



Individual differences in musical training and executive functions: A latent variable approach

Brooke M. Okada¹ · L. Robert Slevc¹

Published online: 11 May 2018
© Psychonomic Society, Inc. 2018

Abstract

Learning and performing music draw on a host of cognitive abilities, and previous research has postulated that musicians might have advantages in related cognitive processes. One such aspect of cognition that may be related to musical training is executive functions (EFs), a set of top-down processes that regulate behavior and cognition according to task demands. Previous studies investigating the link between musical training and EFs have yielded mixed results and are difficult to compare. In part, this is because most studies have looked at only one specific cognitive process, and even studies looking at the same process have used different experimental tasks. Furthermore, most correlational studies have used different “musician” and “non-musician” categorizations for their comparisons, so generalizing the findings is difficult. The present study provides a more comprehensive assessment of how individual differences in musical training relate to latent measures of three separable aspects of EFs. We administered a well-validated EF battery containing multiple tasks tapping the EF components of inhibition, shifting, and working memory updating (Friedman et al. in *Journal of Experimental Psychology: General*, 137, 201–225, 2008), as well as a comprehensive, continuous measure of musical training and sophistication (Müllensiefen et al., in *PLoS ONE*, 9, e89642, 2014). Musical training correlated with some individual EF tasks involving inhibition and working memory updating, but not with individual tasks involving shifting. However, musical training only predicted the latent variable of working memory updating, but not the latent variables of inhibition or shifting after controlling for IQ, socioeconomic status, and handedness. Although these data are correlational, they nonetheless suggest that musical experience places particularly strong demands specifically on working memory updating processes.

Keywords Musical training · Executive functions · Latent variable analysis

When musicians practice or perform music, they engage in a remarkable cognitive feat. They read and decipher sheet music, recall music from memory, and may even improvise something new on the fly. Thus, they produce musical sounds while also planning ahead by keeping notes and rhythms in mind until they are played. Not only do they monitor their own playing, but they must also attend to the auditory streams produced by other performers to flexibly coordinate and adjust their own sound (e.g., pitch, tempo, volume) to match the group. These, among other demands of musical performance, are the types of planning and monitoring processes characteristic of our ability to regulate behavior and cognition according

to task demands more generally—abilities often called *executive functions* (EFs; Diamond, 2013; Miyake & Friedman, 2012).

There has been increasing interest in uncovering the cognitive mechanisms involved in learning, producing, and perceiving music (see, e.g., Collins & Fleming, 2017, for a discussion of a collaboration between the NIH and the Kennedy Center). Although there is little evidence that short-term musical exposure can directly influence other cognitive abilities (i.e., there is little support for the so-called “Mozart effect”; Pietschnig, Voracek, & Formann, 2010), a growing body of work has focused instead on the possibility that regular long-term engagement in music (e.g., learning to play an instrument) influences non-musical abilities (for reviews, see Benz, Sellaro, Hommel, & Colzato, 2015; Schellenberg & Weiss, 2013). These posited relationships with musical experience span a wide range of domains, ranging from social cognition (e.g., Good, Choma, & Russo, 2017; Kirschner & Tomasello, 2010) to cognitive abilities such as intelligence and EFs (e.g.,

✉ Brooke M. Okada
bokada@umd.edu

¹ Department of Psychology, University of Maryland, College Park, MD, USA

Moreno & Bidelman, 2014; Schellenberg, 2004; Slevc, Davey, Buschkuehl, & Jaeggi, 2016).

As suggested above, one non-musical cognitive process that has often been linked to musical experience is EFs. These top-down processes adaptively regulate behavior and cognition and are linked to many types of complex behavior, including music processing (e.g., Hannon & Trainor, 2007; Slevc et al., 2016; Slevc & Okada, 2015). Most models of EFs postulate three related but separable components: *inhibition* (or inhibitory control), *shifting* (or cognitive flexibility or switching), and *updating* (or working memory [WM] updating) (Diamond, 2013; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Logue & Gould, 2014; Miyake & Friedman, 2012; Miyake et al., 2000). Inhibition requires one to override an prepotent response; shifting requires one to switch between task demands; and updating requires one to constantly add, delete, and manipulate items within WM (Miyake et al., 2000). EFs develop through adolescence and have been associated with quality of life and with success in school and in the workplace, and they have also been shown to improve with practice (Diamond, Barnett, Thomas, & Munro, 2007; for a review, see Diamond, 2013), so there has been considerable interest in both predictors of EF abilities and how EF abilities might be improved.

The relationships between musical ability and EFs have mostly been demonstrated in correlational studies, and these links could reflect a few possibilities. One possibility is that musical experience draws strongly on some (or all) aspects of EFs, which leads to collateral benefits, or transfer effects, to those EF abilities more broadly (e.g., Holochwost et al., 2017). This might mean that musical training could serve to improve EFs more broadly (i.e., as a type of relatively enjoyable cognitive training). A second possibility is that these relationships could instead reflect a selection bias, in which individuals with pre-existing skills that are useful for music learning are more likely to pursue and persist in continued music lessons (Elpus, 2013; Mosing, Madison, Pedersen, & Ullén, 2016). Third, these possibilities may both be true, and music lessons may exaggerate pre-existing differences (i.e., these relationships may reflect a gene-environment interaction; Schellenberg, 2015). Although these possibilities are still debated, all three postulate that musical experience draws strongly on certain cognitive abilities, so that the relationships between musical experience and cognitive abilities should reflect the specific abilities that are critical to musical experience. Thus, even if musical experience does not directly produce improvements in EFs (which is plausible, especially given the limits on transfer from WM training paradigms; e.g., Dougherty, Hamovitz, & Tidwell, 2016; Melby-Lervåg & Hulme, 2013), understanding the cognitive abilities that predispose individuals to musical training should inform the cognitive processes that are particularly critical aspects of musical experience.

These possibilities are, of course, based on robust relationships between musical experience and the cognitive abilities in question. Unfortunately, the current state of the literature makes it difficult to know exactly which (if any) EFs are reliably related to musical experience, for reasons discussed below. Nonetheless, there are good reasons to expect musical experience to be related to multiple EF components, given plausible EF demands in the perception, production, and learning of music.

EFs in musical perception, production, and learning

Each EF component (i.e., inhibition, shifting, and updating) has been linked to aspects of musical processing and experience. For example, overriding expectations and unexpected resolutions of musical ambiguity are likely to draw on general inhibitory control mechanisms (Slevc & Okada, 2015). Shifting is likely involved when playing music in an ensemble, where musicians must coordinate their own playing with others in the group (Jentzsch, Mkrtchian, & Kansal, 2014; Palmer, 2013). This requires both flexible shifting between auditory streams (Loehr, Kourtis, Vesper, Sebanz, & Knoblich, 2013) and also dynamically adjusting to other members of the group (e.g., Loehr & Palmer, 2011; Moore & Chen, 2010). Updating is critical in music reading (especially sight-reading), during which musicians need to glance ahead in the score to prepare what to play. In fact, good sight-readers typically look about four notes ahead of where they are playing (Furneaux & Land, 1999; cf. Drake & Palmer, 2000; Goolsby, 1994) and thus, are constantly updating the contents of their WM: holding in mind which notes are being played and which are yet to be played. Correspondingly, sight-reading ability is related to non-musical measures of WM capacity (Meinz & Hambrick, 2010).

Given these types of relationships, it seems likely that musical experience and training would be associated with abilities in each component of EF. This question has been addressed in both correlational studies, investigating whether specific EFs are associated with extent of musical training, and a handful of experimental studies, investigating whether random (or quasi-random) assignment to musical training may improve specific EFs (for a review, see Okada & Slevc, 2018). However, as we discuss below, it remains difficult to draw firm conclusions from the current studies in the literature given their reliance on individual tasks, the variety of tasks and definitions of musical experience across studies, and, in many cases, their relatively limited statistical power. Nevertheless, this past work suggests that musical experience may indeed relate to multiple aspects of EFs.

Inhibition Bialystok and DePape (2009) reported that musicians performed better than non-musicians on two tests of inhibition, a Simon arrows task and an auditory Stroop task (see also Amer, Kalender, Hasher, Trehub, & Wong, 2013; Travis, Harung, & Lagrosen, 2011). However, note that the scoring used by Bialystok and DePape differed from traditional ways of measuring costs (i.e., musicians were faster overall than non-musicians, but did not show the traditional smaller cost for incongruent relative to neutral trials). In contrast, other work has not found a relationship between performance on those same tasks and (continuous) measures of musical ability and experience (Slevc et al., 2016) or between both child and adult musicians and non-musicians on similar Stroop tasks (Zuk, Benjamin, Kenyon, & Gaab, 2014). Furthermore, in work examining differences between types of musicians, percussionists were found to outperform vocalists and non-musician controls on another inhibition test (the integrated visual and auditory plus continuous performance test; Slater, Azem, Nicol, Swedenborg, & Kraus, 2017).

