16 Music Training, Individual Differences, and Plasticity

E. Glenn Schellenberg

Funded by the Natural Sciences and Engineering Research Council of Canada.

An exciting and relatively recent avenue of scientific investigation focuses on *plasticity*, which refers to a living organism's capacity to change due to experience. *Neuroplasticity* refers more specifically to changes in brain structure and/or function. This phenomenon became popularized after the publication of *The Brain That Changes Itself* (Doidge, 2007). Although the book was written for a general audience, it was well received, garnering positive reviews from the popular press (e.g., *The Guardian*, *The New York Times*), neuroscientists (e.g., Michael M. Merzenich, Vilayanur S. Ramachandran), and intellectuals and artists (e.g., Yoko Ono, Jeanette Winterson). It provided an accessible review of research, documenting primarily how the brain works to heal and reorganize itself in the face of trauma or atypical development, sometimes in response to specific interventions, but sometimes more or less on its own.

Because neuroplasticity has profound implications for health in general and rehabilitation in particular, it remains a focus of much neuroscientific research. Most of the available literature includes typically developing individuals as participants, however, which raises a different question: do the neuroplastic consequences that are observed after brain trauma extend to typical brains? For example, it is one thing to document how the brain reorganizes itself after left hemisphere damage, such that language use becomes more of a right hemisphere function. It is quite another thing to speculate that individual differences in experience and learning, such as differences in amount of music training, influence brain structure and function in a systematic way among typically developing individuals.

The word *systematic* is crucial. From a personal perspective, I have no doubt that my development, behavior, and brain structure were affected by waking up early before school as a child to practice the piano, on a daily basis, from the age of 5 until I was 16. It is much less clear, however, that a similar history of childhood experience would engender similar effects for another person. After all, development is the result of an interaction between genes and the environment, such that the consequences of years of piano training would almost certainly be influenced by preexisting traits and behaviors (e.g., Ullén, Hambirck, & Mosing, 2016).

The focus of the present chapter is on music training, and whether it has systematic consequences that extend beyond musical knowledge and ability, which are obvious outcomes, to nonmusical cognitive abilities, which are far less obvious. The overarching thesis is that a focus on plasticity, particularly by neuroscientists, has led to an imbalance between the relative emphasis placed on nature and nurture—a kind of radical environmentalism. I use the term "radical" to describe a tendency to interpret correlational findings as evidence of causation, specifically that music training causes systematic effects on brain development, which then extend to behavior. This interpretation, which ignores the role of preexisting individual differences, is further belied by (1) an apparent obliviousness about the genetic contribution to most human behaviors and traits, and (2) centuries of evidence that near-transfer effects tend to be relatively small, whereas far transfer is virtually nonexistent.

Genetics

In this section, I argue that musicians are as much born as they are made. Claims that music training represents a good or an ideal model for the study of plasticity (e.g., Herholz & Zatorre, 2012; Hyde et al., 2009; Jäncke, 2009; Münte, Altenmüller, & Jäncke, 2002; Schlaug, 2001; Strait & Kraus, 2014) rest on the assumption that musically trained and untrained individuals differ *only* in music training, which is untrue. Rather, music training is at least partly confounded with other variables, including cognitive abilities, personality, and demographics (Corrigall, Schellenberg, & Misura, 2013). In other words, whether one becomes a musician is not akin to random assignment. Consequently, comparisons of musically trained and untrained individuals cannot lead to clear interpretations of a causal role for the training, unless researchers (1) assign individuals randomly to music training and appropriate control conditions for many years, which is unadvisable because of attrition and artificiality, or (2) measure all possible confounding variables with perfect accuracy, hold them constant in the statistical analyses, and argue persuasively that the reverse causal direction is implausible. Some scholars claim that evidence of an association between "dose" (years of music training) and "response" (performance on a nonmusical variable, size of a brain region) allows for inferences of causation. This is also untrue because confounding variables have a similar dose-response association with music training (Corrigall et al., 2013).

General cognitive ability, which is typically measured with tests of IQ, has a strong genetic component (Deary et al., 2012; Mackintosh, 2011; Plomin & von Stumm, 2018). Contrary to what one might expect, the genetic contribution increases as individuals age, with heritability reaching over 60% by old age (Deary et al., 2012; McClearn et al., 1997). This increase—or genetic *amplification* (Plomin & DeFries, 1985)—is thought to be due to the fact that as individuals age, they are more likely to be found in environments that match their genetic potential (i.e., a gene-environment correlation; Scarr & McCartney, 1983). In other words, as we get older, genes increasingly determine the environments we are in, which in turn magnify our genetic predispositions.

Music training is likely to work similarly, with predispositions (re: general cognitive ability, personality, and music aptitude) influencing who takes music lessons, which then, potentially, magnify these predispositions. Socioeconomic status (SES) also plays a role, because music lessons cost money, and parents need to be supportive and cooperative. Although SES seems likes a prime example of an environmental influence, IQ predicts many markers of SES, such as years of education, income, occupational status, and lifetime achievement (for review see Mackintosh, 2011; Wai, Worrell, & Chabris, 2017). In short, because SES co-varies with IQ, it is a variable that incorporates influences of genes *and* the environment.

