The Structure and Mechanism of the Relation between Music Making, Executive Functions, and Sensory Discrimination

By:

Christ Billy Aryanto

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Science

Department of Psychology

August 2024

Acknowledgements

Praise be to God, Jesus, and the Holy Spirit, for Thou bless me so I can finish my thesis timely. Starting the PhD journey during the pandemic was challenging, but the support of many amazing people made my research journey possible. There is a saying, “It takes a village to raise a child”, and I am sure it applies to PhD too. I would like to express my gratitude to people who have been part of my academic journey.

Firstly, I would like to give special thanks to Dr Claudia C. von Bastian, Dr Emma Blakey, and Prof. Renee Timmers. Claudia, thank you so much for replying to my email back in 2019 when I proposed you as my supervisor. You have been supportive even before I started my PhD. I cannot imagine doing a PhD without you! Emma, thank you for your support and for always reminding me of my strength and believing in me. Renee, thank you for the whole journey from when I did my master and for the critical perspective on my studies. I am lucky to have you all as my supervisors!

Second, I want to extend my thanks to Indonesia Endowment Fund for Education, Ministry of Finance, Republic of Indonesia (Lembaga Pengelola Dana Pendidikan, Kementerian Keuangan, Republik Indonesia) for the PhD scholarship and Atma Jaya Catholic University of Indonesia for the research grant. I am gratefully indebted for the chance to pursue the higher education abroad and conducted high quality research.

Third, I want to thank Cognitive Ability and Plasticity Lab, Indonesians at Sheffield, and St. Andrew’s United Reformed Church congregations for making my PhD journey so colourful. I also want to acknowledge my PhD cohort, who were once called “forgotten students” because of our struggle during the pandemic. Good luck for future endeavours!

Last but not least, I want to thank my family, especially my dad and my mom. Thank you for the good morning and the weekend call. You once told me you dreamt for a son as a medical doctor. Now, I am a Doctor of Philosophy, close enough.

Declaration and Note on Inclusion of Published Work

I, Christ Billy Aryanto, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university.

This thesis is in a publication format, and contains the following published work:

Chapter 2: Findings from this chapter have been presented in the following conferences:

Aryanto, C.B., Aisyah, A.R.K., Blakey, E., Timmers, R., & von Bastian, C.C. *How are active music making and executive functions related? A systematic review and meta-analysis* [Poster]. Psychonomic Society 62nd Annual Meeting, Online.

Aryanto, C.B., Aisyah, A.R.K., Blakey, E., Timmers, R., & von Bastian, C.C. *How are active music making and executive functions related? A systematic review and meta-analysis* [Talk]. Learning and Plasticity (LaP) Meeting, Finland.

Aryanto, C.B., Aisyah, A.R.K., Blakey, E., Timmers, R., & von Bastian, C.C. *How are active music-making and executive functions related? A systematic review and meta-analysis* [Poster]. The 22nd European Society for Cognitive Psychology (ESCoP) Conference, Lille, France.

Chapter 3: Findings from this chapter have been presented in the following conference:

Aryanto, C.B., Blakey, E., Timmers, R., Hyde, E.R.A., Reinhartz, A., Jiang, S., Bellevile, S., Strobach, T., & von Bastian, C.C. *Does the Frequency of Music Making Relate to Executive Functions? An Individual Differences Investigation Across the Adult Lifespan* [Talk]. 12th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM12).

Chapter 4: Research plan from this chapter have been presented in the following conference:

Aryanto, C.B., Blakey, E., Timmers, R., & von Bastian, C.C. *Research plan - Why are active music making and executive functions related? Testing the roles of sensory discrimination and fluid intelligence* [Poster]. Experimental Psychology Society (EPS) Meeting, Online. Access to the video presentation: https://www.youtube.com/watch?v=td2ElPWcgyM, and poster presentation: https://eps.ac.uk/wp-content/uploads/2021/06/Christ-Billy-Aryanto.pdf

Chapter 4: Findings from this chapter have been presented in the following conferences:

Aryanto, C.B., Blakey, E., Timmers, R., & von Bastian, C.C. *Does sensory discrimination ability explain the relation between active music-making and working memory?* [Talk]. The 23rd European Society for Cognitive Psychology (ESCoP) Conference, Porto, Portugal.

Aryanto, C.B., Blakey, E., Timmers, R., & von Bastian, C.C. *Does sensory discrimination ability account for the relation between active music-making and working memory?* [Poster]. Psychonomic Society 64th Annual Meeting, San Francisco, USA.

Thesis Abstract

Music making, people's lifetime engagement with actively making music, has been hypothesised to be associated with better executive functions, which are higher-order cognitive skills underpinning goal-directed behaviour. There were many empirical studies have shown a relationship between music making and executive functions, but no clear explanation of why there was a relationship between them. This thesis aims to investigate why and how are music making and executive functions related. For this purpose, Study 1 (Chapter 2) assessed the strength of the relationship between music making and three factors of executive functions (inhibition, shifting, and updating) through a systematic review and meta-analysis and found the largest average association with updating. Next, Study 2 (Chapter 3) investigated the relationship between music making and executive functions by assessing the relationship between frequency of music making in everyday life and executive functions across the adult lifespan. The results showed no significant relationship between frequency of music making and executive functions after accounting for age. Finally, Study 3 (Chapter 4) tested the hypothesis that music making and working memory are related through better sensory discrimination in people with stronger music skills. The results showed that sensory discrimination mediated the relationship between music skill and auditory working memory, but not visual working memory, suggesting that the relationship is specific to the auditory modality. Taken together, the findings of this thesis give a new insight on the underlying mechanism between music making and executive functions and for future longitudinal studies to investigate the causal relationship, especially in different modalities.

Keyword: Executive functions, music making, music skills, sensory discrimination

Table of Contents

List of Tables

List of Figures

Chapter 1 - General introduction

Contributions:

Christ Billy Aryanto (conceptualisation, writing – original draft)

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

Renee Timmers (conceptualisation, supervision)

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

1.1.Introduction

Active engagement in music making has been hypothesised to enhance cognitive skills (Schellenberg & Lima, 2023, Schellenberg & Weiss, 2013) and, specifically, executive functions (Okada & Slevc, 2020). Executive functions are higher-order cognitive skills underpinning goal-directed behaviour (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), and 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). To date there has been mixed findings on the relationship between music making and executive functions. Some meta-analyses did not find a relationship between music making and cognitive skills (Sala & Gobet, 2017a, 2017b, 2020), but other meta-analyses found the relationship (Bigand & Tillmann, 2022), including executive functions (Hernández et al., 2020). It is interesting to investigate music making as opposed to other types of leisure activities, for example, playing video games or playing chess, because music is ubiquitous and can be done anywhere and anytime. Making music can be as simple as singing and clapping hands or complex, such as playing musical instruments or composing music. Moreover, a growing body of literature has tried to investigate the effect of music making on executive functions (Alain et al., 2019; D’Souza & Wiseheart, 2018), showing that making music is an activity that can stimulate one’s cognition (von Bastian et al., 2024). However, the question of why music making and executive functions are related is yet to be answered. Therefore, this thesis will investigate the relationship between music making and executive functions to gain insight into the structure and underlying mechanisms of their relationship.

In the remainder of this chapter, I will introduce the concepts of executive functions and music making and how these are typically measured. Next, I will summarise existing findings regarding their relationship and discuss mixed findings from previous studies, which will lead up to the research gaps this thesis aims to fill and my research questions.

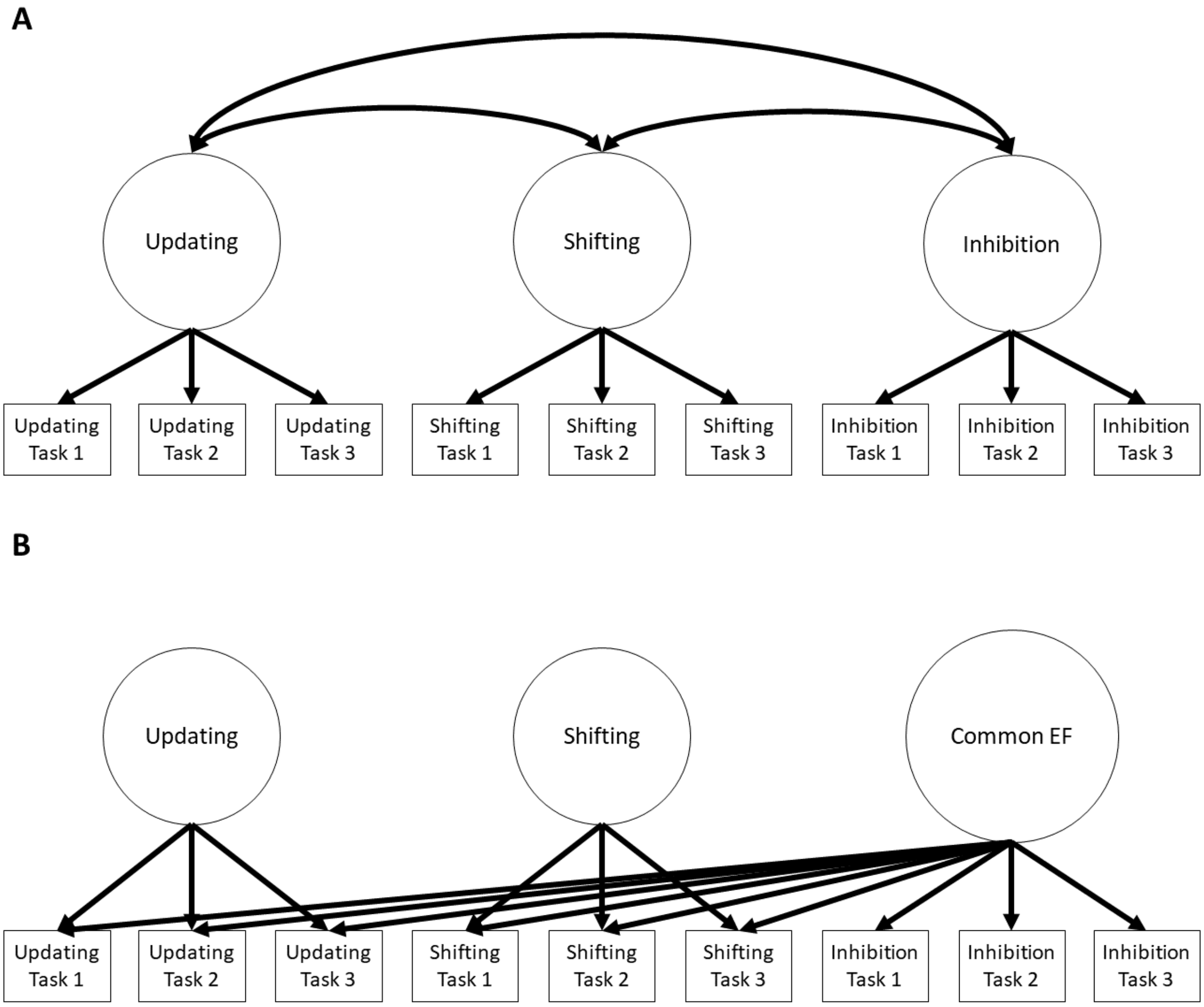
1.2. What are Executive Functions?

Past research has established three factors of executive functions: inhibition, shifting, and updating (Miyake et al., 2000; Miyake & Friedman, 2012). Inhibition is a deliberate, intended, and controlled action to override dominant but incorrect responses. Shifting is the ability to switch from one task to another, activating the currently relevant task set and inhibiting the currently irrelevant task set. Finally, updating refers to replacing no longer relevant information in working memory with new information. Working memory refers to maintaining and actively manipulating ongoing information processing (Cowan, 2017). As updating is one aspect of the working memory, this thesis will employ a broadened definition of the updating factor of executive functions and focus on working memory.

The unity-diversity model (Friedman & Miyake, 2017; Miyake & Friedman, 2012) suggests that while executive functions share common underlying mechanisms (unity), they also exhibit unique, component-specific characteristics (diversity). The notion of unity and diversity originated from previous neuropsychology studies. On one hand, executive functions were associated with the activation of prefrontal cortex, and it was known that patients with frontal lobe problem had deficient executive functions, suggesting a unitary mechanism of executive functions (Miyake et al., 2000). On the other hand, other evidence found that patients could excel in one task measuring shifting, but not inhibition, suggesting that shifting and inhibition may not share a single common mechanism (Friedman & Miyake, 2017; Miyake et al., 2000; Miyake & Friedman, 2012). Furthermore, based on the results of previous latent variable studies on healthy participants, Friedman and Miyake (2012) created two models of executive functions, as depicted in Figure 1.1. The first model (Figure 1.1.A) is a three-factor model in which inhibition, shifting, and updating are three separate yet related latent variables. The second model (Figure 1.1.B) is a bi-factor model in which the three factors are united in one common executive function, with unique variance in updating-specific and shifting-specific factors.

Figure 1.1.

Three-factor and bi-factor models of executive functions



Both three-factor and bi-factor models showed the unity and diversity of executive functions in different way. The three-factor model showed the unity of executive functions through the correlation of each factor of executive functions, indicating shared cognitive resources between factors (Miyake et al., 2000). Conversely, the bi-factor model had one latent factor including all executive functions tasks, showing that executive function has a common factor (Friedman et al., 2008). Furthermore, the diversity of the three-factor model was shown through three different latent variables which have three specific roles (Miyake et al., 2000). Meanwhile, the bi-factor model had updating-specific and shifting specific latent factors unrelated with each other because they are nested models, indicating the unique factors (Friedman et al., 2008). In summary, the three-factor model focuses on identifying distinct components of executive functions, whereas bi-factor model incorporates a higher-order general factor while acknowledging both shared and unique contributions to executive functions (Friedman & Miyake, 2017; Miyake & Friedman, 2012).

A common difficulty in executive functions research is the task impurity problem. The tasks measuring executive functions may have systematic variance measuring non-executive functions processes (Miyake & Friedman, 2012), that is, lower-level cognitive processes such as colour processing in Stroop tasks, and visuospatial processing in spatial n-back tasks (Friedman & Miyake, 2017; Snyder et al., 2015). Therefore, it is problematic to measure executive functions with only one task because performance in that task may reflect all these other processes above and beyond 'purely' executive functions. The executive functions model was better to be developed with latent-variable modelling, as a low score in one particular task does not imply a low executive functions ability (Miyake et al., 2000).

1.3.What is Music Making?

There are two ways in which music can relate to cognitive abilities: actively making music (Hetland, 2000a; Schellenberg, 2011) and receptively listening to music (Hetland, 2000b; Pietschnig et al., 2010). In this thesis, I will focus only on music making, in which people actively engage in musical activity as opposed to listening to music. I choose to use the term *music making* to investigate individual differences in musical experience in making music and music skills. I deliberately avoid using the term *music training* as it typically implies formal learning of musical instruments or vocal techniques and may not fully encompass the broader spectrum of music-making experiences that can be related to executive functions.

Past studies in music making and executive functions varied in their definition and operationalisation of music making, possibly contributing to inconsistent results of these past studies. Four approaches to measuring music making can be identified in the literature. First, many studies have classified participants as musicians and non-musicians by years of music training (e.g. Bialystok & DePape, 2009; Criscuolo et al., 2019; D'Souza et al., 2018; Moradzadeh et al., 2015; Suárez et al., 2016). For example, one study classed participants as musicians when they reported having had eight or more years of music training (e.g. D'Souza et al., 2018), whereas other studies set the threshold to five years (e.g. Criscuolo et al., 2019; Suárez et al., 2016). Another method to classify the participants was by separating musicians and non-musicians according to the score on the questionnaire. For instance, Porflitt and Rosas-Díaz (2019) asked the participants to fill a questionnaire, then classified participants above the score threshold as musicians and below the score threshold as non-musicians. A drawback in dichotomising participants based on continuous measures (e.g. years of music training) and comparing extreme groups (e.g. professional musicians and non-musicians) was that this might ignore people who had average performance which could inflate the effect sizes and then yield a misleading result (MacCallum et al., 2002; Preacher et al., 2005).

Second, studies have used self-report measures of active music-making engagement, such as the Goldsmith Musical Sophistication Index (GOLD-MSI; Müllensiefen et al., 2014) and the Ollen Musical Sophistication Index (OMSI; Ollen, 2006). Other studies used a single question to assess years of music training (e.g. der Nederlanden et al., 2020; Lu & Greenwald, 2016) or years of playing a musical instrument (e.g. Gray & Gow, 2020) and then used them as a continuous variable to correlate with executive functions. Another possible measure is the frequency of music making which asked participants how often they made music in daily life in the certain amount of times (Jopp & Hertzog, 2010). It is more common in the leisure activity studies that ask participants the frequency of engaging in certain activities, including music activities (e.g. Guye et al., 2020; Paggi et al., 2016). Furthermore, it is noteworthy that active music engagement is not limited to self-identified musicians; individuals who do not consider themselves musicians may still participate in music-making activities and have high musical test scores (Correia et al., 2023). This method is used in Chapter 3, in which the frequency of music-making is assessed using a single self-report question.

Third, some studies use computer-based tasks to measure musical skills objectively. One of the earliest influences in the history of measuring music skills was by Seashore (1915), who distinguished three elements of musical sounds: pitch, time (rhythm), and intensity of the tone. Nowadays, the tasks commonly measuring musical skills consist of a listening test in which participants are asked to decide whether the two melodies are similar or choose which rhythms are on the beat. Examples of computerised tasks measuring music skill are the Musical Ear Test (MET; Wallentin et al., 2010), Melodic Discrimination Testing (MDT; Harrison et al., 2017), Profile of Music Perception Skills (PROMS; Law & Zentner, 2012) and the Computerised Adaptive Beat Alignment Test (CA-BAT; Harrison & Müllensiefen, 2018).

Lastly, some studies define music making broadly by combining self-report measures and computer-based tasks as composite scores (e.g. Slevc et al., 2016). Besides combining them as composite scores, another option was to consider those measures as manifest variables to measure music making indirectly as a latent variable in a structural equation modelling (Kline, 2023). The advantage of using this method is reducing the social desirability bias of completing a self-report questionnaire (see Brenner & DeLamater, 2016) by complementing the questionnaire with a more objective task. In this thesis, this method is used to measure music making by complementing a questionnaire with two computer-based tasks (Chapter 4).

1.4.The Relationship between Music Making and Executive Functions

Many reviews and meta-analyses have shown that music-making is related to executive functions (e.g. Degé & Frischen, 2022; Hernández et al., 2020; Román-Caballero et al., 2018; Schellenberg & Lima, 2023; Schellenberg & Weiss, 2013). However, most of the past research on music-making and executive functions focuses on children and adolescents, which supported by previous review in executive functions that found 25 % of studies including samples of adults and only 7% including older adults in 106 studies reviewed (Baggetta & Alexander, 2016). It can be summarised that, in general, research on executive functions in adults is under-explored compared to children and adolescents. Gaining a comprehensive understanding of these connections in adults could offer significant knowledge on the development of executive functions and possible strategies for aiding executive function in adults in various age ranges.

Existing past research in adults found that music-making associated with each of the executive function components, that is, shifting (Moradzadeh et al., 2015), inhibition (D'Souza et al., 2018), and working memory (D'Souza et al., 2018; George & Coch, 2011). Much of past research on the relationship between music making and executive functions compared dichotomous groups of musicians and non-musicians, but they differed in how they classified. For example, Clayton et al. (2016) found that musicians, who were mostly students with music majors with at least ten years of formal music training, outperformed non-musicians in auditory working memory tasks but no significant differences regarding their inhibition and shifting performance. Another study classified musicians and non-musicians by comparing their Ollen Musical Sophistication Index (OMSI) score, and those who scored below 500 were considered as non-musicians (Porflitt & Rosas-Díaz, 2019). The result of the study was that musicians performed significantly better than non-musicians in all inhibition tasks and 1 out of 2 working memory tasks, but no differences were found for shifting. Aside from comparing musicians and non-musicians, other studies utilised correlational design to find the relationship between music making and executive functions. Criscuolo et al. (2019) investigated the relationship between music making, as measured using years of music training and music practice, and executive function, which was measured solely using the Stroop task, in adults. The results showed a positive relationship between music-making and executive function after accounting for individual differences in intelligence.

Based on the previous explanation, it can be summarised that there was a relationship between music making and executive functions, especially inhibition and working memory. However, due to different ways of operationalising music-making in previous studies, it is difficult to conclude why music-making is related to inhibition and working memory in particular**.** Furthermore, these past studies measured executive functions only with a single task and not using a latent-variable approach, which raises the task-impurity problem (Friedman & Miyake, 2017; Miyake et al., 2000; Miyake & Friedman, 2012).

Okada and Slevc (2018) addressed the task impurity problem by assessing all three factors of executive functions with three different measures for each factor and related these to the Gold-MSI. The results showed that music making was correlated with working memory, but not inhibition and shifting. The study also found that the correlation between music-making and working memory was still apparent even after accounting for individual differences in fluid intelligence. Thus, music-making may be related to specific executive functions instead of executive functions in general.

In addition to addressing the task impurity problem by using several measures to assess factors of executive functions, it is important to consider using different ways of measuring music-making and executive functions to capture these concepts more comprehensively. Slevc et al. (2016) measured music making by creating a composite score from a self-report questionnaire using the Ollen Musical Sophistication Index (OMSI) and an objective test using the Musical Ear Test (MET). Each factor of executive functions was assessed with two tasks with different modalities each: auditory and visual. The results were that the music-making composite score correlated significantly to working memory in both modalities but not to inhibition and shifting. This study showed the importance of measuring different modalities of executive functions because music making, which is commonly associated with auditory modality, may benefit visual modality, too.

Taken together, most of the studies found that music making, as measured using musical sophistication, music skill, and years of music training in the previous studies, was mostly correlated with working memory and inhibition when executive functions were measured using Miyake and Friedman's (2012) three-factor model. However, some past studies measured executive functions using a single measurement, which measured only one or two factors of executive functions (e.g. inhibition and working memory; Criscuolo et al., 2019; D'Souza et al., 2018). Therefore, in this thesis, I will measure executive functions using the three-factor model of executive functions with at least three tasks for each factor (Miyake et al., 2000; Miyake & Friedman, 2012).

1.5. Why are Music Making and Executive Functions Related?

Many studies have investigated the relationship between music making and executive functions. However, it was uncertain why they are related. There are two possible explanations for their relationship. One potential explanation is that the skill acquired from music making may benefit executive functions (Criscuolo et al., 2019; Okada & Slevc, 2018). It can also be other way around in which people who have high executive functions may be more likely to make music (Okada & Slevc, 2018). Importantly, however, it should be noted that previous studies were correlational only and, thus, cannot imply causation between music making and executive functions.

Another possible explanation is that the relationship between music making and executive functions are mediated by another variable. The potential variable is sensory discrimination, the ability to make fine discriminations in different sensory stimuli, such as different colour hue (visual), loudness (auditory), and weight (tactile; Spearman, 1904). Sensory discrimination is a good candidate as a mediator because making music is known to be a multisensory activity, including auditory, visual, and somatosensory processing, which is related to the brain plasticity (Rogenmoser et al., 2018). In this thesis, I investigate whether music making, an activity closely associated with sound perception and production, is related specifically to auditory discrimination or it can be generalised to another type of discrimination, namely visual discrimination.

Previous studies have found that sensory discrimination was related to both music making and cognitive abilities. Sensory discrimination was known to be enhanced in musicians, especially auditory discrimination (Hyde et al., 2009; Kraus & Chandrasekaran, 2010; Okada & Slevc, 2018). Although another study also found that musicians were better in auditory, visual, and temporal discriminations compared to non-musicians (Kubaszek et al., 2021). In addition, sensory discrimination was also related to cognitive abilities (Acton & Schroeder, 2001; Deary et al., 2004), including inhibition (Burgoyne et al., 2024) and working memory (Tsukahara et al., 2020; Voelke et al., 2013). In separate studies, it can be summarised that music making was related to sensory discrimination and sensory discrimination was related to executive functions. However, to date, these three variables were not measured and accounted for simultaneously. Therefore, I will investigate whether sensory discrimination account for the relation between music making and executive functions.

1.6.Research Gap

Although there is evidence indicating a correlation between music-making and executive functions, three critical gaps remain that warrant further investigation. First, the variability in how music making and executive functions have been operationalised across different studies may give different conclusions on how they are related. Previous studies often measured only one specific aspect of executive functions but drew conclusions for executive functions in general. Moreover, more investigation is required to ascertain whether there is a distinction when assessing music making as a discrete (e.g., musicians vs. non-musicians) and continuous variables. This research gap will be addressed in Chapter 2 by reviewing both group comparison and correlational studies, and what task paradigms were used. Furthermore, Chapters 3 and 4 will address the task-impurity problem by assessing executive functions as latent variables, using at least three measures for each factor of executive functions.

Second, prior investigations have primarily focused on children and adolescents with limited exploration of the relationship between music-making and executive functions in adults and older adults, especially in healthy participants (Bagetta & Alexander, 2016). This discrepancy implies that, overall, research on the relationship between music-making and executive functions in adult populations remains under-explored in comparison to studies involving younger individuals. This research gap will be addressed in Chapter 2 by focusing on past studies with adult samples, Chapter 3 by including participants with an age range spanning the adult lifespan, and Chapter 4 by including young adult participants. Addressing these critical research gaps will not only contribute to a more comprehensive understanding of the relationship between music-making and executive functions but also offer recommendations for a more refined methodology in this research area.

Lastly, past studies rarely investigated why music-making and executive functions are related. As previously mentioned, sensory discrimination, which was known to be enhanced in musicians (Kubaszek et al., 2021) and related to working memory (Tsukahara et al., 2020; Voelke et al., 2013), was a prime candidate to be the mediator for the relationship. Understanding the underlying mechanism between them may answer whether music making has direct or indirect relationship with executive functions. Therefore, Chapter 4 will address this research gap by measuring music making, sensory discrimination, and working memory in a single study.

1.7. Research Questions and Aims of This Thesis

The overarching research question of this thesis is **how and why music making and executive functions are related**. The first aim of this thesis is to investigate how strong the relationship between music-making and executive functions in typical adults. In Chapter 2, I conducted a systematic review and meta-analysis to investigate how strongly music-making relates to inhibition, shifting, and working memory. In Chapter 3, to test the hypothesis in an empirical study, I investigated the relationship between the frequency of music making and each executive functions factor across the lifespan. The results of Chapter 2 showed a stronger relationship between music making and working memory, but Chapter 3 found no relationship between the frequency of music making and working memory. Therefore, the second aim is to investigate what is the mechanism underlying the relationship between music making and working memory. The second aim is addressed in Chapter 4, where I conducted a latent-mediation study to investigate how music-making relates to working memory and sensory discrimination in adults when measured and accounted for simultaneously. Bringing together all the studies, Chapter 5 will highlight the main findings of this thesis, theoretical and methodological implications, limitations, and future outlook of music-making and executive functions study.

Chapter 2 - How Is Music Making Related to Executive Functions? A Systematic Review and Meta-Analysis

Interim summary:

As previous studies found mixed results on the relationship between music making and executive functions, we first conducted a systematic review and meta-analysis to systematically search the previous literature. We investigated how music making and three factors of executive functions are related by finding the strength of the relationship between them. This chapter would inform us whether all factors or specific factors of executive functions should be considered further.

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)

Submission status: Under review

Findings from this chapter have been presented in the following conferences:

Aryanto, C.B., Aisyah, A.R.K., Blakey, E., Timmers, R., & von Bastian, C.C. *How are active music making and executive functions related? A systematic review and meta-analysis* [Poster]. Psychonomic Society 62nd Annual Meeting, Online.

Aryanto, C.B., Aisyah, A.R.K., Blakey, E., Timmers, R., & von Bastian, C.C. *How are active music making and executive functions related? A systematic review and meta-analysis* [Talk]. Learning and Plasticity (LaP) Meeting, Finland.

Presentation in the following conference based on research in this chapter has received the Grindley Grant from Experimental Psychology Society (EPS):

Aryanto, C.B., Aisyah, A.R.K., Blakey, E., Timmers, R., & von Bastian, C.C. *How are active music-making and executive functions related? A systematic review and meta-analysis* [Poster]. The 22nd European Society for Cognitive Psychology (ESCoP) Conference, Lille, France.

Acknowledgement:

This chapter 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. 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.

Abstract

A person's lifetime involvement with music making has been hypothesised 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 music making relates to each of the three executive functions. The results showed significant medium associations between music making and executive functions (g = 0.43), with small to medium associations between music making 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 music making and inhibition, and the paradigm used to assess executive functions moderated the association between music making 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 music making. 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 music making.

Public Significance Statement

Regularly playing music is thought to relate to executive functions, that is, one's ability to plan and engage in goal-directed behaviour, based on the assumption that there is a functional overlap between music making and executive functions. This systematic review and meta-analysis found that playing music is related to three executive functions. Music making relates significantly, although weakly, to the ability to switch between tasks and control

impulses, and significantly and more strongly to the ability to maintain and process information in memory.

Keywords: Music making, executive functions, shifting, inhibition, updating, meta-analysis

2.1. Introduction

Executive function is an umbrella term for higher-order cognitive skills underpinning goal-directed behaviour (Friedman & Miyake, 2017; Miyake & Friedman, 2012). Past research has established three factors of executive functions: inhibition, shifting, and working memory updating (Miyake et al., 2000; Miyake & Friedman, 2012). *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. 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).

Playing music and involvement in musical training has been shown to be related to cognitive function (Schellenberg & Weiss, 2013; Schellenberg & Lima, 2023) and, specifically, executive functions (Okada & Slevc, 2020). Here, we use the term *music making* 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. Music making 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 music making 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 music making. Specifically, some studies found that music making correlated with inhibition and working memory but not with shifting (Okada & Slevc, 2018; Porflitt & Rosas-Díaz, 2019). Another study found that music making was related only to working memory, but not inhibition and shifting (Slevc et al., 2016).

The present review systematically synthesises these different findings to examine the relation between music making and inhibition, shifting, and working memory updating across the body of the existing literature. Moreover, we aimed to clarify the role 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 music making is assessed. Some past research used only self-report questionnaires to assess musicianship (Criscuolo et al., 2019; D'Souza et al., 2018; Okada & Slevc, 2018), whereas other studies combined self-report questionnaires with objective measures such as melodic and rhythm test scores (Hansen et al., 2013; Slevc et al., 2016; Talamini et al., 2016). 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. 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.

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 music making and executive functions may only emerge with a certain amount of music making experience.

Third, a related but more fundamental issue concerns whether considering music making 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 music making 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 artificial 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 music making 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 music making 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 music making, and demographic characteristics (age, gender, socioeconomic status, and education).

2.2. 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 music making 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.