Paralleling the correlational data, findings from experimental longitudinal studies have been similarly mixed. Moreno and colleagues quasi-randomly assigned children to receive computerized music training or visual arts training and found that the musical-training group showed better behavioral performance as well as a greater index of brain plasticity on no-go trials in a go/no-go task (Moreno et al., 2011; see Moreno & Farzan, 2015, for discussion). Similarly, children who enrolled in after-school music lessons also showed improvement on multiple tasks of inhibitory control (go/no-go, Stroop, and flanker tasks), as compared to a waitlist (no-contact) control group (Holochwost et al., 2017). In contrast, Bugos and DeMarie (2017) found more equivocal results, in which pre-school children assigned to music lessons outperformed children assigned to Lego-building lessons on one test of inhibitory control (a matching-familiar-figures test), but not on another (a day/night Stroop task).

Shifting Two previous correlational studies have shown that musicians had smaller switch costs than non-musicians on a standard shifting task, the trail making test (TMT; Reitan & Wolfson, 1985), which was found for both older adults (Hanna-Pladdy & MacKay, 2011) and children (Zuk et al., 2014). College-aged musicians also showed better performance than non-musicians on another task-switching paradigm, in which participants shifted between indicating the quantity and the value of numbers on a screen (Moradzadeh, Blumenthal, & Wiseheart, 2014). Interestingly, however, younger adult musicians did not seem to show an advantage on the TMT as compared to non-musicians (Bialystok & DePape, 2009; Zuk et al., 2014), nor on other auditory and visual categorization shifting tasks (Slevc et al., 2016).

In experimental work, Bugos, Perlstein, McCrae, Brophy, and Bedenbaugh (2007) found that older adults who received

piano lessons performed better than a no-treatment control on the TMT. Similarly, Bugos and Kochar (2017) found that older adults who had short-term piano lessons also improved on category switching in a verbal fluency task.

Updating In correlational studies, musicians have been shown to outperform non-musicians on auditory and visual n-back tasks (Oechslin, Van De Ville, Lazeyras, Hauert, & James, 2013; Pallesen et al., 2010; Slevc et al., 2016), and adult musicians outperformed non-musicians on the backward digit span¹ (Zuk et al., 2014). However, work has also shown that musicians do not outperform non-musicians on the backward digit span among either adults (Hansen, Wallentin, & Vuust, 2012) or children (Zuk et al., 2014).

Experimental studies that have investigated the effects of music training on WM updating also revealed mixed findings. In a quasi-experimental study, Roden, Grube, Bongard, and Kreutz (2014) found that children who self-selected into music lessons outperformed children who received science lessons on two tasks tapping updating ability: a counting span test and a complex span test. Mehr, Schachner, Katz, and Spelke (2013) found that the children who had music lessons showed better performance on a map use/navigation task (in which children had to hold in mind and rotate a map/locations in order to find a toy) than did children who received visual arts lessons, but Mehr et al. did not find differences on an omnibus test of all four cognitive tasks assessed. Moreover, in a follow-up study comparing a new group of children randomly assigned to either music lessons or a no-treatment control, the children who had music lessons did not outperform those in the control group (Mehr et al., 2013).

Only a few studies have assessed multiple aspects of EFs, and these correlational studies have also yielded mixed results. Degé, Kubicek, and Schwarzer (2011) found that months of music lessons in children correlated with performance on inhibition and shifting tasks; in contrast, Schellenberg (2011) found that music lessons in children were not related to inhibition or shifting tasks. Slevc et al. (2016) found that musical ability in college-aged adults predicted performance on auditory and visual updating tasks, but not on inhibition or shifting tasks. Zuk et al. (2014) similarly found that adult musicians were better than non-musicians at updating, but not at inhibition or shifting; however, they found that child musicians were better at shifting but not at inhibition or updating.

In sum, the literature is mixed: while there is evidence that musical training is related to (or has an effect on) inhibition, shifting, and updating, this evidence is highly variable.

¹ Here we focus on working memory updating measures, and so do not include work using simple short-term memory tasks (e.g., musician advantages in forward digit span; Talamini, Carretti, & Grassi, 2016).

Why are the results so mixed? Obstacles to understanding the link between musical training and EFs

One hurdle in interpreting the literature on the cognitive benefits of music training is that each study typically investigates only one specific cognitive process. For example, Moreno et al. (2011) only measured inhibition, and Bugos et al. (2007) only measured shifting ability. Even studies looking at the same process use different tasks to measure that process. For example, inhibition might be measured with a Stroop task (e.g., Bialystok & DePape, 2009; Schellenberg, 2011) or with a go/no-go task (e.g., Moreno et al., 2011). Thus, it is unclear whether the mixed results reflect the underlying construct (inhibition) or some other difference between tasks.

A related problem is that many studies use only a single task to measure a given process. Even tasks that are, in fact, good measures of the construct of interest also involve other processes (the “task impurity problem”; Miyake et al., 2000), so the extent to which a relationship with performance on any one task actually reflects a relationship with the intended construct is unknown. For example, a go/no-go task (in which one needs to inhibit responding when given a cue) certainly involves inhibitory ability, but presumably it also requires other processes, such as sustained attention and the memory for maintenance of task instructions and goals.

Another limitation to the generalizability of the correlational studies described above is how they operationalize musical experience. For example, some of the varying definitions of musician groups that have been used are: students or graduates of the Sibelius Music Academy (Pallesen et al., 2010), people who have taken music lessons for at least half of their life (Bialystok & DePape, 2009), and children with three or more years of lessons (for 11- and 12-year-olds; Schellenberg, 2011). Other studies did not rely on group comparisons, but rather on continuous measures of musical training, such as months or years of music lessons. Yet other studies asked participants to self-report their duration of formal music training (a continuous measure) but then used seemingly arbitrary dichotomizations to create musician and non-musician groups for comparison. (Note that this last approach may be especially problematic because dichotomizing a truly continuous latent variable may result not only in low statistical power, but also in low reliability and spurious correlations; DeCoster, Iselin, & Gallucci, 2009; MacCallum, Zhang, Preacher, & Rucker, 2002.)

A similar limitation characterizes the handful of experimental studies on musical training and EFs: each musical training regimen has differed not only in the type of lessons (e.g., computerized training in Moreno et al., 2011, vs. private piano lessons in Bugos et al., 2007), but also in the duration of training (e.g., six weeks in Mehr et al., 2013, vs. 1.5 years in Roden et al., 2014), and the frequency of training (e.g., five

days a week in Moreno et al., 2011, vs. once a week in Bugos et al., 2007). Relatedly, some experimental studies have compared groups that had music lessons with no-treatment control groups, so these results need to be interpreted with caution (Shadish, Cook, & Campbell, 2002; cf. Boot, Simons, Stothart, & Stutts, 2013). These varying definitions between studies further preclude the ability to glean general conclusions.

Finally, the differences in the literature might reflect other factors (not controlled for) that are related to both musical training and the cognitive ability of interest. For example, *socioeconomic status (SES)* correlates with both musical participation (Corrigall, Schellenberg, & Misura, 2013; Kaushal, Magnuson, & Waldfogel, 2011; Norton et al., 2005; Southgate & Roscigno, 2009) and EF abilities (Hackman & Farah, 2009). *Handedness* is associated with aspects of musical ability (Kopiez, Galley, & Lee, 2006) and with EF performance (Beratis, Rabavilas, Kyprianou, Papadimitriou, & Papageorgiou, 2013), and *intelligence (IQ)* is linked with both musical training (Schellenberg, 2004; Schellenberg, 2006) and EF abilities (Friedman et al., 2006). Because these potentially confounding factors have not consistently been assessed, it is not yet clear whether specific cognitive abilities are related to musical training or whether these relationships reflect differences in other factors.

The present study

The goal of this study was to more thoroughly investigate if and how musical training is associated with EFs by improving on the shortcomings above. Specifically, to avoid the “task impurity problem,” we used multiple tasks to assess each EF component (cf. Miyake et al., 2000). This allowed us to use a latent variable approach, in which we estimated what was common between multiple tasks of each EF component, removed from task-specific effects. Additionally, to provide better comparability to other work, we used a battery of EF tasks that has previously been validated and used in other research on individual differences in EF abilities (Friedman et al., 2006; Friedman, Miyake, Robinson, & Hewitt, 2011; Friedman et al., 2008; Ito et al., 2015; Miyake et al., 2000). To avoid problems associated with binary categorizations of “musicians” and “non-musicians,” as well as any ambiguity regarding how participants define “music lessons” when estimating their lifetime musical experience, we used a continuous measure of musical training from a well-validated questionnaire on musical sophistication (Müllensiefen, Gingras, Musil, & Stewart, 2014). Finally, we assessed a number of covariates in order to better evaluate the specific relationship between musical training and EF abilities.

