The effects of musical instrument training on fluid intelligence and executive functions in healthy older adults: A systematic review and meta-analysis

Fionnuala Rogers*, Claudia Metzler-Baddeley
Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Maindy Road, Cardiff University, Cardiff, United Kingdom

**Abstract**

Intervention studies combining cognitive and motor demands have reported far-transfer cognitive benefits in healthy ageing. This systematic review and meta-analysis evaluated the effects of music and rhythm intervention on cognition in older adulthood. Inclusion criteria specified: 1) musical instrument training; 2) healthy, musically-naïve adults (≥60 years); 3) control group; 4) measure of executive function. Ovid, PubMed, Scopus and the Cochrane Library online databases were searched in August 2023. Data from thirteen studies were analysed (N = 502 participants). Study quality was assessed using the Cochrane Risk of Bias tool (RoB 2; Sterne et al., 2019). Random effects models revealed: a low effect on inhibition (d = 0.27, p = .0335); a low-moderate effect on switching (d = -0.39, p = .0021); a low-moderate effect on verbal category switching (d = 0.39, p = .0166); and a moderate effect on processing speed (d = 0.47, p < .0001). No effect was found for selective visual attention, working memory, or verbal memory. With regards to overall bias, three studies were rated as “high”, nine studies were rated as having “some concerns” and one was rated “low”. The meta-analysis suggests that learning to play a musical instrument enhances attention inhibition, switching and processing speed in ageing.

1. Introduction

An increasingly ageing population has prompted a vast body of research on cognitive training programmes with the aim to maintain or restore cognitive functions in later adulthood (Corbett et al., 2015; Melby-Lervåg et al., 2016; Jaeggi et al., 2008; Holmes et al., 2019). Typical cognitive ageing is characterised by a slowing of “fluid” intelligence, which refers to the ability to process and learn new information, generate new solutions, and interact with novel environments (Harada et al., 2013), compared with “crystallised” intelligence, which refers to familiar skills and knowledge that accumulate throughout the lifespan (e.g., vocabulary and general knowledge; Park et al., 2001). Age-related decline in fluid abilities hampers other cognitive functions, notably executive functions (EF), i.e., higher-level cognitive processes involved in planning, problem-solving, and multi-tasking abilities, that largely (but not exclusively) rely on fluid abilities (Roca et al., 2014).

Fluid intelligence and EF have received particular attention within the cognitive training literature of ageing because they play an important role in an individual’s competence to live independently, carry out daily functions, have a good quality of life and engage in social and meaningful activities (Diamond & Ling, 2016; Salthouse, 2012). For example, EF have been found to be significantly related to daily functioning (Vaughan & Giovanello, 2010) as measured with the Instrumental Activities of Daily Living (IADL; Fillenbaum, 1985), notably to performance-based rather than self-report IADL.

Thus, the primary focus of this meta-analysis was on the assessment of training effects on cognitive domains that are known to recruit fluid abilities and to be impacted by age, notably EF but also processing speed, visuo-spatial attention, and episodic memory. Processing speed refers to time taken to receive, process and respond to information in the environment (Salthouse, 2000). This ability begins to decline from midlife and continues to decrease with age, and a reduction in processing speed is a contributing factor to declines in other cognitive domains (Salthouse, et al., 2003, 2009), and is also associated with the need for assistance in daily activities (Bezdicek et al., 2016). Visuo-spatial attention refers to the ability to identify specific environmental information and ignore irrelevant information, and is impacted in ageing (Lezak et al., 2012). Episodic memory refers to recollection of events or information of a specific time and place (Harada et al., 2013). Episodic memory shows lifelong decline (Rønnlund, et al., 2005), and is

* Corresponding author.
E-mail address: RogersF2@cardiff.ac.uk (F. Rogers).

https://doi.org/10.1016/j.bandc.2024.106137
Received 19 September 2023; Received in revised form 29 January 2024; Accepted 29 January 2024
Available online 9 February 2024
0278-2626/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
often measured in terms verbal memory, which is an individual’s ability to encode, store and recall or recognize verbal information within a test session.

Fluid intelligence and EF abilities are thought to be negatively affected by ageing because they rely on optimal functioning of the prefrontal cortex (PFC) and its connected regions. According to the retrogenesis (first in last out) hypothesis (Raz, 2005), the PFC, as one of the late-maturing brain regions, is particularly susceptible to age-related grey and white matter atrophy (Resnick et al., 2003; Salat et al., 2004; Douaud et al., 2014). PFC structure and function (Liston et al., 2009; Cerqueira et al., 2007) and hence EF and fluid abilities, are also disproportionately impacted when a person is experiencing adverse life events, feeling sad, lonely, or in poor physical health (Diamond & Ling, 2016).

Attention switching/divided attention, inhibitory control of distracting information, and updating of information stored in working memory have been proposed as core EF components (Miyake et al., 2001; Spieler et al., 1996; Braver & West, 2011). In a large-scale longitudinal study of 354 number of participants between the ages of 10–46 years old, age-related EF decline was shown to commence in middle adulthood, with working memory capacity and inhibitory control showing reductions as early as the third decade (Ferguson et al., 2021).

Previous attempts to enhance EF (most commonly working memory capacity) were often targeted through computerised training formats. Commercially available “brain training” programmes such as Lumosity (Lumsos Labs Inc., San Francisco), Cogmed (Klingberg et al., 2005) and BrainHQ (Posit Science) have been widely used in an attempt to reduce or prevent cognitive decline in older adults. However, there is now a general consensus among researchers that such computerised training programmes have limited transfer effects (Holmes et al., 2019; Melby-Lervåg et al., 2016; Kelly et al., 2013; Dahlén et al., 2008), effect sizes of benefits on global cognition are low, and there is little evidence for long-term benefits (Gates et al., 2020). Computerised programmes may be unsuccessful because they tend to be unimodal or single-domain (Suctcliffe et al., 2020), and occur in “impoverished” computer learning environments, which do not mimic the complexity of real-world problems and therefore lack ecological validity (Moreau & Conway, 2014). Furthermore, computerised trainings tend to encourage single-domain strategies that cannot be transferred to other problems (i.e., do not encourage fluid abilities).

It has been argued that the brain has evolved for movement, rather than for thinking alone, and that EF have evolved as an extension of the motor control system to facilitate optimal interaction with the environment (Koziol et al., 2012). Besides the PFC and connected cortical parietal areas, the cerebellum and basal ganglia, which have been traditionally associated primarily with motor control, are now widely recognised as playing an important role in EF (Stoodley, 2012; Strick et al., 2009; Baillieux et al., 2008; Hazy et al., 2006, 2007). Thus, the learning of new cognitive-motor skills which require a combination of multi-sensory-motor integration and interaction between complex cognitive processes may provide a more ecologically valid and hence a more effective approach in delivering far transfer effects than computerised programmes. This may be because such an approach may mimic the complex and fluid planning and problem-solving abilities needed in everyday life. In response, there have been calls for more “ecological” and holistic approaches to cognitive training which combine both cognitive and motor demands in real-life settings (Moreau & Conway, 2014; Bernard & Seidler, 2014; Diamond & Ling, 2016).