2.3. Present Meta-Analysis

Previous systematic review on children's music making and executive functions had been done several times, yet studies in adults' population were limited to older adults instead of the entire adulthood. This meta-analysis will fill the gap by focusing on the adults aged 18 and above. The few existing previous meta-analyses focusing on adults suggest that music making 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 music making and executive functions are related in adulthood when considering the degree of experience in music making 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 music making and executive functions to address the following research questions:

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

2.4. Method

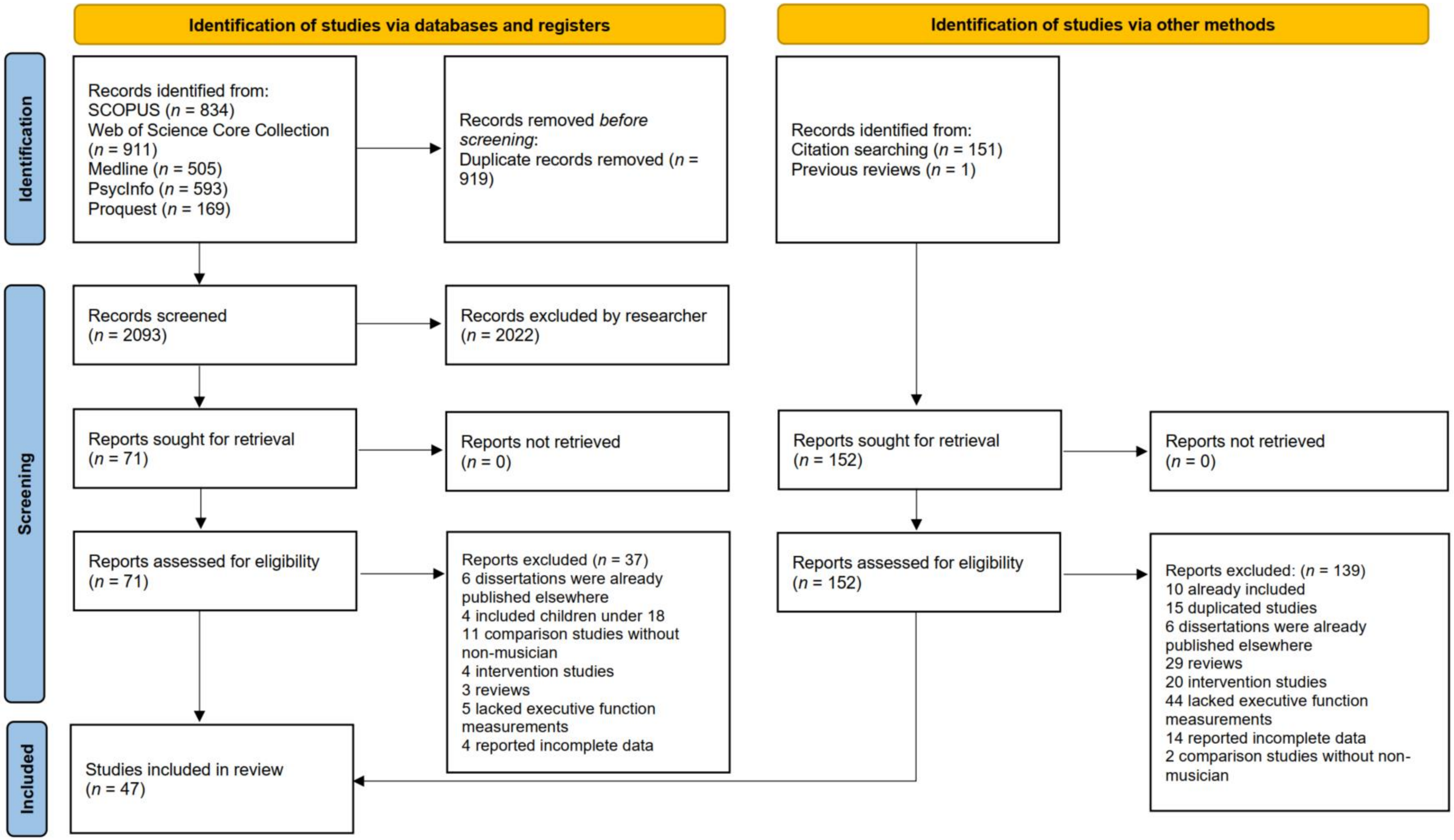
This meta-analysis followed the PRISMA statement 2020 (Page et al., 2021). The study protocol was preregistered on the Open Science Framework on January 19, 2021 (https://osf.io/evj5d).

2.4.1. Search Strategy

The flow diagram depicted in Figure 2.1 summarises 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 2.1

PRISMA Flowchart of Systematic Search and Study Selection



2.4.2. 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 music making) between the measures of at least one executive function (i.e., inhibition, shifting, updating/working memory) and music making (e.g., musicality, musicianship, music training, music playing). Furthermore, we included only samples of healthy adults (at least 18 years old). Forty-seven of the 223 full texts assessed met these criteria.

2.4.3. 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/?view_only=1e92639423da4c9ab8df0ea3a4f1702f). 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.

2.4.4. 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 music making with other variables such as participants socio-economic background. Measures of exposure referred to performance biases caused by inadequate measures for assessing cognitive and music-making 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 music making). 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).

2.4.5. 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.

2.4.6. Meta-Analytic Procedure

2.4.6.1. 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 the code provided by Fernández-Castilla et al. (2020).

2.4.6.2. Moderator Analysis

We considered study type, the sum of the risk of bias score, gender distribution, age, socioeconomic status, years of education, and task paradigm as potential moderators. The gender distribution was counted based on the proportion of females to males. Each task paradigm was considered a separate category if at least 10 cases were available; otherwise, the paradigm was categorised as ‘others’.

2.4.6.3. 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.

2.4.6.4. 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 (α = 1.23, β = 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 based on the t-statistics and computed using BayesFactor package (version 0.9.12-4.4; Morey & Rouder, 2015) and the BF against a null region of -∞ 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 (see Table 2.1).

Table 2.1

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 favour of the alternative hypothesis, BF_{01} = evidence in favour of the null hypothesis.

2.4.6.5. Publication Bias

We assessed publication bias with funnel plots and Egger's and Begg’s and Mazumdar's tests. Egger's test aims to analyse the relationship between the standardised 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 analysed using the *metafor* package (Version 3.4-0, Viechtbauer, 2010).

2.5. Results

Table 2.2 provides descriptions of the included studies. Of the 47 included studies included in the meta-analysis, 33 reported between-group comparisons (k = 124) and 14 correlational findings (k = 111) from 4651 participants (M age = 33.39 years, SD = 20.01

years). 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).

Table 2.2

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	Gender	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			-0.05 [-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), gender (percentage of females relative to males), and education (years) scores are given in means. *N* is the total sample size.

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 favour of the alternative hypothesis; comp = comparison.

2.5.1. Three-Level Meta-Analysis

Before conducted three-level meta-analysis, we conducted outlier analyses to identify influential cases across all executive function factors and for each factor of executive functions. 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 two influential cases in executive functions and inhibition, and one influential case in shifting and working memory. The value of the different influence diagnostics of each executive functions factor and their plots can be seen in Appendix A. Here, we reported the result without influential cases and, for the sake of completeness, the result with influential cases can be found in Appendix B.

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 music making 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 suggested notable differences in the strength of associations between music making and single executive functions factors. Indeed, the association between music making 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. Table 2.3 summarises 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 2.2 shows a forest plot of pooled effect sizes for the association between music making 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 2.3

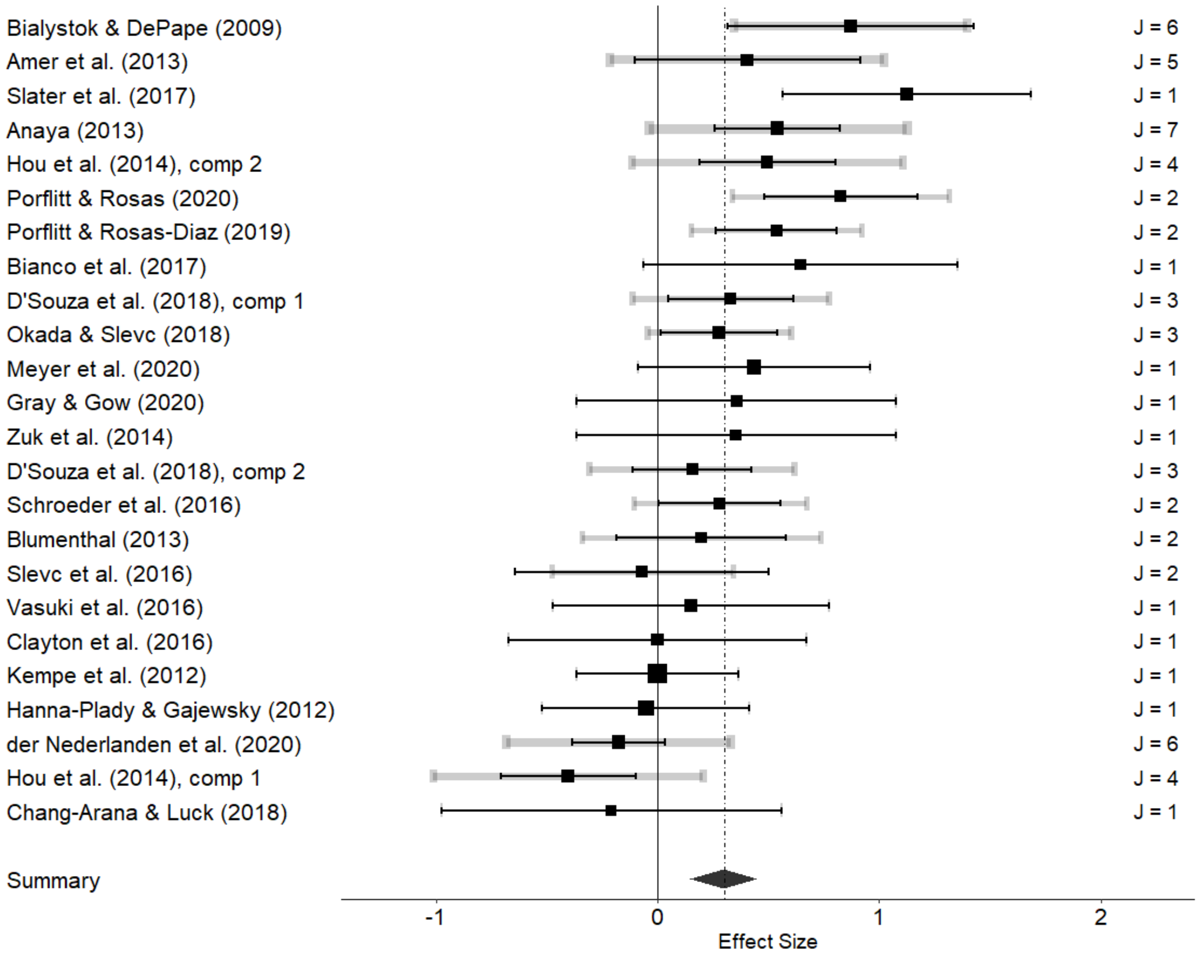
Average Effect Sizes for the Relation between Music Making and Executive Functions

						I^2		τ^2	
Measures	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 favour of the alternative hypothesis.

Figure 2.2

Multilevel Meta-Analysis of Music Making and Inhibition

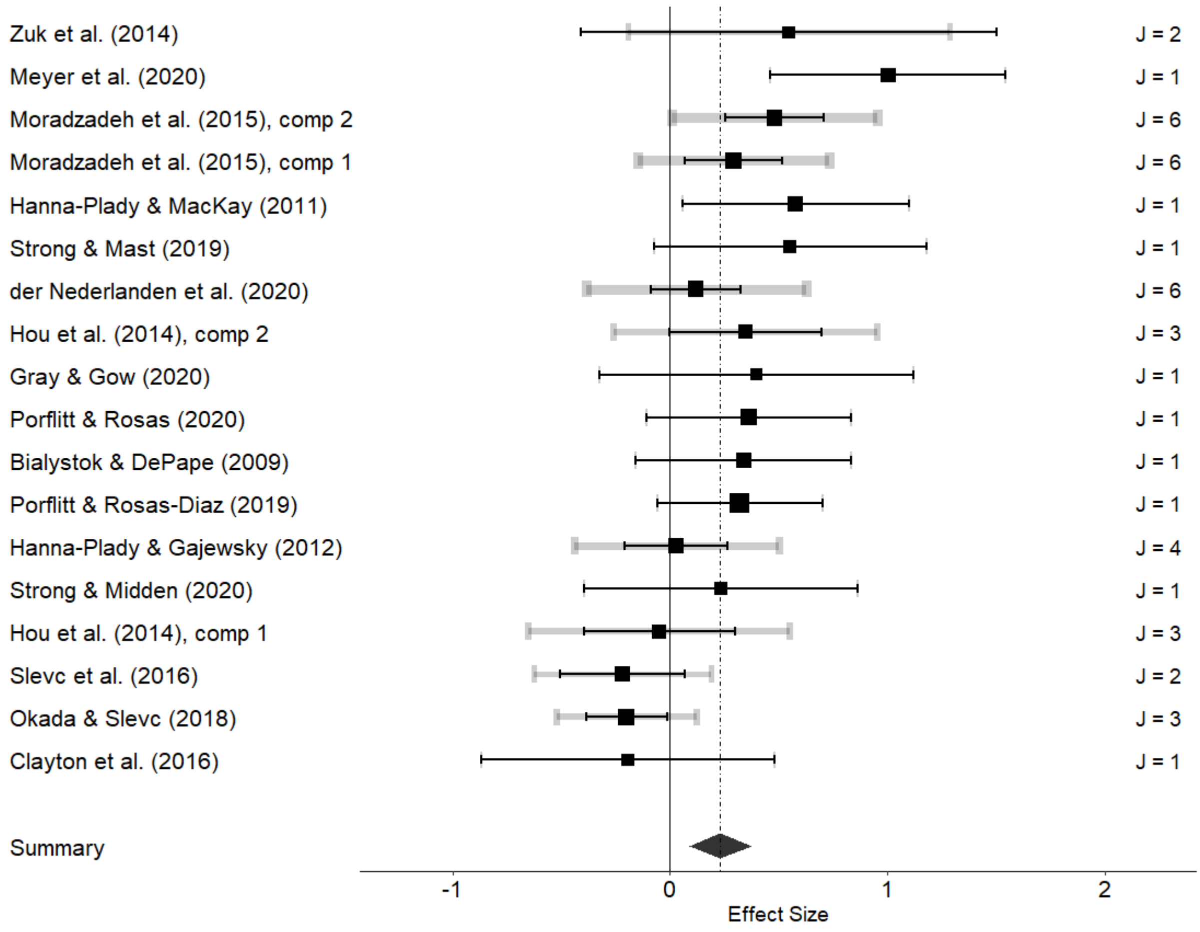


Note. Effect sizes are weighted averages and can include multiple effect sizes *J*. Black error bars represent 95% confidence intervals of the weighted averages. Gray error bars reflect the sampling variance of individual observed effect sizes of each study. The thickness of the grey error bars is proportional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

Music making was also significantly related to the shifting factor (see Figure 2.3), 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 2.3

Multilevel Meta-Analysis of Music Making and Shifting

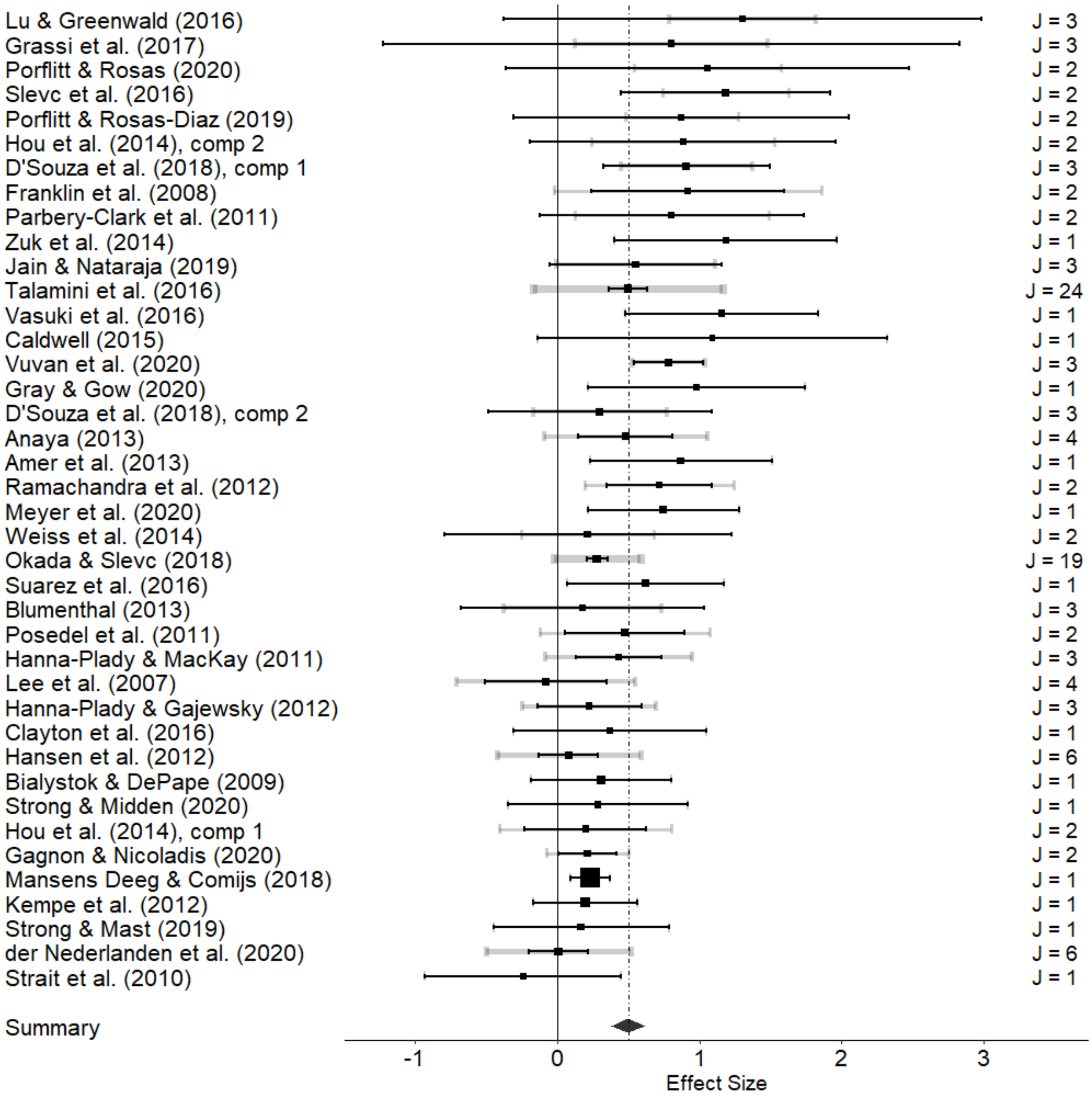


Note. Effect sizes are weighted averages and can include multiple effect sizes *J*. Black error bars represent 95% confidence intervals of the weighted averages. Gray error bars reflect the sampling variance of individual observed effect sizes of each study. The thickness of the grey error bars is proportional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

Finally, we also found a significant medium association between music making and working memory (see Figure 2.4), 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, $Q(df = 125) = 426.82$, $p < .001$, $\tau^2_{level\ 3} = 0.05$, $\tau^2_{level\ 2} = 0.12$, $I^2_{level\ 3} = 20.98\%$, $I^2_{level\ 2} = 55.08\%$.

Figure 2.4

Multilevel Meta-Analysis of Music Making 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 grey error bars is proportional to the number of effect sizes within individual studies.

2.5.2. Risk of Bias Assessment

Figure 2.5 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 as these studies relied only on self-report measurement to measure music making, for example asking participants' years of music training. For the other studies, the risk of bias for music assessment was unclear because they used both self-report and objective measures of music making (e.g. both a self-report questionnaire and melodic testing) but combined into a composite score measuring musical ability. The risk of bias from blinding of both participants and experimenters was rated to be largely unclear, because most studies did not explicitly explain their blinding procedures. The risk of bias was considered low for all other criteria (i.e. selection of participants, confounding variable, cognitive measurement, incomplete outcome, and selective outcome).

Figure 2.5

Risk of Bias Ratings

2.5.3. Moderator Analysis

Table 2.4 lists the results of the moderator analyses in regard to participant characteristics (gender distribution, age, and years of education), study type (correlational vs. comparison studies), cognitive task used to assess executive functions (task paradigm, see Table 2.5 for paradigms included in this moderator analysis), and study risk of bias. 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). In addition, no moderator analysis was run for years of education on the association between music making and shifting, because less than 10 studies reported the relevant data.

The association between music making 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 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 favoured the null hypothesis, $BF_{10} = 1/4.09$. Furthermore, we found a significant moderation effect of study type, $F(1, 42) = 5.50$, $p = .024$, $BF_{10} = 2.38$, 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, $BF_{10} = 2.38$. Finally, the association between music making and working memory was significantly moderated by the task paradigm used, $F(2, 123) = 3.76$, $p = .026$, 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 2.4

Effects of Moderators on the Relationship between Music Making and Executive Functions

Moderator	*k*	*F*	DF	*p*	BF_{10} [sensitivity]
			Inhibition		
Study type	61	0.46	1, 59	.499	1/3.41 [1/1.85, 1/5.16]
Risk of bias score	61	16.80	1, 59	<.001	6.65 [6.56, 4.99]
Gender distribution	34	3.92	1, 32	.056	1/9.59 [1/3.16, 1/16.14]
Age	49	0.03	1, 47	.860	1/59.13 [1/20.07, 1/99.21]
Years of education	15	0.22	1, 13	.649	1/3.23 [1/1.66, 1/5.03]
Task paradigm	61	1.82	1, 59	.183	1/2.58 [1/1.41, 1/3.85]
			Shifting		
Study type	44	5.50	1, 42	.024	2.38 [2.75, 1.79]
Risk of bias score	44	0.31	1, 42	.583	1/3.42 [1/1.67, 1/5.47]
Gender distribution	21	6.46	1, 19	.020	1/4.09 [1/1.41, 1/6.80]
Age	41	0.70	1, 39	.410	1/69.73 [1/22.49, <1/100]
Task paradigm	44	0.97	2, 41	.387	1/8.55 [1/2.78, 1/18.22]
			Working Memory		
Study type	126	0.19	1, 124	.659	1/5.55 [1/2.66, 1/8.36]
Risk of bias score	126	0.07	1, 124	.796	1/4.48 [1/1.98, 1/7.00]
Gender distribution	91	1.01	1, 89	.316	1/56.94 [1/19.38, 1/93.72]
Age	116	<0.01	1, 114	.988	<1/100 [1/39.68, <1/100]
Years of education	52	0.22	1, 50	.643	1/9.91 [1/3.67, 1/16.30]
Task paradigm	126	3.76	2, 123	.026	1/1.28 [1.40, 1/2.40]

Note. DF = degrees of freedom; BF_{10} = Bayes factor in favour of the alternative hypothesis.

Table 2.5

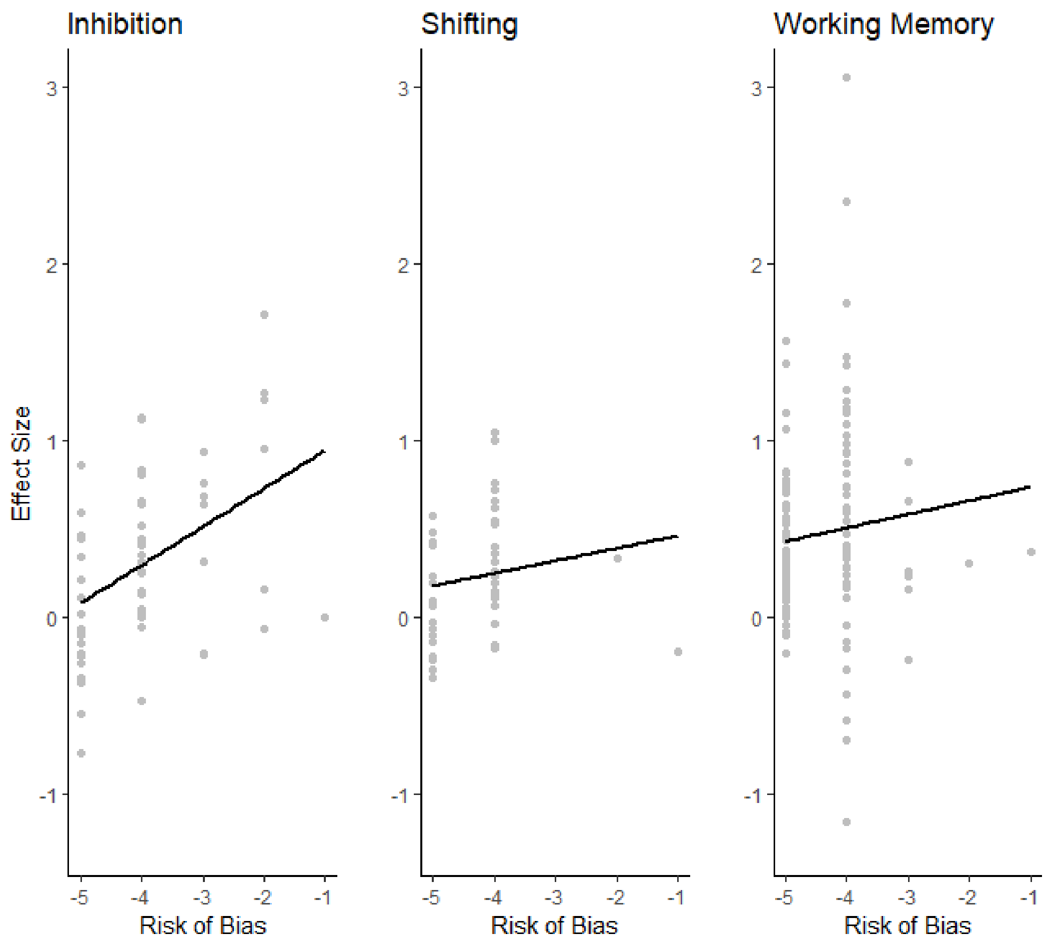
List of paradigms included in the systematic review

Factor	Paradigm	Description
Inhibition	Stroop	Participants were shown colour words (like "red," "blue," etc.) printed in different colours. They must name the colour of the ink, not the word itself.
	Simon	Participants responded to the colour of a stimulus, which appears on the left or right side of the screen.
	Go/No-Go	Participants responded to certain stimuli (Go) and withheld their response to others (No-Go). This task measures response inhibition.
	Continuous Performance Task (CPT)	Participants responded to certain target stimuli and ignored non-targets over a period of time.
	Flanker	Participants required to focus on a central target stimulus and ignored flanking stimuli that can either be congruent or incongruent.
	Stop signal	Participants responded to a Go signal but must stop their response when a stop signal appears.
	Antisaccade	Participants inhibited the reflexive saccade (eye movement) to a suddenly appearing stimulus and instead make a saccade in the opposite direction.
	Tower Test	Participants moved stack of disks from a pre-determined starting position to another position in a certain amount of moves.
Shifting	Card sorting	Participants sorted cards according to different rules (colour, shape, number) which change periodically.
	Trail Making	Participants drew lines to connect a sequence of numbered and/or lettered circles in a specific order, alternating between numbers and letters.
	Task switching	Participants switched between different tasks or rules, often involving different stimulus-response mappings.

	Design fluency	Participants generated as many unique designs as possible within a given time, following specific rules.
	Reversal learning	Participants learnt to respond to certain stimuli to receive rewards and then must adapt when the stimulus-reward contingencies are reversed.
Working memory	Complex span	Participants remembered a series of items (like letters or words) while simultaneously performing another task (e.g., solving math problems).
	Digit span	Participants repeated a sequence of numbers in the same order (forward span) or reverse order (backward span).
	N-Back	Participants monitored a sequence of stimuli and must indicate when the current stimulus matches the one from 'n' steps earlier in the sequence.
	Letter Number Sequencing	Participants were given a mixed sequence of letters and numbers and must reorder them by first stating the numbers in ascending order, then the letters in alphabetical order.
	Visual Pattern Test	Participants were shown a complex visual pattern and must reproduce it after a delay.
	Keep track	Participants kept track of multiple streams of information and update their memory when specific target items appear.
	Binding	Participants remembered combinations of features (e.g., colour and shape) and later recall these combinations.

Figure 2.6

Risk of Bias as Moderator of the Association between Music Making and Executive Functions



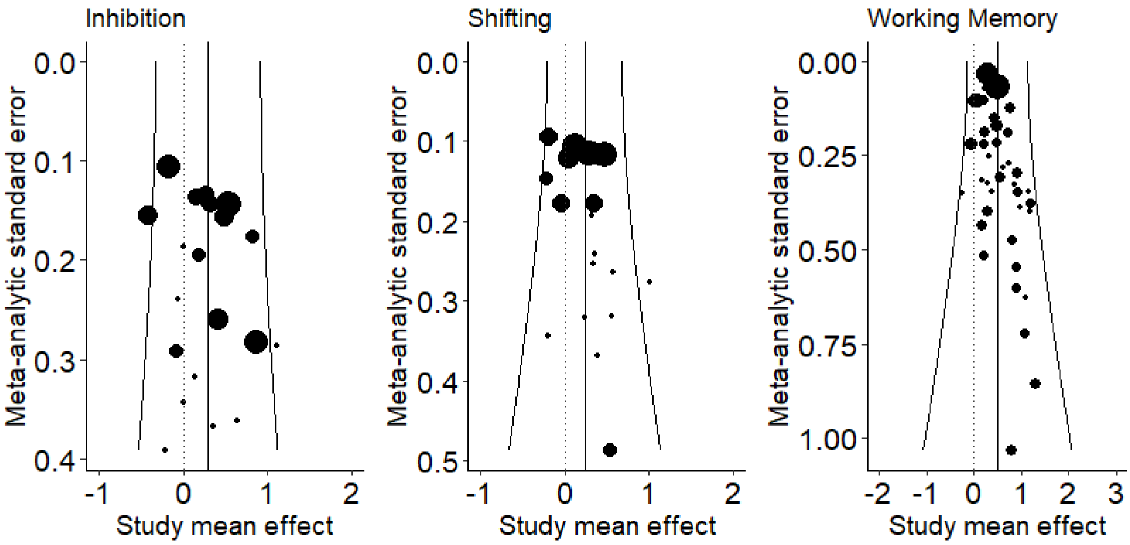
Note. Scatterplots relating pooled effect sizes of the association between music making and each factor of executive functions with the summed risk of bias scores for each study.

2.5.4. Publication Bias

Funnel plots relating standard errors and effect sizes estimates for each study were inspected for each factor of executive functions (see Figure 2.7). 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, $p = .394$, but were significant for shifting ($p = .006$) and working memory ($p = .001$). 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 from shifting and inhibition.