Table 1 Summary of the executive function task battery

Inhibition	Updating	Shifting
Antisaccade: See a cue, then respond to a briefly presented target that appears on the opposite side of the screen.	Keep track: See words presented serially and remember the most recently presented word belonging to each of 2–5 target categories.	Number–letter: Categorize letter–number pairs by letter (vowel/consonant) when they are on one side of the screen, and by number (odd/even) when they are on the other.
Stop-signal: See square or circle and indicate shape, but inhibit this response when a beep sounds (on 25% of trials).	Letter memory: See letters presented serially (with variable list lengths) and maintain the last four letters.	Color–shape: Categorize colored shapes by either color (red/green) or shape (circle/triangle), as indicated by a cue.
Stroop: Name the font color of either letter strings that are potentially incongruent color words (e.g., “blue” in red font) or strings of asterisks.	Spatial n-back: See a series of spatial locations indicated serially and indicate whether the location is the same location indicated <i>n</i> series earlier.	Category switch: Categorize words as living or non-living or as larger or smaller than a soccer ball, as indicated by a cue.

Method

Participants

A total of 161 participants² (total $N = 150$ after list-wise deletion of missing data) were recruited from the University of Maryland via flyers and the undergraduate research pool, and they received either course credit or monetary compensation (\$10/h) for their participation. All participants reported having normal hearing, not being color-blind, and being native English speakers.³ Prior to the study, all participants gave written informed consent, and all procedures were approved by the University of Maryland Institutional Review Board.

Measures

EF measures

Participants completed a battery of nine EF tasks measuring inhibition, switching, and updating abilities (see Table 1 for a summary). This battery of tasks has previously been validated and used in other research looking at individual differences in EF abilities (Friedman et al., 2006; Friedman et al., 2011; Friedman et al., 2008; Ito et al., 2015; Miyake et al., 2000), and the specific implementations of the tasks used here were the same used by Ito et al.

Inhibition tasks These three tasks all required the inhibition of different prepotent responses, despite otherwise distinct task demands.

² The final sample size of 150 was set a priori on the basis of similar participant numbers in other studies examining individual differences in cognitive processes (137 in Miyake et al., 2000; 133 in Engle, Tuholski, Laughlin, & Conway, 1999; 120 in Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; 215 in Shipstead, Lindsey, Marshall, & Engle, 2014).

³ Only 71% of participants supplied data on gender; of those, 33% identified as male.

Antisaccade: In this task (originally adapted from Roberts, Hager, & Heron, 1994), participants first saw a fixation cross in the center of the screen, then saw a cue to either the left or the right side of the fixation cross. After this cue, a numeric target (a number from 1 to 9) appeared for 150 ms, and the participant was asked to verbally indicate the target number. In the first prosaccade block, the cue and target appeared on the same side of the screen in order to create a prepotent response to this stimulus. Next, participants completed three antisaccade blocks, in which the target appeared on the opposite side of the screen of the cue. The antisaccade blocks required participants to inhibit looking toward the cue in order to see the target. Performance was measured by the proportion of correct responses in the antisaccade blocks.

Stop signal: This task (Verbruggen, Logan, & Stevens, 2008; based on Logan, 1994), required participants to fixate on a cross in the center of the screen, which was replaced by either a square or a circle. They were instructed to push left as quickly as possible if they saw a square, and to push right as quickly as possible if they saw a circle. On a quarter of the trials, participants saw the shape then heard a stop signal (a beep from the computer), and were instructed to withhold any response upon hearing a stop signal. On each trial in which the participant received a stop signal, the onset of the stop signal was adjusted until participants could correctly inhibit 50% of the responses. Performance was measured as the stop signal reaction time (SSRT), an estimate of how long it would take to inhibit an “already-initiated response” (Logan, 1994). This was calculated by finding the difference between the median reaction time for identifying the shape and the average onset time of the stop signal. This provided a measure of how much time each participant needed to accurately inhibit responding.

Stroop: In this task, participants read color words (i.e., red, blue, or green) presented on a screen, and were instructed to say the color of the font aloud (Stroop,

1935). There were congruent trials, in which the color of the font was the same as the word (e.g., the word “blue” written in blue ink); incongruent trials, in which the color of the font did not match the word (e.g., the word “blue” written in red ink); and neutral trials, in which a string of asterisks appeared (e.g., “*****” written in blue ink). The Stroop effect was the difference in means between incongruent trials and neutral trials for correct responses.

Updating tasks These three tasks all required that items in WM are constantly being added, deleted, or manipulated according to different task demands.

Keep track: In this task, adapted from Yntema (1963), participants kept track of exemplars from six different categories (i.e., relatives, countries, colors, animals, metals, and distances). In each trial, participants were presented with two to five categories (for a total of four different difficulty levels), then 15–25 exemplars belonging to those categories were shown one at a time. Participants verbally recalled the most recent exemplar they saw from each of the categories presented in the trial. Performance was calculated as the proportion of correct exemplars recalled.

Letter memory: In this task, (adapted from Morris & Jones, 1990), participants saw a string of letters (consonants only) appear one at a time. The strings were either nine, 11, or 13 letters long. For all trials, participants were instructed to say aloud only the last four letters in the string after each letter appeared. Performance was calculated as the proportion of accurate strings said aloud.

Spatial n-back: In this task (from Friedman et al., 2008), one of 12, stationary boxes flashed black, and participants were instructed to indicate whether or not that same box had flashed N trials previously. Participants completed a two-back and a three-back condition. Performance was calculated as the proportion of correct responses across both conditions.

Shifting tasks These three tasks all required set shifting between two types of binary categorization with distinct task demands.

Number–letter: In this task (adapted from Rogers & Monsell, 1995), participants saw a letter–number pair appear in one of four quadrants on the screen. If the pair appeared in one of the two top quadrants, the participant was instructed to categorize the number as odd or even, and if the pair appeared in one of the two bottom quadrants, the participant was instructed to categorize the letter as a consonant or vowel. Performance was calculated as

the switch cost, which is the difference in mean reaction time between switch trials (in which participants switched between what they were categorizing) and repeat trials (in which participants categorized numbers or letters twice in a row) on accurate trials only.

Color–shape: In this task (from Miyake, Emerson, Padilla, & Ahn, 2004), participants first saw a cue (“C” for color or “S” for shape), then saw a red or green circle or triangle. If a “C” appeared, the participant needed to indicate if the color was red or green, and if an “S” appeared, the participant needed to indicate if the shape was a circle or a triangle. Performance was calculated as the switch cost.

Category switch: In this task (adapted from Mayr & Kliegl, 2000), participants were again asked to categorize stimuli by one of two dimensions. Participants first saw a cue (i.e., a heart or crossed arrows), then a word stimulus (e.g., “alligator,” “coat,” or “knob”). If a heart appeared, the participant needed to indicate if the stimulus was living or non-living, and if crossed arrows appeared, the participant needed to indicate if the stimulus was larger or smaller than a soccer ball. Performance was calculated as the switch cost.

Musical training measures

Gold-MSI Musical training was assessed with the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014). The Gold-MSI is a self-report questionnaire that measures “musical sophistication” (i.e., a “construct that can refer to musical skills, expertise, achievements, and related behaviours across a range of facets”; Müllensiefen et al., 2014, p. 2) with questions in five subscales: musical training, active engagement, perceptual abilities, singing abilities, and emotions. This inventory has shown high internal consistency and has been validated through comparisons with a standard musical ability discrimination test and another musical self-report inventory (the Musical Engagement Questionnaire; Werner, Swope, & Heide, 2006).

The Gold-MSI musical training subscale contains seven questions regarding musical training, which include: years of instrument training, years of music theory training, regular daily practice, the number of hours practiced at peak of interest, the number of instruments played, whether compliments about music performances have been received, and whether they considers themselves a musician. Since this measure takes into account how long one has taken music lessons as well as the intensity of practice, participants’ scores from the musical training subscale provide a continuous and robust measure of

musical training (as compared to only looking at duration of music lessons).

Covariate measures Measures of socioeconomic status (SES), handedness, and intelligence (IQ) were collected because these variables have been associated with both music training and EF abilities.

Socioeconomic status: Because musical participation is unevenly distributed across SES (Corrigall et al., 2013; Kaushal et al., 2011; Norton et al., 2005; Southgate & Roscigno, 2009) and SES is a predictor of EF ability (Hackman & Farah, 2009), participants completed the MacArthur Scale of Subjective Social Status (Adler & Stewart, 2007), in which they indicated where they believed they stood (in terms of money, education, and job status) relative to others in the U.S. by placing an X on a ladder with ten rungs (i.e., on a scale of 1 to 10).⁴ This subjective measure of SES has been shown to better predict outcomes (e.g., health status) than objective measures of SES such as income or occupational status (Singh-Manoux, Marmot, & Adler, 2005).