1.1. Music-based approaches to cognitive training

Accumulating evidence suggests that music-based approaches to cognitive intervention, that combine complex cognitive and motor demands, may hold potential as interventions to enhance EF. Therapeutic instrumental music performance (TIMP; Elliott, 1982), for instance, is a form of neurologic music therapy (NMT) that uses musical instruments to exercise movement patterns. In stroke rehabilitation, TIMP-based therapies have been shown to improve motor abilities and EF (mental flexibility) (Rodriguez-Fornells et al., 2012; Haire et al., 2021; Koshimori & Thaut, 2019).

Rhythmic auditory stimulation (RAS) interventions, which use metronomes or music as external triggers for movements (Koshimori & Thaut, 2023), have been used to improve motor symptoms and quality of life in Parkinson’s disease patients (Pantelyat et al., 2016; see Burrai et al., 2021; Lee & Ko, 2023 for reviews), and in stroke patients (Street et al., 2020). There is also preliminary evidence in Huntington’s disease (HD), that rhythmic stimulation (drumming) may improve EF and white matter microstructure in the genu of the corpus callosum which connects the PFC of both hemispheres (Metzler-Baddeley et al., 2014) and in connections between the right Supplementary Motor Area (SMA) and the putamen (Casella et al., 2020). A systematic review found that benefits to global cognition were associated with microstructural changes in HD following rhythm training (Schwartz et al., 2019). The key mechanism of these rhythmic-based trainings which led to cognitive enhancement is thought to be the stimulation of basal ganglia and cerebellar networks and their connections to fronto-parietal regions associated with EF (Casella et al., 2020).

Playing a musical instrument is considered a complex multimodal task which requires the integration and synchronization of auditory, visual, motor, and somatosensory information and places high demands on fluid EF processes (Herholz & Zatorre, 2012). Learning to play a musical instrument, such as the piano, also involves learning to read musical notation, known as sight-reading. Sight-reading necessitates the coordination of bimanual fine motor responses based on complex visual stimuli, the simultaneous monitoring of auditory information produced, and the adjustment of motor responses based on auditory and somatosensory feedback (Jäncke, 2009). Furthermore, playing music involves modulation of neuronal plasticity via the reward circuitry, including the basal ganglia (Herholz & Zatorre, 2012), which involves neurochemistry which may reinforce learning (Chanda & Levitin, 2013; Ferreri et al., 2019). Multi-sensory integration, EF demands, and reward feedback may all be important contributors to neuroplastic effects observed due to training (Herholz & Zatorre, 2012). Taken together, learning to read and play music could be an optimal task for targeting feedforward mechanisms associated with cortico-cerebellum network involved in EF.

All of these mechanisms may contribute to differences in cognition and brain structure and function that have been reported between musicians and non-musicians, which are reviewed elsewhere (Herholz & Zatorre, 2012, Jäncke, 2009; Schlaug, 2001). Notably, cross-sectional studies suggest that learning to play an instrument can enhance cognition in later life (e.g., Verghez et al., 2003). In a correlational study, older musicians were found to perform better in nonverbal memory as well as naming and executive processing tasks compared with non-musicians whilst controlling for age, education, history of physical exercise, age of instrument acquisition and number of years of formal training (Hanna-Pladdy & MacKay, 2011). Further, Mousard et al. (2016) found that older musicians out-performed a group of age- and education-matched non-musicians on a Go/NoGo task, with task accuracy linked with N2 amplitude measured with electroencephalography (EEG) in musicians, which suggests an executive control advantage. In addition to the variability of the definition used for “musician” across these studies (e.g., various degrees of expertise, professionalism, practice hours, and age of commencement), cross-sectional research presents challenges in establishing cause-and-effect relationships due to possible confounding effects such as education, socio-economic status and amount of music experience. This systematic review and meta-analysis was conducted because recent years have seen an increase in randomised controlled intervention studies investigating the potential of short- or long-term musical training interventions to benefit cognition in older music novices.
1.2. Previous meta-analytical work and the current study

A recent systematic review and meta-analysis conducted by Ma et al., (2023) investigated the effects of interventions involving rhythmic exercise or physical activities performed to music, (e.g., dancing), on physical and cognitive function. They concluded that such rhythmic movement interventions improved physical function, global cognition, and quality of life in healthy older adults. However, no effect was found on executive function. The meta-analysis included only three rhythmic music studies, and did not include any studies using TIMP or learning to play an instrument. A key difference between exercise to rhythm interventions and learning to play a musical instrument is the manipulation of an object using fine or gross motor control, and the consequential auditory and somatosensory feedback which can continuously shape motor behaviour (Herholz & Zatorre, 2012). As outlined above, taxing complex sensory-motor integration and feedback loops may be a crucial training component for enhancing EF.

A previous meta-analysis on musical instrument training suggested that learning to play the piano may protect against cognitive decline in ageing (Roman-Caballero et al., 2018). A subsequent meta-analysis conducted on older adults and adults with mild cognitive impairment (MCI) concluded that musical training has promise as a cognitive intervention to deliver far-transfer benefits in processing speed, attention control and working memory capacity in healthy older adults due to active, rather than receptive engagement involved in the learning of a new instrument (Kim & Yoo, 2019). It was also found that cognitive demands varied depending on the instrument trained. Percussion interventions which emphasized improvisation and multitasking were categorized as “immediate involvement” and were found to be more common in populations with MCI. “Sustained engagement” interventions involved playing the piano and the application of memo-"rized information (music theory) to concurrent reading and playing of a score. However, at time of publishing, only four intervention studies in healthy older adults were available for inclusion in these two meta-analytic papers (Roman-Caballero et al., 2018; Kim & Yoo, 2019).

The current systematic review and meta-analysis is timely and novel as recently larger-scale, higher quality studies in older adults without any diagnosis of cognitive impairment have been published, which provide sufficiently large sample sizes to review the impact of learning to play an instrument on EF in healthy older adults. EF were chosen as the primary measures of interest (specifically, switching, inhibition and working memory capacity and updating) due to the important role they play in daily functioning and the hypothesized underlying neural links between motor skill learning and EF outlined above. Other fluid abilities that are negatively impacted by age were also examined: processing speed was analyzed given that age-related response slowing tends to impact performance in other cognitive domains (Saltzhouse, 2010; Revie & Metzler-Baddeley, 2023) and previous meta-analytic evidence suggests that musical training may help improve processing speed (Kim & Yoo, 2019). Visuospatial attention and verbal memory were also examined given the potential for transfer of playing an instrument to these domains as evinced by cross-sectional research (Sluming et al., 2007; Franklin et al., 2008).

