults. *Australian Journal of Psychology*, *68*(1), 38–46. https://doi.org/10.1111/ajpy.12087

Swaminathan, S., Kragness, H. E., & Schellenberg, E. G. (2021). The Musical Ear Test: Norms and correlates from a large sample of Canadian undergraduates. *Behavior Research Methods*, *53*(5), 2007–2024. https://doi.org/10.3758/s13428-020-01528-8

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

*Talamini, F., Carretti, B., & Grassi, M. (2016). The Working Memory of Musicians and Nonmusicians. *Music Perception*, *34*(2), 183–191. https://doi.org/10.1525/mp.2016.34.2.183

Unsworth, N., Redick, T. S., McMillan, B. D., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2015). Is Playing Video Games Related to Cognitive Abilities? *Psychological Science*, *26*(6), 759–774. https://doi.org/10.1177/0956797615570367

Van Erp, S., Verhagen, J., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2017). Estimates of Between-Study Heterogeneity for 705 Meta-Analyses Reported in *Psychological Bulletin* From 1990–2013. *Journal of Open Psychology Data*, *5*(1), 4. https://doi.org/10.5334/jopd.33

*Vasuki, P. R. M., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research*, *342*, 112–123. https://doi.org/10.1016/j.heares.2016.10.008

Viechtbauer, W. (2010). Conducting Meta-Analyses in *R* with the **metafor** Package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Vincenzi, M., Correia, A. I., Vanzella, P., Pinheiro, A. P., Lima, C. F., & Schellenberg, E. G. (2022). Associations between music training and cognitive abilities: The special case of professional musicians. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000481

von Bastian, C. C., Reinhartz, A., Udale, R. C., Grégoire, S., Essounni, M., Belleville, S., & Strobach, T. (2022). Mechanisms of processing speed training and transfer effects across the adult lifespan: Protocol of a multi-site cognitive training study. *BMC Psychology*, *10*(1), 168. https://doi.org/10.1186/s40359-022-00877-7

*Vuvan, D. T., Simon, E., Baker, D. J., Monzingo, E., & Elliott, E. M. (2020). Musical training mediates the relation between working memory capacity and preference for musical complexity. *Memory & Cognition*, *48*(6), 972–981. https://doi.org/10.3758/s13421-020-01031-7

*Weiss, A. H., Biron, T., Lieder, I., Granot, R. Y., & Ahissar, M. (2014). Spatial vision is superior in musicians when memory plays a role. *Journal of Vision*, *14*(9), 18–18. https://doi.org/10.1167/14.9.18

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064. https://doi.org/10.3758/s13423-012-0295-x

*Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and Neural Correlates of Executive Functioning in Musicians and Non-Musicians. *PLoS ONE*, *9*(6), e99868. https://doi.org/10.1371/journal.pone.0099868