Figure 2.7

Funnel Plots of Effect Sizes of the Association between Music Making 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.

2.6. Discussion

This preregistered systematic review and meta-analysis examined the relationship between music making 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 music making was significantly related to 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 throughout ($BF_{10} > 100$). The moderator analyses we conducted revealed mostly non-significant (age and education) or ambiguous effects (gender distribution, study type, and paradigm). The only exception was that a larger risk of bias increased the relationship between music making 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 music making 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.

2.6.1. A Critical Role of Working Memory in Music Making?

The findings of the present meta-analysis suggest that working memory plays an important role in making music. 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 music making 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 music making may not be a direct relationship but could be mediated by auditory discrimination ability, a hypothesis warranting further investigation (see Aryanto et al., chapter 4).

2.6.2. Impact of Methodological and Publication Biases on the Association between Music Making 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 music making and inhibition, suggesting that biases may contribute to inflating meta-analytic effect sizes. The highest risk of bias was identified for the assessment of music making. Specifically, most studies included in the current meta-analysis assessed music making with self-report questionnaires, thereby relying exclusively on participants' memory and perception of their own music making 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 music making and executive functions – or other cognitive abilities, objective measures can assess music making more precisely and on a continuous spectrum. Objective measures such as the beat alignment (P. M. C. Harrison & Müllensiefen, 2018) or the melodic discrimination test (P. M. C. 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, complementing the assessment of music making with objective measures in addition to self-report questionnaires can enhance our understanding of the relationship between music making and executive functions.

The analysis of publication bias suggested a potential small-sample bias on the relationship between music making, 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 music-making measures (MacCallum et al., 2002; McClelland et al., 2015; Preacher et al., 2005). Thus, future studies should instead consider operationalizing music making as a continuous variable and follow sample size recommendations for assessing robust and stable correlations (e.g., Schönbrodt & Perugini, 2013).

2.6.3. Methodological Recommendations for Future Research

Taken together, future research investigating the relationship between music making 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 music making by objective, performance-based measures of music making. Preregistering research and reporting the results transparently would further decrease the risk of biases.

2.6.4. 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 relating to study design (correlational vs. group-comparison designs) or relating to assessment of music making (self-report vs. objective instruments). After discussion, raters agreed to rather the criterion based on the assessment of music making 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 analyses.

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).

It also needs to be acknowledged that there were some potential moderators not identified in this study, for example the modality of executive functions (e.g., visual and auditory). Previous meta-analysis showed there was difference on musicians and non-musicians tonal, verbal, and visuospatial short term, long term, and working memory, suggesting that the musicians maybe better in specific modality (Talamini et al., 2017). Other moderators that could be added are duration of musical training and type of music making. Although, like socioeconomic status measures, previous studies varied in measuring musical training, for example by using a Likert scale and asking directly about the frequency and intensity of musical training. In terms of the type of music making, musicians can be defined as professional, amateur, serious amateur, or self-learned musicians, which not always been reported in the papers. Future studies could have a consensus on how to measure musical training and type of music making to be comparable among different studies.

2.6.5. Conclusions

This pre-registered systematic review and meta-analysis confirmed that music making 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 music making in larger, correlational samples. Furthermore, this review suggests exciting avenues for further research to explore the mechanisms underpinning the relation between music making and executive functions, in particular working memory.

Chapter 3 - Is the Frequency of Music Making Related to Executive Functions?

Interim summary:

In Chapter 2, we found that music making was related significantly to all three factors of executive functions, with the largest and the most decisive evidence for an association with working memory. In this chapter, I will further test these associations with executive function in an empirical study focusing on the frequency of music making. We measured frequency of music making in the past 12 months by self-report, and executive functions using three different tasks for each factor of executive function in a sample spanning the adult lifespan.

Contributions:

Christ Billy Aryanto (conceptualisation, data curation, formal analysis, validation, investigation, data curation, visualisation, writing — original draft)

Emma Blakey (conceptualisation, supervision, writing – review & editing)

Renee Timmers (conceptualisation, supervision)

Claudia C. von Bastian (conceptualisation, funding acquisition, methodology, resources, software, data curation, supervision, writing – review & editing)

Submission status: Manuscript in preparation for publication

Findings from this chapter have been presented in the following conference:

Aryanto, C.B., Blakey, E., Timmers, R., Hyde, E.R.A., Reinhartz, A., Jiang, S., Bellevile, S., Strobach, T., & von Bastian, C.C. *Does the Frequency of Music Making Relate to Executive Functions? An Individual Differences Investigation Across the Adult Lifespan* [Talk]. 12th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM12).

Acknowledgement:

This chapter was supported by the Economic and Social Research Council (ESRC) to Claudia C. von Bastian (ES/V013610/1).

Abstract

Frequency and expertise in music making has been related to executive function ability, although the literature has reported mixed findings. This study took a latent variable approach to testing whether the frequency of music making across the lifespan is related to each of the three factors of executive functions (inhibition, shifting, and working memory) and whether there is a difference between people who make music with high and low frequency. Adults (n = 405) ranging in age from 18 to 85 years completed a leisure activity questionnaire and three tasks each measuring inhibition, shifting, and working memory. Age and years of education were also analysed as covariates. While age was significantly related to executive functions, frequency of music making was not significantly related to any executive function factor. Comparing participants grouped into high frequency music makers, who indicated making music at least once a week, and low frequency music makers, who reported making music once a month or less often, yielded different results, with significant differences between groups in shifting, inhibition, and age. Therefore, this study suggests that the frequency of music-making per se is not related to executive functions, indicating that previously reported positive associations may have been driven by musical skill rather than the frequency of making music.

Keywords: Frequency of music making, executive functions, lifespan

3.1 Introduction

Executive functions refer to higher-order cognitive skills which underpin goal-directed behaviour (Miyake et al., 2000). Three key factors of executive functions have been identified: inhibition, shifting, and updating (Friedman & Miyake, 2017; Miyake et al., 2000; Miyake & Friedman, 2012). Inhibition is a deliberate, intended, and controlled suppression of pre-potent responses that overrides dominant responses. Shifting is the ability to switch flexibly between mental sets in order to engage in the currently relevant task and disengage from a previous, no longer relevant task. Updating requires actively replacing irrelevant information in working memory with new information. Working memory refers to the ability to maintain and manipulate actively ongoing information (Cowan, 2017). Executive functions have been linked to academic and career success, physical and mental health, marital satisfaction, quality of life, decision making, planning, and problem solving (Diamond, 2013; Ferguson et al., 2021).

Given the importance of executive functions to broader functioning, much research has focused on identifying possible factors that might improve executive functions over the lifespan. Cognitive training, while initially thought to hold promise, emerged with more robust studies as not effective in improving broader un-trained skills (for meta-analysis see Sala et al., 2019). Therefore, the spotlight has been placed on identifying modifiable life factors that might shape executive functions. One such factor has been music making – a term which encompasses individual differences in musical engagement in making music. Executive functions have been hypothesised to be related with music making, but previous studies reported mixed findings. Sala and Gobet (2017a; 2017b; 2020) even questioned whether music making is related to cognitive skills at all. Yet, several studies taking both a correlational approach or a group-based comparison approach (musicians vs non-musicians) have demonstrated a relationship of music making to executive functions (Aryanto et al., in prep.; Hernández et al., 2020; Román-Caballero et al., 2018, 2022; Talamini et al., 2016, 2017). Previous meta-analyses found that musical training may benefit children's and adolescents' executive functions in general (Román-Caballero et al., 2022) and older adults' inhibition (Román-Caballero et al., 2018). Other meta-analyses suggested that musicians outperformed non-musicians in working memory tasks (Talamini et al., 2017) and executive functions tasks when all three factors were combined into a single score (Hernández et al., 2020). Our meta-analysis combining correlational and comparison studies found that music-making is related to all three factors of executive functions, with the largest average association with working memory (Aryanto et al., in prep.). However, previous studies had several limitations.

First, some empirical studies in music making and executive functions assessed executive functions with only a single measure (e.g. Criscuolo et al., 2019; Zuk et al., 2014) Using a single measure to assess executive functions gives rise to the task-impurity problem. Specifically, performance in any single task includes systematic variance of measuring lower-level cognitive processes in addition to executive functions. Not accounting for this measurement error lowers the reliability of the executive functions measure (Miyake et al., 2000; Miyake & Friedman, 2012). Latent-variable modelling addresses this issue by triangulating the shared variance between at least three tasks, with the non-shared variance being explicitly modelled as task-specific measurement error (Friedman & Miyake, 2017). Based on the unity and diversity model of executive functions (Friedman & Miyake, 2017; Miyake & Friedman, 2012), it was suggested that while executive functions share common underlying mechanisms (unity), they also exhibit unique, component-specific characteristics (diversity). Therefore, in this study, two models were tested to investigate which can explain the relationship between music making and executive functions better: a three-factor model (separated inhibition, shifting, and working memory facets) or a bi-factor model (separated common executive functions factors with shifting-specific and updating-specific functions).

Second, the relationship between music making and executive functions may be confounded by demographic factors. One potential confounding variable is age, as even within adults who have well developed skills, executive functions are known to decline in older adulthood (Ferguson et al., 2021; Loaiza, 2024). Another possible factor that may confound the correlation is education, as musicians and non-musicians may come from different socio-economic and educational backgrounds, and education and socioeconomic status are related to higher cognitive abilities (Wang, 2022). A handful of studies have attempted to control for these potential confounds, finding that age and education did not account for the relationship between music making and cognitive abilities (Arndt et al., 2023; Criscuolo et al., 2019; D'Souza et al., 2018; Okada & Slevc, 2018; Slevc et al., 2016). However, the characteristics of the participants from those studies were mostly undergraduate students or young and middle-aged adults. Previous studies with participants across the adult life span (i.e. between 18 and 84 years old) found weaker correlations between musical skills and cognitive abilities after accounting for age and education (Correia et al., 2023; Okely et al., 2022; Vincenzi et al., 2022). Therefore, conducting a study across the adult lifespan and in a sample beyond just university students would be beneficial to elucidate the role that age and education may play in the relation between music making and executive functions.

Third, previous studies operationalise music making by the duration of musical training or music lessons, but not the frequency of making music in daily life. Platz et al. (2014) argued that research in music making should distinguish between merely making music and deliberately practising music to acquire a music skill. Most of the studies then operationalise people's music skill by comparing musically trained (musicians) and non-musically trained (non-musicians) and found that musicians had higher inhibition and working memory performance compared to non-musicians (Criscuolo et al., 2019; Hansen et al., 2013), though another study found non-significant difference (Meyer et al., 2020). Yet, all those studies had different criteria for identifying who were considered as musicians and non-musicians, which were arbitrary and with no consensus between studies yet. Another way studies have operationalised music making is by using a continuous measure such as years of musical training and/or duration of musical practice (e.g., Okada & Slevc, 2018; Slevc et al., 2016). However, none of the previous studies measure the frequency of music making in daily life. A further investigation is needed to answer whether everyday music making may be related to executive functions regardless of the music maker's musicianship or musical background. By investigating the relationship between frequency of music making and executive functions, it may inform music intervention studies to further understand whether engaging in music in daily life may offer a way to increase executive functions.

In summary, previous studies showed that people who engage in music making may have better inhibition, shifting, and working memory skills compared to non-musicians or people who never make music. However, many studies made the comparison between extreme groups of participants such as 'musicians' and 'non-musicians' instead of examining if the relation holds with frequency of music making. Conducting a study across the adult lifespan and measuring the frequency of music making in daily life rather than the duration of training could provide valuable insights into the relationship between music making and executive functions. Specifically, it will help us to better theoretically understand any relation to determine if executive functions can only be shaped by deliberate or skilled music practice, or whether they are amenable to general music engagement. From an applied perspective, the research will also help inform interventions aimed at enhancing executive functions through music-related activities, regardless of the individual's level of musical expertise or background. To move beyond prior research, we took a latent-variable approach and examined the relationship between the frequency of music making and executive functions across the lifespan, accounting for age and years of education. Furthermore, this study will clarify whether different results are found when both correlational and comparative analysis are done in the same sample.

3.2. Method

Data were drawn from a larger multi-site project examining the efficacy of cognitive training (von Bastian et al., 2022). Here, the baseline data collected at the Canada, Germany, and UK sites was analysed.

3.2.1. Participants

Participants (n = 405) were adults (228 women, 176 men, 1 prefer not to say) between 18 and 85 years old (M_{age} = 48.34, SD_{age} = 18.24). Participants were reimbursed with £25 or course credits after completing the pre-test assessments. Stratified sampling was employed to ensure an even age distribution across the adult lifespan. On average, participants had 16.72 years of education (SD = 3.72).

3.2.2. Measures

In the original study (von Bastian et al., 2022), participants' cognitive abilities were assessed with 22 computer-based tasks before, after, and three months after a cognitive training intervention. Before participants came to the lab, participants completed questionnaires regarding their health, personality, motivation, and leisure activities. In this study, we analysed baseline performance only from tasks measuring executive functions (i.e. inhibition, shifting, and working memory/updating) and used frequency of music making in the leisure activity questionnaire as the predictor variable. Each facet of executive function was assessed by three tasks, resulting in a total of nine tasks analysed in the present study.

3.2.2.1. Frequency of Music Making

Frequency of music making was assessed using a single question in the leisure activity questionnaire asking participants “Please indicate on average, how often you have played music (Examples include: Playing a musical instrument, being in a choir, playing in an orchestra or band) in the last 12 months”. Participants indicated their answer on a 6-point Likert scale: (1) Once a month or less, (2) 2-3 times a month, (3) 1-2 times per week, (4) 2-3 time per week, (5) 5-6 times per week, and (6) every day.

3.2.2.2. Inhibition

Inhibition was assessed by Go/No-Go, Number Stroop, and Simon task. Each of these tasks required the ability to suppress prepotent responses and filter distracting information. In the Go/No-Go task, participants were instructed to press space when a square appeared (“go” trial) and to withhold the key press when a diamond appeared (“no-go” trial). Following a 250 ms display of a fixation cross, stimuli are displayed in the middle of the screen for a maximum of 2000 ms, or until a response is given. The outcome variable for this task was calculated using the parameter *d’*, which was the difference between *z*-transformed hit rates to go trials and *z*-transformed false alarm rates to no-go trials. In the Number Stroop task, participants were asked to indicate the number of digits shown while inhibiting the predominant response to instead respond with digits presented. In the Simon task, participants were asked to indicate the colour of a green or red circle which is presented on the left or right of the screen by pressing a left response key for green circles, and a right response key for red circles. The reaction time cost (RT cost) of the Number Stroop task and the Simon task was computed by calculating the difference in log-transformed RTs between incongruent and congruent trials. Only RTs to correct responses were used, and RT was trimmed by excluding any RTs more than 3 median absolute deviations away from the overall median (determined per participant and condition).

3.2.2.3. Shifting

Shifting was assessed by two choice reaction time task sets which asked participant to switch between *animacy-size* (drawing), *colour-shape* (shape), and *magnitude-parity* (number). Using a task-cueing paradigm, the currently relevant task set is indicated by a visual cue that is presented for 150 ms before stimulus onset and until a response is given. Participants needed to react as quickly as possible to classify the stimuli according to varying rules across trials. In the animacy-size task, participants switched as either animate or inanimate, or smaller or larger than a football. In the colour-shape task, participants switched between green or blue colour, or curvy or angular shape, whereas in the magnitude-parity task, participants needed to classify as either odd or even, or less than or greater than 5. Switching costs served as outcome variable and were computed by calculating the difference between log-transformed switch trials RT and log-transformed repetition trials RT. Only RTs to correct responses were be used, and RT was trimmed by excluding any RTs more than 3 median absolute deviations away from the overall median (determined per participant and condition).

3.2.2.4. Working Memory/Updating

Working memory was assessed with digit keep-track, binding, and continuous reproduction tasks. In the digit keep-track task, participants were asked to memorise the most recent number that appears in each of 3 boxes. During the memory phase, the initial set of single digits presented simultaneously for 3750 ms in three boxes on the screen, followed by a 250 ms blank interval. In the following updating phase, these digits are substituted by new digits and displayed for 1250 ms each, with a 250 ms blank interval between substitutions. Then, participants reported the most recent number of the probed box. Accuracy served as the outcome variable in this task. In the binding task participants memorised a set of sequentially presented associations between coloured triangles and their locations in a 4 x 4 grid, in which

3 to 5 triangles are presented sequentially for 900 ms, followed by a 100 ms blank interval Then participants were asked to recognise whether triangles were shown at the respective locations. For the binding task, *d*' was computed by calculating the difference between *z*-transformed hit rates to match probes and *z*-transformed false alarm rates to intrusion probes. Continuous reproduction task asked participant to memorise a set of 5 oriented triangles presented simultaneously and spaced equally in a circle on the screen for 1200 ms, and then re-orient one of the triangles (cued by location) with the mouse after a 900 ms retention interval. Recall error was calculated as outcome variable by finding the difference between the angle by the participant and target's angle, then corrected it to the smallest difference between angle to have the range between 0 and 180 degrees.

3.2.3. Analysis

Data pre-processing and analysis were performed in R using RStudio (version 4.2.0). We pre-processed all data similar to the procedure described by von Bastian et al. (2022) and all variables were z-standardised during the analysis. We used the lavaan package to conduct a confirmatory factor analysis (CFA), multigroup confirmatory factor analysis (MGCFA), and structural-equation modelling (SEM; Rosseel, 2012, version 0.6-16). To evaluate the model's fit, we used several indices, including the $\chi 2$ statistic of the model, comparative fit index (CFI), root mean squared error of approximation (RMSEA) and its 90% confidence interval, and the standardised root mean squared residual (SRMR). For the $\chi 2$ of the model, a *p*-value that is not significant is indicative of a good fit. However, this measure is affected by sample size, necessitating the use of other indices to evaluate model fit accurately. A CFI higher than 0.90 is deemed acceptable, while above 0.95 is considered as a strong fit. An RMSEA and SRMR value below 0.06 suggests the fit is satisfactory (Hu & Bentler, 1999).

After analysing the fit of the model based on the CFA, we conducted SEM with frequency of music making as the predictor variable, each executive functions factors as the outcome variable, and age and years of education as covariates. Lastly, we split the sample into high and low frequency music makers based on the answer in the leisure activity question. Participants who answered 0 (i.e. once a month or less of music making) were considered as low frequency music makers and participants who answered 2 or above (i.e. at least one week of music making) were considered as high frequency music makers. We then conducted independent sample t-tests comparing low and high-frequency music makers, using composite scores for each factor of executive functions.

3.3. Results

Descriptive statistics for all measures are listed in Table 3.1. We performed correlational analysis before conducting CFA to analyse the model fit and SEM to confirm whether the frequency of music making predicted each factor of executive functions. The correlation matrix between demographics, frequency of music making, and executive functions tasks is presented in Table 3.2. The results showed that age, but not education, was significantly related to all executive functions tasks. Still, both age and years of education were included as covariates in the following analysis. Tasks measuring the same executive functions factor were significantly correlated. The frequency of music making was negatively related to performance in two of the inhibition tasks (Stroop and go/no-go tasks) and one of the shifting tasks (Drawing shifting task). Furthermore, the frequency of music making had significant relationship with one of the working memory updating tasks (continuous reproduction task) but not to any other tasks.

Table 3.1

Descriptive statistics

Measure	M	SD	Range	Skewness	Kurtosis	Reliability
Frequency of music making	0.96	1.66	0 – 5	1.53	0.85	-
Inhibition						
Simon (RT cost)	-0.17	0.09	-0.89 – 0.02	-1.90	11.47	0.88
Number Stroop (RT cost)	-0.16	0.07	-0.48 – 0.11	-0.25	1.49	0.83
Go/No-Go (*d*')	4.22	0.56	1.88 – 5.29	-0.45	1.02	0.97
Shifting						
Drawing (RT cost)	-0.32	0.15	-0.85 – 0.15	0.29	1.04	0.93
Shape (RT cost)	-0.30	0.15	-0.78 – 0.13	0.03	0.37	0.94
Number (RT cost)	-0.31	0.14	-0.78 – 0.06	-0.23	0.35	0.91
Working Memory						
Digit keep-track	0.94	0.12	0.06 – 1.00	-3.49	14.17	0.99
Binding (*d*')	1.53	0.95	-1.94 – 4.62	0.11	0.45	0.99
Continuous reproduction (recall error)	-60.09	15.73	-99.25 – -20.15	0.22	-0.51	0.88

Notes. All scores were coded so that higher value means better performance. RT = reaction time. All reliabilities were calculated using split-half reliability corrected with Spearman-Brown prophecy formula.

Table 3.2

Correlation matrix between demographics, frequency of music making, and executive functions

Variable	1	2	3	4	5	6	7	8	9	10	11
Demographic and Music Background											
1. Age											
2. Education	-.03										
3. Music making	**-.16**	-.08									
Inhibition											
4. Simon[1]	**.15**	-.01	-.01								
5. Stroop[1]	**.31**	-.05	**-.11**	**.32**							
6. Go/No-Go	**.29**	-.03	**-.11**	**.27**	**.32**						
Shifting											
7. Drawing shifting[1]	**.28**	-.03	**-.10**	.09	**.17**	**.13**					
8. Shape shifting[1]	**.40**	-.03	-.08	**.12**	**.14**	**.12**	**.59**				
9. Number shifting[1]	.07	-.03	-.03	**.12**	.09	.09	**.45**	**.50**			
Working memory											
10. Continuous reproduction[1]	**-.52**	**.10**	**.13**	**.12**	**-.11**	.04	**-.19**	**-.28**	**-.10**		
11. Binding task	**-.51**	.06	.04	**.11**	**-.12**	.02	**-.18**	**-.29**	-.03	**.65**	
12. Updating task	**-.20**	**.14**	.03	**.32**	.02	**.12**	**-.28**	**-.37**	**-.13***	**.36**	**.40**

Note. Significant correlations are displayed in bold ($p < .05$); Age and education are given in years; [1]The value for these tasks were reversed, so positive values mean better performance.

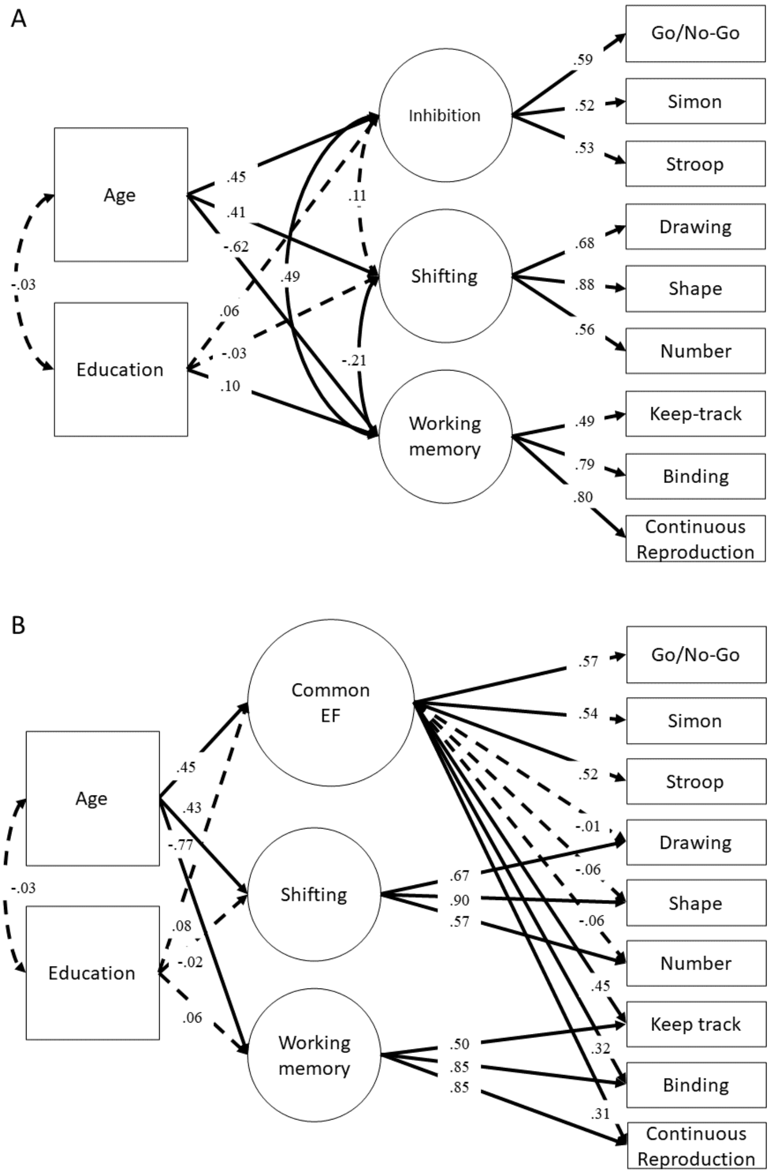
3.3.1. Measurement model

CFA was performed to determine the shared variance within each factor of executive functions, and then the three-factor model and the bi-factor model were tested. The model estimation reported in this section included the covariates and the models without the covariates are in Appendix C. Figure 3.1 illustrates these models. The three-factor model had moderate fit, $\chi^2(36) = 166.46$, $p < .001$, CFI = .88, RMSEA = .09 [.08, .11], SRMR = .07. Similarly, the bi-factor model also had moderate fit, $\chi^2(33) = 160.50$, $p < .001$, CFI = .89, RMSEA = .10 [.08, .11], SRMR = .07. The chi-square difference test between the three-factor and the bi-factor models was non-significant, $\Delta\chi^2(3) = 5.96$, $p = 0.113$. Therefore, the more restricted model was preferable and we continued the SEM analysis with the three-factor model.

The standardised factor loadings and variances for all latent factors of the three-factor model were all significant with all *p*s < .001. Inhibition had a positive significant relationship to working memory, $r = .49$, $p < .001$, but not shifting, $r = .11$, $p = .177$. A different direction was found between shifting and working memory, with a significant negative relationship, $r = -.21$, $p = .003$. In the present study, we reversed the switch cost score such that the higher value means better performance. The same pattern was also found in Draheim et al.'s (2016) study in which working memory was related negatively to number and category switch costs. Furthermore, after accounting for the age covariate, the results showed that age significantly increased inhibition ($\beta = .45$, $p < .001$) and shifting ($\beta = .41$, $p < .001$), and significantly decreased working memory ($\beta = -.62$, $p < .001$). Therefore, in the study across the adult lifespan, older participants had better inhibition and shifting abilities but lower working memory ability.

Figure 3.1

CFA of Three-factor Model (A) and Bi-factor Model (B)



3.3.2. Group Differences between High and Low Frequency Music Makers

We conducted a MGCFA to determine whether high frequency (*n* = 101) and low frequency music makers (*n* = 278) have equal measures and structural models. MGCFA was done to test whether scores from high frequency and low frequency music makers were comparable; thus, have the same meaning or interpretation (Kline, 2023). Two CFAs were conducted and showed moderate fit for high frequency music makers, $\chi^2(36) = 58.59$, $p = .010$, CFI = .89, RMSEA = .08 [.04, .11], SRMR = .08 and for low frequency music makers,

$\chi^2(36) = 132.81$, p = <.001, CFI = .89, RMSEA = .10 [.08, .12], SRMR = .07. Next, we tested for measurement invariance with configural (freely estimated), metric (constrained the factor loadings), scalar (constrained the factor loadings and intercept) and strict (constrained the factor loadings, intercept, and covariances) models as shown in Table 3.3 for the fit indices. There was no significant difference between the metric and scalar models, but the strict model showed significant different between high and low frequency music makers. Therefore, the model between high and low frequency music makers was comparable on the level of scalar measurement invariance, but not strict measurement invariance. This finding indicates that the results between high and low frequency music makers had unequal covariances.

Table 3.3

Fit coefficients of multigroup confirmatory factor analysis between musicians and non-musicians

Model	CFI	RMSEA [90% CI]	SRMR	χ^2	df	*p*-value
Configural	.883	.09 [.08, .11]	.08	202.18	78	
Metric	.883	.09 [.07, .10]	.08	208.22	84	.419
Scalar	.881	.09 [.07, .10]	.08	216.59	90	.212
Strict	.877	.09 [.07, .10]	.09	224.42	93	.049

Before we conducted an independent t-test to compare high and low-frequency music makers' composite scores of each executive function factor, we first compared the groups for each task separately. High frequency music makers were better at one working memory task, but had lower score at two inhibition and two shifting tasks, with all other comparisons being non-significant (see Appendix C for detailed statistical results). Next, we tested for group differences for each composite executive functions factor composite score. As can be seen in Figure 3.2, there was a significant difference between high and low frequency music makers in inhibition, $t(377) = -2.28$, $p = .023$, and shifting, $t(377) = -2.76$, $p = .006$, but not working memory, $t(377) = 1.36$, $p = .174$. In order to confirm that the difference was due to the difference in the frequency of music making instead of demographic factors, we compared the age and years of education of both group. Figure 3.3 showed that high frequency music makers were significantly younger than low frequency music makers, $t(377) = -2.69$, $p = .007$, yet no difference in years of education was found, $t(376) = -1.21$, $p = .227$.