2. Method

2.1. Search strategy and Inclusion criteria

This systematic review and meta-analysis was conducted in line with recommendations from Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA, Page et al., 2021). A completed PRISMA checklist is available in the Supplementary Material. Electronic databases Ovid, PubMed, Scopus and the Cochrane Library were searched for intervention studies involving training on a musical instrument in healthy older adults, published in English. References lists were also hand-searched for relevant studies. Inclusion and exclusion criteria for studies to be included in the meta-analysis are presented in Table 1.

The following search equation was used in the databases search in August 2023: (ageing OR aging OR older*) AND (music training OR piano training OR keyboard training OR music intervention) AND (cogniti* OR executive function*) without any time restrictions. The review was not pre-registered. EndNote referencing software was used to organise the literature from the four databases and to remove duplicates. After 621 duplicates were removed, titles and abstracts of the remaining 1,581 articles were screened by the first author (FR) to exclude articles which did not meet the inclusion criteria (Fig. 1).

Fifteen studies met inclusion criteria for meta-analyses. Where information was missing from a paper, corresponding authors were emailed to request the data (MacRitchie et al., 2020; Bugos et al., 2022; Lister et al., 2023). Two studies could not be included because descriptive statistics were unavailable (Lister et al., 2023; Bugos et al., 2022). Data were collected from eligible studies by FR and transferred into an Excel spreadsheet. See Table 2 for study design information. The domains of EF (switching, inhibition, working memory capacity and updating), verbal fluency, processing speed, visuospatial attention, and verbal memory were investigated. With regards to verbal fluency, three sub-domains were investigated: letter fluency, category fluency and category switching. Letter fluency refers to the ability to generate words beginning with a given letter within a time limit (e.g., words beginning with “F”). Category fluency refers to the number of words generated which belong to a given category (e.g., supermarket items). Category switching refers to the number of words generated when alternating between two given categories (e.g., alternating between vegetables and furniture items each time a word is spoken). Verbal memory was examined in three ways: 1) total number of words recalled after five trials where the same word-list was presented for each trial; 2) immediate/short-delay recall, where verbal information was tested shortly after it was presented or 3) long-delay recall, where verbal information was requested following an approximate 20-minute delay. Specific tests used to assess EF, processing speed, visuospatial attention and verbal memory in each study are reported in Table 3.

Table 1. Inclusion and exclusion criteria for studies to be entered in meta-analysis.

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention studies which required the production of music using a pitched and/or percussion instrument(s)</td>
<td>Music-based interventions which did not require training in a pitched or percussion instrument (e.g., singing, dancing or exercise to music interventions)</td>
</tr>
<tr>
<td>Cognitively healthy adults as assessed by screening for cognitive impairment and/or neurological disorders over the age of 60</td>
<td>Studies that exclusively examined participants with diagnoses of cognitive and/or neurological impairment or younger adults, adolescents or children</td>
</tr>
<tr>
<td>In addition to the above, participants who were musically-naïve, defined as having &lt; 3 years of formal training on a musical instrument and were not involved in any other music-based activities</td>
<td>Inclusion of participants with &gt; 3 years formal music training or those who were actively involved in regular music-based activities</td>
</tr>
<tr>
<td>Inclusion of a control group (either RCT or matched control groups)</td>
<td>Correlational studies or within-subjects designs without a comparative control</td>
</tr>
<tr>
<td>Inclusion of at least one outcome measure of EF from a test designed to measure one of the following domains: distractor inhibition (Stroop test); attention switching (Trail-Making Test Part B); working memory capacity (digit span tests) and updating (N-back); verbal fluency (letter, category, and category switching tests)</td>
<td>Lack of EF outcome measure</td>
</tr>
</tbody>
</table>

EF, executive functions; RCT, randomised controlled trial.
2.2. Risk of bias assessment

The Cochrane Risk of Bias Tool (RoB2; Sterne et al., 2019) was used to assess the validity of the included studies. RoB2 requires the reviewer to answer a series of signalling questions in order to elicit information about specific study features which are relevant regarding risk of bias, and then provides risk of bias judgment which is generated by a computer algorithm. Studies are evaluated as “low”, having “some concerns” or “high” on the following six criteria: random sequence generation; allocation concealment; blinding of participants; blinding of outcome assessment; incomplete outcome data; and selective outcome reporting. Overall study bias was judged as “low”, “some concerns” or “high” based on combined scores of these six categories. Judgements reported in this meta-analysis were based on evaluations by FR, combined with RoB2 algorithm. Overall bias was determined as low only when all six domains were scored as low.

2.3. Statistical analysis

Random-effects models were chosen for the meta-analyses because they control for differences in study methods and sample characteristics, which could introduce heterogeneity among the true effects (Field & Gillett, 2010; Viechtbauer, 2010). This was relevant given the heterogeneous intervention parameters across the thirteen studies (see Table 2).

Effect sizes were calculated from the descriptive statistics reported in each study. Pooled Cohen’s $d$ estimates were analysed for each of the cognitive domains measured using the metafor package in R (Viechtbauer, 2010). Effect sizes were considered as follows: low (0.2–0.4); moderate (0.5–0.7); or large (>0.8) (Cohen, 2013). Heterogeneity across studies was calculated using the $I^2$ statistic (Higgins & Thompson, 2002), and was determined as low, medium or high at 25, 50, or 75 %, respectively.

Additional exploratory analyses were undertaken for each cognitive domain. A “leave-one-out” sensitivity analysis (Viechtbauer & Cheung, 2010), whereby one study was removed from the analysis at a time, whilst pooled effect size was recalculated, was undertaken to determine the influence of each study on overall effect size for each cognitive domain.

Cohen’s $d$ effect sizes were also calculated for the two individual studies which reported three-month follow-up data (Bugos et al., 2007; Bugos & Wang, 2022).

Fig. 1. PRISMA flowchart of screening strategy for studies to be included in the meta-analysis.