Handedness: Because handedness has been associated with aspects of musical ability (Kopiez et al., 2006) and performance in EF tasks (Beratis et al., 2013), participants completed the Edinburgh handedness inventory (Oldfield, 1971), in which they indicated which hand they preferred to use during various activities (e.g., writing, drawing, or using scissors). This provided a continuous measure of laterality, scored on a scale from -100 (completely left-handed) to +100 (completely right-handed).

IQ: Since IQ is linked with both musical training (Schellenberg, 2006) and EF abilities (Friedman et al., 2006), the Cattell Culture Fair Intelligence Test (Scale 3B) was administered to measure fluid intelligence (Cattell & Cattell, 1960). This test contained four subtests (each with its own time constraints), in which participants completed a sequence of drawings (3 min), classified which images were different from others (4 min), completed a matrix of patterns (3 min), and chose which option of geometric drawings satisfied a given rule (2.5 min). Scores

were calculated as the proportion of correct answers given within the 12.5-min time limit of the test.

Procedure

Participants completed two sessions lasting 1.5 h each, separated by at least a day. The order of the tasks was fixed as follows: stop-signal, spatial two-back, category switch, Stroop, keep track, color-shape, letter memory, antisaccade, number-letter, and spatial three-back (following Ito et al., 2015). In this way, no sequential tasks targeted the same EF component, and any order effects were equivalent across participants. After the first five EF tasks were completed during the first session, participants filled out questionnaires (Gold-MSI, Müllensiefen et al., 2014; SES, Adler & Stewart, 2007; handedness, Oldfield, 1971), and after all the EF tasks had been completed in the second session, participants completed the IQ test (Cattell & Cattell, 1960).

Data scoring and analyses

All scores were standardized (*z*-scored). To facilitate interpretation, the scores for each task were adjusted so that higher scores would indicate better performance (specifically, the *z*-scaled scores for all three shifting tasks, the stop-signal task, and the Stroop task were multiplied by -1). For all reaction time (RT) measures, the data from accurate trials were trimmed to exclude RTs under 200 ms and above 3,000 ms. Then, RTs that were more than two standard deviations away from each participant's mean were excluded.⁵ For all three shifting tasks, trials following an incorrect response were excluded because a previous error reflected that a switch in categorization had not been achieved (following Friedman et al., 2008).

Our analysis proceeded through three stages. First, we used confirmatory factor analysis (CFA) to assess whether our data fit as expected with our theoretical models for both EFs (Miyake & Friedman, 2012; Miyake et al., 2000) and musical sophistication (Müllensiefen et al., 2014). Second, once the CFA had indicated we were capturing the expected factor structures for EFs and musical training in our sample, we used linear mixed-effects models to determine whether music training could predict the latent factors of each EF component, after controlling for the assessed covariates. Third, although the main focus and a priori predictions of our study were based

⁴ We also asked participants to self-report their parents' income and education levels (on 9-point scales) as measures of objective SES. However, multiple participants indicated that they did not know or did not wish to provide this information (specifically, 24 participants did not report their father's income, 18 participants did not report their mother's income, and two did not report their father's education level). Reassuringly, for those participants who did provide these ratings, our subjective measure of SES correlated significantly with both parental education [$r(146) = .40, p < .001$] and parental income [$r(118) = .51, p < .001$].

⁵ Note that this trimming criterion differs from that used previously for this battery (in which RTs above 3,000 ms were not excluded, any score above $3.32 \times$ a participant's median for each condition was excluded, and the accuracy measures were arcsine-transformed). However, an analysis on the untrimmed RT and arcsine-transformed accuracy measures yielded the same pattern of results, suggesting that these results are robust to the specific method of data trimming.

on relationships with musical experience, we additionally conducted a set of exploratory analyses investigating if and how EF components related to other measures of musical sophistication, as assessed by the Gold-MSI.

Results

Descriptive statistics for participant demographics, covariate measures, and the musical training measure as well as descriptive statistics for each of the executive function tasks are provided in Table 2.

Model estimation

Executive functions One strength of the present study is that each EF component of interest was measured with three different tasks in order to more accurately assess EF abilities. To ensure that our nine EF tasks were indeed tapping their intended latent constructs (i.e., inhibition, updating, and shifting), our first step in the analysis was to conduct a CFA. We then ran a similar CFA on the Gold-MSI to determine whether our musical training questions accurately reflected their intended latent construct to make sure we could meaningfully investigate how EFs and musical training were related.

Figure 1 shows a CFA on the EF tasks, with standardized factor loadings (one-headed arrows) pointing from the latent factors (circles) to each measured task (squares). A three-factor model (based on a typical division of EF into inhibition, updating, and shifting components; Friedman et al., 2006; Friedman et al., 2008; Ito et al., 2015; Miyake et al., 2000) was estimated using R (v. 3.2.4; R Core Team, 2016) with the lavaan package (v. 0.5-23.1097; Rosseel, 2012). All nine tasks loaded significantly onto the expected EF constructs, yielding a good model fit [$\chi^2(24, N = 150) = 31.18, p = .15$; CFI = .96, RMSEA = .04, SRMR = .05, AIC = 3,699.09]. A similar CFA based on a more recent model of EFs that assumes a common EF factor along with updating-specific and shifting-specific factors (Miyake & Friedman, 2012) led to essentially the same model fit [$\chi^2(21, N = 150) = 26.35, p = .19$; CFI = .97, RMSEA = .04, SRMR = .06, AIC = 3,700.59], but with a slightly higher Akaike information criterion (AIC). Furthermore, only nine of the 15 factor loadings were significant in the model with the common EF factor. Because these models seemed to capture our data equally well, and for compatibility with most previous work on music and EFs, we maintained the more parsimonious three-factor distinction in our analyses below.

Musical sophistication We conducted a similar CFA on the responses from the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014) to assess whether our participants' data fit with the distinct components of

musical sophistication proposed in previous work (i.e., musical training, general sophistication, active engagement, perceptual abilities, singing abilities, and emotions). The fit of this six-factor model showed only an acceptable model fit [$\chi^2(647, N = 150) = 1,398.99, p < .001$; CFI = .84; RMSEA = .067, SRMR = .077]. Importantly, however, all seven items from the Music Training subscale loaded significantly onto the Music Training factor. Because these data showed the expected relationships between EF components and showed a good fit for the musical training measures, we confidently relied on these measures for further analyses.

Linear mixed-effects modeling

We addressed our main question of interest—whether musical training predicts performance differentially for EF components—with a set of linear mixed-effects (LME) models. Analyzing individual differences in LME models is roughly analogous to doing so in structural equation models (SEM), but without requiring such large sample sizes (cf. von Bastian & Oberauer, 2013). These models estimate latent EF components by specifying each EF *task* as a random effect nested within *EF type* (inhibition, shifting, and updating). This essentially treats task as a sample drawn from a population of the tasks measuring each construct of interest, so the effects for any given task are assumed to deviate randomly from the group (task-population) mean.

Specifically, to assess whether musical training could predict performance differentially for EF components, we first conducted an omnibus LME model with *subjects* and *tasks* as random effects, and *EF type* (inhibition, shifting, and updating), *musical training*, and our three *covariate measures* (SES, handedness, and IQ)⁶ as fixed effects. The random effects structure included random intercepts for *subjects* and by-subject random slopes for *EF type*. *EF type* was contrast coded so that the first contrast (EF Type 1) compared shifting scores with the mean of inhibition and updating scores, and the second contrast (EF Type 2) compared the inhibition and updating scores only. Analyses were run in R (v. 3.2.4; R Core Team, 2016) with the lme4 package (v. 1.11.1; Bates, Maechler, Bolker, & Walker, 2015), and statistical significance was evaluated with the lmerTest package (v. 2.0-33; Kuznetsova, Brockhoff, & Christensen, 2016).

In this omnibus analysis, IQ was a significant predictor of task performance overall. Although there was no main effect

⁶ Handedness and SES did not significantly improve model fit, but because they were predicted a priori to be associated with musical training and were planned as covariates, they are nonetheless included in the models reported below. To guard against spurious effects produced via the inclusion of these factors (i.e., residual confounding; Kahneman, 1965; Westfall & Yarkoni, 2016), we also re-ran all models without including these covariates, which did not substantively change the pattern of results.