Figure 3.2

The Difference in Composite Scores between High and Low Frequency Music Makers

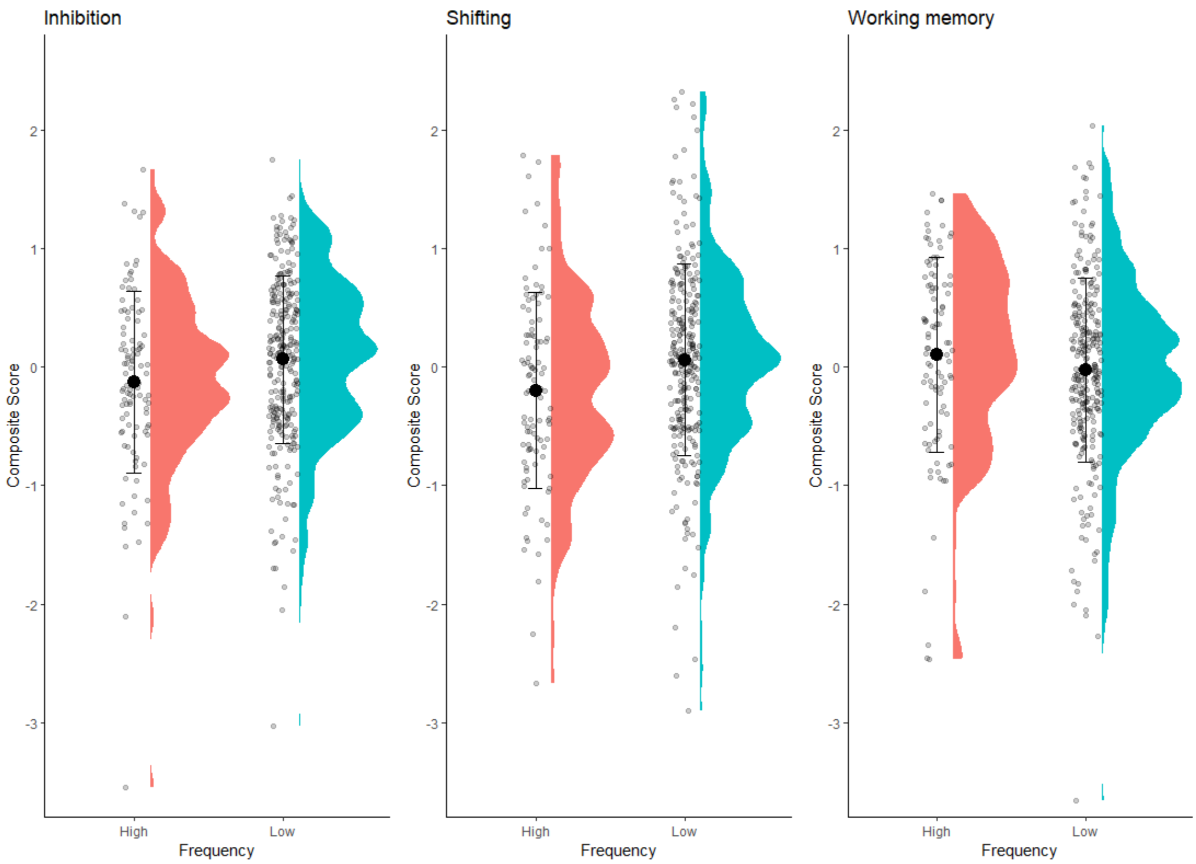
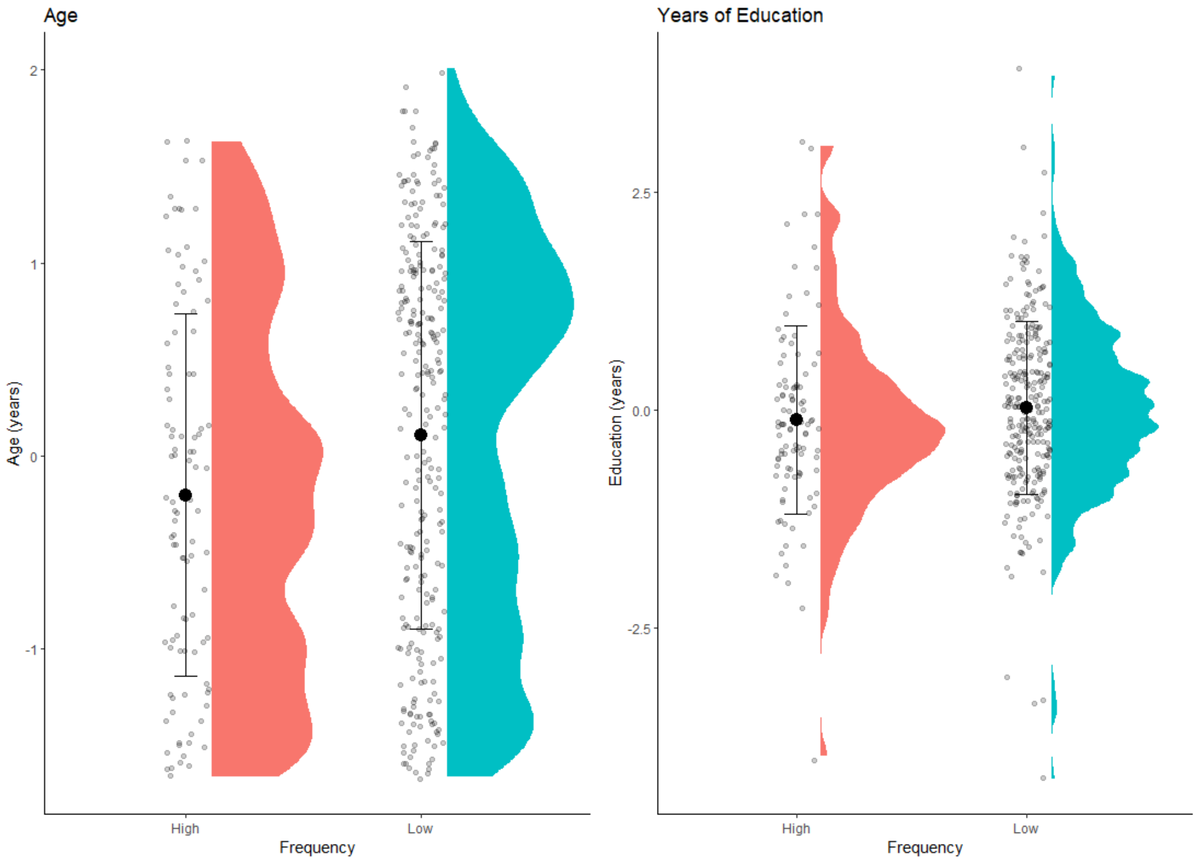


Figure 3.3

The difference in age and years of education between high and low frequency music makers



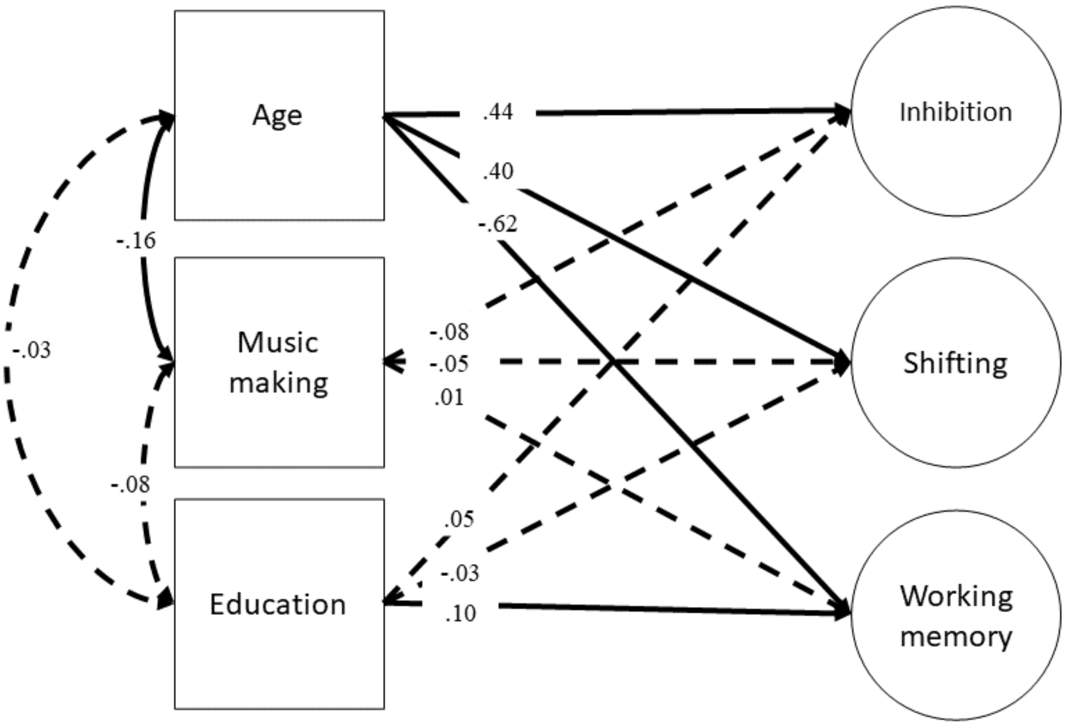
3.2.3. Relationship between Frequency of Music Making and Executive Functions

Finally, we ran an SEM using the three-factor model with the frequency of music making as a continuous predictor variable, each factor of executive functions as outcome, and age and years of education as covariates. For the sake of completeness, Appendix C showed the results without covariates and showed that frequency of music making was negatively related to inhibition, but not shifiting and working memory. Figure 4 illustrates the results with the covariates. There was no significant relationship between the frequency of music making and any executive function factors (inhibition: β = -.08, 95% CI [-.12 - .03], *p* = .220, shifting: β = -.05, 95% CI [-.10 - .04], *p* = .377, and working memory: β = .01, 95% CI [-.06 - .08], *p* = .746). Age was significantly related to all executive functions factors, β = .44, 95% CI [.18 - .35] for inhibition, β = .40, 95% CI [.20 - .35] for shifting, β = -.62, 95% CI [-.57 - -

.41] for working memory, *ps* < .001, and frequency of music making, *r* = -16, 95%, CI [-.26 - -.06], *p* = .002. Years of education, on the other hand, was not significantly related to inhibition, β = .05, 95% CI [-.04 - .10], shifting, β = -.03, 95% CI [-.09 - .05], and frequency of music making, *r* = -.08, 95% CI [-.18 - .02] (all *ps* non-significant), but significantly related to working memory, β = .10, 95% CI [.01 - .15], *p* = .035.

Figure 3.4

Schematic summary of the results of the structural equation modelling



Note. Significant relationships are indicated by solid arrows, non-significant relationship by dashed arrow.

3.4. Discussion

This study aimed to take a latent-variable approach to examine if frequency of music making in everyday life (as both a continuous and categorical variable) relates to executive functions across the lifespan, after accounting for age and years of education. This study revealed that the frequency of music making was not related to inhibition, shifting, or working memory. Indeed, only age was significantly related to executive functions. However, the different pattern of results was found when comparing high and low frequency music makers using composite scores of inhibition and shifting but not working memory. This pattern occurred because there was a significant age difference between high and low frequency music makers which was confirmed by the results of the SEM.

At first glance, the results of this study seem to contradict previous meta-analysis, in which making music was related to all executive function factors, particularly working memory (Aryanto et al., 2024; Talamini et al., 2017). Critically, however, the present study focused on the frequency of music making, whereas many of the previous studies included in the meta-analysis used years of formal music training, duration of music practice, and musical skills as predictors. This study supported the argument of the previous study that making music did not necessarily mean deliberate music practice which in turn may enhance the musical skill (Platz et al., 2014). This suggests that people who play music everyday do not necessarily have higher musical skill. Based on this notion, there was a possibility that music training or music skills may be the variable that is related to executive functions instead of merely making music in daily life. Therefore, this study highlights the crucial distinction between frequency of music making and formal music training/skills as predictors of executive functions. This difference would be informative for any future music interventions, specifically, suggesting that focusing only on frequency alone would be unlikely to benefit executive functions. It suggests that musical skill and formal training may be the critical factor that shapes executive functions.

The results of the correlational analysis showed that the frequency of music making was related significantly to performance in several executive functions measures, and the group comparisons of high and low frequency music makers showed similar patterns. However, the results of SEM found no relationship with any executive function factors when accounting for the age and years of education, and the composite scores comparing high and low frequency music makers showed significant differences in inhibition and shifting due to the age difference. Thus, the significant results may be due to task-specific measurement error and/or multiple comparisons. This study specifically used a latent-variable approach to address the task impurity problem (Friedman & Miyake, 2017; Miyake et al., 2000; Miyake & Friedman, 2012). Therefore, we deem the latent variable model's results more reliable than those from single executive functions measures.

We found an entirely unexpected negative relationship between working memory and shifting, which was the opposite findings from the original Miyake et al.'s (2000) study that found these skills were positively related. However, recent study found the similar pattern in which working memory and shifting showed negative relationship (Draheim et al., 2016; Löffler et al., 2024). This results may be explained because the outcome variable of working memory was based on the accuracy (i.e. *d'* for binding task and recall error for continuous reproduction task), whereas shifting was measured by the reaction time (i.e. switching cost). Draheim et al. (2016) found a negative relationship was found between working memory capacity and shifting when switching cost was calculated as the outcome variable. This may occur because only the reaction time in the correct response was calculated for the switching cost. One possible solution for this issue is to combine accuracy and reaction time in future studies (Draheim et al., 2016).

This study showed that age was positively correlated with inhibition and shifting, and negatively correlated with working memory. A wealth of prior research demonstrates ageing is related to working memory decline (Ferguson et al., 2021; Loaiza, 2024). However, it is interesting and unexpected that age would be positively associated with inhibition and shifting. Ferguson et al. (2021) found that inhibition began to decrease at the age of 30 and 40 years old, which contrary to what was found in this study, yet adolescents and young adults have difficulties in shifting between tasks compared to middle and older adults, and this notion was supported by this study. The outcome variable for inhibition in the present study combined the difference in reaction time (Simon and Stroop tasks) and *d'* (Go/No-Go task) which may explain the different findings compared to previous study.

Furthermore, age was found to be a significant covariate in the relation between music making and executive functions. This result is contrary to the previous study, which did not find age to be a moderator between music and working memory (Aryanto et al., in prep). This could be explained because most of the previous studies included in the meta-analysis had a narrow range of ages as the participants of past studies were either younger adults or older adults (e.g. Grassi et al., 2017; Gray & Gow, 2020; Okada & Slevc, 2018; Talamini et al., 2017). In this study, participants' age was stratified from 18 to 85 years and previous studies with participants across the lifespan also found the same pattern where age accounting for the relationship between music making and cognitive abilities (Correia et al., 2023; Okely et al., 2022; Vincenzi et al., 2022). The results of this study were also supported by another study on the individual differences in executive functions across the lifespan that found inhibition and working memory start to decline at the age of 30-40 years old (Ferguson et al., 2021). Thus, the results emphasise the importance of considering age as a possible confounding variable, especially in the study across adults life span.

There were two limitations in the current study. First, one of the working memory measures, digit keep-track task, had a ceiling effect; thus, there were limited variability in that particular task. The set size three of the task was considered easy for most of the participants of any age and resulted a highly negative skewed distribution. Future studies may consider employing an increased set size to the task or have increasing difficulty to capture more variation among the participants. Another limitation was the frequency music making measured with only a single self-report question. The lowest possible choice of the question was once a month or less, which can possibly be chosen by people who never play music, musicians who rarely play music, or former musicians. Musicians were found to have different cognitive ability compared to non-musicians (Schellenberg & Lima, 2023), but the question about frequency of music making cannot be identified if low frequency music makers were musically trained or not. While it was sufficient to measure frequency of music making using a single question, the validity and reliability of this question cannot be further inspected and a latent factor cannot be formed with only a single question.

In conclusion, this study contributes to the understanding of the relationship between the frequency of music-making and executive functions across the adult lifespan. The results of SEM showed that none of the executive function factors was related significantly to the frequency of music making. However, there was a difference in inhibition and shifting between high and low frequency makers due to the age difference between groups. This study shows that the frequency of music making may be different from musical skills, and there is a possibility that musical skill are related to executive functions. This may inform future music intervention studies to focus on practising the musical skill instead of increasing the duration of daily music making.

Chapter 4 - Does Sensory Discrimination Ability Account for the Relation Between Music Skill and Working Memory?

Interim summary:

In Chapter 2, we found that music making was particularly strongly related to working memory. In Chapter 3, we found that the frequency of music making was not related to working memory when accounting for age. This finding suggests that the association between music making and working memory observed in Chapter 2 may be driven by musical skill rather than the mere frequency of music making. Therefore, the study reported in the present chapter focused on the relationship between musical skill and working memory. Moreover, the present chapter investigates why these two abilities are related. As discussed in Chapter 2, previous studies have shown that musicians have better ability to discriminate between two auditory stimuli compared to non-musicians (Wang, 2022). Furthermore, auditory discrimination ability has been found to be related to working memory capacity (Tsukahara et al., 2020). Therefore, sensory discriminations and, in particular, auditory discrimination, may explain the relationship between musical skill and working memory. However, no study yet has related all three variables – musical skill, working memory, and sensory discriminations – simultaneously in the same sample. In the present chapter, we addressed this gap and tested whether sensory discriminations mediate the relationship between musical skill and working memory. Furthermore, we investigated whether any relationship between the three variables was specific to the auditory modality or modality general.

Contributions:

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

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

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

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

Submission status: Manuscript in preparation for publication

Research plan from this chapter have been presented in the following conferences:

Aryanto, C.B., Blakey, E., Timmers, R., & von Bastian, C.C. *Research plan - Why are active music making and executive functions related? Testing the roles of sensory discrimination and fluid intelligence* [Poster]. Experimental Psychology Society (EPS) Meeting, Online. Access to the video presentation: https://www.youtube.com/watch?v=td2ElPWcgyM, and poster presentation: https://eps.ac.uk/wp-content/uploads/2021/06/Christ-Billy-Aryanto.pdf

Findings from this chapter have been presented in the following conference:

Aryanto, C.B., Blakey, E., Timmers, R., & von Bastian, C.C. *Does sensory discrimination ability explain the relation between active music-making and working memory?* [Talk]. The 23rd European Society for Cognitive Psychology (ESCoP) Conference, Porto, Portugal.

Presentation in the following conference based on research in this chapter has received the 2023 J. Frank Yates Student Conference Award from Psychonomic Society:

Aryanto, C.B., Blakey, E., Timmers, R., & von Bastian, C.C. *Does sensory discrimination ability account for the relation between active music-making and working memory?* [Poster]. Psychonomic Society 64th Annual Meeting, San Francisco, USA.

Acknowledgement

This research was supported by a PhD scholarship 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 research grant by Atma Jaya Catholic University of Indonesia (No. 1533/II/SK/Dir.SDM-KP.100.03/07/2020) to Christ B. Aryanto. We thank Hakan Atis for his collaboration in the data collection. We are grateful to all the participants who participated in the study.

Abstract

Musical skill has been shown to be related to better working memory, but it is yet unclear why. This study tested whether this relationship is specific to visual or auditory working memory, and whether enhanced sensory discrimination accounts for this relationship. Healthy young adults (N = 263) completed a music questionnaire, musical skill tasks, and a series of auditory and visual working memory and sensory discrimination tasks. Measurement models were fitted to the data to analyse the relations between musical skill, and both modality-general and modality-specific (auditory and visual) latent factors of working memory and sensory discrimination. The results showed that musical skill was related to sensory discrimination, and sensory discrimination was related to working memory regardless of modality. Moreover, sensory discrimination accounted for the relation between musical skill and auditory and modality-general working memory, but not visual working memory. Therefore, this study indicates that sensory discrimination may indeed explain the relation between musical skill, suggesting that musical skill may be associated with better working memory through refined auditory representations.

Keywords: Musical skill, working memory, sensory discrimination

4.1. Introduction

Playing music, be it as an instrumentalist or a singer, requires a significant amount of information processing in working memory, the ability to maintain and update information in the present moment (Cowan, 2017; Oberauer et al., 2018). The terms to explain how people actively engage in music were varied in the previous studies, such as musical training, musical ability, music playing, music engagement, and musical sophistication. In this study, the term “musical skill” was used to describe individual differences associated with active engagement with music with a particular focus on perceptual skills.

Indeed, previous studies have found that working memory and musical skill are related (for reviews see Okada & Slevc, 2020; Swaminathan & Schellenberg, 2019; for meta-analysis, see Aryanto et al., in prep.; Talamini et al., 2017). The positive relationship between musical skill and working memory has been demonstrated for different modalities (i.e. visual and auditory working memory, Slevc et al., 2016), with typically larger benefits in working memory tasks with tonal than visuospatial stimuli (Talamini et al., 2017). However, some previous studies found mixed results, with these inconsistencies likely due to the way musical skill or working memory were measured. When musical skill was assessed with self-report measures (e.g. musical sophistication questionnaires, self-rated musical ability), previous studies often found no association with working memory (e.g. Kempe et al., 2012; Slevc et al., 2016), but objective measurements of musical skill (i.e. melodic and beat test) generally resulted in a positive correlation (e.g. Silas et al., 2022; Slevc et al., 2016). Furthermore, past studies generally found positive relationships to auditory and verbal working memory, but not visuospatial working memory (e.g. Arndt et al., 2023; Lad et al., 2022; Talamini et al., 2017). Overall, past studies provide mixed evidence depending on how musical skill was measured and what working memory modalities (i.e. auditory, verbal, visual) were measured.

Moreover, it is still unclear why musical skill and working memory are related. Playing music or singing puts demands on working memory as musicians need to maintain, update, retrieve, and coordinate the music in their mind to perform (Okada & Slevc, 2020). Therefore, musical skill and working memory may be directly related. Alternatively, it is also possible that musical skill and working memory are mediated by another underlying mechanism. Making music requires constant discriminating between stimuli such as pitch, rhythm, and other features of music, making sensory discrimination a prime candidate mediator. Sensory discrimination is the ability to make fine discriminations in different sensory stimuli, such as different colour hue (visual), loudness (auditory), and weight (tactile). Critically, sensory discrimination was found to be positively related to cognitive abilities (Acton & Schroeder, 2001; Deary et al., 2004), including working memory (e.g. Troche et al., 2014; Voelke et al., 2013), and to musical skill (e.g. Hyde et al., 2009; Kubaszek et al., 2021).

Past studies showed that musicians have better sensory discrimination, especially auditory discrimination (Hyde et al., 2009; Kraus & Chandrasekaran, 2010; Okada & Slevc, 2018), but sometimes also *across* modalities (auditory, visual, and temporal; Kubaszek et al., 2021). There is also evidence that better sensory discrimination skills are related to enhanced working memory performance both in adults (Troche et al., 2014; Tsukahara et al., 2020) and children (Voelke et al., 2013). Prior research, however, did not distinguish between auditory and visual modalities within working memory (Troche et al., 2014; Tsukahara et al., 2020; Voelke et al., 2013, 2014). By assessing auditory and visual modalities in *both* working memory and sensory discrimination, it is possible to determine whether the relationship between musical skill, sensory discrimination, and working memory is generalisable (modality-general) or specific to certain modalities (modality-specific) to understand the possibility of the cross-modal relationship.

Previous studies often compared musicians to non-musicians. However, similar to other extreme-groups designs (see Unsworth et al., 2015), musicians may differ from non-musicians on more dimensions than just their musicianship status, which may affect the relationship between musical skills, working memory, and sensory discrimination. Wang (2022) argued that musicians and non-musicians differ in their socio-economic status and education, thus rendering these as possible confounding variables. Previous studies found weaker correlations between musical skills and cognitive abilities after accounting for demographic characteristics such as age, gender, education, and socioeconomic status (Correia et al., 2023; Okely et al., 2022; Vincenzi et al., 2022), although some other studies did not find the same results (Arndt et al., 2023; Criscuolo et al., 2019; D'Souza et al., 2018; Okada & Slevc, 2018; Slevc et al., 2016). In the current study, we therefore measured musical skills of healthy adults regardless of their musicianship status.

Overall, musical skill has the potential to be related to sensory discrimination which in turn contributes to higher working memory performance. Previous studies had examined the effect of musical skill on working memory (e.g., Degé & Frischen, 2022; George & Coch, 2011; Marie et al., 2023), implying that musical skill may affect people's working memory. However, whether the relationship between musical skill and working memory is direct or mediated by another variable needs to be further explored. Furthermore, no study yet has assessed musical skill, working memory, and sensory discrimination simultaneously in the same sample. Moreover, there were mixed findings regarding musicians' sensory discrimination and working memory ability in different modalities and, thus, there is a need to clarify whether common cognitive processes between musical skill, sensory discrimination, and working memory are specific to the auditory modality or generalised.

The aim of this current preregistered study was to investigate how musical skill, working memory, and sensory discrimination are related in healthy adults when measured and accounted for simultaneously. Working memory and sensory discrimination were measured in auditory and visual modalities. In order to avoid the task impurity problem, where the tasks may have systematic variance measuring other cognitive processes (Miyake & Friedman, 2012), this study employed a latent-variable approach in which musical skill, and auditory and visual working memory and sensory discrimination were measured with three different tasks each. Only one previous study had employed this approach but did so only for working memory but not musical skill (Okada & Slevc, 2018); thus, this study expanded previous study by employing latent variable approach on both musical and cognitive skills.

This study has three pre-registered hypotheses: (1) If the relationship between musical skill and working memory is modality-general, modality-general sensory discrimination will mediate the positive relationship between musical skill and modality-general working memory. (2) If the relationship between musical skill and working memory is modality-specific, auditory sensory discrimination will mediate the positive relationship between musical skill and auditory working memory. (3) Furthermore, visual sensory discrimination will mediate the positive relationship between musical skill and visual working memory.

4.2. Method

This study and its hypotheses were pre-registered on the Open Science Framework (https://osf.io/bdj9g) on 5th October 2022. All computerised tasks were administered on Bang & Olufsen computers and using Audio Technica ATH-AVC200 headphones. This study received ethical approval from the University of Sheffield Research Ethics Committee.

4.2.1. Participants

The target sample was 250 participants, based on Schönbrodt and Perugini's (2013) recommendations by assuming the true correlation (ρ) between .1 and .2 with the confidence level of 80% and the corridor of stability width of .1. Participants were included if they were

(1) between 18 – 35 years old, (2) fluent in English, (3) had normal hearing and no hearing impairment (e.g. tinnitus), (4) no current diagnosis of neurological, psychological, or psychiatric illness, and (5) were not currently using recreational drugs (e.g. cannabis, cocaine, or methamphetamine). Of the 267 participants recruited, the data from four had to be excluded from the analysis due to technical issues. Table 4.1 presents demographic information of the remaining 263 participants. Interestingly, from 228 participants who perceived themselves as a non-musician, 44 participants (19.3%) made music according to the music activity question, showing that people who did not perceive themselves as non-musicians may still make music.

Table 4.1

Participant Demographics

Demographics	Total	Range
Age (years)	23.59 (5.29)	18 - 35
Gender (f/m/n)	171/91/1	
Years of education	15.28 (3.34)	1 – 26
Socioeconomic status (1-10 scale)	6.25 (1.37)	3 – 10
Musicianship (musician/non-musician)	35/228	
Music activities (1-6 scale)	1.63 (1.24)	1 - 6

Note. SDs are given in parentheses where applicable. The higher value in socioeconomic status showed the higher socioeconomic status relative to others. The higher value in music activities showed the more frequency of music making.

4.2.2. Measures

The measures are described below. Working memory and sensory discrimination were assessed in both visual and auditory modalities.

4.2.2.1. Demographic questionnaire

The demographic questionnaire was adapted from von Bastian et al. (2022) and assessed age, gender, ethnicity, education, occupation, handedness, socioeconomic status, music activity, and musicianship. Socioeconomic status was measured using MacArthur

Scale of Subjective Social Status (Adler et al., 2000) in which participants indicated where they believed they stood relative to others in their home country in terms of money, education, and employment status by selecting a number from 1 (low) to 10 (high) on a ten-rung ladder. Music activity was measured with a single item on frequency of making music which was the same measure in Chapter 3, and the question was “Please indicate on average, how often you have played music (Examples include: Playing a musical instrument, being in a choir, playing in an orchestra or band) in the last 12 months”. Participants responded on a 6-point Likert scale: (1) Once a month or less, (2) 2-3 times a month, (3) 1-2 times per week, (4) 2-3 time per week, (5) 5-6 times per week, and (6) every day. The musicianship question asked whether participants perceived themselves as a musician or non-musician.

4.2.2.2. Musical skill measures

Musical skill was assessed using a self-report measure to evaluate active engagement with music, which was complemented with two listening tasks to measure musical skill objectively. Previous studies have found a moderate relationship between self-reported musical sophistication, melodic memory, and beat perception (Correia et al., 2022, 2023; Müllensiefen et al., 2014; Slevc et al., 2016). Therefore, we utilised these three measures to assess individual differences in their musical skill.

Goldsmith Musical Sophistication Index (Gold-MSI) (Müllensiefen et al., 2014). This questionnaire uses a 7-point Likert scale with 39 items and provides scores for five subscales (i.e. active engagement, musical training, perceptual abilities, singing abilities, and emotions) and a general musical sophistication by totalling the scores for questions for each subscale. For this study, the general musical sophistication score, which was the sum score of 18 items taken from the five subscales (for scoring, see Müllensiefen et al., 2014), was used as the outcome variable.

Melodic Discrimination Testing (MDT) (Harrison et al., 2017). This is an adaptive test to measure the skill to discriminate melodies with 3-alternate forced choice paradigm. In each trial, participants are presented three variations of an unknown melody that is transposed one semitone higher in pitch with each iteration. One of these versions is the "odd one out" because a note has been changed in it. The participant's task is to identify which melody was different. The outcome variable was based on item-response theory and ranged between -4 to +4.

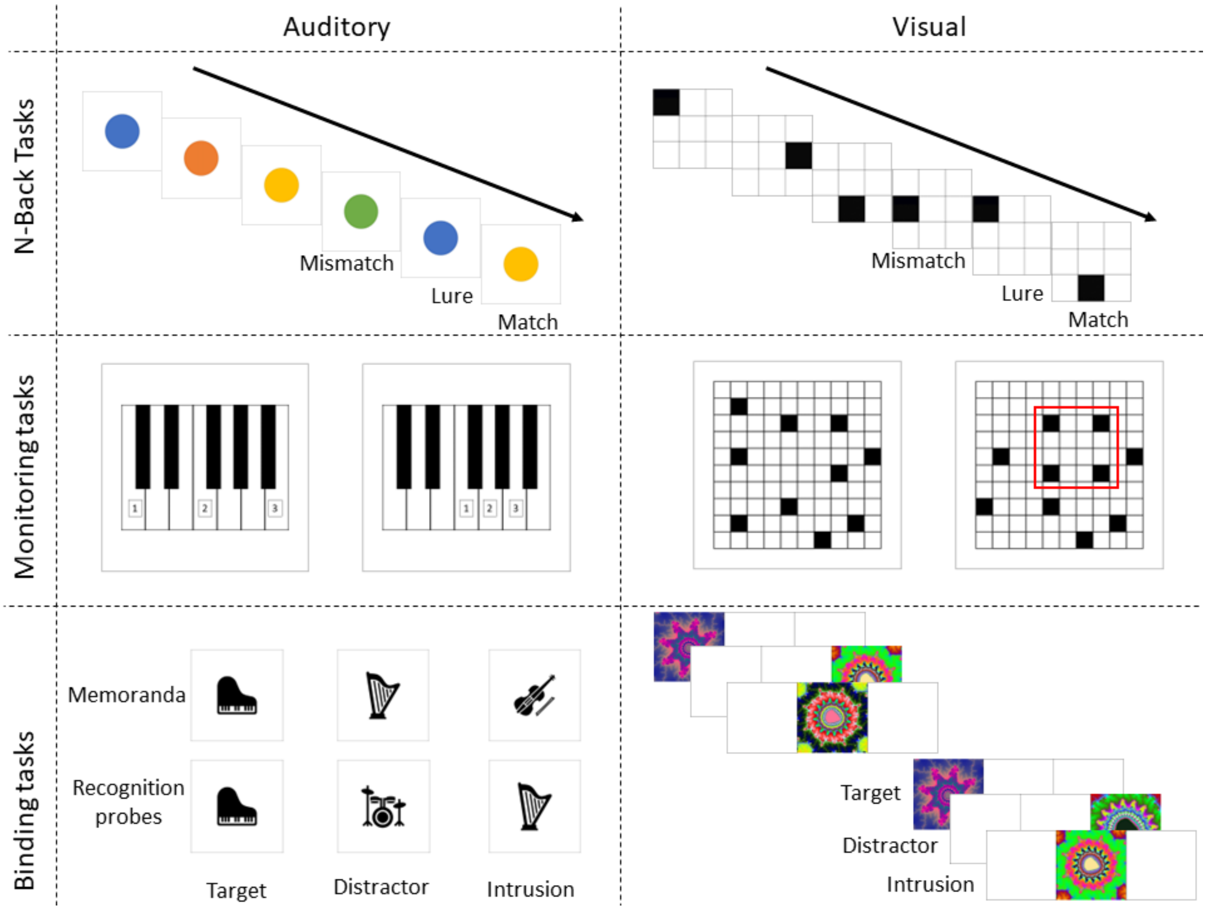
Computerised Adaptive Beat Alignment Test (CABAT) (Harrison & Müllensiefen, 2018). This is an adaptive test to measure the skill to perceive beats. Each trial presents the participant two different musical tracks that are both layered with a beeping sound that resembles a metronome. One version of the target has a beep track that precisely matches the places of the musical beats. The other beep track's version is always behind or ahead from the true musical beat placements, called the lure. The participant's task is to determine which extract is the target. Similar to MDT, the score based on item-response theory ranged between -4 to +4 was used as the outcome variable.