### Table 2

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Study Design</th>
<th>N (Int/ Con)</th>
<th>Mean Age (Int/ Con)</th>
<th>Cognitive Screening tool</th>
<th>Intervention Type</th>
<th>Control Group(s)</th>
<th>Intervention intensity, frequency, duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>Bugos et al.</td>
<td>RCT</td>
<td>16/15</td>
<td>71.4/ 69.6</td>
<td>TICS</td>
<td>Individualised piano training</td>
<td>No treatment</td>
<td>30 min, 1/week, 6 months (min. 3 h practice per week)</td>
</tr>
<tr>
<td>2010</td>
<td>Bugos</td>
<td>Age- and IQ-matched active control</td>
<td>24/22</td>
<td>69.3/ 67.7</td>
<td>MMSE</td>
<td>Group piano playing</td>
<td>Music Listening Intervention</td>
<td>45 min, 1/week, 16 weeks; (30 mins daily practice)</td>
</tr>
<tr>
<td>2013</td>
<td>Seinfeld et al.</td>
<td>Cluster sampling – Active control</td>
<td>13/16</td>
<td>69.3/ 69.6</td>
<td>MMSE</td>
<td>Group piano playing</td>
<td>Other leisure activities</td>
<td>90 min, 1/week, 4 months; 45 mins practice 5 days per week (min. 4 h practice per week)</td>
</tr>
<tr>
<td>2015</td>
<td>Thorne</td>
<td>RCT</td>
<td>10/10</td>
<td>71.5/ 71.7</td>
<td>TICS</td>
<td>Group piano playing</td>
<td>Music Listening intervention</td>
<td>30 min, 1/week, 6 months (30 mins practice per day)</td>
</tr>
<tr>
<td>2017</td>
<td>Bugos &amp; Kocher</td>
<td>Within-subjects (Subjects as own control)</td>
<td>34</td>
<td>70.79</td>
<td>TICS</td>
<td>Short-term, intense piano training</td>
<td>Subjects as own control (pre-test and then second pre-test after 2-week period)</td>
<td>3 h per day for 2 weeks (total 30 h)</td>
</tr>
<tr>
<td>2018</td>
<td>Biasutti &amp; Mangiacotti RCT</td>
<td>(participants with MCI also included)</td>
<td>18/17</td>
<td>83.39/ 83.76</td>
<td>MMSE</td>
<td>Cognitive music training</td>
<td>Gymnastic activities</td>
<td>12 bi-weekly 70 min</td>
</tr>
<tr>
<td>2018</td>
<td>Dégé &amp; Korkievís RCT</td>
<td></td>
<td>8/7/9</td>
<td>77.9/ 76.6/ 77.5</td>
<td>MMSE</td>
<td>Group drumming and singing training</td>
<td>Literature programme; no treatment</td>
<td>60 min, 1/week, 15 weeks</td>
</tr>
<tr>
<td>2019</td>
<td>Bugos</td>
<td>Age and IQ-matched active control groups</td>
<td>49/ 38/48</td>
<td>69.13/ 68.83</td>
<td>TICS</td>
<td>Group piano playing</td>
<td>Percussion group; Music listening course</td>
<td>45 min 1/week, 16 weeks (30 min practice per day or 3 h per week)</td>
</tr>
<tr>
<td>2020</td>
<td>Guo et al.</td>
<td>Pseudorandomized (fMRI)</td>
<td>27/26</td>
<td>73.3/ 72.9</td>
<td>MMSE</td>
<td>Group piano training: ensemble playing, Figure notes</td>
<td>No treatment</td>
<td>60 min, 1/week, 4 months</td>
</tr>
<tr>
<td>2020</td>
<td>MacRitchie et al. RCT</td>
<td></td>
<td>8/7</td>
<td>70.9</td>
<td>ACE-III</td>
<td>Group piano training: ensemble playing, Figure notes</td>
<td>Waitlist (no treatment)</td>
<td>60 min, 1/week, 10 weeks (30 mins practice per day)</td>
</tr>
<tr>
<td>2021</td>
<td>Dos Santos et al. RCT</td>
<td></td>
<td>15/12</td>
<td>68.4/ 67.3</td>
<td>MMSE</td>
<td>Improvisation and percussion activities</td>
<td>Choir singing</td>
<td>60 min. 1/week, 8 weeks</td>
</tr>
<tr>
<td>2021</td>
<td>Kim &amp; Yoo RCT</td>
<td></td>
<td>10/10</td>
<td>78.8/ 70.2</td>
<td>MMSE</td>
<td>Rhythm-motor dual tasks</td>
<td>No treatment</td>
<td>30 min. 2/week, 8 weeks</td>
</tr>
<tr>
<td>2022</td>
<td>Bugos &amp; Wang</td>
<td>RCT</td>
<td>30/ 35/50</td>
<td>67.2/ 69.29/ 67.64</td>
<td>TICS</td>
<td>Group piano playing</td>
<td>Computer-assisted cognitive training; no treatment</td>
<td>90 min, 2/week, 16 weeks (no outside practice)</td>
</tr>
</tbody>
</table>

ACE-III, Addenbrooke’s Cognitive Examination-III (Hsieh et al., 2013); Con, control; fMRI, functional magnetic resonance imaging; Int, intervention; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination (Folstein et al., 1975); RCT, randomised controlled trial; TICS, Telephone Interview for Cognitive Status (Brandt et al., 1988).

### 3. Results

Thirteen studies were included in the analyses, one of which was a PhD dissertation, and it should be noted that this study was not published and peer-reviewed (Thorne, 2015). Following attrition in each study, a total of 502 participants were analysed. As not all thirteen studies examined all five of the cognitive domains of interest, the number of total samples for the meta-analyses for each cognitive domain ranged between 211 and 360 participants. For this reason, funnel plots were not created to assess publication bias for any cognitive domain tested, given that statistical power was too low to detect meaningful asymmetry in fewer than ten studies (Higgins et al., 2011). The pooled mean age across all studies was 71.95 years (SD = 4.83), ranging from 67.5 to 83.7 years old. All studies which were included in the analysis were conducted on healthy older adults (see Table 2 for cognitive screening tools used). One study included in the meta-analyses also included participants with MCI (Biasutti & Mangiacotti, 2018).

In three studies, piano training was compared with a music listening intervention which provided training in music appreciation and active listening skills (Bugos, 2010; Bugos, 2019, Thorne, 2015). For five studies, music intervention was compared with a passive, no-treatment control group (Bugos et al., 2007; Guo et al., 2021; Kim & Yoo, 2020; Bugos & Wang, 2022). One study used a waitlist design where one group was tested pre- and post- piano training, while the control group remained inactive during the intervention period. The control group later received the intervention after the second testing and were tested a third time after they had received the intervention (MacRitchie et al., 2020). In one study, the intervention group acted as their own control, with testing occurring pre- and post- a two week no-treatment period, and then again after the piano training. However, the order was not counterbalanced across participants because of potential long-term effects of the intervention (Bugos & Kochar, 2017). Other studies compared music training with other leisure activities (Seinfeld et al., 2013), gymnastic activities (Biasutti & Mangiacotti, 2016) or choir participation (Dos Santos et al., 2020). For studies which included two control groups, effect sizes were calculated from control groups requiring no instrument training – a passive control group (Bugos & Wang, 2022; Dégé & Korkievís, 2018) or music listening intervention group (Bugos & Kochar, 2019), because these controls did not require learning a motor skill and were therefore comparable with control groups used in the other studies.

#### 3.1. Risk of bias

Of the thirteen studies, nine studies used randomization in assigning participants to a music intervention group or a control group. Only five studies explicitly stated that outcome assessors were blind to group allocation. Attrition rates ranged from 20 to 34 %. Overall bias was rated as “high” for three studies and as “some concerns” for nine studies. One study was rated as having “low” overall bias. Overall percentages of studies with “high”, “some concerns” or “low” risk of bias are displayed in Fig. 2. Summary of risk of bias within each study for six domains of
Table 3
Cognitive domains tested and outcome measures used in each music intervention study.