Another variable that co-varies with music training is personality, particularly the trait called *openness-to-experience*. Openness refers to intellectual curiosity, or an interest in new ideas, novelty in general, and aesthetics and the arts. The genetic contribution to openness is substantial, although slightly smaller than it is for general cognitive ability (Bouchard & McGue, 2003; Power & Pluess, 2015; Vernon, Martin, Schermer, & Mackie, 2008). Moreover, twin studies confirm that the association between openness and cumulative duration of music practice is higher among monozygotic than dizygotic twins (Butkovic, Ullén, & Mosing, 2015). Finally, and perhaps most importantly, music aptitude is a marker of general intelligence (Swaminathan & Schellenberg, 2018a; Swaminathan, Schellenberg, & Khalil, 2017; Swaminathan, Schellenberg, & Venkatesan, 2018). This correlation appears to be explained by genetics but not by shared or nonshared environment (Mosing, Pedersen, Madison, & Ullén, 2014). A different genetic component provides additional but independent explanatory power of associations among different tests of aptitude, specifically those that measure the discrimination of tone sequences based on rhythm, melody, or pitch (Mosing, Pedersen et al., 2014). Heritability estimates for these tests range between 12% and 59%, with shared-environment effects evident only on the pitch task, and only for males (Ullén, Mosing, Holm, Eriksson, & Madison, 2014). In short, unless we assume that music aptitude is unrelated to music training, which is nonsensical, taking music lessons has at least two genetic components: one related to general intelligence, the other to music-specific listening skills.

The contribution of genetics to music training is further documented by findings showing that musical skill and achievement are much more than just practice (Hambrick & Tucker-Drob, 2015; Macnamara, Hambrick, & Oswald, 2014), as some scholars used to claim (Ericsson, Krampe, & Tesch-Römer, 1993; Howe, Davidson, & Sloboda, 1998). Notably, twin studies reveal that the link between practicing music and musical ability is stronger among monozygotic than among dizygotic twins (Mosing, Madison, Pedersen, Kuja-Halkola, & Ullén, 2014). In fact, individual differences in practicing music—typically considered to represent an environment influence—are actually heritable to a substantial degree (i.e., 40%–70%; Hambrick & Tucker-Drob, 2015; Mosing, Madison et al., 2014). Moreover, when differences between monozygotic twins are analyzed on their own (thereby ruling out a role for genetics), practice is unrelated to musical ability as measured by an aptitude test (Mosing, Madison et al., 2014). Nevertheless, for individuals who have the genetic potential, practice is essential to becoming musically accomplished (Hambrick & Tucker-Drob, 2015).

Other findings from twin studies suggest that the link between amount of music training and fluid intelligence "is mostly due to shared genetic influences" (Mosing, Madison, Pedersen, & Ullén, 2016, p. 504). This result raises doubts about the proposal that music training causes increases in general cognitive ability. Rather, higher-functioning individuals may be more likely than other individuals to take music lessons and practice music, particularly for long durations of time. Recent results reveal, moreover, that identical twins are more likely than fraternal twins to play the same instrument and the same genre of music (Mosing & Ullén, 2018). In other words, genetics appears to play a role in the instrument one plays, and the genre of music one chooses to play.

Transfer

Transfer refers to situations in which learning and knowledge in one domain lead to faster learning or better performance in a different domain. *Near* transfer occurs between domains that are closely related, such as if learning and improvement on one test of working memory (e.g., n-back) lead to better performance on a different test of working memory (e.g., counting span). *Far* transfer, by contrast, occurs between two domains that differ substantially, such as if working memory training leads to better performance on a test of fluid intelligence. Claims that music training leads to benefits in nonmusical cognitive domains are claims of far transfer.

The concept of far transfer is central to the belief in a liberal-arts education. Most people in the developed world think that completing a university degree has cognitive benefits that extend beyond the actual courses one takes. For more than 100 years, however, results from laboratory studies indicate that far transfer is much less likely than near transfer, and that transfer is most likely when the learning and transfer domains have considerable overlap (Thorndike & Woodworth, 1901). In applied contexts, interventions such as *Head Start* are based on the idea that far transfer can ameliorate poor cognitive abilities that are often evident in young children from poor families. Nevertheless, these sorts of interventions have only modest success at best, and there is almost no evidence of long-term cognitive benefits (Love, Chazan-Cohen, Raikes, & Brookes-Gunn, 2013; U.S. Department of Health and Human Services & Administration for Children and Families, 2010).

In the present climate of enthusiasm for plasticity, commercial software developers and scholars have resurrected interest in the possibility of far-transfer effects and stimulated much scholarly debate. An overview of publicly available "brain-training" programs (e.g., *CogMed*, *Lumosity*) came to a conclusion, however, that closely matched the one reached by Thorndike and Woodworth 115 years earlier (Simons et al., 2016). Training clearly improved performance on the actual task that was trained, but the effects were much weaker for closely related transfer tasks, and virtually nonexistent for distantly related tasks.

Basic research on far transfer has focused on working memory training in the laboratory. In a recent meta-analytic review (Melby-Lervåg, Redick, & Hulme, 2016), working memory training led to improvements on non-trained tests of working memory, but these were short lived. There was no evidence of far-transfer effects, however, particularly for studies that had *active* control groups. (With *passive* control groups, observed effects could be due to other aspects of the training program.) Finally, the magnitude of improvements in working memory was not related to the magnitude of far-transfer effects, which undermines the possibility that far transfer actually occurred.

Some findings suggest that positive results from working memory training are evident with particular learning protocols (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Buschkuehl, Jonides, & Shah, 2011). For example, one meta-analysis concluded that n-back tasks lead to improvements in fluid intelligence (Au et al., 2015). A more detailed follow-up meta-analysis, which included a larger number of original studies, reached a different conclusion. N-back training led to improvements on novel n-back tasks, but transfer effects to other working memory tasks and to fluid intelligence were minimal and independent of the amount of training (Soveri, Antfolk, Karlsson, Salo, & Laine, 2017). In a meta-analysis of studies with typically developing children from 3–16 years of age, working memory training transferred to other tests of working memory, but far-transfer effects to fluid intelligence and academic skills were very small (Sala & Gobet, 2017c). The authors concluded, therefore, that "far transfer rarely occurs and its effects are minimal" (p. 671). This result is germane to the present chapter because music lessons are usually taken by typically developing children.