4.2.2.3. Working memory measures

Figure 4.1 illustrated the working memory stimuli in the current study.

Figure 4.1

Illustration of the Stimuli for Auditory and Visual Working Memory



Notes. The example for n-back task used the 3-back paradigm.

N-Back. For the auditory n-back task (adapted from Rinne et al., 2009), participants were instructed to press the left arrow key when the test tone was exactly the same tone as the one presented three (3-back) or four (4-back) trials before, or otherwise the right arrow key. In the visual n-back task (adapted from Schmiedek et al., 2009), participants were asked to press the left arrow key when the highlighted square was exactly the same as the one presented three or four trials before, and otherwise the right arrow key. In the experiment, participants always did the 3-back task first and then 4-back. The stimuli of both auditory and visual n-back tasks were presented for 500 ms each (with an interstimulus interval of 3000 ms). Each task included 141 trials. There were three types of probes in this task: match trial (presented at exactly 3- or 4-back), mismatch trial (entirely new stimulus), and lure trial (stimulus presented previously but not exactly 3- or 4-back). The outcome variable for this task was *d*', which was the difference between *z*-transformed hit rates (based on match trials) and *z*-transformed false alarm rates (based on lure trials).

Monitoring. For the auditory monitoring task (Schwärzler, 2015), participants were instructed to press the spacebar as soon as they heard three tones in a row that exactly differed a whole tone from each other (e.g. notes C, D, and E). The individual tones were presented for 500 ms each (with an interstimulus interval of 2000 ms). For the visual monitoring task (adapted from Oberauer et al., 2003), participants were asked to watch independently changing dots presented in 10 x 10 matrix on the screen and press the space key whenever the dots formed a square. The dots were changing every 2000 ms. Each task included 85 trials. The outcome variable for this task was *d*', which was the difference between *z*-transformed hit rates to target trials (i.e. press spacebar when three tones were heard or dots formed a square) and *z*-transformed false alarm rates to intrusion trials (i.e. press spacebar when three tones were not heard or dots were not formed a square).

Binding. Binding tasks for both modalities were adapted from Schwärzler (2015). In the auditory binding task, participants were asked to memorise a sequence of musical instrument sounds presented one after the other. After memorizing the sequence, they listened to another set of musical instrument sounds which may be similar or different from the first one. Participants were then asked to press left arrow key when the same music instruments have been played before at the same order and press right arrow key when the music instruments have not been played before at the same order. In the visual binding task, participants memorised a set of sequentially presented associations between fractal pictures and their locations in a grid. Participants were then asked to press the left arrow key when the pictures were shown at the same locations and right arrow key when the picture was not shown at the same locations. The individual music instruments and individual pictures were presented for 900 ms each (with an interstimulus interval of 100 ms). Participants completed 16 trials each for set sizes 3-5 (i.e. the number of stimuli heard or shown in the screen during memoranda phase). There were three types of probes during recognition probe phase: target trial (probes presented at exactly the same order or place), distraction trial (probes were never presented), and intrusion trial (probes presented in memoranda, but at different order or place). The outcome variable for this task was *d*', which was the difference between *z*-transformed hit rates (based on target trials) and *z*-transformed false alarm rates (based on intrusion trials).

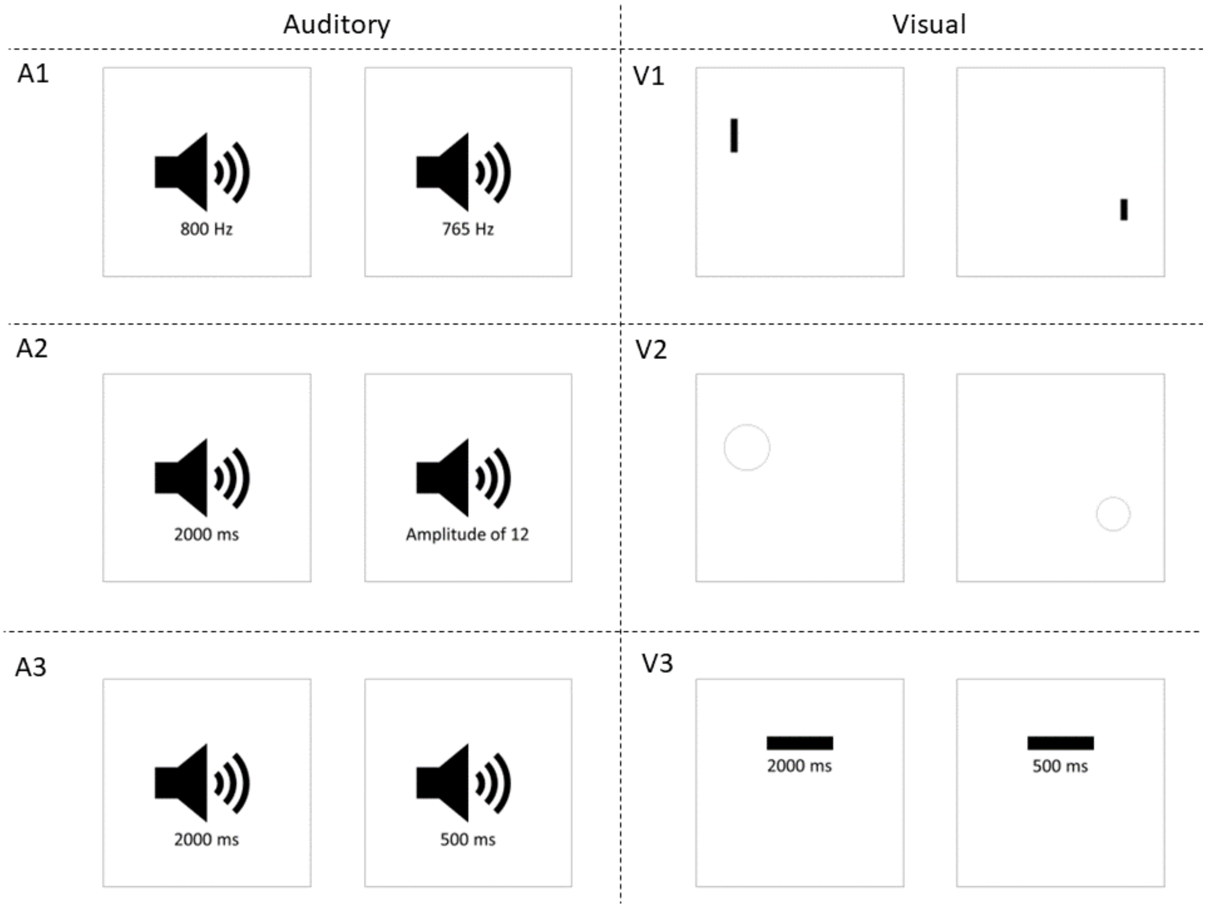
4.2.2.4. Sensory discrimination measures

Three auditory discrimination tasks (pitch, loudness intensity, and auditory duration) and three visual discrimination tasks (line, circle, and visual duration) were used in this study (adapted from Troche et al., 2014; Tsukahara et al., 2020). Figure 4.2 illustrates the stimuli used in this study. Participants were asked to discriminate between two sensory stimuli. In each trial, a standard and a comparison stimulus were presented successively with an interstimulus interval of 500 ms. Then participants were required to determine whether the first or the second stimulus was of longer duration (auditory and visual duration discrimination), higher frequency (pitch discrimination), louder (loudness discrimination), longer size (line discrimination), or bigger (circle discrimination). Participants had unlimited time to respond by pressing left arrow key to choose the first stimulus and right arrow key to choose the second stimulus. The comparison stimulus was randomly set to have a longer, higher, and bigger or shorter, lower, and smaller stimulus compared to the standard stimulus.

Each discrimination task was measured using a 2-alternate forced choice weighted up and down method as an adaptive psychophysical procedure with 64 trials (Kaernbach, 1991). With increasing levels, stimuli decreased in their difference. A correct answer increased the level by one, and an incorrect answer decreased the level by three. There were total of 32 levels with the first six levels having bigger step sizes than the rest of the levels. If the participants responded incorrectly during level 1-3, they went back to level 1. If the participants responded correctly during level 32, they would do level 32 again until they made a mistake or the end of the trial. The outcome variable of all sensory discrimination tasks was accuracy (proportion correct responses).

Figure 4.2

Illustration of the Stimuli for Auditory and Visual Discriminations



Notes. A1 = Pitch discrimination; A2 = Loudness discrimination; A3 = Auditory duration discrimination; V1 = Line discrimination; V2 = Circle discrimination; V3 = Visual duration discrimination. In this illustration, participants needed to press left arrow key because the first stimulus was of higher frequency (A1), louder (A2), longer duration (A3 and V3), longer size (A1), and bigger (A2).

4.2.2.4.1. Auditory discrimination

All auditory discrimination stimuli were created using Audacity.

Pitch discrimination. Participants were asked to decide which of two sequentially presented sine wave tones had a higher pitch. The two tones were of the same duration (500 ms) and intensity (68 dB). The standard frequency of the tone was 800 Hz, and the initial comparison stimulus tone was 762 Hz and 838 Hz. The first six levels had a pitch difference step size of 2 Hz, and the rest of the levels had a pitch difference step size of 1 Hz.

Loudness discrimination. Participants were asked to decide which tone had a louder sound from two sine wave tones of the same duration (500 ms) and pitch (440 Hz) that were played sequentially. The standard tone was produced based on Audacity's amplitude of 50 and the initial amplitudes of the comparison stimulus tones were 12 and 88. The first six levels had an amplitude difference step size of 2, and the rest of the levels had an amplitude difference step size of 1.

Auditory duration discrimination. Participants were asked to decide which tone had a longer duration from two sine wave tones of the same pitch (400 Hz) and intensity (68 dB) that were played sequentially. The standard duration of the tone was 1000 ms and initial comparison stimulus interval was either 65 ms or 1935 ms. The first six levels had a tone duration difference step size of 50 ms, and the rest of the levels had a tone duration difference step size of 25 ms.

4.2.2.4.2. Visual discrimination

All visual discrimination stimuli were created using PowerPoint.

Line discrimination. Participants were asked to decide which line had longer length from two black lines that were presented successively for 1000 ms each in the upper left and in the lower right quadrants of the screen. The lines were 20 cm apart from each other horizontally. The standard length of the line was 5 cm and the comparison stimulus for the first trial was either 3.15 cm or 6.85 cm. The first six levels had a length difference step size of 0.1 cm, and the rest of the levels had a length difference step size of 0.05 cm.

Circle discrimination. Participants were asked to decide which circle had larger size from two black circles that are presented successively for 1000 ms each (with an interstimulus interval of 500 ms) in the upper left and in the lower right quadrants of the screen. The standard and comparison circles diameter, and the diameter difference step size were similar to the line discrimination task.

Visual duration discrimination. Participants were asked to decide which line had a longer duration from two black horizontal lines (5 cm) that are presented successively in the top quadrant of the screen with different duration. The standard and comparison line durations, and the line duration difference step size were similar to the stimuli and procedure for testing auditory duration.

4.2.3. Procedure

Participants completed a questionnaire before attending the laboratory session, which took 15 minutes to complete. The laboratory session lasted for 2 hours and 15 minutes. Participants also completed a Need for Cognition Scale (NFC; Cacioppo, Petty, and Kao, 1984) and two cognitive tasks (i.e. simple reaction time and numerical complex span task), but these measures were not part of the current study and form part of another study. The order of the tasks was fixed as follows: Melodic Discrimination Testing (MDT), Computerised Adaptive Beat Alignment Test (CABAT), numerical complex span tasks, simple reaction time, first break, visual n-back, pitch discrimination, visual monitoring, loudness discrimination, visual binding, auditory duration discrimination, second break, visual duration discrimination, auditory binding, circle discrimination, auditory monitoring, line discrimination, and auditory n-back. The order of the tasks was fixed to have no subsequent tasks targeting the same type of task and modality, and any order effects were equivalent across participants (following Slevc et al., 2016). All tasks were performed on a Windows 7 computer with an LED-backlit LCD display, while participants wore headphones. After the laboratory session, the participants were debriefed and received a £20 Amazon voucher. Students who were psychology undergraduate students at the University of Sheffield could also choose if they wished to receive 10 credits instead. The questionnaire was administered using Qualtrics, whilst MDT and CABAT were administered using psychtestR package in R (Harrison, 2020). All sensory discrimination and working memory tasks were administered using Tatool Web (von Bastian et al., 2013; https://www.tatool-web.com).

4.2.4. Analysis

Analyses were conducted in R using RStudio (version 4.2.0). We conducted confirmatory factor analysis (CFA) and structural equation modelling (SEM) using the lavaan package (Rosseel, 2012, version 0.6-16). All variables were z-standardised during the analysis. We assessed model fits using the model's χ^2 statistic, the comparative fit index (CFI), the root mean squared error of approximation (RMSEA) and its 90% confidence interval, and the standardised root mean squared residual (SRMR). A non-significant *p*-value for the model χ^2 suggests the model fits well, but the sample size can affect the result. Thus, we examined other indices to ensure that the model's fit is satisfactory. A CFI greater than 0.90 is considered acceptable, and beyond 0.95 indicates an exceptionally strong fit. Models with RMSEA and SRMR values below 0.06 are considered to have a satisfactory fit (Hu & Bentler, 1999). After the fit was analysed with a CFA, a SEM was used for the mediation analysis, with musical skill as the predictor variable, working memory as the outcome variable, and sensory discrimination as the mediator.

4.3. Results

4.3.1. Descriptive Data

Data, analysis scripts, and results output are available on the Open Science Framework: https://osf.io/y67ar/?view_only=7ce71905511e4cf7844a436be84fea9f. The descriptive statistics for each task are presented in Table 4.2. The correlation matrix between tasks is presented in Table 4.3. The correlation between sensory discrimination and working memory in both modalities showed small to medium correlations, with the highest significant correlation emerging between auditory monitoring and pitch discrimination, r = .52. Musical skill measures were significantly intercorrelated, with small to medium correlations with the cognitive measures. Demographic factors correlated only with performance on a few of the tasks administered in this study.

Table 4.2

Descriptive statistics

Measure	M	SD	Range	Skewness	Kurtosis	Reliability
Musical skill						
General Sophistication	66.32	17.38	28 – 107	0.19	-0.71	.87[a]
MDT	-0.19	0.97	-3.59 – 2.01	-0.31	0.54	0.55[b]
CABAT	-0.25	1.08	-4.00 – 2.07	-0.97	1.44	0.72[b]
Auditory						
Monitoring	1.81	1.31	-1.3 – 4.31	-0.55	-0.45	.91[c]
Binding	0.92	0.44	-0.53 – 2.54	0.28	0.63	.45[c]
N-Back	0.14	0.53	-1.90 – 1.81	0.02	0.70	.49[c]
Loudness	0.84	0.04	0.69 – 0.94	-1.14	2.37	.95[c]
Pitch	0.80	0.11	0.26 – 0.97	-1.47	3.12	.99[c]
Duration	0.83	0.04	0.62 – 0.89	-2.55	8.27	.96[c]

Visual						
Monitoring	2.21	0.69	-0.99 – 3.36	-0.63	0.98	.62[c]
Binding	1.09	0.54	-0.37 – 3.29	0.34	0.76	.69[c]
N-Back	0.47	0.77	-2.90 – 2.52	-0.81	2.34	.75[c]
Circle	0.86	0.03	0.58 – 0.92	-3.53	23.41	.87[c]
Line	0.85	0.03	0.61 – 0.91	-4.03	27.69	.93[c]
Duration	0.73	0.15	0.34 – 0.88	-1.02	-0.51	.99[c]

Notes. [a]Cronbach's alpha. [b]Standard error of measurement, note that lower value means better reliability. [c]Split-half reliability corrected with Spearman-Brown prophecy formula

Table 4.3

Correlation matrix between demographics, working memory, sensory discrimination, and musical skill

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Demographic Background																	
1. Age																	
2. Education	**.62**																
3. SES	**.14**	.12															
Auditory Working Memory																	
4. Monitoring	**-.23**	-.07	-.01														
5. Binding	-.03	-.05	.04	**.30**													
6. N-back	-.12	-.10	.01	**.34**	**.24**												
Visual Working Memory																	
7. Monitoring	-.01	.12	.05	**.23**	.12	**.27**											
8. Binding	**-.16**	-.10	.01	**.21**	**.16**	**.21**	**.12**										
9. N-back	-.05	.02	.05	**.25**	**.19**	**.31**	**.26**	**.21**									
Auditory Discrimination																	
10. Duration	.04	.10	**.13**	**.33**	**.32**	**.17**	**.20**	**.34**	**.30**								
11. Loudness	.01	.01	.04	**.24**	**.25**	**.19**	**.18**	**.21**	**.29**	**.53**							
12. Pitch	**-.20**	**-.13**	-.04	**.52**	**.25**	**.23**	.06	.08	**.14**	**.38**	**.29**						
Visual Discrimination																	
13. Duration	.03	.04	-.01	.02	**.18**	**.15**	.09	**.13**	**.14**	**.25**	**.16**	**.13**					
14. Circle	.11	**.16**	.05	**.19**	**.20**	.09	**.12**	**.19**	.10	**.37**	**.20**	**.19**	**.19**				
15. Line	.06	.04	.11	**.15**	.10	.08	**.21**	**.22**	**.14**	**.28**	**.26**	**.13**	.10	**.27**			
Musical Skill																	
16. Gold-MSI	.01	.01	.02	**.30**	**.20**	.05	.07	-.02	.01	**.14**	.06	**.34**	.04	.06	.04		
17. MDT	.11	.11	**.16**	**.30**	**.30**	.08	-.01	.06	.10	**.15**	**.14**	**.34**	.07	**.13**	.04	**.39**	
18. CA-BAT	-.05	-.05	.07	**.29**	**.14**	**.14**	.09	**.19***	**.15**	**.27**	**.22**	**.28**	.11	**.15**	**.13**	**.15**	**.23**

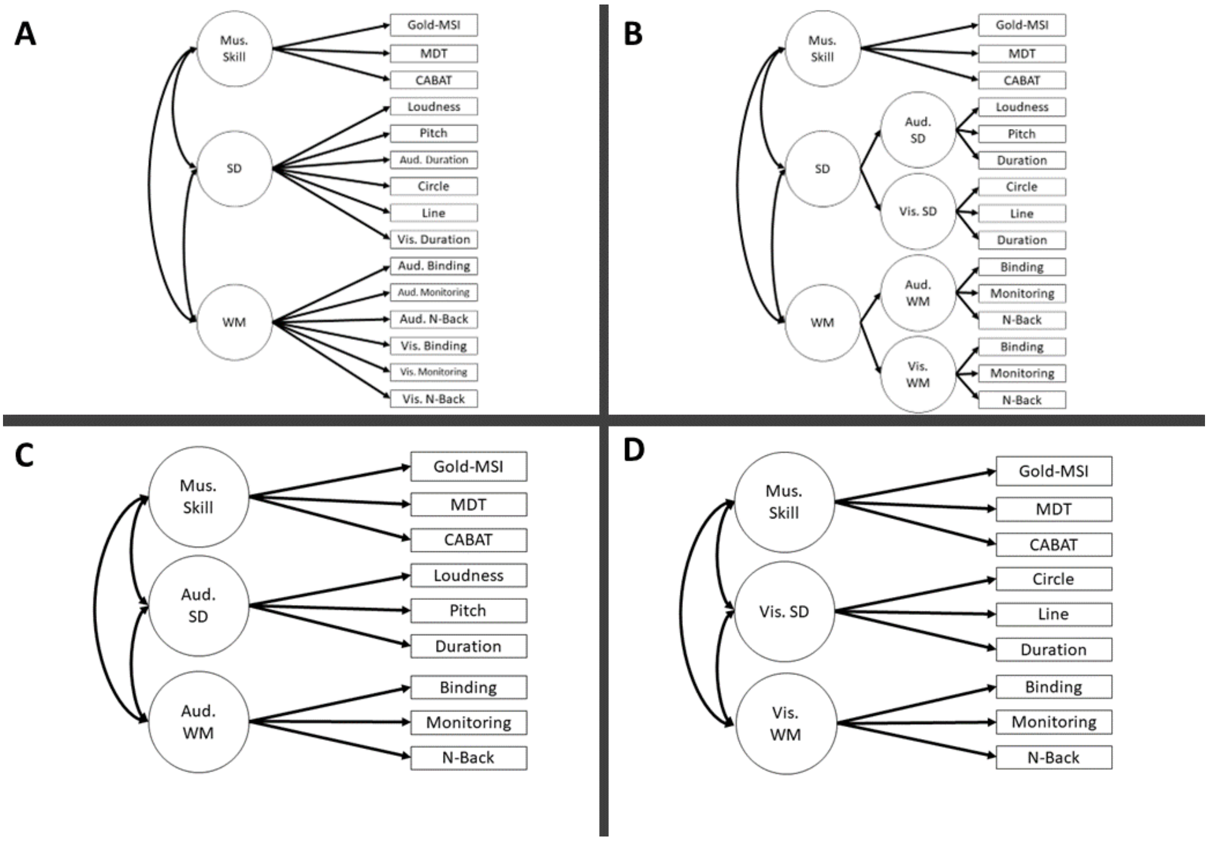
Note. Significant correlations are displayed in bold ($p < .05$); Age and education are given in years. SES = Socioeconomic status; Gold-MSI = Goldsmith Music Sophistication Index; MDT = Melodic Discrimination Testing; CA-BAT = Computerised Adaptive Beat Alignment Test.

4.3.2. Measurement models

Four models were compared in their ability to account for the variance across variables, using confirmatory factor analysis (Figure 4.3). We first conducted confirmatory factor analysis for two modality-general models: Figure 4.3A shows a three-factor model consisting of musical skill, sensory discrimination, and working memory as three separate latent variables. Figure 4.3B shows a hierarchical factor model separating the two modalities as an additional layer. The three-factor model had a poor fit, $\chi^2(87) = 193.81$, $p < .001$, CFI = .83, RMSEA = .07 90% CI (confidence interval) [.05, .08], SRMR = .07, with moderate correlations between musical skill and sensory discrimination and working memory, $r = .51$ and $r = .60$ (ps < .001), respectively, and a strong relationship between sensory discrimination and working memory, $r = .76$, $p < .001$.

Figure 4.3

Models Tested for the Confirmatory Factor Analysis



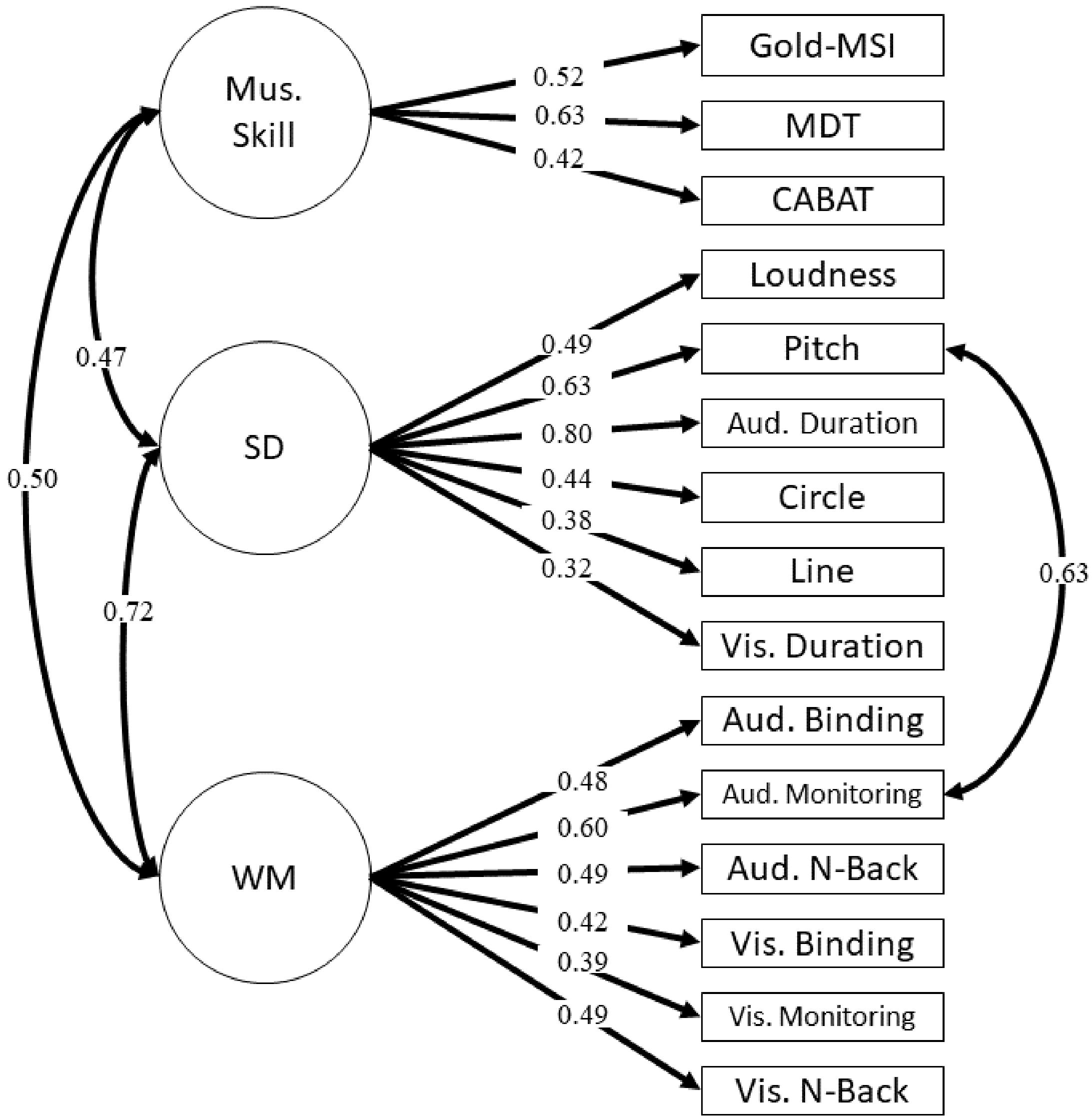
Note. (A) Three-factor model, (B) Hierarchical model, (C) Auditory modality model, (D) Visual modality model. MS = Music skill; SD = Sensory discriminations; WM = Working memory; ASD = Auditory sensory discriminations; VSD = Visual sensory discriminations; AWM = Auditory working memory; VWM = Visual working memory.

The initial hierarchical factor model did not converge because of negative variance estimates (“Heywood cases”) of the auditory discrimination and auditory working memory latent variables. Negative variances for latent factors can occur if there was a misspecification of the model. To explore the reasons Heywood cases, we investigated the modification indices. These indices provide information as to which modifications to the model specifications (i.e. added paths) improve the overall fit (Kline, 2023). Auditory monitoring and pitch discrimination showed the highest modification indices. Further inspection revealed that auditory monitoring and pitch discrimination correlated moderately, $r = .52$. In hindsight, the auditory monitoring and pitch discrimination tasks were highly similar, as participants needed to distinguish between high and low pitch in order to determine whether the three tones they heard were whole tones. To account for this similarity, we added a correlation between the residual variances of these two tasks, which resolved both Heywood cases. The confirmatory factor analysis for a hierarchical model with this added residuals correlations converged and showed a moderate fit, $\chi^2(82) = 148.16$, $p < .001$, CFI = .89, RMSEA = .05 90% CI [.04, .07], SRMR = .07, with moderate correlations between musical skill and sensory discrimination and working memory, $r = .47$ and $r = .51$ (*p*s < .001), respectively, and a strong relationship between sensory discrimination and working memory, $r = .74$, $p < .001$.

To facilitate comparability between models, we re-analysed the three-factor modality-general model with the added correlation between the residual variances of auditory monitoring and pitch discrimination. The results showed the three-factor model had a moderate fit, $\chi^2(86) = 152.77$, $p < .001$, CFI = .89, RMSEA = .05 90% CI [.04, .07], SRMR = .07. This model fit the data better than the model without the added residual correlations, $\Delta\chi^2(1) = 41.04$, $p < .001$. There was no significant difference between the three-factor model and the hierarchical model, $\Delta\chi^2(4) = 4.60$, $p = .330$. Therefore, the more restricted three-factor model was retained for further analysis and the model is presented in Figure 4.4.

Figure 4.4

Confirmatory Factor Analysis for the Three-factor Model



Next, we assessed the two modality-specific three-factor models (see Figures 4.3C and 4.3D). Again, we added the correlation between the residual variances of auditory monitoring and pitch discrimination in the auditory-modality model. Figure 4.5 presents the auditory-modality model and the results showed a moderate fit, $\chi^2(23) = 70.14$, $p < .001$, CFI = .89, RMSEA = .09 90% CI [.06, .11], SRMR = .07, with high correlations among the three latent variables, r = .52 – .73, ps < .001. The visual-modality model is shown in Figure 4.6 and it had a good fit, $\chi^2(24) = 30.29$, p = .175, CFI = .96, RMSEA = .03 90% CI [<.01, .06], SRMR = .05, with relatively lower correlations between musical skill and visual sensory discrimination, r = .29, p = .038, but non-significant relationship with visual working memory, r = .21, p = .125, and a strong relation between visual sensory discrimination and visual working memory, r = .74, p = .001.