<table>
<thead>
<tr>
<th>Study</th>
<th>Executive Functions</th>
<th>Verbal Memory</th>
<th>Visuospatial Attention</th>
<th>Processing Speed</th>
<th>Visuospatial Attention</th>
<th>Visuospatial Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Switching Inhibition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bugos et al. (2007)</td>
<td>TMT-B</td>
<td>n/a</td>
<td>n/a</td>
<td>Digit span</td>
<td>Digit Symbols</td>
<td>TMT-A</td>
</tr>
<tr>
<td>Bugos (2010)</td>
<td>TMT-B</td>
<td>Stroop (error rates)</td>
<td>n/a</td>
<td>D-KEFS (LF, CF, CS)</td>
<td>PASAT</td>
<td>TMT-A</td>
</tr>
<tr>
<td>Seinfeld et al. (2013)</td>
<td>TMT-B</td>
<td>Stroop (correct responses)</td>
<td>n/a</td>
<td>Digit span</td>
<td>SDMT</td>
<td>TMT-A</td>
</tr>
<tr>
<td>Thorne (2015)</td>
<td>TMT-B</td>
<td>Stroop (error rates)</td>
<td>n/a</td>
<td>n/a</td>
<td>Digit symbols</td>
<td>TMT-A</td>
</tr>
<tr>
<td>Bugos and Kochar (2017)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>D-KEFS (LF, CF, CS)</td>
<td>WAIS-IV (coding and symbol search index scores)</td>
<td>n/a</td>
</tr>
<tr>
<td>Biasutti and Mangiacotti (2018)</td>
<td>n/a</td>
<td>n/a</td>
<td>LF</td>
<td>n/a</td>
<td>TMT-A</td>
<td>n/a</td>
</tr>
<tr>
<td>Deg and Kerkovius (2018)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>Digit span</td>
<td>n/a</td>
<td>Immediate word-list recall</td>
</tr>
<tr>
<td>Bugos (2019)</td>
<td>TMT-B</td>
<td>Stroop (error rates)</td>
<td>n/a</td>
<td>D-KEFS (LF, CF, CS)</td>
<td>PASAT</td>
<td>TMT-A</td>
</tr>
<tr>
<td>Guo et al. (2021)</td>
<td>TMT-B</td>
<td>LF, CF</td>
<td>n/a</td>
<td>Digit span</td>
<td>TMT-A</td>
<td>Immediate story recall (WMS-LM I)</td>
</tr>
<tr>
<td>MacRitchie et al. (2020)</td>
<td>TMT-B</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>TMT-A</td>
<td>n/a</td>
</tr>
<tr>
<td>Dos Santos et al. (2021)</td>
<td>n/a</td>
<td>n/a</td>
<td>LF</td>
<td>n/a</td>
<td>TMT-A</td>
<td>n/a</td>
</tr>
<tr>
<td>Kim and Yoo (2020)</td>
<td>TMT-B</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>TMT-A</td>
<td>n/a</td>
</tr>
<tr>
<td>Bugos and Wang (2022)</td>
<td>TMT-B</td>
<td>Stroop (error rates)</td>
<td>n/a</td>
<td>D-KEFS (LF, CF, CS)</td>
<td>N-back</td>
<td>PASAT</td>
</tr>
</tbody>
</table>

CF, category fluency; CNS, Computerised Neurocognitive Assessment (CNS Vital Signs, Inc: Morrisville, NC); CS, category switching; D-KEFS, Delis-Kaplan Executive Function Systems (Delis et al., 2001); EF, executive functions; LF, letter fluency; PASAT, Paced Auditory Serial Addition Test (Gronwall, 1977); RAVLT, Rey Auditory Verbal Learning Test (Rey, 1941); TMT-A, Trail-Making Test (Part A); TMT-B, Trail-Making Test (Part B; Reitan & Wolfson, 2009); WAIS-IV, Wechsler Adult Intelligence Scale IV (Wechsler, 2008); WMS-LM, Wechsler Memory Scale – Logical Memory (Wechsler, 1987).

Fig. 2. Percentage risk of bias summary from Cochrane Risk of Bias tool (Stern et al., 2019).
Cochrane RoB 2 is presented in Fig. 3.

### 3.2. Main findings

With regards to EF, statistically significant effect sizes were found for attention switching (low-moderate effect size), distractor inhibition (low effect size), and the category switching verbal fluency subtest (low-moderate effect size). No statistically significant effect size was found for working memory capacity/updating or the other verbal fluency subtests (letter and category).

A significant moderate effect size was found for processing speed, but no significant effects were found for visuospatial attention, or any other cognitive domain.

---

**Fig. 3.** Summary of risk of bias within each study for six domains of Cochrane RoB 2.

D1: Randomisation process; D2: Deviations from the intended interventions; D3: Missing outcome data; D4: Measurement of the outcome; D5: Selection of the reported result

Cochrane RoB 2 is presented in Fig. 3.
measure of verbal memory (immediate, short- or long-delay measures).

Sensitivity analyses for all cognitive domains tested are reported in detail in the following subsections. Table 4 reports the full statistical models for each domain with statistically significant effect sizes indicated (*p < .05, **p < .01 *** p < .001). Forest plots for each domain are presented in Fig. 4 with statistically significant effect sizes indicated (*p < .05, **p < .01 *** p < .001).

3.2.1. Executive functions

3.2.1.1. Switching. Nine studies used TMT-B completion times or delta scores as a measure of attentional switching. However, one study (Bogus, 2010) did not report descriptive statistics. A random-effects model conducted on the remaining 8 studies revealed a moderate effect size (d = -0.42, p = .0143), with moderate heterogeneity (\(I^2 = 51.62\)%). One study was rated as having overall “high” bias (Seinfeld et al., 2013), while the other six were rated as “some concerns”.

Notably, when the study rated as “high bias” (Seinfeld et al., 2013) was removed from the analysis, the effect size dropped to a low effect (d = -0.28, p = .0202) with low heterogeneity (\(I^2 = 4.53\)%), thus suggesting a considerable impact on the model, and therefore this study was omitted.

In a final exploratory synthesis, one study (Guo et al., 2021) which required training in keyboard harmonica (an instrument played with one hand only and therefore did not involve practice in bimanual coordination) was removed from the model. The final model, which included 6 studies, revealed a small-moderate effect size (d = -0.3868, p = .0021), with 0% heterogeneity. This suggests that musical instrument interventions which require a switching attentional component from one hand to the other may show strongest transfer to TMT-B performance. This final model is presented in Table 4 and Fig. 4 (A).