One might wonder about the applied relevance of transfer effects. Perhaps working memory training is more effective in this regard among people who need it the most. If this were the case, targeted interventions could be important for ameliorating cognitive deficits, even if they are less important more generally. Although this is a reasonable hypothesis, the evidence actually suggests otherwise. For example, in a study that included 23 training sessions, adult participants who began the training with good working memory abilities actually improved the most, and there was no evidence of far transfer to different tasks (Foster et al., 2017).

In short, although tests of fluid intelligence (e.g., Raven's matrices) place demands on working memory, actual training in working memory does not seem to influence performance reliably. Even near-transfer effects to other tests of working memory are transient. Learning and performing music also place demands on working memory, but the training process is much less focused. If intensive laboratory-based procedures fail to produce far transfer, one has to question why music training would lead to particularly *distant* transfer effects. Perhaps the long timescale (i.e., year of lessons compared to weeks of lab-based training) is implicated, or the fact that much of nonmusical learning (re: working memory and other executive functions) is implicit rather than explicit. Music training also involves sensory-motor integration and goal-oriented decision making, which could improve performance on tests of fluid intelligence. In any event, causal evidence regarding these hypotheses is difficult to obtain.

Music Training and Nonmusical Cognitive Abilities

Music training is associated positively with performance on a wide variety of nonmusical tasks. Consequently, positive findings are plentiful and research in this area has been rambunctious for several years. For example, a search of *PsycINFO* (29 May 2018) with keywords "music training" or "music lessons" revealed 455 sources published since the year 2000. My colleagues and I have previously provided detailed reviews of the documented associations (Schellenberg, 2016; Schellenberg & Weiss, 2013; Swaminathan & Schellenberg, 2016, 2018b). In each case, we concluded that music training had moderate positive associations with performance on tests of general cognitive ability (e.g., IQ, working memory), language ability (speech perception, vocabulary), and visuospatial skills (visual search, mental rotation).

We also concluded that the evidence for a causal role for music training is very weak. Recent reviews from other research teams reached similar conclusions (Benz, Sellaro, Hommel, & Colzato, 2016; Dumont, Syurina, Feron, & van Hooren, 2017), specifically that there is suggestive evidence that music training improves cognitive abilities, but that the jury is still out regarding the causal role of music training and the underlying mechanisms. By contrast, Costa-Giomi (2012, 2015) considers the evidence showing that music training confers intellectual benefits to be *convincing* in the short-term (after 1 or 2 years of lessons), but she also notes that nonmusical individual differences complicate the issue of longer-term effects by influencing who takes music lessons and practices for years on end. A notable exception is that music training, particularly when it focuses on rhythm perception, appears to improve listening skills that are required for perceiving and isolating the sounds of speech (i.e., phonological awareness), at least for some populations (e.g., young children, children with dyslexia).

In the review of the literature that follows, I focus on evidence from longitudinal studies published since 2000. Before I begin, let me summarize the issues that inform my critique. Some scholars argue that longitudinal research allows for inferences of causation even when participants are not assigned randomly to the music training and control conditions (e.g., Hyde et al., 2009; Tierney, Krizman, & Kraus, 2015). Their point is that if group differences are absent before the intervention begins, any differences that are evident afterward must be due to the different experiences. This view ignores the possibility of genetic *innovation*, specifically that some genetically determined behaviors emerge later in development (Plomin, DeFries, Knopik, & Neiderhiser, 2016). For example, gene-influenced individual differences in a nonmusical ability might be evident at 6 years of age but not at 5 years. Moreover, other environmental effects that are correlated with music training, such as SES, could affect phenotypical behavior differently at different points in time. Affluence, for example, could have little to no effect on a toddler's personality but a large effect for a teenager. In short, it is cavalier to assume that children and their families who opt to take music lessons are identical to other families, except for the decision to enroll in music training.

A related complication when self-selection is involved concerns the way group equivalence is determined before the intervention. Simply documenting that groups do not differ significantly (with $p > .05$) on some variables is not the same thing as "matching" groups. This is a problem in longitudinal designs (e.g., Norton et al., 2005; Habibi et al., 2014), but even more so for cross-sectional designs when potential confounding variables are not held constant in the analysis (e.g., Mongelli et al., 2017).

A separate issue concerns how participants are assigned to the intervention and control groups. Rather than assigning children randomly and individually (e.g., Schellenberg, 2004), it is often more convenient or practical to provide an intervention to preexisting *groups* of children (e.g., some kindergarten classes), while assigning other groups (other kindergarten classes) to the control condition (e.g., Gromko, 2005; Jaschke, Honing, & Scherder, 2018; Portowitz, Lichtenstein, Egorova, & Brand, 2009; Rauscher & Zupan, 2000). Even though these groups may be assigned randomly to the conditions, the design is sub-optimal because other factors that distinguish the groups (e.g., teaching quality, intragroup dynamics) could influence whether an intervention is successful. Thus, studies with group assignment designs are not considered further.

Another complicating factor is the choice of an appropriate control condition. Often, the control group does nothing in place of the music training intervention. Such *passive* control groups preclude the possibility of a clear interpretation of subsequent group differences, which could have stemmed from nonmusical aspects of the experience (e.g., more contact with an adult, more time spent in a structured learning environment). Ideally, the control group should be involved in some nonmusical training that is as similar as possible to music lessons, but without the music. For example, an *active* control group could involve painting training, drama lessons, or instruction in a foreign language, taught at approximately the same time of day as the music lessons, at a similar location, with similarly qualified instructors.