Table 2 Descriptive statistics for the participant demographics, covariate measures, musical training measure, and executive function tasks

Measure	Mean	SD	Min	Max	Skewness	Kurtosis	Reliability
Age	19.26	1.11	17.00	22.00	0.48	- 0.62	–
Cattell IQ	27.85	4.61	11.00	38.00	- 0.45	0.34	–
SES	6.44	1.52	3.00	9.00	- 0.27	- 0.60	–
Handedness	68.40	43.39	- 100.00	100.00	- 2.51	6.17	–
Musical training	25.29	10.25	7.00	47.00	0.10	- 0.92	.88 [@]
Antisaccade ^a	0.72	0.13	0.34	0.94	- 0.62	- 0.12	.90*
Stop-signal ^b	269.05	36.05	177.30	420.60	0.63	1.90	.72 [@]
Stroop ^c	120.27	62.46	- 37.18	320.34	0.69	0.68	.94*
Keep track ^a	0.73	0.09	0.45	0.91	- 0.47	0.12	.62 [@]
Letter memory ^a	0.77	0.14	0.38	1.00	- 0.26	- 0.79	.93 [@]
Spatial n-back ^a	0.80	0.07	0.53	0.96	- 0.49	1.06	.75*
Number-letter ^d	179.27	112.62	- 27.47	628.29	1.17	2.15	.77*
Color-shape ^d	151.51	136.93	- 69.02	760.37	1.75	4.31	.80*
Category switch ^d	120.22	92.58	- 39.49	432.45	0.88	0.55	.75*

$N = 150$ for all measures except Age ($N = 145$). ^a Proportion accuracy. ^b Stop-signal reaction time (SSRT) in milliseconds. ^c RTs for incongruent trials minus RTs for neutral trials in milliseconds. ^d RTs for switch trials minus RTs for stay trials in milliseconds. [@] Cronbach's alpha. *Brown–Peterson corrected split half.

of musical training, musical training did interact with EF type, which indicates that the extent to which musical training predicted task scores differed across EF components. Table 3 summarizes the omnibus results.⁷

To unpack the significant interaction found in the omnibus model, separate pre-planned LME models were conducted for each of the three EF components (Fig. 2). Each model used the same model structure as the omnibus analysis, except without random slopes for EF type by subjects (since only one EF type was present in each model).

Latent inhibition scores were not significantly related to musical training in the models both with and without covariates. In terms of the individual tasks, we did find a significant pairwise correlation between musical training and performance on the Stroop task ($b = -1.43$, $SE = 0.48$, $t = -2.95$, $p = .004$; see Appx. 1); however, this relationship disappeared when controlling for covariates ($b = -0.96$, $SE = 0.52$, $t = -1.86$, $p = .065$). Similarly, shifting scores (at the latent and individual task levels) were not related to musical training with or without covariates included in the model. Interestingly, updating scores (at the latent and individual task levels) were related to musical training both with and without covariates included in the model.

Exploratory analyses: EF's relationship with other aspects of musical sophistication

Although the primary goal of this study was to investigate the relationship between musical training and EFs (fitting

with the focus on experience in the extant literature), we additionally examined other types of musical sophistication that might be linked with EFs. Specifically, we conducted a set of exploratory analyses in which we ran the same LME models as above, but used other subscales from the Gold-MSI: General Sophistication, Active Engagement, Perceptual Abilities, Singing Abilities, and Emotions (Müllensiefen et al., 2014). Note that there are some limitations to these exploratory analyses. First, the CFA on the Gold-MSI had only an acceptable fit to our data. However, note that each factor showed good or excellent reliability (i.e., the Cronbach's alphas for general sophistication, active engagement, perceptual abilities, singing abilities, and emotions were .90, .83, .86, .77, and .84, respectively). Thus, it is not entirely clear whether these subscales reflect the true multidimensionality of the data. Second, these subscales are all self-report measures, which may not accurately reflect actual perceptual or singing abilities; for example, the Perceptual Abilities subscale of the Gold-MSI is only moderately correlated ($r = .51$) with perceptual abilities as measured by the Advanced Measures of Musical Audiation (AMMA; Gordon, 1989; Müllensiefen et al., 2014).

These limitations aside, some subscales might relate more to musical aptitude (e.g., the Perceptual and Singing Abilities scales), in contrast to musical experience we investigated above; therefore, the different subscales might differentially relate to aspects of EFs. Although it is not clear how much these different subscales capture variance distinct from musical aptitude and musical training, the differential relationships of the Gold-MSI subscales with EFs may nonetheless suggest interesting avenues for future research.

⁷ Variance inflation factors (VIFs) were calculated in order to address the potential concern of multicollinearity between the predictors; reassuringly, all predictors had VIFs below 1.2.

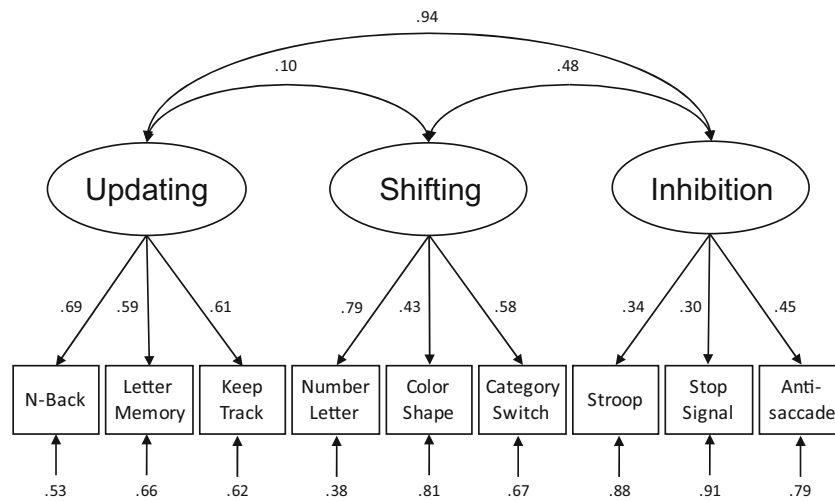


Fig. 1 Three-factor model of executive functions. Unobserved latent factors are in circles, and the measured tasks are in squares. The single-headed arrows from latent factors to the measured tasks are standardized factor loadings, which are all significant ($p < .05$). The short arrows at

the bottom represent measurement error; squaring this number gives the amount of variance in each task left unaccounted for by the latent factor. The double-headed arrows at the top represent correlations between the latent factors.

Interestingly, these exploratory results do suggest that the subscales of the Gold-MSI relate differentially to at least one aspect of EF. Table 4 below reports the beta coefficients for each Gold-MSI subscale and each latent EF factor, after controlling for SES, handedness, and IQ. The relationships between the other Gold-MSI subscales and updating appear to reflect the same positive relationship reported for the Musical Training subscale. Similarly, there are no significant relationships between any of the

Gold-MSI subscales and shifting ability. However, the relationship between musical sophistication and inhibition differs across the subscales, with stronger relationships (reaching conventional significance levels for confirmatory tests) for the General Sophistication, Active Engagement, and Singing Abilities scales, and weaker (non-significant) relationships with the Musical Training, Perceptual Abilities, and Emotions scales. This suggests that, although inhibition abilities might not be related to

Table 3 Mixed-effects model examining the effects of musical training and EF type on EF performance

Parameters	Fixed Effects				Random Effects			
	Estimate	SE	t	p	SD	EF1	EF2	SD
Intercept	-.01	.04	-0.15	.88	.36	-.19	.09	–
Musical Training	.03	.04	0.72	.47	.01	–	–	.01
EF Type 1	.01	.07	0.144	.89	.61	–	-.79	–
EF Type 2	<.01	.03	<0.01	.90	.13	–	–	–
IQ	.21	.04	5.1	<.001	–	–	–	–
Handedness	-.04	.04	-1.04	.30	–	–	–	–
SES	.01	.04	0.23	.82	–	–	–	–
Musical Training × EF Type 1	.30	.07	4.29	<.001	–	–	–	–
Musical Training × EF Type 2	-.06	.03	-2.12	.07	–	–	–	–

Factors were contrast coded as follows: EF Type 1 (–1 = Shifting, .5 = Inhibition, –.5 = Updating), EF Type 2 (0 = Shifting, .5 = Inhibition, –.5 = Updating). Under random effects, values to the right of the SD columns indicate estimated correlations between random effects. Model formula for correlated random effects model: Task Performance ~ Musical Training × EF type + IQ + Handedness + SES + (0 + Musical Training | Task) + (EF Type | Subject). * Bold p values indicate $p < .05$.

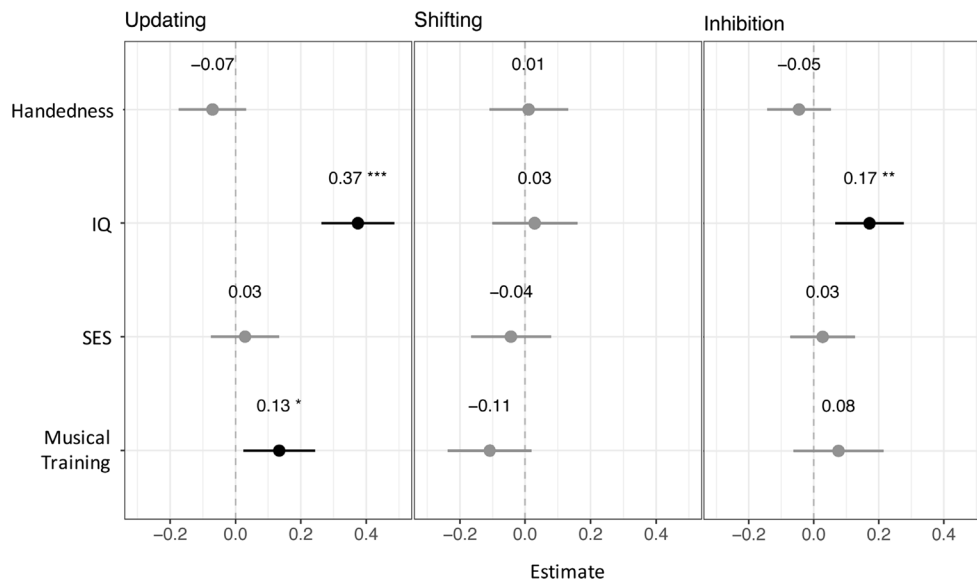


Fig. 2 The x -axis indicates the fixed effects estimates from the three linear mixed-effects models and their associated 95% confidence intervals. The y -axis contains the predictors included in each model. * $p < .05$, ** $p < .01$.

training per se (as reported above), they may still relate to other aspects of musical sophistication.