3.2.1.2. Inhibition. Five studies measured inhibitory control using the incongruent condition of the Stroop test. Two of these studies were rated as “high” for overall bias (Bogus, 2010; Seinfeld et al., 2013), while the other three were rated as ‘some concerns’. The model which included these 5 studies revealed no effect, and was non-significant (d = 0.03, p = .9236), with high heterogeneity (\(I^2 = 76.63\)%).

During the leave-one-out analysis, Thorne (2015) found to have a substantial impact on the overall effect size and therefore omitted from the final model in Fig. 4 (B). The final model suggests a small, but statistically significant, effect size for inhibition (d = 0.27, p = .0353, \(I^2 = 0\)%), but should be interpreted with caution given that two of the four studies were rated high for overall bias.

3.2.1.3. Working memory capacity and updating. Four studies utilised digit span tasks as a measure of working memory capacity, and one study measured updating using the N-back task (see Table 3). One study was rated high for risk of bias (Seinfeld et al., 2013). A random effects model using data from these five studies revealed a moderate effect size which was not statistically significant (d = 0.4729, p = .1374), and contained high heterogeneity (77.17%).

When only working memory capacity was investigated (i.e. when Bugos & Wang, 2022 was removed), a moderate effect which approached statistical significance (d = 0.6420, p = .0882) was detected, however heterogeneity remained high (74.17%).

When the high bias study was removed, random effects model of the other four studies revealed a drop in the effect size to below the “low” threshold and was non-significant. (d = 0.1576, p = .4112), with moderate heterogeneity (31.24%). This is the final model presented in Table 4 and Fig. 4 (C).

3.2.1.4. Verbal fluency (letter fluency, category fluency and category switching subtests). Three studies employed the category switching subtest of verbal fluency. Category switching is a subtest of the D-KEFS verbal fluency assessment which requires alternating between words of two categories as many times as possible within 60-second limit. Of the three studies which used this measure, one of which was rated as having “high” risk for overall bias (Bugos & Kochar, 2017), and the other two were rated as “some concerns”. A statistically significant small-to-moderate effect was found for category switching (d = 0.39, p = .0166, \(I^2 = 36\)%). Fig. 4 (F).

Analyses of other verbal fluency tests (letter fluency and category fluency) did not reveal effect sizes larger than 0.20, and none of which were statistically significant. However, effect sizes for letter and category fluency approached the 0.2 “low” effect size threshold (d = 0.19, p = .0795) and (d = 0.18, p = .1144) respectively (see Figures (D) and (E)).

3.2.2. Processing speed

A random-effects model for processing speed was conducted on 7 studies which included a measure of processing speed (see Table 3 for specific tests used). A moderate effect size was found, which was statistically significant (d = 0.60, p = .0004), with moderate heterogeneity (\(I^2 = 54\)%).

Notably, during the exploratory “leave-one-out” analysis, pooled effect size for processing speed dropped to low-moderate range when

Table 4

<table>
<thead>
<tr>
<th>Model</th>
<th>d</th>
<th>P</th>
<th>CI lower</th>
<th>CI upper</th>
<th>Q</th>
<th>df</th>
<th>p</th>
<th>I2</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switching</td>
<td>-0.39**</td>
<td>0.0021</td>
<td>-0.6325</td>
<td>-0.1402</td>
<td>1.31</td>
<td>5</td>
<td>0.93</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inhibition</td>
<td>0.27**</td>
<td>0.0335</td>
<td>0.0212</td>
<td>0.5237</td>
<td>1.97</td>
<td>3</td>
<td>0.58</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Working memory capacity/updating</td>
<td>0.16</td>
<td>0.4112</td>
<td>-0.2182</td>
<td>0.5333</td>
<td>4.33</td>
<td>3</td>
<td>0.23</td>
<td>31.24</td>
<td>0.05</td>
</tr>
<tr>
<td>Verbal fluency (letter fluency)</td>
<td>0.19</td>
<td>0.08</td>
<td>-0.0220</td>
<td>0.3970</td>
<td>4.51</td>
<td>5</td>
<td>0.48</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Verbal fluency (category fluency)</td>
<td>0.18</td>
<td>0.11</td>
<td>-0.0429</td>
<td>0.3982</td>
<td>3.44</td>
<td>4</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Verbal fluency (category switching)</td>
<td>0.39**</td>
<td>0.0166</td>
<td>0.0715</td>
<td>0.7158</td>
<td>3.11</td>
<td>2</td>
<td>0.21</td>
<td>36</td>
<td>0.03</td>
</tr>
<tr>
<td>Processing speed</td>
<td>0.47***</td>
<td>&lt;0.0001</td>
<td>0.2513</td>
<td>0.6850</td>
<td>3.21</td>
<td>5</td>
<td>0.67</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Visuospatial attention</td>
<td>-0.05</td>
<td>0.70</td>
<td>-0.3110</td>
<td>0.2101</td>
<td>15.63</td>
<td>9</td>
<td>0.08</td>
<td>29.16</td>
<td>0.05</td>
</tr>
<tr>
<td>Verbal Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total recall after 5 trials</td>
<td>0.1955</td>
<td>0.3310</td>
<td>-0.1987</td>
<td>0.5896</td>
<td>3.7893</td>
<td>2</td>
<td>0.1504</td>
<td>31.87</td>
<td>0.0394</td>
</tr>
<tr>
<td>Immediate/ short-delay recall</td>
<td>0.2327</td>
<td>0.1689</td>
<td>-0.0988</td>
<td>0.5641</td>
<td>0.3599</td>
<td>2</td>
<td>0.8353</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Long-delay recall</td>
<td>0.2681</td>
<td>0.1536</td>
<td>-0.1001</td>
<td>0.6363</td>
<td>1.0517</td>
<td>1</td>
<td>0.3051</td>
<td>4.91</td>
<td>.0594</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01 *** p < .001.
Seinfeld et al. (2013), a study which had been rated as high bias, was removed \((d = 0.47, p < .0001, I^2 = 0)\). Therefore, Seinfeld et al. (2013) was omitted from the final forest plot for processing speed depicted in Fig. 4 (G) to minimise risk of inflating effect sizes. It should also be noted that, of the six remaining studies which were included in the final model, two were rated as high for bias (Bugos, 2010; Bugos & Kochar, 2017),
and four were rated as having some concerns. Removal of the two high bias studies did not impact effect size and so were left in the final model. However, this result should be interpreted with caution given that the Bugos group constitute five of the six studies, which may introduce bias given that the majority of the data was collected by the same lab.

### 3.2.3. Visuospatial attention

Ten studies employed the TMT-A as a measure of visuospatial attention. Two studies were rated as having high risk for overall bias (Bugos, 2010; Seinfeld et al., 2013), one was rated as having low overall bias (Dos Santos et al., 2020), and the remaining 7 studies were rated as some concerns. No effect size was found ($d = -0.0505$, $p = .7042$).