In the selective overview that follows, I first review studies with more-or-less optimal designs, followed by those with designs that are less than optimal (i.e., passive controls, self-selection into music training).

Longitudinal, Random Assignment, Active Controls

Let us first consider studies that included random assignment and an active control group. Schellenberg (2004) recruited 144 6-year-olds who were assigned randomly and individually to a year (36 weeks) of weekly keyboard, vocal (Kodály method), drama (active control), or no lessons (passive control). All lessons were provided free of charge and taught in groups of six children, at the same location, with similarly qualified instructors. Attrition over the course of the study was moderate (8.3%), leaving 132 children for the data analysis. Pre- to post-test improvements in IQ were greater for the children in the two music groups, who did not differ, compared to children in the two control groups, who did not differ. More detailed analyses revealed that the two music groups had larger improvements than children in the no-lessons group, but direct comparisons with the drama group led to inconsistent results (i.e., null, marginal, or significant), depending on the analysis (Schellenberg, 2005–2006). At post-test, it became apparent from parent reports that the children practiced minimally between lessons, which raises questions of ecological validity. The same children were invited back to take a test of their ability to decode the emotions conveyed by prosody in speech (Thompson, Schellenberg, & Husain, 2004). For one comparison (anger vs. fear), the keyboard and drama children outperformed the control children. It is unclear why the keyboard and vocal group performed differently, but attrition was substantial (only 30% of the original 144 participated) so the findings are equivocal.

In a large-scale attempt to replicate and extend the original Schellenberg (2004) findings to academic achievement (mathematics and literacy), 909 2nd-graders from 19 different schools in the UK were assigned randomly to string lessons (violin or cello), singing lessons based on the Kodály method, or drama lessons (Haywood et al., 2015). Children were pretested with standardized tests at the end of 1st grade, and post-tested a year later after taking 32 weeks of weekly, 45-min lessons, in groups of approximately 10 children. Attrition was modest (10.5%), leaving 814 children in the sample at post-test. Children from all three groups performed similarly at post-test (controlling for pre-test scores), and neither music group had larger improvements in mathematics or literacy compared to the drama group. In fact, effect sizes were close to 0 ($ds < .05$), even when the two music groups were collapsed and compared to the drama group. The findings were identical when the analyses were limited to children from low-SES families. Unfortunately, details of the study were published in an "evaluation report and executive summary"[1], but not in an academic journal. Nevertheless, the information provided in the report suggests that the design, method, and analysis were meticulous. Although one can never prove the null hypothesis, the power afforded by the large sample size implies that if music confers nonmusical, cognitive benefits that extend to academic achievement, such effects are very small indeed.

Besson and her colleagues assigned children individually to music or painting training. Instead of true randomization, the authors used pseudo-randomization, to ensure that the two groups were equivalent at pre-test on the measures of interest. In the first study, after 6 months of two 75-min lessons per week, 8-year-olds in the music group could read irregularly spelled words better than children in the painting group, but on two other reading tests, the groups had similar improvement (Moreno et al., 2009). The music group was also better at detecting subtle pitch anomalies in speech (one word in a sentence shifted in pitch), but not in music (one note in a melody shifted in pitch). Finally, the music group had stronger electrophysiological (ERP) responses to pitch anomalies in speech and music. Because the music and speech tasks involved changes in pitch, these results are best considered as examples of near transfer. The reading result appears to provide evidence of far transfer, but the findings are far from conclusive because of the small sample sizes (n = 16 per group).

In a second study, the design was similar except that 8-year-olds took music or painting lessons for two years, six months per year. At the end of the study, the children in the music group had larger mismatch negativity (MMN) responses to syllables that were altered in duration or voice-onset time, but not in vowel frequency (i.e., pitch), which seems counter-intuitive (Chobert, François, Velay, & Besson, 2014). After the first year, the music group was also better at a task that required them to identify whether sequences of three syllables were similar to those heard during an exposure phase (François, Chobert, Besson, & Schön, 2013). This advantage was even greater by the end of the second year, and ERP responses paralleled the behavioral results. During the exposure phase of the task, however, individual syllables were matched one-to-one with different tones, which likely provided a better learning cue for the music group than for the painting group. In other words, the data provide no behavioral evidence for far transfer, and the electrophysiological data are confusing. As in the first study, the sample sizes were small (n = 12 per group).

Another pair of studies tested whether six weeks of child-centered music or visual-arts training leads to cognitive benefits among 4-year-olds (Mehr, Schachner, Katz, & Spelke, 2013). Lessons were provided to groups of 7–8 children, each of whom was accompanied by a parent. In the first study, children in the music group (n = 15) had marginally higher performance at post-test on one measure of spatial abilities (map use/navigation), whereas children in the visual-arts group (n = 14) had marginally higher performance on a second measure (visual form analysis). The second study was the same except that children in the music group (n = 23) were compared to a passive control group (n = 22), and no effects were found. Null findings also emerged when children from the two studies were combined. The null result could be due to the small sample sizes, or because children had only 4.5 hours of training in total. Nevertheless, if there is an effect of music training on nonmusical cognitive abilities, it appears to be relatively small.

Two other studies used pedagogies that were markedly different from those of typical music lessons. Degé and Schwarzer (2011) asked whether phonological awareness could be improved by music training. Phonological awareness is an important prerequisite for learning to read. The authors assigned 5- and 6-year-olds to 20 weeks of daily 10-minute training in music (primarily listening), explicit training in phonological awareness, or sports (participant numbers per group, ns = 13–14). Improvement from pre- to post-test was virtually identical for the music and phonological-awareness groups, but no improvement was evident for the sports group. These results are some of the most clear-cut in the literature, but the samples were small, and attrition was substantial (25%). Nevertheless, the authors successfully replicated the findings with a new sample of children from immigrant families (Patscheke, Degé, & Schwarzer, 2016).