Discussion

Considerable evidence links musical training to a range of cognitive processes, including executive functions. However, it has been difficult to assess the strength of this evidence, given that most previous studies have used single tasks to measure a given cognitive process and have relied on different definitions of “musicians” and “non-musicians.” To provide a clearer picture of the relationship between musical training and EFs, we sought to improve upon these problems by measuring multiple components of EFs (i.e., inhibition, shifting, and updating) with multiple measures of each EF component, and by using a well-validated continuous measure of musical training.

Overall, these data show a positive relationship between individual differences in musical training and WM updating ability (a latent construct estimated from scores on the keep track, letter memory, and n-back tasks), but no such relationship with inhibition ability (estimated from scores on the Stroop, stop-signal, and antisaccade tasks) or shifting ability (estimated from scores on the color–shape, category switch, and number–letter tasks). The relationship between musical training and updating scores, even after accounting for the (significant) variance predicted by IQ, fits with previous

evidence that musicians outperform non-musicians on n-back tasks (Oechslin et al., 2013; Pallesen et al., 2010; Slevic et al., 2016) and that musical training leads to improvements in WM tasks (Roden et al., 2014).

In contrast, the lack of a relationship between musical training and the latent construct of shifting suggests that the relationships between musical training and shifting tasks found in previous studies (e.g., Hanna-Pladdy & MacKay, 2011; Zuk et al., 2014) may have resulted from the use of binary versus continuous measures of musical training and/or might be limited to the specific task chosen to measure shifting ability.

These findings are also inconsistent with previous evidence for a relationship between musical training and

Table 4 Beta coefficients for each Gold-MSI subscale for each EF component

Gold-MSI Subscale	Inhibition	Shifting	Updating
Musical Training	.08	– .11	.13 [†]
General Sophistication	.15 [†]	– .05	.17 [†]
Active Engagement	.13 [†]	.02	.12 [†]
Perceptual Abilities	.14	.04	.16 [†]
Singing Abilities	.15 [†]	– .03	.10
Emotions	.11	– .03	.11 [†]

Model formula for the correlated random effects model: EF Task Performance \sim GoldMSI Subscale + IQ + Handedness + SES + (0 + Gold-MSI Subscale | Task) + (1 | Subject). [†] Beta coefficient that would reach significance in conventional (pre-planned) analyses for that EF component after controlling for SES, handedness, and IQ.

inhibition. This difference might be due to IQ mediating or confounding musical/inhibition relationships, as many of the previous studies supporting a relationship between musical training and inhibition did not control for IQ (Degé et al., 2011; although note that some such studies did find that IQ did not differ across groups and therefore did not use it in subsequent analyses, Bialystok & DePape, 2009; Moreno et al., 2011; Schellenberg & Moreno, 2009). In fact, the inclusion of IQ as a covariate in the data presented here may have eliminated an otherwise positive association between musical training and tasks measuring inhibition ability: the significant pairwise correlation between music training and performance on the Stroop task (see Appx. 1) fits with some previous findings (Amer et al., 2013; Bialystok & DePape, 2009; Travis et al., 2011), but this relationship disappeared when controlling for covariates. The significant correlation between musical training and IQ emphasizes the need for studies measuring the relationship between musical training and cognitive abilities to hold IQ constant. In contrast, we did not find a significant pairwise correlation between musical training and performance on the Stop Signal task (see Appx. 1), which is similar to the go/no-go task that showed a musician advantage in Moreno et al. (2011). This may have been due to the different task designs and outcome measures used: our study used the stop-signal reaction time in a task that adapted difficulty levels based on accuracy, whereas Moreno et al. found a relationship when measuring task accuracy in a non-adaptive task.

Although these results are correlational and cannot provide evidence of directionality, they suggest three possible explanations for the relationship between musical training and WM updating. The first possibility is that musical training leads to benefits in WM updating. If the relationships observed here are transfer effects, this has important implications for the malleability of memory and cognitive abilities and the ways in which processes involved in musical training relate to these abilities. A second possibility is that individuals with high WM updating abilities are more likely to pursue and persist in continued music lessons (i.e., selection bias), in which case these results show a specific cognitive process that underlies successful musical training. And third, music training may exaggerate pre-existing WM updating abilities (i.e., a gene–environment interaction). These possibilities all postulate that musical experience draws strongly on WM updating abilities, which is presumably critical for representing the order of musical events across time (cf. Lashley, 1951). That is, WM may be

critical to maintaining and retrieving the serial order of musical elements, which is needed to represent and perform music (Palmer, 1997; Palmer & Pfordresher, 2003; Pfordresher, Palmer, & Jungers, 2007). This maintenance of serial order might be especially demanding in trained musicians, who plausibly rely on cues from a chain of serial musical elements to recall a piece from memory (Chaffin, Lisboa, Logan, & Begosh, 2009) and look farther ahead in the music while sight-reading (Meinz & Hambrick, 2010). In fact, music might be uniquely dependent on the sustained engagement of WM updating abilities to represent serial order (more so than does, e.g., language processing, in which the representation of meaning is quickly abstracted; Slevc et al., 2016), and so may be a particularly good “test case” to investigate broader questions of how WM updating underlies the representation of serial order.

The relationships (and non-relationships) between EF abilities and musical experience parallels, in some ways, work relating individual differences in EFs and language experience (i.e., bilingualism). Previous work finding EF advantages in bilinguals, particularly for shifting and inhibition (e.g., Bialystok, Craik, & Luk, 2012), contrasts with the lack of relationship with musical experience found here. One explanation for this difference may be that musical training is highly amenable to self-selection (i.e., there is often relatively little pressure to continue with musical training for individuals who do not find it rewarding), which may amplify the opportunity to observe predispositional effects among musicians. In contrast, second language learning is often required (e.g., when moving to a new language community), potentially offering a better opportunity to observe transfer effects. However, note that there is considerable debate about the existence of EF advantages among bilingual populations (e.g., Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; deBruin, Treccani, & Della Sala, 2015), suggesting that the similarly mixed results in both bodies of literature may reflect similar measurement issues and/or the highly variable experiences of both bilinguals and musicians.

These mixed patterns further highlight the need for future studies to control for extraneous variables that are related to both musical experience and other outcomes of interest. However, note that our covariate measures did not all show the relationships we expected; in particular we surprisingly did not find a relationship between musical training and SES (see Appx. 2). This did not appear to reflect a problem with our SES measure, as our subjective measure of SES was significantly correlated with our

objective measures of SES (i.e., parental education and parental income), fitting with previous findings that subjective and objective measures of SES are related, but not perfectly correlated (Adler & Stewart, 2007). Instead, this non-relationship may simply reflect the limited SES variability within our college-aged sample.

Interestingly, other subscales of the Gold-MSI revealed a slightly different picture of the relationships between musical sophistication and EFs. Updating performance was robust across most Gold-MSI subscales (except for Singing Abilities and Perceptual Abilities), and shifting ability was not related to any of the Gold-MSI subscales. However, inhibition abilities were related to some subscales of the Gold-MSI (General Sophistication, Active Engagement, and Singing Abilities) but not to others (Musical Training, Perceptual Abilities, or Emotions). This selective relationship might reflect factors besides musical experience, such as musical aptitude. However, this conclusion is clearly tenuous (e.g., it is not obvious that aptitude is implicated more in the active engagement in music and singing ability than in perceptual ability). In addition, given the modest fit of the Gold-MSI scales to the present data, it is difficult to interpret differential patterns across subscales. Nevertheless, these different patterns do emphasize the need to consider different aspects of musical ability and experience in future work.