![Graph showing effect sizes for visuospatial attention](image)

(D) Verbal fluency (letter fluency)
(E) Verbal fluency (category fluency)

<table>
<thead>
<tr>
<th>Author(s) and year</th>
<th>SMD [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bugos &amp; Kochar (2017)</td>
<td>0.48 [0.00, 0.97]</td>
</tr>
<tr>
<td>Guo et al. (2020)</td>
<td>-0.10 [-0.64, 0.43]</td>
</tr>
<tr>
<td>Bugos (2019)</td>
<td>0.25 [-0.15, 0.65]</td>
</tr>
<tr>
<td>Dos Santos (2021)</td>
<td>0.25 [-0.53, 1.02]</td>
</tr>
<tr>
<td>Bugos &amp; Wang (2022)</td>
<td>-0.01 [-0.47, 0.44]</td>
</tr>
<tr>
<td>RE Model</td>
<td>0.18 [-0.04, 0.40]</td>
</tr>
</tbody>
</table>

(F) Verbal fluency (category switching) *

<table>
<thead>
<tr>
<th>Author(s) and year</th>
<th>SMD [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bugos &amp; Kochar (2017)</td>
<td>0.69 [0.20, 1.18]</td>
</tr>
<tr>
<td>Bugos (2019)</td>
<td>0.13 [-0.27, 0.53]</td>
</tr>
<tr>
<td>Bugos &amp; Wang (2022)</td>
<td>0.43 [-0.03, 0.89]</td>
</tr>
<tr>
<td>RE Model</td>
<td>0.39 [0.07, 0.72]</td>
</tr>
</tbody>
</table>

Fig. 4. (continued).
Heterogeneity was moderate ($I^2 = 29.16\%$). After visual inspection of the forest plot, the analysis was repeated whilst omitting Thorne (2015), but effect size and significance $p$ values remained virtually unchanged ($d = 0.0370, p = .7342, I^2 = 0$). Therefore, Thorne (2015) was left in the final model reported in Table 4 and Fig. 4 (H).

### 3.2.4. Verbal memory

#### 3.2.4.1. Total word-list recall following five trials

Three studies reported total word recall following five successive trials where the same word list was presented (Thorne, 2015; Bugos & Kochar, 2017; Bugos & Wang, 2019).
The analysis indicated an effect size which approached the “low” threshold, but was non-significant (see Table 4; Fig. 4 (I)). However, further research is needed given that only three studies included this measure, one of which had been rated as high bias (Bugos & Kochar, 2017), and one which had not been peer-reviewed (Thorne, 2015).

3.2.4.2. Immediate and short-delay recall. Four studies measured immediate or short-delay verbal memory using different measures (see Table 4). The random effects model revealed a moderate effect size, which was non-significant (d = 0.6064, p = .1265), with high heterogeneity (Q(3) = 9.7770; p = .0206; $I^2 = 79.53\%$).

When Degé and Kerkovich (2018) was removed from the analysis,
effect size was low and heterogeneity dropped to 0%, but remained non-significant. This study was therefore omitted from the final model presented in Table 4 and Fig. 4 (J).

3.2.4.3. Long-delayed recall. Three studies employed measures of long-delay verbal recall (see Table 3). The model indicated a moderate effect size which approached statistical significance ($d = 0.4602$, $p = .0758$) with moderate heterogeneity ($Q(2) = 4.0866; p = .1296; I^2 = 50.42\%$).

When Thorne (2015) was removed from the analysis, effect size dropped to a low effect, with only 4.91% heterogeneity, but remained non-significant (see Table 4; Fig. 4 (K)).

3.3. Three-month follow-up effect sizes

Of the thirteen studies included in the meta-analyses, only two conducted long-term follow-up testing. Effect sizes calculated from descriptive statistics reported in Bugos et al. (2007) indicated large effect sizes at 3-month follow-up testing in processing speed ($d = 1.36$) and switching ($d = -0.77$). Moderate effects were calculated for processing speed ($d = 0.47$), working memory ($d = 0.51$), inhibition ($d = 0.38$) and switching ($d = 0.34$) from 3-month follow-up data in Bugos and Wang (2022). No effect was found for verbal fluency at follow-up.

4. Discussion

The purpose of this systematic review and meta-analysis was to examine the effects of learning to play a musical instrument in older adulthood on cognition. This review is novel and timely given the shift in interest in recent years from computerised training programmes towards more ecologically-valid cognitive interventions which tax multiple sensory and motor modalities. Data from thirteen studies were included in this meta-analysis, which represents nine more studies than the most recent meta-analyses papers which were published when only four musical instrument intervention studies on older adults were available (Kim & Yoo, 2019; Roman-Caballero et al., 2018).

Training in a pitched or rhythm instrument was found to have a positive effect on EF—specifically: a low-moderate effect on switching; and a low effect on inhibition. No effect was found on working memory capacity/updating. With regards to verbal fluency, a low-moderate effect size was found for the category switching subtest, but no statistically significant effects were found for letter fluency or category fluency subtests. With regards to other fluid domains tested, a moderate effect was found for processing speed, but no effects were found for visuospatial attention or verbal memory.

4.1. Effects of musical instrument training on EF and fluid intelligence

The benefits of learning to play an instrument to EF found in this study are in line with previous meta-analyses (Kim & Yoo, 2019; Roman-Caballero et al., 2018). However, the findings are in contrast to a meta-analysis which examined effects of rhythmic exercise interventions (e.g., dancing and exercise to music) on EF (Ma et al., 2023). This suggests that some features of playing an instrument such fine motor control or concurrent auditory feedback could be important for EF beyond the effects of movement to music, although further research is needed given that only three out of the forty-four rhythm studies included in Ma et al. (2023) investigated EF.

The findings are also in agreement with cross-sectional research suggesting an inhibitory control advantage for musicians versus non-musicians (Moreno et al., 2014; Moreno & Farzan, 2015; Moussard et al., 2016; Hanna-Pladdy & MacKay, 2011), as well as with evidence from a neuropsychological study showing that event related potentials (ERPs) were associated with inhibitory control during a GoNoGo task over right frontal regions in older adults following three months of percussion training (Alain et al., 2019).

Inhibition requires the ability to regulate incoming sensory input and behaviour so as to override automatic responses or external distractions to ensure more appropriate actions or behaviours (Diamond, 2013). Similarly, switching involves the disengagement of an irrelevant task and subsequent engagement of a relevant task ( Miyake et al., 2000). Taken together, music training is suggested to improve supervisory mechanisms which act to ensure desired responses and reinforce interference monitoring.

It is unclear as to why music training may impact cognitive mechanisms responsible for controlling automatic responses. Continuous and dynamic suppression of motor responses for incorrect notes in favour of
playing the correct ones, for the correct length of time (e.g. without releasing a piano key too soon), may tax inhibitory processes while playing music. Interestingly, the best-fitting model for switching data (TMT-B completion times) was when only bimanual instruments were included. It may be that inter-hemispheric sensory-motor integration and synchronization which occurs from continuously switching attention between right- and left-hand auditory output may train this particular domain of EF, although further research is needed given that only one study used a unimanual instrument. This could perhaps also explain the far-transfer effect found on verbal category switching, but again, further research is needed with larger trials given that only three studies examined this domain.