Finally, Moreno et al. (2011a) assigned preschoolers to computer-based training in music listening or visual arts, five days per week for four weeks. Children were pre- and post-tested on measures of vocabulary, spatial ability (block design), and attention/inhibition (go/no-go). Improvements were evident for only the music group on the measures of vocabulary and attention/inhibition. In the latter case, ERPs were also correspondingly larger for the music group. Another test that required children to match arbitrary symbols with words showed inconsistent results: The music group had larger improvements in one analysis (ANCOVA) but not in another (mixed-design ANOVA; Moreno, Friesen, & Bialystok, 2011b). In a follow-up study, the visual-art (control) program was replaced with a second language (French) program (Janus, Lee, Moreno, & Bialystok, 2016). Both groups showed similar improvements over the four weeks on tests of verbal and nonverbal executive functions. In other words, the French-language control program was as beneficial as the music program.

Other results suggest that phonological awareness and early reading skills can be enhanced among atypically developing children—children with dyslexia—after they take music lessons that focus specifically on *rhythm*. A core deficit in dyslexia appears to be one of temporal processing (Goswami, 2011), such that the deficit in reading ability is predicted by temporal-processing difficulties in speech (Leong & Goswami, 2014) and in music (Flaugnacco et al., 2014; Goswami, Huss, Mead, Fosker, & Verney, 2013; Huss, Verney, Fosker, Mead, & Goswami, 2011). Flaugnacco et al. (2015) assigned 8- to 11-year-old children with dyslexia to seven months of training in music or painting for two hours per week. The music training was based on the Kodály method, but modified to focus on rhythm and temporal processing. After the intervention, the music group had larger improvements on tests of rhythm skills, as one would expect. More importantly, the music group also had larger improvements in phonological awareness and on tests that required them to read aloud text or pseudo-words.

If we consider these "best designed" studies as a whole, what can we conclude? The Schellenberg (2004) results are weak, without successful replication for almost 15 years. The null results from the large UK study are particularly disheartening (Haywood et al., 2015). If music lessons cause increases in nonmusical cognitive ability, the effect appears to be very small. Besson's studies (Chobert et al., 2014; François et al., 2013; Moreno et al., 2009) suggest that electrophysiological responses to speech become stronger and more reliable as a consequence of music lessons, but the behavioral results are weak, perhaps because of small samples. The most reliable results come from studies of rhythm-based training, which appear to improve phonological awareness among young children who are learning to read, and among children with dyslexia who have difficulty reading. Such improvements may, in turn, lead to improvements in reading.

It remains unclear, however, just how "musical" the music training has to be in order to see these effects. For example, Thomson, Leong, and Goswami (2013) compared a seven-week, rhythm-based intervention to one that used commercial software specifically designed to improve phonological awareness. A third group was a passive control group. Both interventions improved phonological awareness relative to controls. The most musical components of the rhythm intervention involved (1) copying a rhythm on a drum as a warm-up activity, and (2) moving the middle tone of a three-tone sequence forward or backward in time to make the sequence isochronous. In short, music per se may not be necessary to see beneficial effects of rhythm training on phonological awareness, although incorporating music into the training regimen may make the experience more enjoyable.

Finally, some findings stand out as anomalies. For example, one would expect that better phonological awareness leads to improvements in reading aloud nonwords, for which grapheme-phoneme matching is regular, but not necessarily to reading irregularly spelled words (e.g., thyme, cello; Moreno et al., 2009). Moreover, there is no obvious mechanistic explanation that would motivate one to predict that short-term but relatively intense music-listening training improves vocabulary (Moreno et al., 2011a).

Longitudinal, Random Assignment, Passive Controls

The next group of studies included random assignment, which eliminates the role of self-selection, but passive control groups, which make the findings impossible to interpret unequivocally. In one such study, Iranian 5-year-olds were assigned randomly to 13 weeks of music lessons (Orff method) or to no lessons (Kaviani, Mirbaha, Pournaseh, & Sagan, 2014). Both groups (*ns* = 30) were matched for age, gender, and SES, and they took the Farsi version of the Stanford-Binet IQ test before and after the intervention. The children in the music group had larger increases in IQ compared to the control group, which stemmed from greater improvement in visual/abstract and verbal reasoning. Although these results parallel the findings of Schellenberg (2004), the increases in performance cannot be attributed without doubt to music training. Other interventions could have the same effect.

In another study of low-income Hispanic children living in Los Angeles, Kraus and her colleagues recruited families of 6- to 9-year-olds, who were on a waiting list for a community-based music program. Enrollment in the study guaranteed a place in the program either right away (Group 1) or a year later (Group 2), with group assignment determined pseudo-randomly. Thus, Group 2 served as a passive control group during the first year. At the end of the year, Group 2 exhibited a *decline* in age-normed reading level, which is normal in this population, but Group 1 did not (Slater et al., 2014). After the second year, Group 1 had larger improvements in speech-in-noise perception compared to Group 2 (Slater et al., 2015). The authors concluded that music training causes improvement in the perception of speech in noise after two years of training but not after one year. In other words, more training was associated with better performance.

The results are less than compelling, however, because in a sample of low-SES children, the structure and routine of being involved in any extra-curricular activity could improve motivation, self-esteem, and performance on many tests. This perspective helps to explain the widespread popularity of *El Sistema*, and why children who were more engaged in the music program (i.e., with the best attendance and participation) also tended to show the largest increases in reading ability and strongest neural encoding of speech at the end of the study (Kraus, Hornickel, Strait, Slater, & Thompson, 2014). The very small sample sizes (n = 19 per group at the end of the second year in Slater et al., 2015) and the data analysis are also problematic. Instead of using multi-level modeling for the speech-in-noise data, the authors used repeated-measures analysis of covariance, which leads to interpretative problems (i.e., distorted estimates of the within-subject variable, increased Type I error, or reduced power), particularly if the covariates are not centered (Schneider, Avivi-Reich, & Mozuraitis, 2015). Finally, of the 80 participants who were tested at the beginning of the study, fewer than half (47.5%) were included in the final data analyses.