Figure 4.5

Confirmatory Factor Analysis for Auditory-modality Model

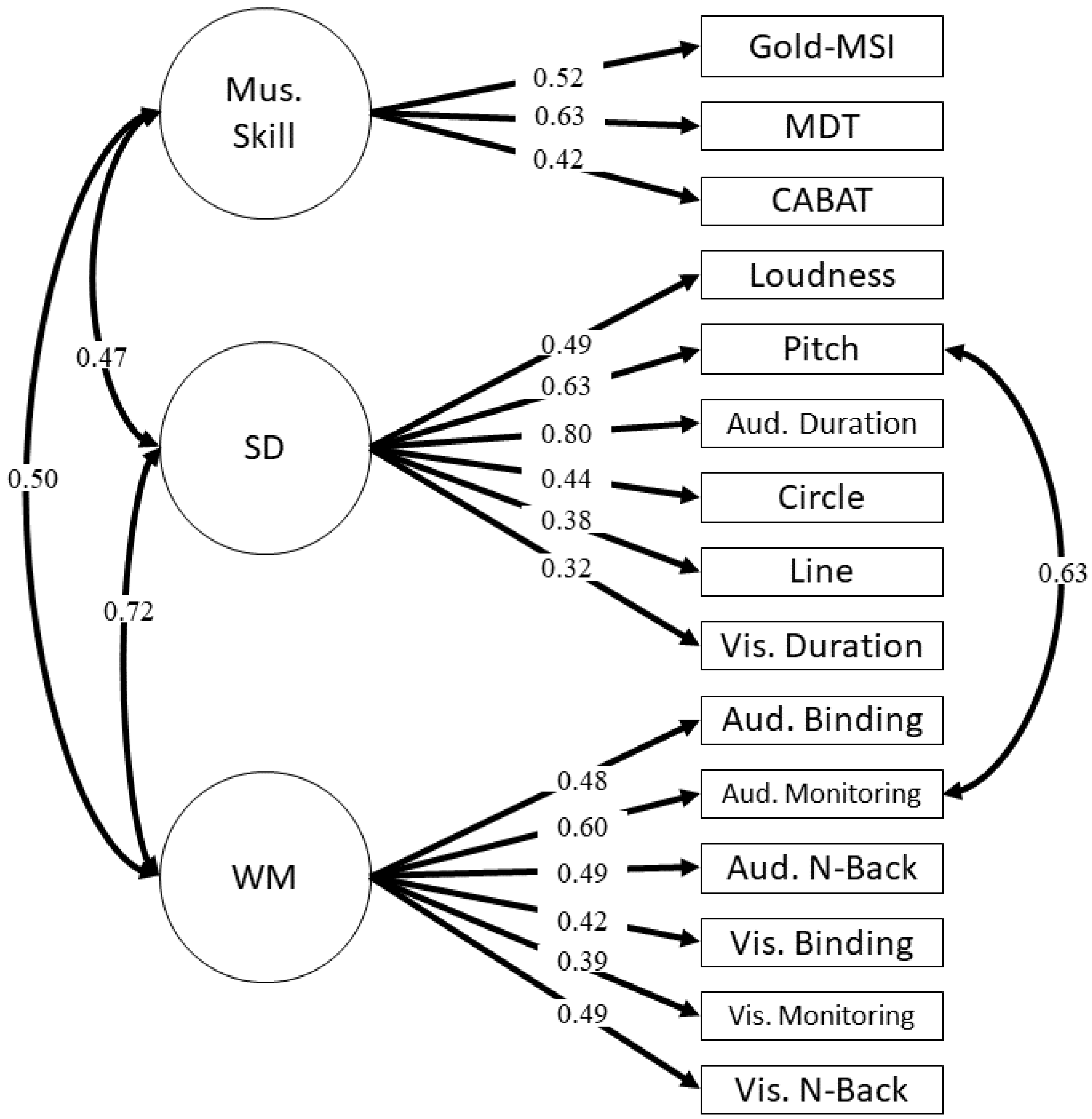
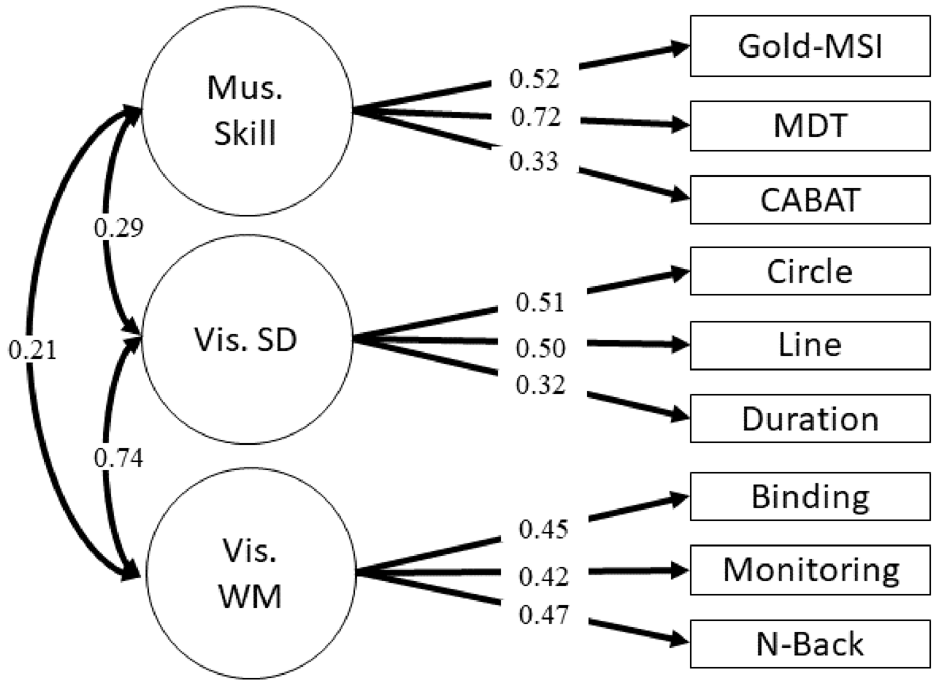


Figure 4.6

Confirmatory Factor Analysis for Visual-modality Model



4.3.3. Latent Mediation Modelling

The hypothesis that sensory discrimination mediates the relationship between musical skill and working memory was tested with a modality-general latent-mediation model and two modality-specific (auditory and visual) models using structural-equation modelling. Figures 4.7, 4.8, and 4.9 illustrate the results. Musical skill was positively related to modality-general discrimination, $a = 0.47$, 95% CI [0.30, 0.63], $z = 5.51$, $p < .001$. This was also the case for each modality, with positive associations between musical skill and auditory discrimination, $a = 0.52$, 95% CI [0.35, 0.69], $z = 6.07$, $p < .001$, and visual discrimination, $a = 0.29$, 95% CI [0.06, 0.52], $z = 2.51$, $p = .025$. Furthermore, the results confirmed positive relationships between modality-general sensory discrimination and modality-general working memory, $b = 0.63$, 95% CI [0.47, 0.79], $z = 7.65$, $p < .001$, which was also reflected in the modality-specific models, with positive associations between auditory sensory discrimination and auditory working memory, *b* = 0.52, 95% CI [0.30, 0.73], *z* = 4.73, *p* < .001, and visual sensory discrimination and visual working memory, *b* = 0.75, 95% CI [0.46, 1.04], *z* = 5.09, *p* = .003. The total effect was significant for the relationship between musical skill and modality-general working memory, *c* = 0.50, 95% CI [0.32, 0.68], *z* = 5.54, *p* < .001, which was also the case for musical skill and auditory working memory, *c* = 0.68, 95% CI [0.50, 0.86], *z* = 7.44, *p* < .001, but not for visual working memory, *c* = 0.20, 95% CI [-0.03, 0.44], *z* = 1.71, *p* = .111.

Finally, we examined whether sensory discrimination mediated the relationship between musical skill and working memory. The results showed indirect effects through sensory discrimination in the general modality, *a*b* = 0.29, 95% CI [0.17, 0.42], *z* = 4.62, *p* = .003, and in the auditory modality, *a*b* = 0.27, 95% CI [0.14, 0.40], *z* = 4.17, *p* = .002. In the visual modality, the indirect effect approached but did not reach significance, *a*b* = 0.22, 95% CI [0.01, 0.42], *z* = 2.09, *p* = .058. After accounting for these indirect effects, the direct effect was still significant for the association between musical skill and auditory working memory, *c'* = 0.41, 95% CI [0.17, 0.65], *z* = 3.40, *p* = .004, but not for the relationship between musical skill and modality-general working memory, *c'* = 0.21, 95% CI [0.01, 0.41], *z* = 2.01., *p* = .086, and visual working memory, *c'* = -0.01, 95% CI [-0.9, 0.26], *z* = -0.10, *p* = .919. Therefore, the results suggest that the relationship between musical skill and general working memory was fully mediated by general sensory discrimination in the general modality, and the relationship between musical skill and auditory working memory was partially mediated by auditory discrimination[1].

[1] We conducted two data re-analyses to further investigate whether musicianship may affect the results. We removed 35 participants who self-identify as musicians for the first analysis and further removed 44 participants who played music according to music activity question yet considered themselves as non-musicians. We found similar results to the main analysis, therefore we did not report it in this thesis.

Figure 4.7

Schematic Results of the Latent-mediation Model Analysis, Testing Modality-general Sensory Discrimination (SD) as a Mediator of the Relationship between Musical Skill (Mus. Skill) and Modality-general Working Memory (WM)

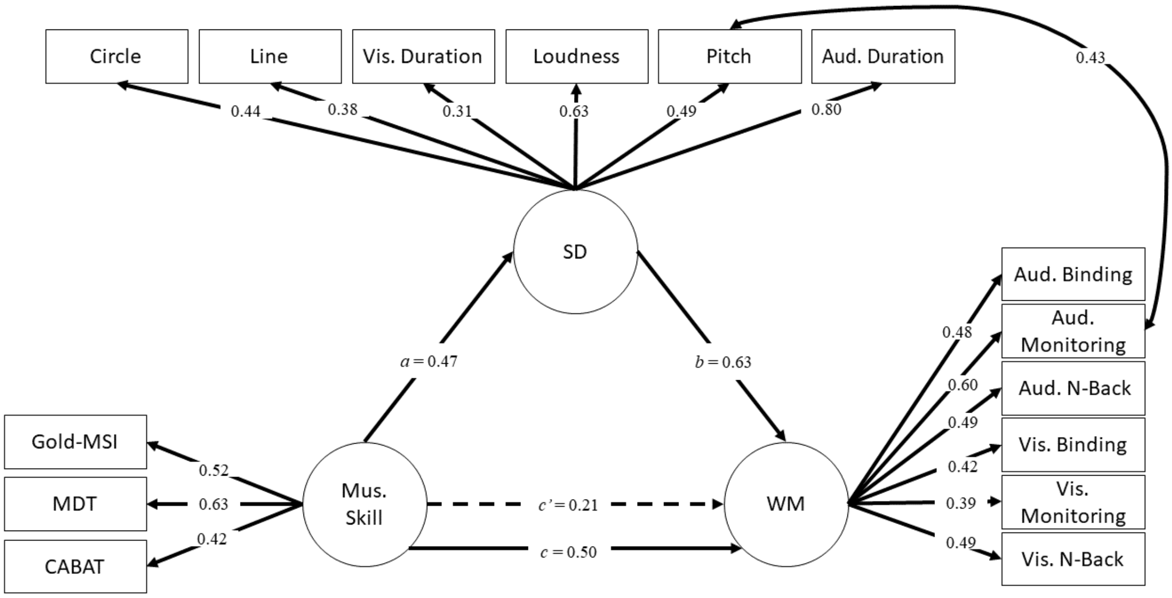


Figure 4.8

Schematic Results of the Latent-mediation Model Analysis, Testing Auditory Sensory Discrimination (Aud. SD) as a Mediator of the Relationship between Musical Skill (Mus. Skill) and Auditory Working Memory (Aud. WM)

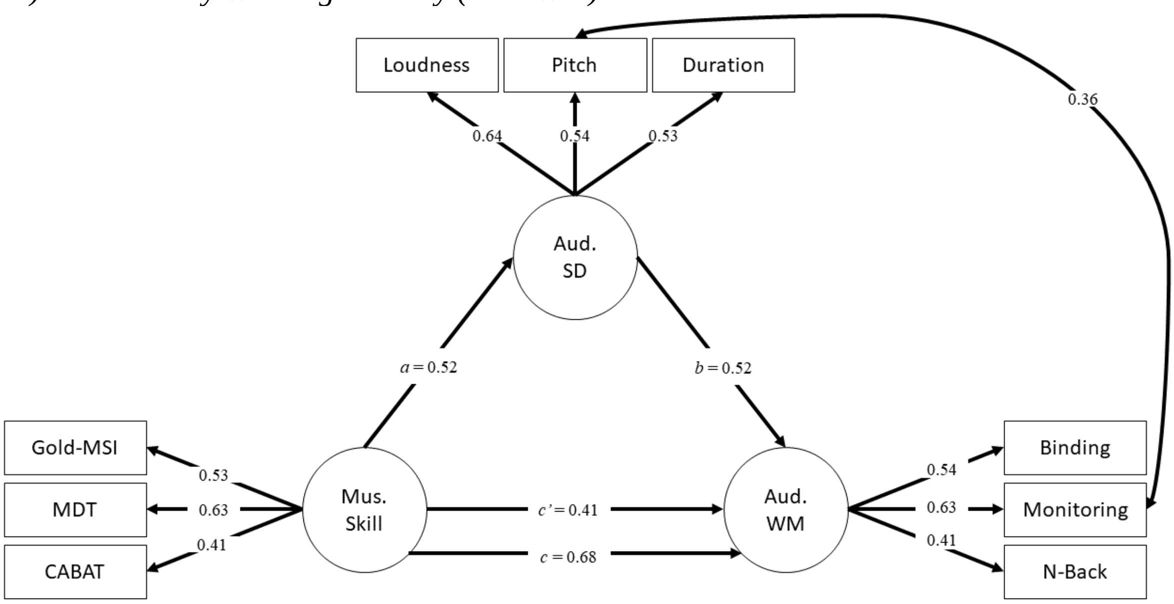
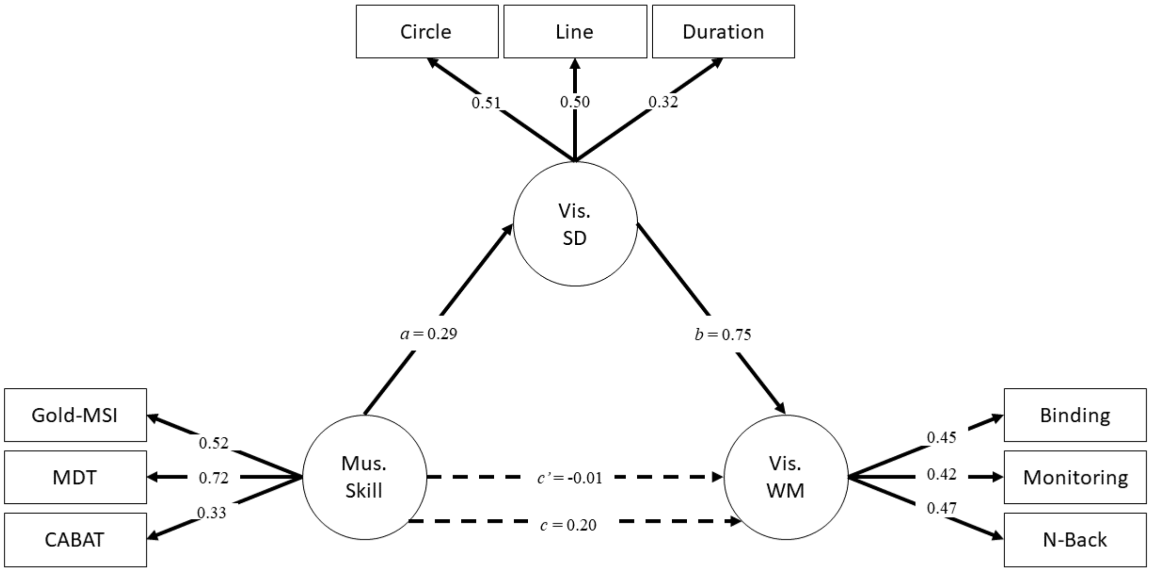


Figure 4.9

Schematic Results of the Latent-mediation Model Analysis, Testing Visual Sensory Discrimination (Vis. SD) as a Mediator of the Relationship Between Musical Skill (Mus. Skill) and Visual Working Memory (Vis. WM)



4.4. Discussion

The present study took a latent-variable approach to examining how musical skill (measured objectively as well as via self-report) relates to working memory and sensory discrimination across modalities. This study revealed that the relationship between musical skill and working memory was mediated by sensory discrimination both in the modality-general and the auditory modality-specific models, but not in the visual modality-specific model.

The results indicate that the relationship between musical skill, sensory discrimination, and working memory may be modality-specific and tied specifically to enhanced auditory sensory discrimination. Individuals with high musical skills are good at sensory discriminations, which in turn enhances the working memory, and this relationship may be attributed to enhanced auditory discrimination. It should be noted that auditory discrimination did not entirely account for the advantage in the auditory domain and may be influenced by other factors. At first glance, this appears to be inconsistent with previous research that found a correlation between musical skill and both auditory and visual working memory (Silas et al., 2022; Slevc et al., 2016), although the correlations between musical skill and visual working memory were generally weaker than those with auditory working memory. The relatively larger correlations also in the visual modality in these past studies may be due to differences in the tasks used. For example, Slevc at al. (2016) administered n-back stimuli in the form of letters and Silas et al. (2022) presented digit in the backward digit span task rather than squares, as we did in the present study. Participants in that study may have processed these letters and digits stimuli in the phonological loop, which taps into verbal working memory, whereas the current study reduced such phonological processing by using nonverbal stimuli (black squares or pictures of abstract fractals). As visually presented verbal material has access to the phonological store (Deschamps et al., 2020), this may suggest previous findings assessed verbal working memory, whereas the present study better measured visual working memory. This notion was supported by previous empirical study by Arndt et al. (2023) and a meta-analysis by Talamini et al. (2017) that showed musicians had significantly higher tonal (auditory) and verbal working memory compared to non-musicians but not visuospatial working memory. Therefore, taken together with the body of literature, our findings provide further support for the hypothesis that musical skill is modality-specific, particularly auditory working memory.

The association between musical skill, auditory discrimination, and working memory may be explained by how people process musical information and the overlap with auditory processing. In the present study, on one hand, the tasks utilised to measure musical skill objectively (i.e. MDT and CA-BAT) were argued that they were heavily relied on working memory since melodies and rhythms must be stored in working memory in order to be contrasted and distinguished (Harrison et al., 2017; Harrison & Müllensiefen, 2018). On the other hand, making music relies heavily on the auditory system more so than on visual processing, resulting in enhanced auditory discrimination ability (Wang, 2022). When learning to play a musical instrument or singing, musicians have to learn distinguishing between fine-grained differences between pitches and loudness during music lessons. The pitch and loudness discrimination tasks – such as those administered in the present study – assess the fundamental processes involved in these musical skills. This explains the moderate relationship between musical skill and sensory discrimination. Additionally, one should maintain information in working memory to observe differences between pitches, musical dynamics, and duration of musical notes. Taken together, musical skill may predict sensory discrimination ability which in turn is associated with a better working memory.

Regarding the way musical skill was assessed, it is possible that sensory discrimination better measures musical skill because the objective tasks asked participants to discriminate between stimuli. In the early musical skill research, the ability to hear music was assessed with tasks associated to auditory discrimination, such as pitch and loudness discriminations, and sense of rhythm (Seashore, 1915). A previous study on sensory discrimination adapted Seashore Measures of Musical Talent to measure the ability to discriminate pitch (Acton & Schroeder, 2001), which highlights the similarity between musical skill and auditory discrimination tasks. The present study extended the way to measure musical skill by complementing MDT and CA-BAT with Gold-MSI. The results of this study showed a moderate task loading between musical skill and auditory sensory discrimination (r = .52), suggesting that musical skill assessed the skill above and beyond discriminating pitch, loudness, and duration. As the present results are correlational and, therefore, cannot provide evidence for a causal relationship, this warrants further investigation on the effect of sensory discrimination on the musical skill or vice versa.

The results of the modality-general mediation model showed no direct relationship between musical skill and working memory once sensory discrimination was included as a mediator of their relationship. Further inspection of the factor loadings showed that the sensory discrimination and working memory latent factors loaded more in the auditory modality than the visual modality; thus, most of the variance in the modality-general sensory discrimination and working memory was explained by the auditory modality. Previous studies that measured both auditory and visual sensory discrimination and combined them into a single general sensory discrimination latent variable found a relationship between sensory discrimination and working memory (Troche et al., 2014; Tsukahara et al., 2020). Similarly, previous latent-variable research that found a positive relationship between musical skill and working memory updating (Okada & Slevc, 2018) combined different modalities of working memory updating into a single latent variable (i.e. verbal, and visuospatial). Therefore, this further supports the idea that working memory is multisensory (Quak et al., 2015), and the relationship between musical skill, sensory discrimination, and working memory was both modality-specific and modality-general.

4.4.1. Methodological Strength

The key strength of the present study was the utilisation of the latent variable approach to avoid the task-impurity problem. The musical skill was measured by complementing both questionnaire and objective measures (i.e. melody and rhythm tasks). Administering both measures may reduce the methodological limitation of a self-report questionnaire, in which it may likely cause measurement noise because participants may give inaccurate responses due to under- or overestimating their musical skills or showing social desirability bias. The current study extended the previous research (i.e. Okada & Slevc, 2018) by using both self-report and objective measurements to measure musical skill and to distinguish between auditory and visual modalities.

Most participants in this study perceived themselves to be non-musicians, although the answer to the music activity question indicated that 19.3% of the non-musicians still engaged in musical activities in their daily life. Non-musicians can have high musical skill despite lacking musical training and, hence, those with high musical skill may also show high cognitive abilities. Hence, by measuring musical skill as a continuous variable rather than dichotomising people in musicians and non-musicians, this study extends previously found results, showing incremental benefits of musical skill and sensory discrimination for working memory, even for participants who do not consider themselves musicians (see also Correia et al., 2023).

4.4.2. Limitations and Future Research

There were several limitations that need to be addressed. First, as this was a cross-sectional investigation, causality cannot be inferred between the variables. Thus, future longitudinal randomised-controlled trial studies that train non-musicians and compare them to an active control group are required to ascertain the causality of sensory discrimination mediating the effect of music skill on working memory, including for different modalities to investigate generalisability of transfer effects of music skill development.

Another limitation was the reliability of the tasks which showed that visual and auditory working memory tasks had relatively lower reliabilities compared to other variables. One possibility of the low reliabilities was that the tasks were deemed difficult for participants as we utilised relatively bigger set size for the binding task (3 – 5 set size) and n-back task (3-back and 4-back) compared to other studies (e.g. Porflitt & Rosas, 2020; Slevc et al., 2016). Due to the low reliability, those tasks may have various results when tested to other samples with the same criteria because of the inconsistency of the measures (Urbina, 2012). Despite the low reliability, the tasks measuring working memory were still valid as they were related to each other and loaded to the same factor.

The limitation was also found in the instruction of the task, especially the visual duration discrimination. The participants were instructed to decide which line was shown longer between two lines presented sequentially, then the procedure was illustrated and had a chance to try during the practice trials. However, some participants mentioned that they thought they were expected to choose which line had longer length. This may explain why the mean of visual duration discrimination was lower compared to other discrimination tasks, and it was weakly related to circle discrimination and not related to line discrimination. Future study should follow the procedure of previous studies by using light emitting diode (LED) or flashing colour on the screen instead of line on the screen to avoid confusion to discriminate length and duration (Troche et al., 2014; Voelke et al., 2014).

4.4.3. Conclusions

In conclusion, this pre-registered latent-mediation study confirmed that sensory discrimination mediated the relationship between musical skill and working memory in the modality-specific auditory and in the modality-general models, but not in the modality-specific visual model. Therefore, the relationship was specific to the auditory modality, and the associations observed in the modality-general model may be driven by the strong relationship within the auditory modality. This study provides further evidence for the relationships between musical skill, auditory discrimination abilities and auditory working memory in the general population and could serve as a foundation for future longitudinal studies that include sensory discrimination as a mediator of the effect of musical skill on working memory.

Chapter 5 – General Discussion

Contributions:

Christ Billy Aryanto (Conceptualisation, writing – original draft)

Emma Blakey (Conceptualisation, supervision)

Renee Timmers (Conceptualisation, supervision)

Claudia C. von Bastian (Conceptualisation, supervision, writing – review & editing)

5.1. Summary of Main Findings

This thesis had two main objectives: To investigate (1) the relationship between music making and executive functions in healthy adults, and (2) the mechanism underlying this relationship. In Chapter 2, we conducted a systematic review and a meta-analysis to find the relationship between music making and three factors of executive functions by including previous comparison and correlation studies. In this chapter, we found that music making was significantly related to inhibition, shifting, and working memory, with the largest average association and the most decisive evidence emerging for working memory. In the meta-analysis, music making was operationalised as comparison between musicians and non-musicians, and mixed different types of music making measures related to executive functions such as music training, musical skill, and musical sophistication. To further test the relationship in an empirical study, we test frequency of music making as the operationalisation of music making. Chapter 3 revealed that the frequency of music making was not related to any of the executive functions factors across the lifespan after accounting for demographic variables, particularly age. Therefore, this result suggested that frequency of music making may be different from musical skill which previously found in the meta-analysis.

The results of Chapter 2 suggested a strong relationship between music making and working memory, and specifically measured musical skill, as Chapter 3 suggested that the frequency of music making may not necessarily enhance musical skill. Therefore, in Chapter 4, we focused on the relationship between musical skill and working memory instead. We assessed sensory discrimination and working memory in auditory and visual modalities to investigate whether their relationship with musical skill was modality-specific or modality general. In this chapter, we found evidence that musical skill was associated with general-modality and auditory working memory but not visual working memory, and their relationship was mediated by sensory discrimination. Therefore, the relationship between music making and working memory was both modality-general and auditory modality-specific. This finding shed the new light that sensory discrimination explained the underlying mechanism on the relationship between music making and working memory.

5.2. Theoretical and Methodological Implications

The results of this thesis showed that the relationship between music making and executive functions is likely due to the stronger relationship with working memory compared to other two factors. However, this does not mean that other factors of executive functions are irrelevant to music making. Arguably, working memory is also needed in inhibition and shifting tasks as previous studies also showed positive relationships between them (e.g. Miyake and Friedman, 2012; Friedman & Miyake, 2017). The strong positive relationship between music making and working memory suggests the importance of working memory in musical activities, and supports the finding from previous study that people who make more music had higher working memory compared to people who did not make music (e.g. Grassi et al., 2017; Talamini et al., 2016, 2017).

The results of Chapter 4 showed the interplay between music making, sensory discrimination, and working memory which were measured separately in previous studies, such music making and sensory discrimination (Hyde et al., 2009), sensory discrimination and working memory (Troche et al., 2014; Tsukahara et al., 2020), and music making and working memory (Talamini et al., 2016). There are two possible reasons indicated by the results. First, sensory discrimination explained the underlying mechanism between music making and working memory. The original creation of MDT (Harrison et al., 2017) and CA-BAT (Harrison & Müllensiefen, 2018) argued those tasks relied heavily on auditory working memory, as melodies must be held in working memory if they are to be compared and discriminated. The results of this study found that the relationship between music making and working memory was mediated by the sensory discrimination, although direct relationship was still pertinent between musical skill and auditory working memory. Therefore, there may be another variable aside from sensory discrimination which was not measured in this thesis that may explain the relationship. Second, the relationship between music making and auditory working memory may suggest that people who can update and maintain auditory information in their mind may find it easier to make music. This notion needs further exploration to confirm whether people with higher auditory working memory would be better musicians or more likely to make music than people who have lower auditory working memory.

Another suggestion based on the triarchic relationship was that sensory discrimination may contribute to working memory performance. A possible explanation for that may be because the tasks measuring sensory discrimination asked participants to keep the information in the working memory as the stimuli were presented sequentially. Previous studies found a direct relationship between general-modality sensory discrimination and working memory capacity on children (Voelke et al., 2014) and adults (Troche et al., 2014), although Tsukahara et al. (2020) found that the relationship between them was mediated by inhibition. The results of present thesis extended previous studies and found that the relationship between sensory discrimination and working memory was both modality-general and modality-specific.

Furthermore, the results of thesis showed that people who had high musical skill have better executive functions compared to people who had lower musical skill. This could expand the idea of capacity-efficiency model of cognitive training (von Bastian et al., 2022) which have been proposed to explain how working memory can be enhanced. According to the capacity-efficiency model, there are two, not mutually exclusive, strategies to enhance the working memory. The first strategy was by increasing the number of representations that can be actively held in working memory and, thus, increasing the capacity of working memory. The second strategy was by enhancing the efficiency with the current working capacity limit as a person became more expert in a particular task. In this thesis, it can be argued that music making was related to the efficiency of cognition as musical skill related to the higher accuracy in sensory discrimination and working memory. A recent study has shown that visual working memory training can increase the precision of visual working memory representations but not working memory capacity (Jiang et al., 2023). It can be argued that music making is comparable to visual working memory training, as this thesis suggested that people who were trained in music would have better musical skills and then they may use their cognitive capacity more efficiently than people who did not train in music, and this efficiency was shown in auditory modality. Therefore, this may suggest that both visual and auditory working memory were possible to be trained with a specific type of training targeting the specific modality. Further investigation should be undertaken to determine the causality of the association between music making and cognitive abilities.

The modality-specific relationship between music making and working memory may suggest a potential relationship between different types of art and different modalities of working memory. For example, painting is related to visual processing, and it may be related to visual working memory, whereas poetry is related to verbal processing, and it may be related to verbal working memory. To the best of my knowledge, there was no research that directly studies the relationship between visual art and writing and visual and verbal working memory yet. There was evidence that visual working memory related to art appreciation (Sherman et al., 2015), and verbal working memory related to language production (Kellogg et al., 2007) and expressive writing (Klein & Boals, 2001). Research on different types of art and working memory could be further investigated to understand whether different art forms may be related to working memory or if it was a special case for music making.

The different modalities of working memory studied in this thesis supported the argument of domain specificity of working memory. Recent review by Nozari and Martin (2024) explained that working memory processed the information both domain-generality, as information may need to be temporarily maintained before they are acted upon, and domain-specific, as information may be tied to specific representations (e.g., visual or verbal). They argued that working memory was mostly domain-specific in term of daily application, which is supported by the results of this thesis that the relationship between music making and working was specific to the auditory modality. Furthermore, the results of Chapter 4 showing that working memory can be processed in both auditory and visual may be extrapolated to the potential different modalities of other factors of executive functions. Previous studies in music making and executive functions mostly used visual stimuli, either verbal or non-verbal (e.g., Criscuolo et al., 2019; Okada & Slevc, 2018; Porflitt & Rosas-Díaz, 2019; Porflitt & Rosas, 2020). Slevc et al. (2016) studied the relationship between music making and each factor of executive functions on both visual and auditory modalities. However, they did not conduct latent variable modelling so the model fit of the executive functions in different modalities cannot be identified. Therefore, further studies are needed to investigate the domain-general and domain-specific of executive functions, especially exploring different modalities of each executive function; not only the domain-general and domain-specific of working memory, but also inhibition and shifting.