The moderate effect on processing speed was the largest detected effect size in this meta-analysis, and this finding is in line with Kim and Yoo (2019), and musician-versus-nonnusician cross-sectional research (Pathbauer et al., 2015). The finding is also in line with functional magnetic resonance imaging (fMRI) data from Guo et al. (2021) which found reduced left putamen and right superior temporal gyrus functional activity post-music training which was associated with greater improvement in memory performance post-training, suggesting increased neural efficiency. Processing speed is associated with general cognitive performance (Śliwinski & Buschke, 1999). It is possible that the improvement in processing speed may have contributed to the statistically significant effect sizes in other cognitive domains (switching and inhibition).

In contrast to Kim and Yoo (2019), no effect was found for visuospatial attention (TMT-A completion times) when more recent large-scale studies were included. However, this lack of finding may be due to the populations sampled, as Kim and Yoo (2019) also included older adults with MCI, whereas this study focused only on healthy older adults without cognitive impairment. No effect was found on working memory capacity/updating, or in verbal memory in contrast to the cognitive benefits reported in cross-sectional studies comparing musicians and novices (Jäncke, 2009). This suggests that these abilities may take longer to train, or may be more dependent on sensitive development periods in childhood. It is also possible that further data are needed, given that only a small number of studies investigated working memory and verbal performance.

4.2. Limitations

A limitation of this meta-analysis is the relatively small sample size of studies (N = 13), particularly given that not all studies examined the same cognitive abilities, which limits the extent to which publication bias can be investigated. Furthermore, one research group is over-represented across studies. Five out of the thirteen studies (38.4 %) were conducted by Bugos and colleagues which potentially increases the risk of bias in the analyses, particularly for processing speed and inhibition.

A number of methodological limitations were identified in the literature. Three studies ranked as “high”, and nine ranked as “some concerns” for overall bias. Only three studies utilised both an active and a passive control group. Three studies compared music intervention to a no-treatment passive control only. Absence of an active control group presents difficulties in ruling out possible placebo effects and in determining the mechanisms of action in the intervention. Due to the nature of the intervention, managing participant expectations through blinding is impossible, but only five studies reported that assessors were blind to group allocation. Taken together, this raises concerns about the possibility of inflated effect sizes through participant and assessor expectations. Furthermore, study sample sizes were relatively small, with only two studies including samples with over one hundred participants.

Music training studies may be subject to participant selection bias. For example, older adults with upper-limb weakness, or back pain may not have the physical capacity to sit and practice an instrument, such as the piano, for extended periods of time. Educational background and socio-economic status may also impact an individual’s interest, adherence and retention to training.

Attrition rates were relatively high amongst studies. Notably, when Bugos and Wang (2022) examined low-dosage participants, they found significantly lower baseline full-scale IQ compared to the group which completed training, suggesting that pre-existing ability may affect adherence. Consistent with global “g” intelligence theories (Deary et al., 2010) individuals with higher IQ are more likely to have good aptitude for music, when controlling for training duration and socio-economic status (Swaminathan et al., 2017). Whereas high-functioning individuals may be more likely to complete training, it is most effective in those with lower baseline cognitive ability (Bugos & Wang, 2022), indicating a potential mismatch between acceptability and need in previous studies.

Additionally, not all studies explicitly stated how participants accessed an instrument for practice. In some studies, instruments were provided (Guo et al., 2021) or were accessed in community buildings (Bugos et al.), but other studies did not indicate how participants could fulfill practice requirements. Inclusion of participants who already had access to instruments (e.g., piano) increases the risk of confounding factors regarding actual level of experience, pre-exposure to music or socio-economic status. This also highlights the potential costliness of running large-scale “ecologically valid” music RCTs, where instruments and instructors need to be funded – an issue which does not apply to computerized programmes.

4.3. Future directions

This meta-analysis suggests that musical instrument training in later adulthood can benefit EF and processing speed. However further research is needed to address the specific processes involved in musical training which enhance cognition.

Firstly, the specific intervention parameters which lead to improved cognition could be explored. In the present meta-analysis there was not a sufficient number of studies to conduct meaningful moderation analyses to investigate the variables that may have impacted on the size of training effects such as intervention frequency, duration, nature of the control group, effects of pitched versus percussion instrument training. Kim and Yoo (2019) previously reported that percussion or “immediate” involvement interventions tend to be more common in MCI and clinical populations, whereas “sustained” involvement instruments, which require the reading and playing of music, are more common in healthy populations. The extent to which different elements such as finger dexterity training, sight-reading or ear-training play a role in far-transfer therefore in healthy older adults still requires further investigation. Furthermore, near-transfer measures are lacking in the studies included in this meta-analysis, and relationships between improvements in instrument-playing and cognitive improvements could be examined in future studies.

Secondly, more longitudinal studies are needed to determine longevity of effects. Long-term effects of training in middle-aged participants could also be examined, given that some EF begin to decline from 30 to 40 years of age (Ferguson et al., 2021), and playing an instrument has recently been shown to protect against age-related brain atrophy in both grey (Marie et al., 2023) and white (Jüinemann et al., 2022) matter. Training interventions which begin in middle adulthood may also help address acceptability and attrition issues highlighted above.

Thirdly, although there have been some studies showing training-induced neural plasticity (e.g., Worschech et al., 2022), the neural underpinnings of training-induced EF enhancements are still not well understood. For example, one study found that both music listening training and piano training lead to increased volume in the caudate nucleus, Rolandic operculum and inferior cerebellum (Marie et al., 2023), indicating that ear training alone must also be taken into account when examining the impact of music training on cognition and
neuroplasticity. Further research is needed to examine how music-induced neural changes are related to cognition.

5. Conclusion

In conclusion, there is promising evidence that engagement in a musical instrument training positively impacts EF and processing speed in healthy older adults. In addition, motivation can be maintained over time with learning an instrument, and may be more likely to lead to long-term benefits. Whereas computerized cognitive training may be discontinued, there are a limitless number of musical pieces or songs which can be learned after an intervention has ended (Sutcliffe et al., 2020). Given the capacity for plasticity into later adulthood, the potential for motor activities to engage networks important for cognition, and the benefits of music to well-being, music training may be an ecologic and enjoyable activity which benefits cognitive function in older adulthood.

CRediT authorship contribution statement

Fionnuala Rogers: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. Claudia Metzler-Baddeley: Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by an Open PhD studentship from the School of Psychology Cardiff University and a National Institute for Health Research (NIHR) and Health and Care Research Wales (HCRW) Advanced Fellowship [NIHR-FS(A)-2022]. We would like to thank Dr. Jennifer MacRitchie and colleagues for providing data from their study to include in this meta-analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2024.106137.

References