A group of Spanish researchers compared the development of phonological awareness among 4-year-olds, who were randomly assigned to eight weeks of phonological training, combined phonological *and* music training, or no training (Herrera, Lorenzo, Defior, Fernandez-Smith, & Costa-Giomi, 2011). Improvements in phonological awareness were greater in the two intervention groups than in the control group. The group with phonological and music training, however, had the best performance on tasks that required rapid naming or identifying word-final sounds. In this instance, incorporation of music into the intervention may have made the phonological training more engaging. Thus, nonmusical pedagogical improvements could have a similar effect.

Longitudinal, Self-selection

Another longitudinal design involves following children from families who choose to begin taking music lessons at some point in time. In one study of this sort, researchers from the Boston area recruited 70 5- to 7-year-old children, approximately half of whom were just beginning to take weekly, private music lessons. The children were tested twice: once at the beginning of the study (Time 1) and again 15 months later (Time 2). The testing battery included a variety of cognitive tests, tests of listening ability and music aptitude, and structural MRI scans. At Time 1, the children in the music group came from higher-SES families compared to children in the control group, and they were also slightly older (Norton et al., 2005). After controlling for age and SES, the children did not differ on any other measure, and the groups were matched for handedness and gender. Nevertheless, the means were higher in absolute terms for the music group for all seven of the behavioral tests (Table 16.2), which is significant with a two-tailed binomial test ($p = .016$). As one might expect, music aptitude was correlated with measures of general cognitive ability and with phonological awareness.

At Time 2, fewer than half of the original children (31 of 70) were included in the analyses (Hyde et al., 2009). Structural brain differences now distinguished the two groups, and these were correlated with changes in motor skills and music aptitude. Although the authors attributed these results to experience-dependent plasticity, it is also possible that pre-existing group differences, which were evident at Time 1, became exaggerated over a period of 15 months, such that they became evident in the brain scans. At the very least, the results highlight the interpretive problems that arise from longitudinal designs when groups are formed naturally.

In a similar study conducted in Los Angeles, researchers compared the development of 5- to 6-year-olds who enrolled in a community music program modeled on El Sistema (music group), to same-age children who registered for swimming or soccer classes (sports group). A third, control group did not have any intensive extra-curricular activity. All children were recruited from low-SES areas of the city. At the beginning, the three groups did not differ on a variety of cognitive, neural, or social measures (Habibi et al., 2014). After one year, near transfer was evident in the sense that the music group had improved pitch perception and production ability, but the control group showed *poorer* singing and pitch discrimination, which is difficult to interpret (Ilari, Keller, Damasio, & Habibi, 2016). After two years, the music group had structurally different brains than the other children (Habibi et al., 2017), which the authors attributed to the different training experiences. The music group was also better than the sports group at detecting an altered tone in a melody, but only marginally different from the control group (Habibi, Cahn, Damasio, & Damasio, 2016). Electrophysiological responses (ERPs) to musical notes were also more mature among the music group. In short, in the absence of clear behavioral evidence, even for near transfer, it is impossible to interpret structural and functional changes in the brain. The small samples sizes at the end of two years ($ns \leq 13$ and 20 in Habibi et al., 2016, 2017, respectively) undoubtedly contribute to this interpretive difficulty. Moreover, the different sample sizes across four reports from the same study are disconcerting.

In a study of 66 Finnish 5- to 6-year-olds, children were recruited from 26 different kindergartens that also functioned as daycare centers (Linnavalli, Putkinen, Lipsanen, Huotilainen, & Tervaniemi, 2018). Some of the kindergartens offered music playschool classes during daycare hours, others offered dance lessons, and some offered neither music nor dance training. All of the children in the control group (no music, no dance) came from kindergartens that offered neither program, which meant that the intervention was confounded with the specific kindergarten. Children were tested four times over a two-year period on measures of phonological awareness, vocabulary, perceptual reasoning (block design and matrix reasoning), and inhibition. Because children could enter and leave the programs at will, or take private music or dance lessons outside of school, these predictors were treated as continuous variables (i.e., duration of time in music or dance classes). As noted earlier, duration of music training is correlated with SES, cognitive abilities, and personality (Corrigall et al., 2013).

Improvement over time on the measures of phonological awareness and vocabulary were larger as duration of music training (compared to no music training) increased, but there was no association with perceptional reasoning or inhibition. Other findings indicated that children from high-SES families with higher vocabulary scores—across time points—were more likely to take dance lessons for longer durations of time. Similarly, high-SES children with higher perceptual reasoning scores—across time points—were also more likely to take music *and* dance training for longer periods of time. These data clarify that self-selection into arts lessons plays a major role in the outcomes, despite the authors conclusion that "music playschool *enhances* children's linguistic skills" (emphasis added).