Limitations and future directions

Despite the relatively large sample size and comprehensive measures used in this study, there are a few limitations. One type of limitation stems from our participant sample of undergraduates, who may have provided a restricted range of cognitive ability and SES. Our undergraduate participants also chose to participate in our study and could plausibly have assumed that the purpose of the study was to study music and cognition. These sampling biases limit our ability to generalize the relationships found here, and it would be beneficial for future work to replicate these results with a more representative sample. A second type of limitation comes from the task design and structure of the data. For example, due to the considerable testing time (total of 3 h), participants may have experienced fatigue and/or learned strategies to perform these tasks throughout the two sessions even though we tried to alleviate fatigue with mandatory breaks during both sessions. We also used a fixed order of task administration, which could have led to some consistent effects of order. Note, however, that this

limitation does suggest that any such effects should have occurred for all participants, allowing individuals' scores to be more directly comparable (and was chosen following other protocols investigating individual differences in EFs; e.g., Friedman et al., 2008; Ito et al., 2015).

A potential concern with the data has to do with the factor structure for both EFs and musical sophistication. Although our three-factor model of EFs showed good fit, the latent factor of updating was closely related to inhibition, which may indicate that the tasks used in this battery may be capturing updating ability better than inhibition ability. Additionally, our model of musical sophistication (Gold-MSI) showed only an acceptable model fit, so our exploratory analyses of the other factors of musical sophistication should be interpreted with caution.

Finally, because this is a correlational study, it provides no evidence as to the directionality of the relationship between musical experience and updating abilities. It may be that music training affects WM updating abilities (i.e., transfer effects), possibly through demands exerted from reading music or planning multiple aspects of performance more generally. Alternatively, individuals with higher WM updating ability may be more likely to start music lessons and choose to continue them, reflecting a selection bias (Elpus, 2013). A third possibility is that music lessons might exaggerate pre-existing differences by drawing upon (and perhaps also improving) WM updating mechanisms (i.e., a gene–environment interaction; Schellenberg, 2015).

In any case, the relationship between musical training and WM updating observed here (along with the non-significant relationships with inhibition and shifting) indicate appropriate avenues to explore in experimental studies that randomly (or pseudo-randomly) assign participants to musical training or a well-matched control training regimen. These findings also suggest specific types of pre-existing differences to explore in order to better understand who does and does not pursue music lessons and what characteristics predict who is likely to excel in musical training. More generally, these data suggest that musical experience relies on WM updating, which will hopefully motivate future work investigating exactly how and why WM updating is involved in musical experience.

Author note We thank Naomi Friedman for providing the EF test battery, and Judith Tsoi, Toria Hawkins, Arcadia Ewell, and McLaine Rich for assistance with data collection and data coding. We also thank Kenneth Elpus and Michael Dougherty for helpful discussion and comments.

Appendix 1

Table 5 Correlation matrix of EF tasks

		Inhibition		Updating			Shifting			
		Anti-saccade	Stop Signal	Stroop	Keep Track	Letter Memory	Spatial N-Back	Number-Letter	Color-Shape	Category Switch
Inhibition	Antisaccade									
	Stop signal	.11								
	Stroop	.26**	-.04							
Updating	Keep track	.24**	.24**	.15						
	Letter memory	.29***	.20*	.28***	.31***					
	Spatial N-back	.27***	.16	.21*	.47***	.39***				
Shifting	Number-letter	.18*	.15	.09	.12	.002	.06			
	Color-shape	-.04	.05	-.03	.04	-.14	.03	.36***		
	Category switch	.16*	.12	.18*	.01	-.01	.10	.44***	.25**	

Scores have been scaled so that larger number indicate better performance (switch costs, stop-signal reaction time, and Stroop effect scores have been multiplied by - 1). * $p < .05$, ** $p < .01$, *** $p < .001$. $N = 150$

Appendix 2

Table 6 Pairwise correlations between musical training, covariates, and EF task scores

		Musical Training	SES	IQ	Handedness
Covariates	SES	-.04			
	IQ	.33***	.17*		
	Handedness	-.13	.06	.01	
Inhibition	Antisaccade	.17*	.09	.29***	.02
	Stop-Signal	.01	.04	.12	-.05
	Stroop	.23**	.03	.19*	-.12
Updating	Keep Track	.23**	-.07	.38***	-.06
	Letter Memory	.30***	.13	.41***	-.15
	Spatial N-back	.27***	.19*	.48***	-.04
Shifting	Number-Letter	-.11	-.02	-.03	.10
	Color-Shape	-.07	0	-.03	-.13
	Category Switch	-.12	-.08	.02	.10

Scores have been scaled so that larger number indicate better performance (switch costs, stop-signal reaction time, and Stroop effect scores have been multiplied by - 1). * $p < .05$, ** $p < .01$, *** $p < .001$. $N = 150$

References

- Adler, N., & Stewart, J. (2007). The MacArthur scale of subjective social status. MacArthur Research Network on SES & Health. Retrieved from www.macses.ucsf.edu/Research/Psychosocial/subjective.php
- Amer, T., Kalender, B., Hasher, L., Trehub, S. E., & Wong, Y. (2013). Do older professional musicians have cognitive advantages? *PLoS ONE*, *8*, e71630:1–8. <https://doi.org/10.1371/journal.pone.0071630>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Benz, S., Sellaro, R., Hommel, B., & Colzato, L. S. (2015). Music makes the world go round: The impact of musical training on non-musical cognitive functions—A review. *Frontiers in Psychology*, *6*, 2023. <https://doi.org/10.3389/fpsyg.2015.02023>
- Beratis, I. N., Rabavilas, A. D., Kyprianou, M., Papadimitriou, G. N., & Papageorgiou, C. (2013). Investigation of the link between higher order cognitive functions and handedness. *Journal of Clinical and Experimental Neuropsychology*, *35*, 393–403.
- Bialystok, E., Craik, F., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*, 240–250.
- Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 565–574.
- Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, *8*, 445–454. <https://doi.org/10.1177/1745691613491271>
- Bugos, J. A., & DeMarie, D. (2017). The effects of a short-term music program on preschool children's executive functions. *Psychology of Music*, *45*, 855–867. <https://doi.org/10.1177/0305735617692666>
- Bugos, J. A., & Kochar, S. (2017). Efficacy of a short-term intense piano training program for cognitive aging: A pilot study. *Musicae Scientiae*, *21*, 137–150. <https://doi.org/10.1177/1029864917690020>
- Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging and Mental Health*, *11*, 464–71. <https://doi.org/10.1080/13607860601086504>
- Cattell, R. B., & Cattell, A. K. S. (1960). *Cattell Culture Fair Intelligence Test: A measure of "g."* Indianapolis, IN: Bobbs-Merrill.
- Chaffin, R., Lisboa, T., Logan, T., & Begosh, K. T. (2009). Preparing for memorized cello performance: The role of performance cues. *Psychology of Music*, *38*, 3–30.
- Collins, F. S., & Fleming, R. (2017). An NIH–Kennedy Center initiative to explore music and the mind. *Journal of the American Medical Association*, *317*, 2470–2471.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183. [https://doi.org/10.1016/S0160-2896\(01\)00096-4](https://doi.org/10.1016/S0160-2896(01)00096-4)
- Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, *4*, 222:1–10. <https://doi.org/10.3389/fpsyg.2013.00222>
- Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*, 135–149. <https://doi.org/10.1016/j.cognition.2009.08.001>
- deBruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*, 99–107.
- DeCoster, J., Iselin, A. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, *14*, 349–366.
- Degé, F., Kubicek, C., & Schwarzer, G. (2011). Music lessons and intelligence: A relation mediated by executive functions. *Music Perception*, *29*, 195–201.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.
- Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*, 1387–1388. <https://doi.org/10.1126/science.1151148>
- Dougherty, M. R., Hamovitz, T., & Tidwell, J. W. (2016). Reevaluating the effectiveness of n-back training on transfer through the Bayesian lens: Support for the null. *Psychonomic Bulletin & Review*, *23*, 306–316. <https://doi.org/10.3758/s13423-015-0865-9>
- Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, *74*, 1–32. [https://doi.org/10.1016/S0010-0277\(99\)00061-X](https://doi.org/10.1016/S0010-0277(99)00061-X)
- Elpus, K. (2013). Is it the music or is it selection bias? A nationwide analysis of music and non-music students' SAT scores. *Journal of Research in Music Education*, *61*, 175–194.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331. <https://doi.org/10.1037/0096-3445.128.3.309>
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172–179.
- Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology*, *47*, 1410–1430. <https://doi.org/10.1037/a0023750>
- 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*, 201–225. <https://doi.org/10.1037/0096-3445.137.2.201>
- Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society B*, *266*, 2435–2440. <https://doi.org/10.1098/rspb.1999.0943>
- Good, A., Choma, B., & Russo, F. A. (2017). Movement synchrony influences intergroup relations in a minimal groups paradigm. *Basic and Applied Social Psychology*, *39*, 213–238. <https://doi.org/10.1080/01973533.2017.1337015>
- Goolsby, T. W. (1994). Eye movement in music reading: Effects of reading ability, notational complexity, and encounters. *Music Perception*, *12*, 77–96.
- Gordon, E. E. (1989). *Advanced measures of music audition*. Chicago, IL: Riverside.
- Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13*, 65–73.
- Hanna-Pladdy, B., & MacKay, A. (2011). The relation between instrumental musical activity and cognitive aging. *Neuropsychology*, *25*, 378–386. <https://doi.org/10.1037/a0021895>
- Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, *11*, 466–472.
- Hansen, M., Wallentin, M., & Vuust, P. (2012). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*, *41*, 779–793.
- Holochwost, S. J., Propper, C. B., Wolf, D. P., Willoughby, M. T., Fisher, K. R., Kolacz, J., ... Jaffee, S. R. (2017). Music education, academic achievement, and executive functions. *Psychology of Aesthetics, Creativity, and the Arts*, *11*, 147–166.

- Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology, 108*, 187–218.
- Jentsch, I., Mkrtchian, A., & Kansal, N. (2014). Improved effectiveness of performance monitoring in amateur instrumental musicians. *Neuropsychologia, 52*, 117–124.
- Kahneman, D. (1965). Control of spurious association and the reliability of the controlled variable. *Psychological Bulletin, 64*, 326–329. <https://doi.org/10.1037/h0022529>
- Kaushal, N., Magnuson, D., & Waldfogel, J. (2011). How is family income related to investments in children's learning? In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 187–206). New York, NY: Russell Sage Foundation & Spencer Foundation.
- Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior, 31*, 354–364. <https://doi.org/10.1016/j.evolhumbehav.2010.04.004>
- Kopiez, R., Galley, N., & Lee, J. I. (2006). The advantage of a decreasing right-hand superiority: The influence of laterality on a selected musical skill (sight reading achievement). *Neuropsychologia, 44*, 1079–1087.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effect models (R package version 2.0-33). Retrieved from <http://cran.r-project.org/package=lmerTest>
- Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York, NY: Wiley.
- Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59–80.
- Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience, 25*, 1049–1061. https://doi.org/10.1162/jocn_a_00388
- Loehr, J. D., & Palmer, C. (2011). Temporal coordination between performing musicians. *Quarterly Journal of Experimental Psychology, 64*, 2153–2167. <https://doi.org/10.1080/17470218.2011.603427>
- Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language*. San Diego, CA: Academic Press.
- Logue, S. F., & Gould, T. J. (2014). The neural and genetic basis of executive function: Attention, cognitive flexibility, and response inhibition. *Pharmacology Biochemistry and Behavior, 123*, 45–54.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40. <https://doi.org/10.1037/1082-989X.7.1.19>
- Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1124–1140. <https://doi.org/10.1037/0278-7393.26.5.1124>
- Mehr, S. A., Schachner, A., Katz, R. C., & Spelke, E. S. (2013). Two randomized trials provide no consistent evidence for nonmusical cognitive benefits of brief preschool music enrichment. *PLoS ONE, 8*, e82007:1–12. <https://doi.org/10.1371/journal.pone.0082007>
- Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science, 21*, 914–919. <https://doi.org/10.1177/0956797610373933>
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*, 270–291. <https://doi.org/10.1037/a0028228>
- Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J.-C. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115*, 123–142. <https://doi.org/10.1016/j.actpsy.2003.12.004>
- 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*, 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*, 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Moore, G. P., & Chen, J. (2010). Timings and interactions of skilled musicians. *Biological Cybernetics, 103*, 401–414. <https://doi.org/10.1007/s00422-010-0407-5>
- Moradzadeh, L., Blumenthal, G., & Wiseheart, M. (2014). Musical training, bilingualism, and executive function: A closer look at task switching and dual-task performance. *Cognitive Science, 39*, 992–1020. <https://doi.org/10.1111/cogs.12183>
- Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science, 22*, 1425–1433. <https://doi.org/10.1177/0956797611416999>
- Moreno, S., & Bidelman, G. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research, 308*, 84–97.
- Moreno, S., & Farzan, F. (2015). Music training and inhibitory control: A multidimensional model. *Annals of the New York Academy of Sciences, 1337*, 147–152.
- Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology, 81*, 111–121.
- Mosing, M. A., Madison, G., Pedersen, N. L., & Ullén, F. (2016). Investigating cognitive transfer within the framework of music practice: Genetic pleiotropy rather than causality. *Developmental Science, 19*, 504–512.
- 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*, e89642:1–23. <https://doi.org/10.1371/journal.pone.0089642>
- Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., & Schlaug, G. (2005). Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain and Cognition, 59*, 124–134. <https://doi.org/10.1016/j.bandc.2005.05.009>
- Oechslin, M. S., Van De Ville, D., Lazeyras, F., Hauert, C. A., & James, C. E. (2013). Degree of musical expertise modulates higher order brain functioning. *Cerebral Cortex, 23*, 2213–2224.
- Okada, B. M., & Slevc, L. R. (2018). Musical training: Contributions to executive function. In M. Bunting, J. Novick, M. Dougherty, & R. W. Engle (Eds.), *An integrative approach to cognitive and working memory training: Perspectives from psychology, neuroscience, and human development*. New York: Oxford University Press. <https://doi.org/10.13016/M2GM81P70>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS ONE, 5*, e11120:1–12. <https://doi.org/10.1371/journal.pone.0011120>

- Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48, 115–138.
- Palmer, C. (2013). Music performance: Movement and coordination. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 405–422). San Diego, CA: Academic Press.
- Palmer, C., & Pfordresher, P. Q. (2003). Incremental planning in sequence production. *Psychological Review*, 110, 681–712. <https://doi.org/10.1037/0033-295X.110.4.683>
- Pfordresher, P. Q., Palmer, C., & Jungers, M. K. (2007). Speed, accuracy, and serial order in sequence production. *Cognitive Science*, 31, 63–98. <https://doi.org/10.1080/03640210709336985>
- Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effect–Schmozart effect: A meta-analysis. *Intelligence*, 38, 314–323.
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>.
- Reitan, R. M., & Wolfson, D. (1985). *The Halstead–Reitan Neuropsychological Test Battery: Theory and clinical interpretation* (Vol. 4). Tucson, AZ: Reitan Neuropsychology.
- Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General*, 123, 374–393. <https://doi.org/10.1037/0096-3445.123.4.374>
- Roden, I., Grube, D., Bongard, S., & Kreutz, G. (2014). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychology of Music*, 42, 284–298. <https://doi.org/10.1177/0305735612471239>
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207–231. <https://doi.org/10.1037/0096-3445.124.2.207>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36.
- Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, 15, 511–514. <https://doi.org/10.1111/j.0956-7976.2004.00711.x>
- Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, 98, 457–468. <https://doi.org/10.1037/0022-0663.98.2.457>
- Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology*, 102, 283–302.
- Schellenberg, E. G. (2015). Music training and speech perception: A gene–environment interaction. *Annals of the New York Academy of Sciences*, 1337, 170–177. <https://doi.org/10.1111/nyas.12627>
- Schellenberg, E. G., & Moreno, S. (2009). Music lessons, pitch processing, and g. *Psychology of Music*, 38, 209–221. <https://doi.org/10.1177/0305735609339473>
- Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *The psychology of music* (pp. 499–550). London, UK: Academic Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, 72, 116–141.
- Singh-Manoux, A., Marmot, M. G., & Adler, N. E. (2005). Does subjective social status predict health and change in health status better than objective status? *Psychosomatic Medicine*, 67, 855–861.
- 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, 952–963.
- Slevc, L. R., Davey, N., Buschkuhl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, 152, 199–211.
- Slevc, L. R., & Okada, B. M. (2015). Processing structure in language and music: A case for shared reliance on cognitive control. *Psychonomic Bulletin & Review*, 22, 637–652. <https://doi.org/10.3758/s13423-014-0712-4>
- Southgate, D. E., & Roscigno, V. J. (2009). The impact of music on childhood and adolescent achievement. *Social Science Quarterly*, 90, 4–21.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. <https://doi.org/10.1037/0096-3445.121.1.15>
- Talamini, F., Carretti, B., & Grassi, M. (2016). The working memory of musicians and nonmusicians. *Music Perception*, 34, 183–191.
- Travis, F., Harung, H. S., & Lagrosen, Y. (2011). Moral development, executive functioning, peak experiences and brain patterns in professional and amateur classical musicians: Interpreted in light of a unified theory of performance. *Consciousness and Cognition*, 20, 1256–1264.
- Verbruggen, F., Logan, G. D., & Stevens, M. A. (2008). STOP-IT: Windows executable software for the stop-signal paradigm. *Behavior Research Methods*, 40, 479–483. <https://doi.org/10.3758/BRM.40.2.479>
- von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language*, 69, 36–58.
- Werner, P. D., Swope, A. J., & Heide, F. J. (2006). The music experience questionnaire: Development and correlates. *Journal of Psychology*, 140, 329–345.
- Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS ONE*, e152719. 1–22. <https://doi.org/10.1371/journal.pone.0152719>
- Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors*, 5, 7–17.
- Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS ONE*, 9, e99868:1–14. <https://doi.org/10.1371/journal.pone.0099868>