This thesis showed that age was an important factor to be accounted for in the study of executive functions. We found that age had significant positive significant relationship to inhibition and shifting tasks and negative significant relationship to working memory tasks in Chapter 3, but only several measures of sensory discriminations and working memory related negatively in Chapter 4 (cf. Table 3.2 and Table 4.3). The negative relationship with working memory and positive relationship with shifting found in this thesis supported previous studies that working memory decline but better at shifting between tasks as people got older (Ferguson et al., 2021; Loaiza, 2024). It was surprising that age had a positive relationship in this thesis previous study showed that inhibition would decline from as early as 30 years old (Ferguson et al., 2021). However, other studies found that whether age increased or decreased inhibition depended on the tasks to assess the inhibition (Rey-Mermet et al., 2018; Rey-Mermet & Gade, 2018). Therefore, this shows the importance of including age as covariate in executive functions studies as it may confound the results of a study.

Another factor that needed to be considered in the study of music making was the operationalisation of music making. We operationalised music making more broadly in Chapter 2 by including all types of measures to assess music making and became more specific in Chapter 3 by measuring frequency of music making and Chapter 4 by measuring musical skill. The evidence from this thesis suggests that there was a difference between merely making music in everyday life and musical skill. This supported previous study that musical skill was enhanced by deliberately practicing music instead of merely making music daily (Platz et al., 2014). It can therefore be assumed that the precision of the operationalisation of music making is important and further studies need to carefully be considered on way to assess music making.

5.3. Strengths of This Thesis

This study has employed two different types of empirical studies, meta-analysis and latent variable modelling. In the meta-analysis (Chapter 2), a three-level meta-analysis was conducted because each executive functions factor could be measured with several tasks, leading to statistical dependency between effect sizes (Harrer et al., 2021). Conducting a three-level meta-analysis accounts for within- and between-studies variability. Furthermore, this thesis showed the benefit of latent variable modelling to address the task-impurity problem both for executive functions and music making variables (Chapter 3 and 4).

Furthermore, music making was assessed using both subjective measures (i.e. self-report questionnaires) and objective measures (i.e. melody and rhythm tasks) which enabled us to measure the musical skill more comprehensively.

This thesis included both comparative and correlation design in the meta-analysis and conducted both analyses in Chapter 3. In Chapter 2, we included both comparison and correlational studies in the meta-analysis and found that the type of study significantly moderated the relationship between music making and shifting, yet it had ambiguous Bayes Factor. The results from the meta-analysis were supported by Chapter 3 in which comparative analysis showed significant difference comparing the inhibition and shifting's composite scores of high and low frequency music makers, but the correlational analysis showed non-significant results for the relationship between frequency of music making and executive functions. However, one of the drawbacks of comparison studies is that they dismissed people who have the average musical experience (McClelland et al., 2015). Therefore, it is recommended to conduct a correlational study in music making study because a different range of music making experience and skills may be measured from different people as shown in Chapter 3 and Chapter 4.

5.4. Limitations and Future Research

A number of limitations needs to be noted regarding this thesis. First, despite criticising the arbitrary criteria in the previous music making and executive functions study, the comparison between high and low frequency music makers in Chapter 3 may also be deemed as arbitrary. In Chapter 3, the comparison was complemented with a correlational analysis using continuous variable to minimise the bias of comparing extreme groups and found similar results. It would be beneficial for future music making and cognition studies to have a consensus if a comparison between musicians and non-musicians would like to be made.

Regarding the music making variable, the musical tasks employed in Chapter 4 were all measuring receptive musical skills, in which participants were asked to listen to the music stimulus and responded to it. For example, in the MDT, participants were asked to choose which melodies were the odd one out compared to the other two melodies. Thus, the musical skill in this thesis measured perception of music and not yet production of music. Musical skill can be measured more comprehensively by measuring productive musical skills tasks in which participants were asked to produce melodies or rhythm. Assessing productive musical skill objectively will help to understand how strong the relation between the skill to produce music to the executive functions. Okada (2020) reported a strong relationship between perceptive and productive musical skills, and both skills were also significantly related to musical training subscale of Gold-MSI. Future studies should explore both types of musical skills and its relationship to cognitive abilities to extend the results of this thesis which only focused on the receptive musical skill.

As all studies in this thesis were cross-sectional, the causality between music making and executive functions still need to be further explored. Future research could measure different modalities of executive functions (e.g. auditory, numerical, verbal, visual-spatial) and address the causality between music making and executive functions. Measuring different modalities would expand the knowledge on how music making trained specifically to the auditory modality or could be generalised to other modalities. Future longitudinal randomised-controlled trial studies that train non-musicians and compare them to an active control group would ascertain the causality of music making on executive functions. An experiment in executive functions training and its effect on musical skill may also be conducted to investigate the opposite direction of the relationship.

5.5. Conclusion

In conclusion, this thesis answered the research question of **how and why music making and executive functions are related** and with a thorough investigation of the relationship between music making and executive functions, with a specific focus on working memory. The findings of the pre-registered meta-analysis study indicate a significant association between music making and key executive functions factors, that is inhibition, shifting, and working memory, with the strongest association evidenced for working memory. Further exploration of the relationship between frequency of music making and executive functions found that these relationships are subject to variations influenced by demographic factors, especially age. We argued that frequency of music making may not enhance musical skill which in turn related to executive functions, particularly working memory. Thus, we tested sensory discrimination as a prime candidate mediator of the association between musical skill and working memory. We found that musical skill was related to modality-general working memory and auditory working memory, and these associations were mediated by sensory discrimination. Taken together, these findings provide important insight that music making was strongly associated with working memory, and sensory discrimination explained the underlying mechanism between them. This thesis builds a foundation for future longitudinal studies to investigate the causal relationship between music making and executive functions, especially in different modalities.

Appendix A

Table A1

The Results of Influence Diagnostics Across All Three Factors of Executive Functions

	rstudent	dffits	cook.d	cov.r	QE.del	hat	weight	infl
Omitting Amer et al. (2013)	-0.088	-0.018	0.000	1.043	267.088	0.019	1.895	
Omitting Anaya (2013)	-0.025	-0.009	0.000	1.045	267.063	0.020	1.977	
Omitting Bialystok & DePape (2009)	0.337	0.044	0.002	1.044	266.084	0.021	2.079	
Omitting Blumenthal (2013)	-0.594	-0.089	0.008	1.038	265.974	0.020	2.028	
Omitting Caldwell (2015)	0.676	0.069	0.005	1.018	266.121	0.011	1.083	
Omitting Clayton et al. (2016)	-0.767	-0.107	0.012	1.029	265.583	0.018	1.823	
Omitting Criscuolo et al. (2019)	5.356	0.906	0.498	0.604	189.133	0.020	1.993	*
Omitting D'Souza et al. (2018), comp 1	0.141	0.015	0.000	1.049	266.693	0.022	2.177	
Omitting D'Souza et al. (2018), comp 2	-0.536	-0.084	0.007	1.042	265.979	0.022	2.153	
Omitting der Nederlanden et al. (2020)	-0.952	-0.140	0.020	1.025	263.368	0.021	2.094	
Omitting Franklin et al. (2008)	0.568	0.066	0.004	1.026	266.134	0.014	1.428	
Omitting Grassi et al. (2017)	0.416	0.052	0.003	1.036	266.274	0.018	1.750	
Omitting Gray & Gow (2020)	0.055	0.002	0.000	1.039	267.020	0.017	1.728	
Omitting Hanna-Plady & MacKay (2011)	-0.119	-0.023	0.001	1.047	267.088	0.021	2.076	
Omitting Hou et al. (2014), comp 1	-1.148	-0.160	0.025	1.014	262.826	0.019	1.935	
Omitting Hou et al. (2014), comp 2	-1.271	-0.172	0.029	1.007	262.352	0.019	1.856	
Omitting Jain & Nataraja (2019)	0.032	-0.001	0.000	1.046	267.007	0.020	1.999	
Omitting Kempe et al. (2012)	-0.790	-0.124	0.016	1.036	262.859	0.023	2.303	
Omitting Lu & Greenwald (2016)	1.305	0.189	0.035	1.001	259.197	0.020	1.957	
Omitting Mansens Deeg & Comijs (2018)	-0.572	-0.097	0.010	1.049	251.772	0.025	2.546	
Omitting Meyer et al. (2020)	0.311	0.040	0.002	1.044	266.240	0.021	2.055	
Omitting Moradzadeh et al. (2015), comp 1	-0.428	-0.069	0.005	1.046	266.408	0.022	2.193	
Omitting Moradzadeh et al. (2015), comp 2	-0.098	-0.020	0.000	1.049	267.088	0.022	2.150	
Omitting Moser (2003)	2.189	0.381	0.127	0.915	193.698	0.024	2.422	
Omitting Moussard et al. (2016)	4.333	0.486	0.209	0.815	235.378	0.012	1.192	*
Omitting Okada & Slevc (2018)	-0.599	-0.097	0.010	1.044	264.298	0.024	2.362	
Omitting Okada (2018)	-0.546	-0.090	0.008	1.046	264.631	0.024	2.384	
Omitting Porflitt & Rosas-Diaz (2019)	0.155	0.018	0.000	1.051	266.497	0.023	2.273	
Omitting Porflitt & Rosas (2020)	0.486	0.067	0.005	1.041	265.169	0.021	2.111	
Omitting Ramachandra et al. (2012)	0.295	0.038	0.001	1.045	266.283	0.021	2.069	
Omitting Schroeder et al. (2016)	-0.459	-0.075	0.006	1.047	266.072	0.023	2.271	
Omitting Slevc et al. (2016)	-0.424	-0.069	0.005	1.047	266.367	0.022	2.226	
Omitting Strong & Mast (2019)	-0.292	-0.046	0.002	1.042	266.947	0.019	1.909	
Omitting Strong & Midden (2020)	-0.456	-0.068	0.005	1.039	266.613	0.019	1.895	
Omitting Suarez et al. (2016)	0.189	0.022	0.000	1.045	266.693	0.020	2.022	
Omitting Vuvan et al. (2020)	0.444	0.065	0.004	1.049	261.546	0.024	2.436	

Omitting Zuk et al. (2014)	0.186	0.020	0.000	1.038	266.854	0.017	1.711
Omitting Lee et al. (2007)	-1.011	-0.141	0.020	1.021	264.000	0.019	1.902
Omitting Bianco et al. (2017)	0.158	0.017	0.000	1.040	266.882	0.018	1.768
Omitting Chang-Arana & Luck (2018)	-1.148	-0.149	0.022	1.012	263.988	0.017	1.676
Omitting Gagnon & Nicoladis (2020)	-0.597	-0.098	0.010	1.045	263.603	0.024	2.408
Omitting Hanna-Plady & Gajewsky (2012)	-0.773	-0.117	0.014	1.034	264.467	0.021	2.150
Omitting Hansen et al. (2012)	-0.781	-0.117	0.014	1.033	264.703	0.021	2.093
Omitting Parbery-Clark et al. (2011)	0.434	0.055	0.003	1.036	266.138	0.018	1.815
Omitting Posedel et al. (2011)	-0.104	-0.020	0.000	1.045	267.089	0.020	1.955
Omitting Slater et al. (2017)	0.976	0.140	0.020	1.020	261.955	0.020	2.008
Omitting Strait et al. (2010)	-1.241	-0.166	0.027	1.008	262.881	0.018	1.802
Omitting Talamini et al. (2016)	-0.049	-0.012	0.000	1.042	267.080	0.018	1.835
Omitting Vasuki et al. (2016)	0.177	0.019	0.000	1.042	266.813	0.019	1.865
Omitting Weiss et al. (2014)	-0.560	-0.087	0.008	1.042	265.834	0.022	2.160

Figure A1

Plots for Influence Diagnostics Across All Three Factors of Executive Functions

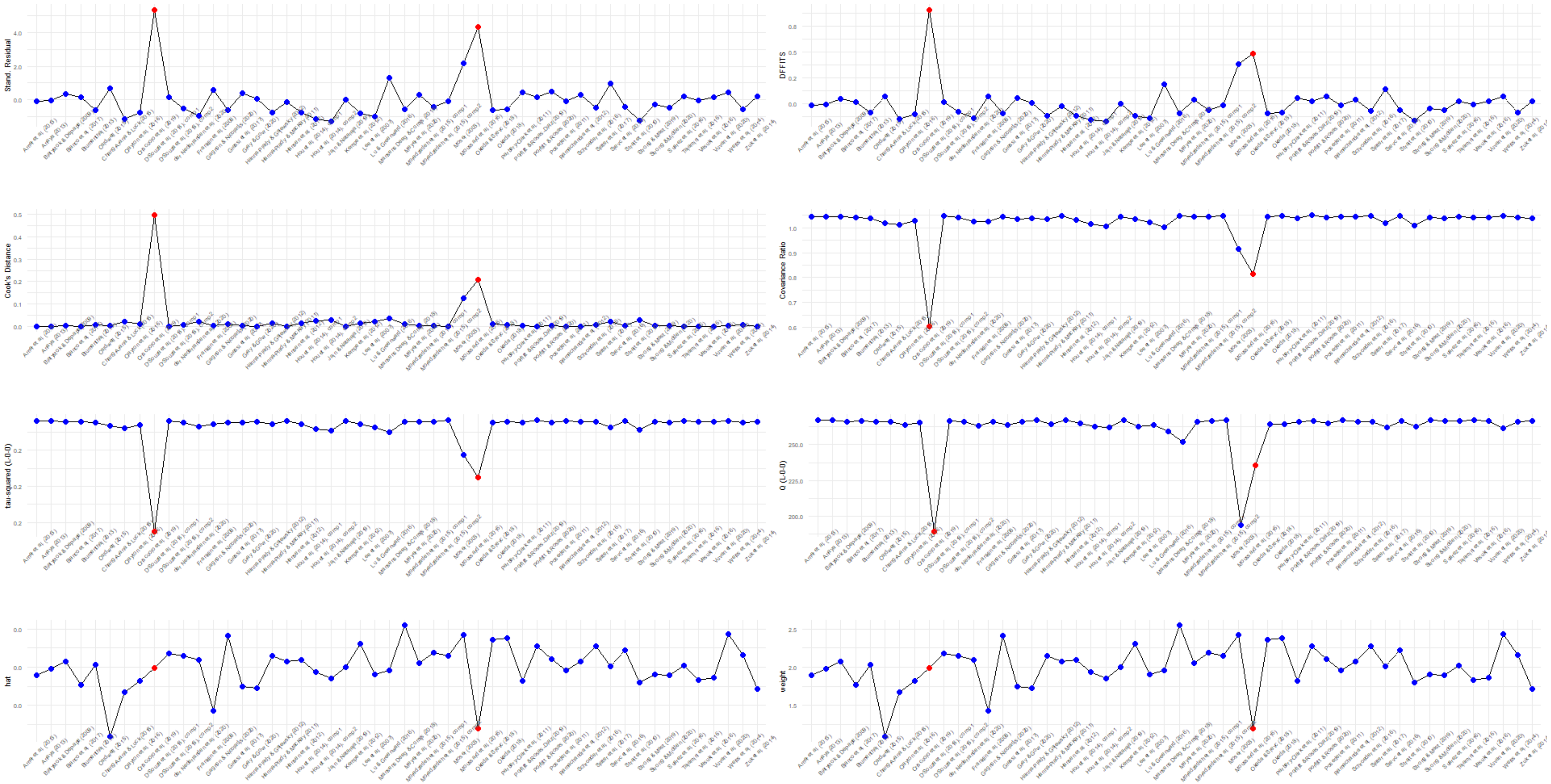


Table A2

The Results of Influence Diagnostics of Inhibition

	rstudent	dffits	cook.d	cov.r	QE.del	hat	weight	infl
Omitting Amer et al. (2013)	-0.063	-0.018	0.000	1.089	202.520	0.038	3.787	
Omitting Anaya (2013)	0.079	0.010	0.000	1.090	202.128	0.038	3.848	
Omitting Bialystok & DePape (2009)	0.432	0.083	0.007	1.081	198.575	0.039	3.919	
Omitting Blumenthal (2013)	-0.288	-0.063	0.004	1.088	202.190	0.039	3.899	
Omitting Clayton et al. (2016)	-0.485	-0.100	0.010	1.077	201.428	0.037	3.727	
Omitting Criscuolo et al. (2019)	5.309	1.234	0.647	0.425	92.958	0.038	3.772	*
Omitting D'Souza et al. (2018)	-0.148	-0.036	0.001	1.094	202.532	0.040	4.010	
Omitting D'Souza et al. (2018).1	-0.333	-0.073	0.006	1.089	201.784	0.040	3.987	
Omitting der Nederlanden et al. (2020)	-0.688	-0.142	0.021	1.069	198.100	0.039	3.938	
Omitting Gray & Gow (2020)	-0.117	-0.028	0.001	1.086	202.547	0.037	3.658	
Omitting Hou et al. (2014)	-0.931	-0.186	0.035	1.048	196.245	0.038	3.811	
Omitting Hou et al. (2014).1	-1.673	-0.320	0.095	0.963	183.296	0.038	3.768	
Omitting Kempe et al. (2012)	-0.507	-0.109	0.012	1.083	198.566	0.041	4.084	
Omitting Meyer et al. (2020)	-0.038	-0.014	0.000	1.093	202.462	0.039	3.921	
Omitting Moussard et al. (2016)	3.484	0.621	0.298	0.732	168.246	0.030	3.042	*
Omitting Okada & Slevc (2018)	-0.208	-0.049	0.003	1.095	202.278	0.041	4.121	
Omitting Porflitt & Rosas-Diaz (2019)	0.071	0.009	0.000	1.095	201.650	0.041	4.068	
Omitting Porflitt & Rosas (2020)	0.375	0.071	0.005	1.085	198.894	0.040	3.959	
Omitting Schroeder et al. (2016)	-0.202	-0.047	0.002	1.094	202.387	0.041	4.062	
Omitting Slevc et al. (2016)	-0.583	-0.124	0.016	1.078	198.013	0.040	4.044	
Omitting Zuk et al. (2014)	-0.119	-0.029	0.001	1.086	202.547	0.037	3.658	
Omitting Bianco et al. (2017)	0.176	0.029	0.001	1.085	201.914	0.037	3.675	
Omitting Chang-Arana & Luck (2018)	-0.690	-0.136	0.019	1.063	200.420	0.036	3.590	
Omitting Hanna-Plady & Gajewsky (2012)	-0.558	-0.118	0.014	1.078	199.435	0.040	3.982	
Omitting Slater et al. (2017)	0.690	0.136	0.019	1.065	195.151	0.039	3.874	
Omitting Vasuki et al. (2016)	-0.334	-0.071	0.005	1.085	202.100	0.038	3.793	

Figure A2

Plots for Influence Diagnostics of Inhibition

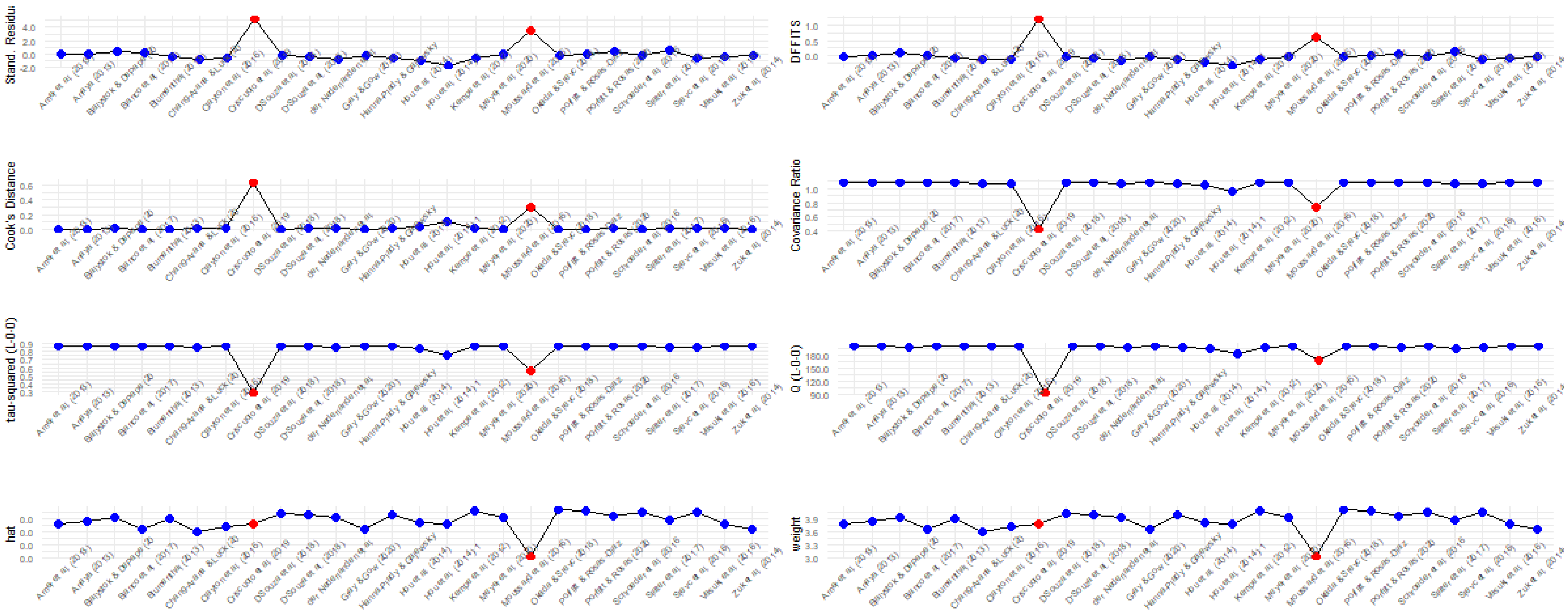


Table A3

The Results of Influence Diagnostics of Shifting

	rstudent	dffits	cook.d	cov.r	QE.del	hat	weight	infl
Omitting Bialystok & DePape (2009)	0.023	0.011	0.000	1.108	120.453	0.054	5.413	
Omitting Clayton et al. (2016)	-0.951	-0.206	0.042	1.049	117.268	0.045	4.476	
Omitting der Nederlanden et al. (2020)	-0.406	-0.092	0.009	1.097	119.136	0.054	5.351	
Omitting Gray & Gow (2020)	0.121	0.030	0.001	1.084	120.567	0.042	4.235	
Omitting Hanna-Plady & MacKay (2011)	0.492	0.120	0.015	1.095	120.217	0.053	5.286	
Omitting Hou et al. (2014)	-0.712	-0.159	0.026	1.071	118.119	0.048	4.844	
Omitting Hou et al. (2014).1	-1.095	-0.240	0.057	1.038	116.255	0.046	4.561	
Omitting Meyer et al. (2020)	1.349	0.311	0.094	1.022	115.994	0.052	5.170	
Omitting Moradzadeh et al. (2015)	-0.065	-0.010	0.000	1.114	120.227	0.057	5.697	
Omitting Moradzadeh et al. (2015).1	0.315	0.081	0.007	1.107	120.505	0.055	5.548	
Omitting Moser (2003)	5.116	1.510	0.748	0.481	30.151	0.065	6.530	*
Omitting Okada & Slevc (2018)	-1.176	-0.310	0.093	1.041	104.217	0.063	6.316	
Omitting Porflitt & Rosas-Diaz (2019)	-0.011	0.003	0.000	1.121	120.270	0.060	6.025	
Omitting Porflitt & Rosas (2020)	0.069	0.022	0.001	1.111	120.502	0.055	5.539	
Omitting Slevc et al. (2016)	-1.172	-0.296	0.086	1.039	110.289	0.059	5.878	
Omitting Strong & Mast (2019)	0.417	0.097	0.010	1.088	120.405	0.047	4.725	
Omitting Strong & Midden (2020)	-0.176	-0.035	0.001	1.092	120.203	0.047	4.705	
Omitting Zuk et al. (2014)	0.398	0.087	0.008	1.077	120.447	0.042	4.151	
Omitting Hanna-Plady & Gajewsky (2012)	-0.600	-0.142	0.021	1.090	117.723	0.055	5.550	

Figure A3

Plots for Influence Diagnostics of Shifting

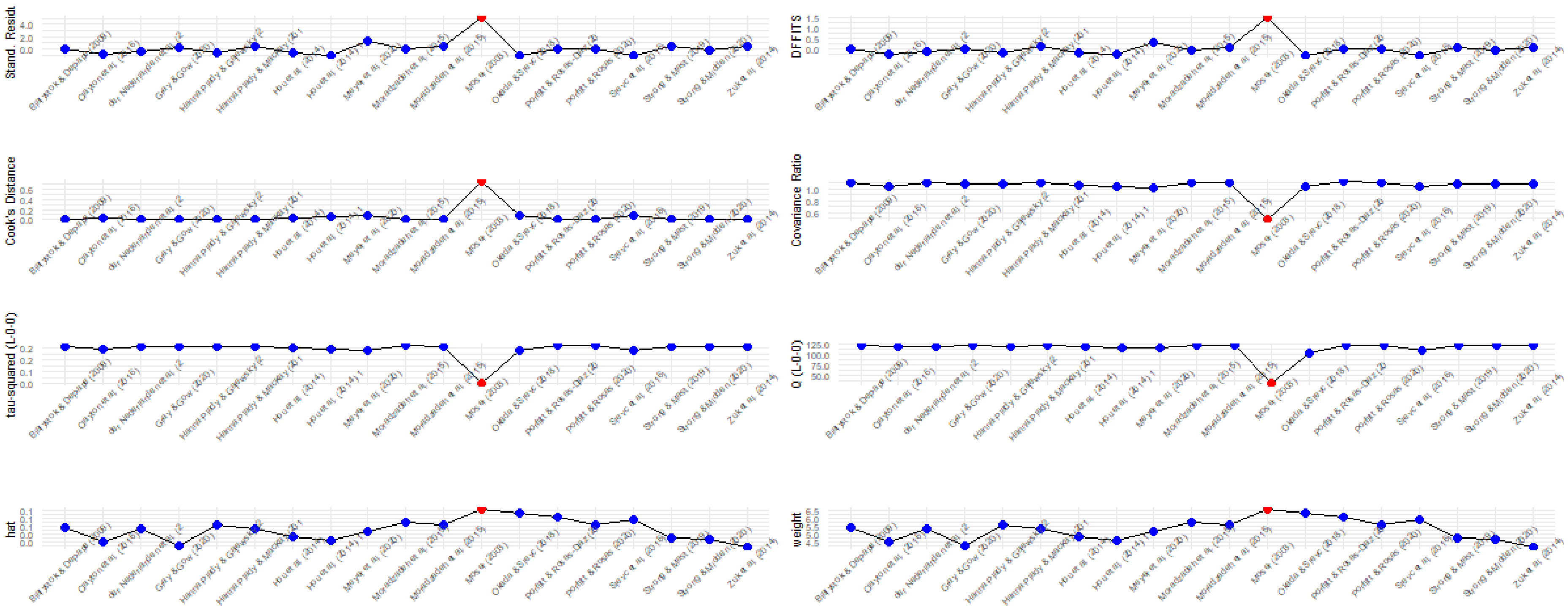


Table A4

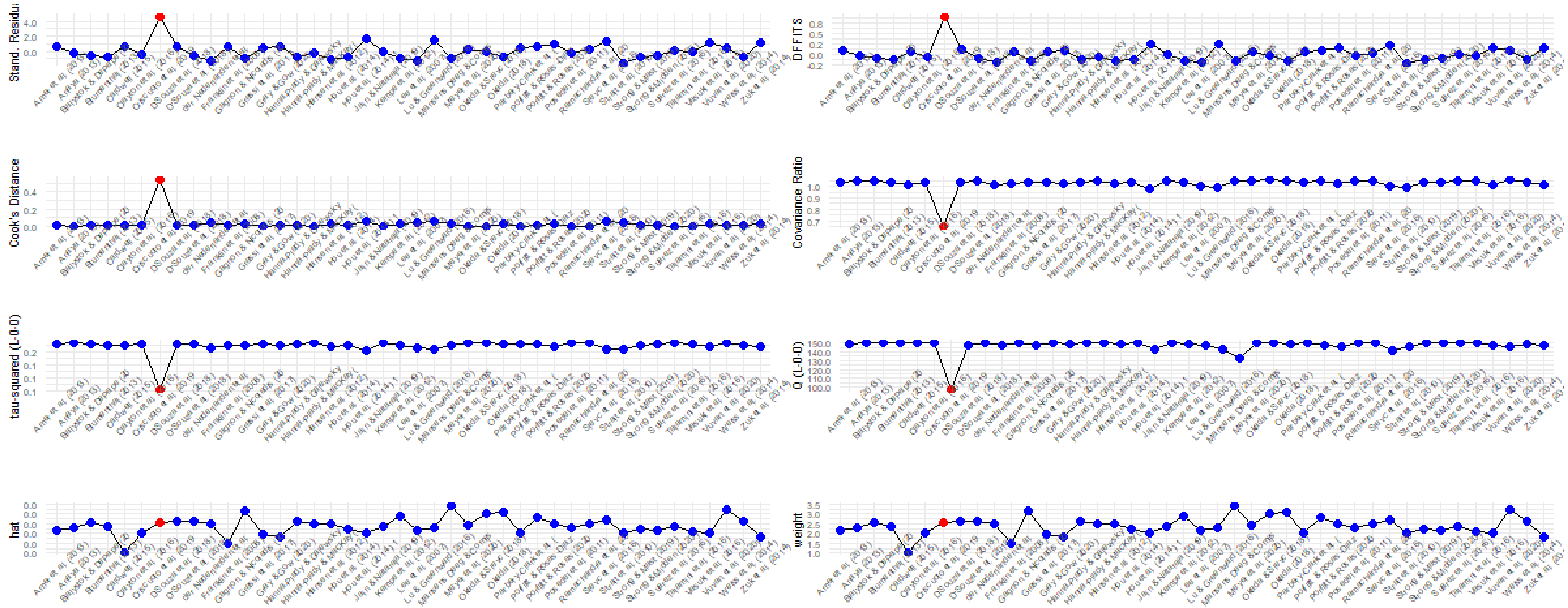
The Results of Influence Diagnostics of Working Memory