Kraus and her colleagues used a similar design to study 68 high-school freshmen in low-SES areas in Chicago (Tierney et al., 2015; Tierney, Krizman, Skoe, Johnston, & Kraus, 2013). As part of the curriculum, students were required to choose between taking a music course (band or choir) or enrolling in the Junior Reserve Officers Training Corps (JROTC), which focused primarily on fitness. The students were tested at the beginning of high school and again two and three years later, but only 43 (after 2 years) and 40 (after 3 years) of 68 were included in the analyses (e.g., many students had moved from music to JROTC or vice versa). After two years, neural encoding of a speech sound presented in noise was more rapid in the music group (Tierney et al., 2013). After three years, outcome measures included subcortical and cortical neural responses to a repeating consonant-vowel syllable, as well as behavioral measures of phonological awareness, rapid naming, and phonological memory. The music group had larger and more adult-like cortical responses at the end of the study, whereas the JROTC group had smaller subcortical responses. The music group also improved more on the test of phonological awareness, but the two groups did not differ at either time on any of the measures. In short, music training may have caused changes in brain development that were more or less independent of behavior. Alternatively, neural-developmental trajectories may have differed between students who chose one program or the other.

In a similar design, Degé and her colleagues examined memory development among German 9- to 11-year-olds, some of whom opted to take an extended music curriculum in school (Degé, Wehrum, Stark, & Schwarzer, 2011). The children were tested when the program began and again two years later. Outcome variables included measures of short-term visual and auditory memory. Unlike children in the control group, children in the music group exhibited improvement on both memory tests over the course of two years. These findings remained evident when confounding variables such as general intelligence, SES, motivation, and music aptitude were held constant, even though the sample comprised only 34 children in total.

Kreutz and his colleagues examined whether an extended music curriculum taught in primary schools influences cognitive development. The program (JeKI—*Jedem Kind ein Instrument*, An Instrument for Every Child) offered German 7- to 8-year-olds a musical instrument and weekly 45-minute lessons. Families chose whether to enroll, so self-selection was an issue, as was finding an appropriate control group. The author's solution to these problems was to measure and control for as many extraneous variables as possible, and to compare children with counterparts who self-selected into extended training in the natural sciences at different schools in another state. Even before the intervention began, however, the children in the music group had higher IQ scores.

The first two studies had small samples (*n*s = 25) and tested children three times over 18 months. In one, the music group had larger improvements in short-term and long-term verbal memory, but not in short-term visuospatial memory, even after controlling for IQ (Roden, Kreutz, & Bongard, 2012). In a second report, visuospatial short-term and working memory were tested, as well as auditory short-term memory (Roden, Grube, Bongard, & Kreutz, 2014). The music group had greater improvement that the natural-sciences group on the auditory tests (i.e., immediate recall of a list of words, or a multisyllabic nonword), and on three measures of visuospatial working memory (span tasks). As in the 2012 paper, the group differences were evident after controlling for IQ, and the groups performed similarly on the tests of short-term visuospatial memory.

In a subsequent paper (Roden, Könen et al., 2014), the same design was used but the authors recruited and tested much larger samples of children (music: $n = 192$, natural sciences: $n = 153$). The outcome variables for tests of far transfer were measures of visual attention and processing speed. Both were paper-and-pencil tasks that required children to cross out items selectively, or to connect consecutive digits, respectively. As with the smaller samples, the music group had higher IQ scores at the beginning of the study. The additional power meant that the groups also differed significantly in SES: Children in the music group came from higher-SES families. For the visual attention task, both groups improved over time but the natural sciences group had larger improvements, with performance exceeding the music group at the end of the study. SES was unrelated to performance or to the rate of change. By contrast, IQ was related to performance but it did not account for improvements across time. For processing speed, both groups again improved over the 18 months. Although the music group had larger improvements from the second to the third testing session, the groups did not differ at any point in time. As with the attention task, SES had no association with performance but IQ did, although IQ could not account for improved speed over time.

Roden, Könen et al. (2014) also included measures of near transfer—performance on a test of music aptitude that included melody and rhythm subtests. For the rhythm subtest, the music group had greater improvements from Time 1 to Time 2, and performance exceeded that of the natural sciences children at Time 2 and Time 3. For the melody subtest (administered only at Times 2 and 3), the music group had higher scores at both time points, but improvement over time was similar between groups. SES differences between groups were unrelated to the findings, but IQ predicted performance on both subtests, and could explain a small proportion of the improvements in performance. In short, when the researchers tested large samples of children from the JeKI program, there was evidence of inconsistent near transfer, but no evidence of far transfer.

Finally, researchers in China used a *retrospective* longitudinal design to study whether private music training could predict 250 children's past academic performance, focusing on native-language (L1) ability, foreign language (L2) ability, and mathematics (Yang, Ma, Gong, Hu, & Yao, 2014). All of the children entered the same school when they were 6.5 years old on average. Eleven semesters (5.5 years) later, they were asked whether they had taken private music lessons—or private painting lessons—since they started. The researchers then compared students with or without music (or painting) training, using scores from standardized tests that each student took at the end of each semester. An interaction between testing time and group indicated that musically trained children improved more than untrained children in L2 development, but not in L1 performance or mathematics. Moreover, although the music group exhibited an advantage at the end of the eleventh semester on all three tests, the group difference disappeared for L1 and mathematics (but not for L2) when IQ and SES were held constant. Painting training had no association with academic performance. It is unclear why there was a "selective" partial association between music training and L2 ability, but personality (e.g., conscientiousness, motivation) could have played a role.

In short, my review of these natural but longitudinal experiments highlights that children who take music lessons are not a random sample but rather a select group who are likely to differ from other children in terms of general cognitive ability, SES, music aptitude, and personality variables. Whether they are similar to (or not significantly different from) other children at the beginning of a longitudinal study on music training does not mean that group differences, which may be observed later on, can be attributed unequivocally to the music training. It is also clear the natural longitudinal studies with low-SES populations are very difficult to conduct, with attrition being a major problem. My hunch is that interpersonal dynamics within small groups also make the data noisier than they would be otherwise. Moreover, in my own experience of dealing with children in El Sistema, which typically runs every day for hours after school, low-income families are eager for after-school childcare that is both free and safe.