	rstudent	dffits	cook.d	cov.r	QE.del	hat	weight	infl
Omitting Amer et al. (2013)	0.540	0.076	0.006	1.037	149.393	0.021	2.136	
Omitting Anaya (2013)	-0.204	-0.037	0.001	1.048	150.761	0.023	2.309	
Omitting Bialystok & DePape (2009)	-0.592	-0.100	0.010	1.045	150.254	0.025	2.549	
Omitting Blumenthal (2013)	-0.847	-0.134	0.018	1.033	149.531	0.024	2.378	
Omitting Caldwell (2015)	0.679	0.068	0.005	1.016	149.826	0.010	1.036	
Omitting Clayton et al. (2016)	-0.419	-0.065	0.004	1.040	150.640	0.020	2.039	
Omitting Criscuolo et al. (2019)	4.676	0.931	0.549	0.654	97.767	0.025	2.532	*
Omitting D'Souza et al. (2018)	0.700	0.112	0.013	1.040	147.471	0.026	2.633	
Omitting D'Souza et al. (2018).1	-0.619	-0.106	0.011	1.045	150.149	0.026	2.607	
Omitting der Nederlanden et al. (2020)	-1.242	-0.197	0.038	1.014	147.231	0.025	2.514	
Omitting Franklin et al. (2008)	0.570	0.067	0.005	1.025	149.849	0.015	1.466	
Omitting Grassi et al. (2017)	0.402	0.053	0.003	1.037	149.997	0.019	1.933	
Omitting Gray & Gow (2020)	0.700	0.094	0.009	1.028	149.152	0.018	1.838	
Omitting Hanna-Plady & MacKay (2011)	-0.326	-0.058	0.003	1.050	150.717	0.025	2.481	
Omitting Hou et al. (2014)	-0.779	-0.121	0.015	1.034	149.862	0.022	2.240	
Omitting Hou et al. (2014).1	1.612	0.241	0.056	0.985	143.350	0.020	2.049	
Omitting Jain & Nataraja (2019)	-0.062	-0.015	0.000	1.049	150.699	0.023	2.343	
Omitting Kempe et al. (2012)	-0.894	-0.157	0.025	1.038	148.286	0.029	2.915	
Omitting Lu & Greenwald (2016)	1.495	0.235	0.054	0.992	143.054	0.023	2.270	
Omitting Mansens Deeg & Comijs (2018)	-0.884	-0.169	0.029	1.045	133.276	0.034	3.449	
Omitting Meyer et al. (2020)	0.317	0.045	0.002	1.048	149.859	0.024	2.441	
Omitting Okada & Slevc (2018)	-0.095	-0.024	0.001	1.064	150.622	0.030	3.033	
Omitting Okada (2018)	-0.824	-0.150	0.023	1.044	148.059	0.031	3.086	
Omitting Porflitt & Rosas-Diaz (2019)	0.637	0.105	0.011	1.046	147.023	0.028	2.830	
Omitting Porflitt & Rosas (2020)	0.995	0.159	0.025	1.024	145.925	0.025	2.490	
Omitting Ramachandra et al. (2012)	0.260	0.036	0.001	1.050	150.025	0.025	2.467	
Omitting Slevc et al. (2016)	1.335	0.229	0.051	1.003	141.020	0.027	2.693	
Omitting Strong & Mast (2019)	-0.834	-0.127	0.016	1.031	149.712	0.022	2.200	
Omitting Strong & Midden (2020)	-0.596	-0.093	0.009	1.038	150.356	0.022	2.163	
Omitting Suarez et al. (2016)	0.066	0.004	0.000	1.050	150.529	0.024	2.385	
Omitting Vuvan et al. (2020)	0.456	0.076	0.006	1.059	145.545	0.032	3.199	
Omitting Zuk et al. (2014)	1.064	0.144	0.021	1.014	147.677	0.018	1.792	
Omitting Lee et al. (2007)	-1.329	-0.194	0.037	1.008	147.547	0.022	2.177	
Omitting Gagnon & Nicoladis (2020)	-0.895	-0.163	0.027	1.041	146.956	0.031	3.139	
Omitting Hanna-Plady & Gajewsky (2012)	-0.781	-0.131	0.017	1.039	149.560	0.026	2.611	
Omitting Hansen et al. (2012)	-1.080	-0.173	0.030	1.024	148.237	0.025	2.511	
Omitting Parbery-Clark et al. (2011)	0.425	0.057	0.003	1.038	149.868	0.020	2.034	
Omitting Posedel et al. (2011)	-0.227	-0.040	0.002	1.047	150.759	0.023	2.267	
Omitting Strait et al. (2010)	-1.587	-0.219	0.047	0.993	146.420	0.020	2.014	

```
Omitting Talamini et al. (2016)          -0.154 -0.028  0.001 1.043 150.757 0.021  2.066
Omitting Vasuki et al. (2016)             1.082  0.158  0.025 1.015 146.972 0.020  2.049
Omitting Weiss et al. (2014)             -0.811 -0.136  0.019 1.038 149.398 0.026  2.635
```

Figure A4

Plots for Influence Diagnostics of Working Memory



Appendix B

Table B1

Average Effect Sizes for the Relation between Music-Making and Executive Functions with Influential Cases

						I^2		τ^2	
Factors	*n*	*k*	*g*	95% CI	BF_{10} [Sensitivity]	Level 2	Level 3	Level 2	Level 3
Inhibition	17/7	41/22	0.51	[0.14 - 0.88]	>100 [48.34, >100]	8.90%	83.53%	0.08	0.74
Shifting	13/5	32/13	0.28	[0.06 - 0.51]	>100 [57.62, >100]	36.05%	43.21%	0.10	0.12
Working memory	25/14	51/76	0.54	[0.41 - 0.67]	>100 [>100, >100]	55.79%	24.39%	0.15	0.06

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 B1

Multilevel Meta-Analysis of Music Making and Inhibition with Influential Cases

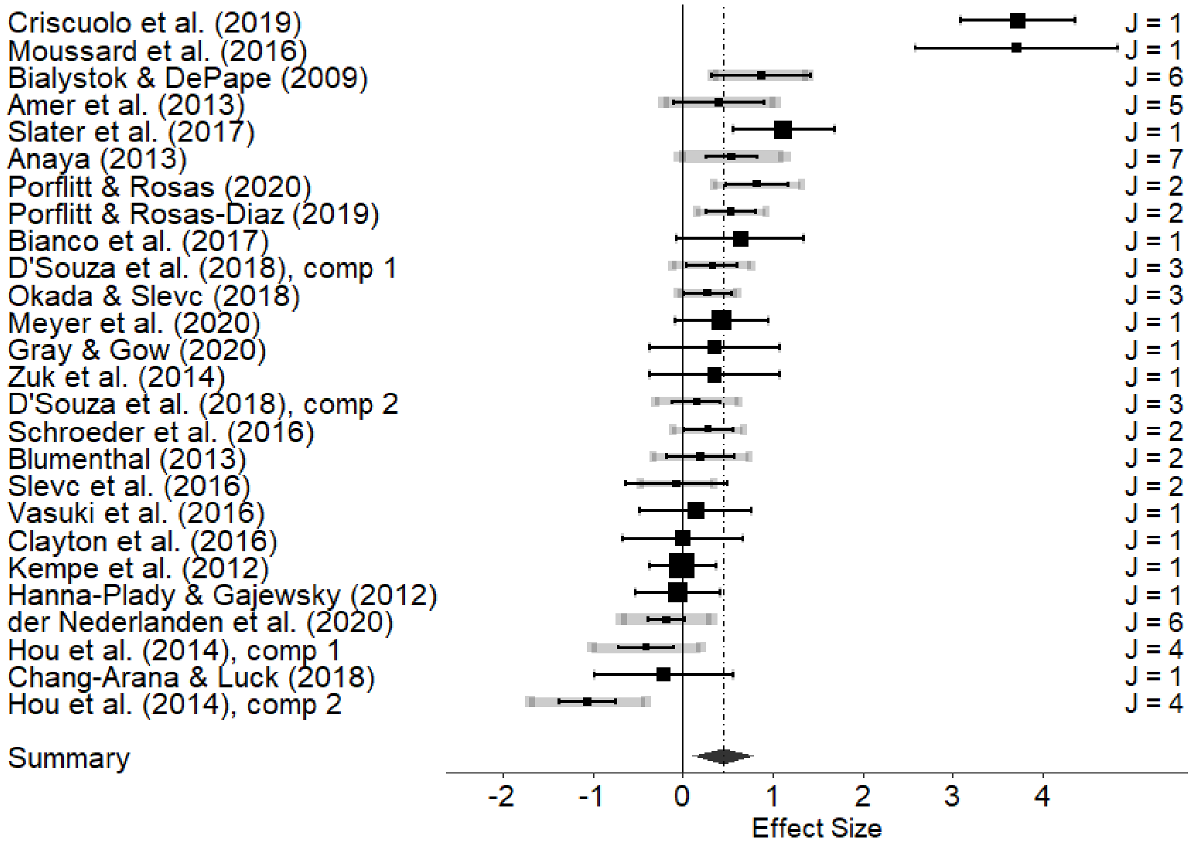


Figure B2

Multilevel Meta-Analysis of Music Making and Shifting with Influential Cases

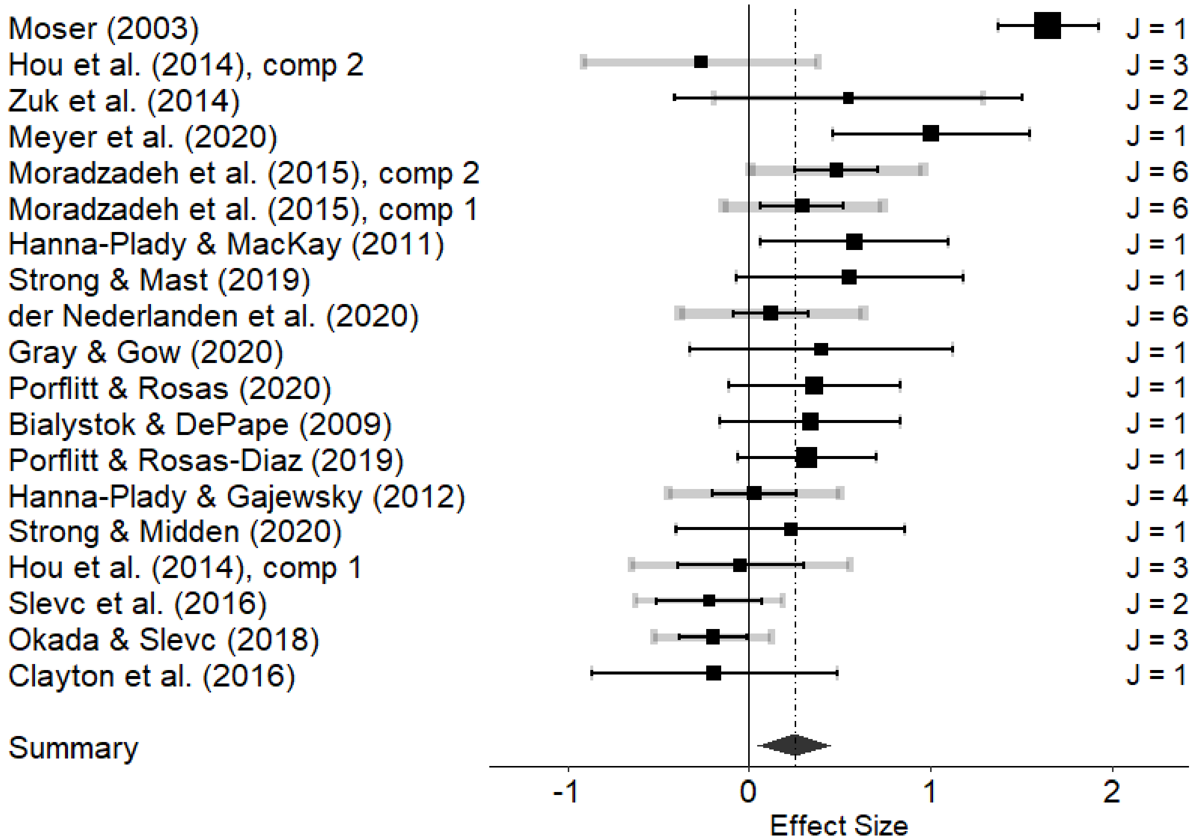
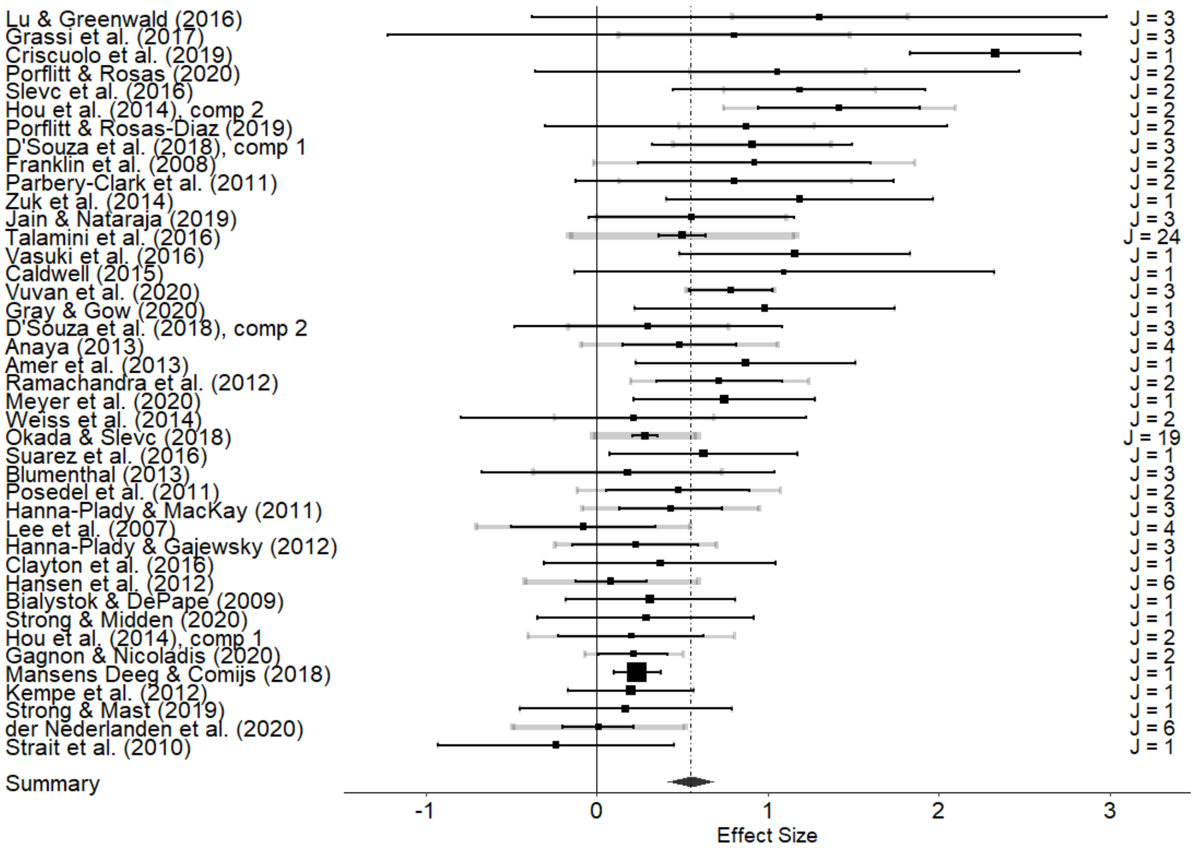


Figure B3

Multilevel Meta-Analysis of Music Making and Working Memory with Influential Cases



Appendix C

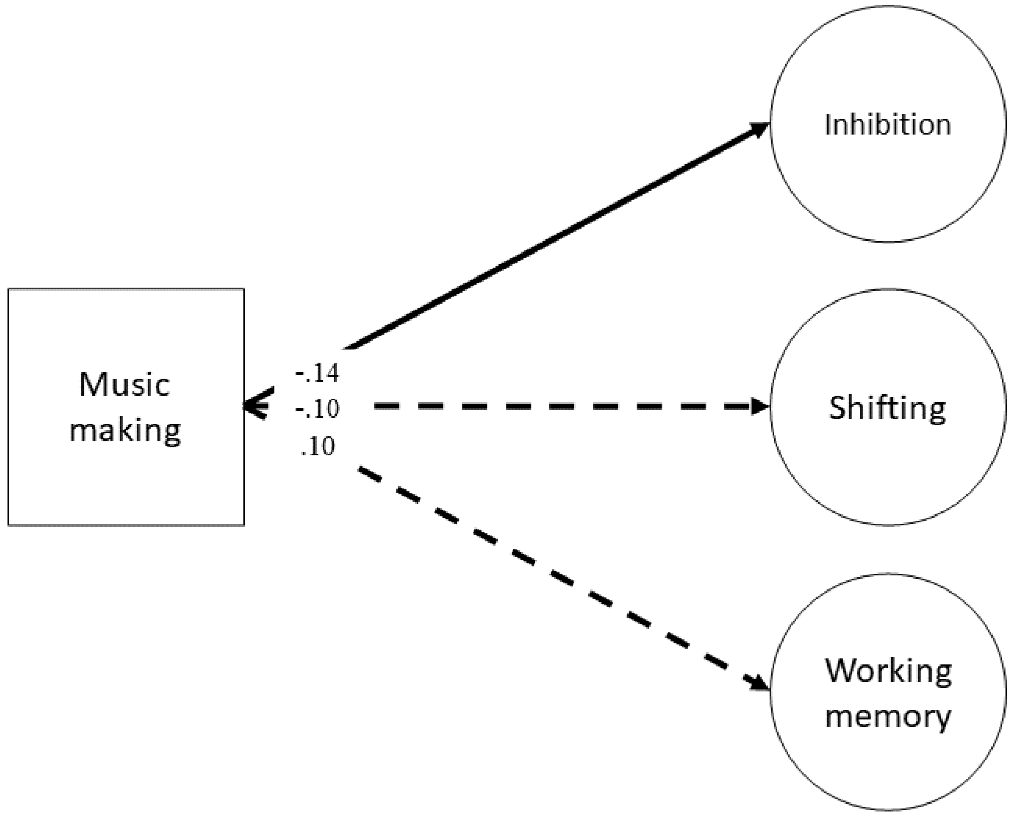
Table C1

Fit Coefficients of Confirmatory Factor Analysis without Covariates

Model	CFI	RMSEA [90% CI]	SRMR	χ^2	df	*p*-value
Three-factor	.878	.10 [.09, .12]	.08	129.67	24	<.001
Bi-factor	.837	.13 [.11, .15]	.12	161.69	21	<.001

Figure C1

Schematic summary of the results of the structural equation modelling without covariates



Note. Significant relationships are indicated by solid arrows, non-significant relationship by dashed arrow.

Table C2

The comparison between high frequency music makers (n = 101) and low frequency music makers (n = 278)

Variable	High frequency		Low frequency		t-test
	M	*SD*	*M*	*SD*	
Simon task	-0.17	0.09	-0.17	0.09	-0.07
Stroop task	-0.18	0.08	-0.16	0.07	-2.38*
Go/No-Go	4.11	0.63	4.28	0.52	-2.63**
Drawing shifting	-0.36	0.16	-0.31	0.14	-2.92**
Shape shifting	-0.34	0.15	-0.29	0.15	-2.63**
Number shifting	-0.32	0.14	-0.30	0.13	-1.17
Continuous reproduction	-57.12	16.93	-61.03	15.26	2.14*
Binding task	1.58	1.01	1.52	0.93	0.54
Updating task	0.95	0.11	0.94	0.12	0.56

Note. *$p < .05$, **$p < .01$, *M* = Mean, *SD* = Standard Deviation.

References

References marked with an asterisk indicate studies included in the Chapter 2 meta-analysis.

Acton, G. S., & Schroeder, D. H. (2001). Sensory discrimination as related to general intelligence. *Intelligence*, *29*(3), 263–271. https://doi.org/10.1016/S0160-2896(01)00066-6

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

Alain, C., Moussard, A., Singer, J., Lee, Y., Bidelman, G. M., & Moreno, S. (2019). Music and Visual Art Training Modulate Brain Activity in Older Adults. *Frontiers in Neuroscience*, *13*, 182. https://doi.org/10.3389/fnins.2019.00182

*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

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and Operationalization of Executive Function. *Mind, Brain, and Education*, *10*(1), 10–33. https://doi.org/10.1111/mbe.12100

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

Bigand, E., & Tillmann, B. (2022). Near and far transfer: Is music special? *Memory & Cognition*, *50*(2), 339–347. https://doi.org/10.3758/s13421-021-01226-6

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.

Brenner, P. S., & DeLamater, J. (2016). Lies, Damned Lies, and Survey Self-Reports? Identity as a Cause of Measurement Bias. *Social Psychology Quarterly*, *79*(4), 333–354. https://doi.org/10.1177/0190272516628298

Burgoyne, A. P., Seeburger, D. T., & Engle, R. W. (2024). Modality matters: Three auditory conflict tasks to measure individual differences in attention control. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02328-6

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

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

*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

Deary, I. J., Bell, P. J., Bell, A. J., Campbell, M. L., & Fazal, N. D. (2004). Sensory Discrimination and Intelligence: Testing Spearman's Other Hypothesis. *The American Journal of Psychology*, *117*(1), 1. https://doi.org/10.2307/1423593

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

Deschamps, I., Courson, M., Dick, A. S., & Tremblay, P. (2020). The phonological loop: Is speech special? *Experimental Brain Research*, *238*(10), 2307–2321. https://doi.org/10.1007/s00221-020-05886-9

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

Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining Reaction Time and Accuracy: The Relationship Between Working Memory Capacity and Task Switching as a Case

Example. *Perspectives on Psychological Science*, *11*(1), 133–155. https://doi.org/10.1177/1745691615596990

*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

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

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

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

*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

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094. https://doi.org/10.1016/j.neuropsychologia.2011.02.001

*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

Guye, S., Röcke, C., Martin, M., & von Bastian, C. C. (2020). Functional Ability in Everyday Life: Are Associations With an Engaged Lifestyle Mediated by Working Memory? *The Journals of Gerontology: Series B*, *75*(9), 1873–1883. https://doi.org/10.1093/geronb/gbz056

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

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. (2020). psychTestR: An R package for designing and conducting behavioural psychological experiments. *Journal of Open Source Software*, *5*(49), 2088. https://doi.org/10.21105/joss.02088

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

Hetland, L. (2000a). Learning to Make Music Enhances Spatial Reasoning. *Journal of Aesthetic Education*, *34*(3/4), 179. https://doi.org/10.2307/3333643

Hetland, L. (2000b). Listening to Music Enhances Spatial-Temporal Reasoning: Evidence for the 'Mozart Effect'. *Journal of Aesthetic Education*, *34*(3/4), 105. https://doi.org/10.2307/3333640

*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

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical Training Shapes Structural Brain Development. *The Journal of Neuroscience*, *29*(10), 3019–3025. https://doi.org/10.1523/JNEUROSCI.5118-08.2009

*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

Jiang, S., Jones, M., & Von Bastian, C. C. (2023). Mechanisms of Cognitive Change: Training Improves the Quality But Not the Quantity of Visual Working Memory Representations. *Journal of Cognition*, *6*(1), 42. https://doi.org/10.5334/joc.306

Jopp, D. S., & Hertzog, C. (2010). Assessing adult leisure activities: An extension of a self-report activity questionnaire. *Psychological Assessment*, *22*(1), 108–120. https://doi.org/10.1037/a0017662

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*(3), 227–229. https://doi.org/10.3758/BF03214307

Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica*, *124*(3), 382–397. https://doi.org/10.1016/j.actpsy.2006.02.005

*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

Klein, K., & Boals, A. (2001). Expressive writing can increase working memory capacity. *Journal of Experimental Psychology: General*, *130*(3), 520–533. https://doi.org/10.1037/0096-3445.130.3.520

Kline, R. B. (2023). *Principles and practice of structural equation modeling*.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Kubaszek, M., Jastrzębski, J., & Chuderski, A. (2021). Auditory, temporal, and visual sensory discrimination advantage of musicians. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43), 3123–3128.

Lad, M., Billig, A. J., Kumar, S., & Griffiths, T. D. (2022). A specific relationship between musical sophistication and auditory working memory. *Scientific Reports*, *12*(1), 3517. https://doi.org/10.1038/s41598-022-07568-8

Law, L. N. C., & Zentner, M. (2012). Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills. *PLoS ONE*, *7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

*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

Loaiza, V. M. (2024). An overview of the hallmarks of cognitive aging. *Current Opinion in Psychology*, *56*, 101784. https://doi.org/10.1016/j.copsyc.2023.101784

Löffler, C., Frischkorn, G. T., Hagemann, D., Sadus, K., & Schubert, A.-L. (2024). The common factor of executive functions measures nothing but speed of information uptake. *Psychological Research*. https://doi.org/10.1007/s00426-023-01924-7

*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

Marie, D., Müller, C. A. H., Altenmüller, E., Van De Ville, D., Jünemann, K., Scholz, D. S., Krüger, T. H. C., Worschech, F., Kliegel, M., Sinke, C., & James, C. E. (2023). Music interventions in 132 healthy older adults enhance cerebellar grey matter and auditory working memory, despite general brain atrophy. *Neuroimage: Reports*, *3*(2), 100166. https://doi.org/10.1016/j.ynirp.2023.100166

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

Nozari, N., & Martin, R. C. (2024). Is working memory domain-general or domain-specific? *Trends in Cognitive Sciences*, S1364661324001645. https://doi.org/10.1016/j.tics.2024.06.006

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Oberauer, K., Suß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*, 167–193. https://doi.org/10.1016/S0160-2896(02)00115-0

*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

Paggi, M. E., Jopp, D., & Hertzog, C. (2016). The Importance of Leisure Activities in the Relationship between Physical Health and Well-Being in a Life Span Sample. *Gerontology*, *62*(4), 450–458. https://doi.org/10.1159/000444415

*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

Pietschnig, J., Voracek, M., & Formann, A. K. (2010). *Mozart effect–Shmozart effect: A meta-analysis*.

Platz, F., Kopiez, R., Lehmann, A. C., & Wolf, A. (2014). The influence of deliberate practice on musical achievement: A meta-analysis. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00646

*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

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00197

*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).

Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, *25*(5), 1695–1716. https://doi.org/10.3758/s13423-017-1384-7

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 501–526. https://doi.org/10.1037/xlm0000450

Rinne, T., Koistinen, S., Salonen, O., & Alho, K. (2009). Task-Dependent Activations of Human Auditory Cortex during Pitch Discrimination and Pitch Memory Tasks. *The Journal of Neuroscience*, *29*(42), 13338–13343. https://doi.org/10.1523/JNEUROSCI.3012-09.2009

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

Rogenmoser, L., Kernbach, J., Schlaug, G., & Gaser, C. (2018). Keeping brains young with making music. *Brain Structure and Function*, *223*(1), 297–305. https://doi.org/10.1007/s00429-017-1491-2

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

Rosseel, Y. (2012). **lavaan**: An *R* Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

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

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. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology*, *102*, 283–302. https://doi.org/DOI:10.1111/j.2044-8295.2010.02000.x

Schellenberg, E. G., & Lima, C. F. (2023). Music Training and Nonmusical Abilities. *Annual Review of Psychology*, *75*, 17.1 - 17.42.

Schmiedek, F., Li, S.-C., & Lindenberger, U. (2009). Interference and facilitation in spatial working memory: Age-associated differences in lure effects in the n-back paradigm. *Psychology and Aging*, *24*(1), 203–210. https://doi.org/10.1037/a0014685

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

Schwärzler, T. (2015). *Der Einfluss von Bilingualismus auf die Arbeitsgedächtnisleistung: Ein Vergleich von tonalen und nicht-tonalen Sprachen.* University of Zurich.

Seashore, C. E. (1915). The Measurement of Musical Talent. *The Musical Quarterly*, *I*(1), 129–148. https://doi.org/10.1093/mq/I.1.129

Sherman, A., Grabowecky, M., & Suzuki, S. (2015). In the working memory of the beholder: Art appreciation is enhanced when visual complexity is compatible with working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 898–903. https://doi.org/10.1037/a0039314

Silas, S., Müllensiefen, D., Gelding, R., Frieler, K., & Harrison, P. M. C. (2022). The Associations Between Music Training, Musical Working Memory, and Visuospatial Working Memory. *Music Perception*, *39*(4), 401–420. https://doi.org/10.1525/mp.2022.39.4.401

*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

Spearman, C. (1904). 'General Intelligence,' Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201–292.

*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

Swaminathan, S., & Schellenberg, E. G. (2019). Music Training and Cognitive Abilities: Associations, Causes, and Consequences. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford Handbook of Music and the Brain* (pp. 644–670). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198804123.013.26

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

Troche, S. J., Wagner, F. L., Voelke, A. E., Roebers, C. M., & Rammsayer, T. H. (2014). Individual differences in working memory capacity explain the relationship between general discrimination ability and psychometric intelligence. *Intelligence*, *44*, 40–50. https://doi.org/10.1016/j.intell.2014.02.009

Tsukahara, J. S., Harrison, T. L., Draheim, C., Martin, J. D., & Engle, R. W. (2020). Attention control: The missing link between sensory discrimination and intelligence. *Attention, Perception, & Psychophysics*, *82*(7), 3445–3478. https://doi.org/10.3758/s13414-020-02044-9

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

Voelke, A. E., Troche, S. J., Rammsayer, T. H., Wagner, F. L., & Roebers, C. M. (2013). Sensory Discrimination, Working Memory and Intelligence in 9-Year-Old and 11-Year-Old Children. *Infant and Child Development*, *22*(5), 523–538. https://doi.org/10.1002/icd.1803

Voelke, A. E., Troche, S. J., Rammsayer, T. H., Wagner, F. L., & Roebers, C. M. (2014). Relations among fluid intelligence, sensory discrimination and working memory in middle to late childhood – A latent variable approach. *Cognitive Development*, *32*, 58–73. https://doi.org/10.1016/j.cogdev.2014.08.002

von Bastian, C. C., Belleville, S., Udale, R. C., Reinhartz, A., Essounni, M., & Strobach, T. (2022). Mechanisms underlying training-induced cognitive change. *Nature Reviews Psychology*, *1*(1), 30–41. https://doi.org/10.1038/s44159-021-00001-3

von Bastian, C. C., Hyde, E. R. A., & Jiang, S. (2024). Tackling cognitive decline in late adulthood: Cognitive interventions. *Current Opinion in Psychology*, *56*, 101780. https://doi.org/10.1016/j.copsyc.2023.101780

von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*(1), 108–115. https://doi.org/10.3758/s13428-012-0224-y 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

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*(3), 188–196. https://doi.org/10.1016/j.lindif.2010.02.004

Wang, L. (2022). Music Aptitude, Training, and Cognitive Transfer: A Mini-Review. *Frontiers in Psychology*, *13*, 903920. https://doi.org/10.3389/fpsyg.2022.903920

*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