Meta-analysis of Longitudinal Studies

In a recent meta-analysis, Sala and Gobet (2017b) reviewed all of the available longitudinal studies that examined whether music training is associated with cognitive and academic skills. Inclusion criteria included (1) participants who were typically developing children or adolescents with no history of formal music training, (2) a control group, and (3) an outcome variable that measured an academic or cognitive skill, which was not music related. They considered four moderator variables, which might influence whether a positive result emerged. These included age, random assignment, whether the control group was active or passive, and the particular outcome variable (e.g., literacy, mathematics, intelligence).

When all studies were considered jointly, the average effect size was small ($d = 0.16$), but statistically significant, and slightly larger for studies that tested intelligence or memory. Nevertheless, moderator analysis revealed that effects tended to be larger for studies with less than optimal designs, specifically a passive control group or no random assignment. By contrast, 95% confidence intervals for the mean effect size for better designs (i.e., active control group *or* random assignment) included 0. In fact, the mean effect size for studies with an active control group *and* random assignment was −0.12; for studies with a passive control group and no random assignment it was 0.33. The authors concluded that "music training does not reliably enhance children and young adolescents' cognitive or academic skills, and that previous positive findings were probably due to confounding variables" (Sala & Gobet, 2017b, p. 55).

Correlation, Causation, Context, and Conclusion

Although many studies have reported that music training is associated with a variety of nonmusical cognitive abilities (for review see Schellenberg & Weiss, 2013), evidence indicating that music training causes the observed associations is very weak. Alternative interpretations include (1) individual differences, including those in cognitive ability, influence who takes music lessons, or (2) unidentified variables are causing cognitive abilities and music training to co-vary in tandem. In my view, the second alternative could only be describing genetic or demographic factors, which contribute to the individual differences in the first alternative. In short, the two alternatives are describing the same thing. Nevertheless, there is an overwhelming bias in the psychology and neuroscience research communities to interpret positive correlations as evidence for a causal role of music training. Even when researchers acknowledge that causation cannot be inferred, a clear bias for interpreting the results remains evident, even though this requires turning a blind eye to the research on genetics and far transfer. This section of the chapter focuses on delineating this problem and trying to explain it.

Consider one example: an article entitled "Music and words in the visual cortex: The impact of musical expertise" (Mongelli et al., 2017). The title infers causation even though the design was quasi-experimental. The justification for this inference appears to be that the musically trained and untrained groups did not differ significantly in terms of age, education, or gender. Consider another example: an article entitled "Different neural activities support auditory working memory in musician and bilinguals" (Alain et al., 2018). Although the title is neutral with respect the role of causation, the first and last sentences of the abstract are not: "Musical training and bilingualism benefit executive functioning and working memory," and "These findings indicate that the auditory WM advantage in musicians and bilinguals is mediated by different neural networks specific to each life experience." In addition to inferring causation, the sentences are problematic because they equate music training with bilingualism. The sample of bilinguals recruited for the study comprised immigrants to Toronto who came from many different countries. Thus, there is no reason to think that these "forced" bilinguals, who needed to speak English in an English-speaking city, had anything in common other than being bilingual, or that they differed systematically from other Torontonians, except for their native tongue and country of birth. The fact that they successfully immigrated to Canada may be a marker of relatively high SES and cognitive ability compared to compatriots in their home country, but many immigrants in Canada are economic refugees, to which this point is less likely to apply. People with music training, by contrast, differ from musically untrained individuals in many ways, even before the training begins.

Table 16.1 provides examples of 16 articles published since 2013. Each has a title that infers causation from correlational designs that were not longitudinal

Table 16.1 Titles of Example Journal Articles, Published since 2013, That Inferred Causation From Correlation (causal terminology in **bold**)

Title	*Year*
Pitch and Time Processing in Speech and Tones: The **Effects** of Musical Training and Attention	2018
Musical literacy **shifts** asymmetries in the ventral visual cortex	2017
Music training **enhances** the automatic neural processing of foreign speech sounds	2017
Musical training **shapes** neural responses to melodic and prosodic expectation	2016
Tuning the mind: Exploring the connections between musical ability and executive functions	2016
Bilingualism and Musicianship **Enhance** Cognitive Control	2016
Investigating the **effects** of musical training on functional brain development with a novel melodic MMN paradigm	2014
Inhibitory control in bilinguals and musicians: Event related potential (ERP) evidence for experience-specific **effects**.	2014
Degree of musical expertise **modulates** higher order brain functioning	2013
Biological **impact** of preschool music classes on processing speech in noise	2013
Effects of music learning and piano practice on cognitive function, mood and quality of life in older adults	2013
Musical expertise **modulates** early processing of syntactic violations in language	2013
Musical training **heightens** auditory brainstem function during sensitive periods in development	2013
Early musical training and white matter **plasticity** in the corpus callosum: evidence for a sensitive period	2013
Musical training **enhances** neural processing of binaural sounds	2013
Musical experience **influences** statistical learning of a novel language	2013

and had no random assignment. The table is not meant to be exhaustive or even representative. Rather, the articles were selected solely because the title displayed a logical failure, and to document that the problem is common. Note, however, that the problem appears to be particularly acute in neuroscience, perhaps because neuroscientists believe that they are looking at the underlying mechanism for an "established" phenomenon, which blinds them to the limitations of their correlational data. This speculation led to the formation of a hypothesis that I tested recently (Schellenberg, 2018), specifically that *in the published literature on music training, inferring causation from correlation is more common among neuroscientists than it is among psychologists*. The sample included 114 published articles (in English) that reported results from correlational or quasi-experimental studies, each of which examined associations between music training and nonmusical abilities (including brain structure/function). Raters who were blind to the hypothesis analyzed the titles and abstracts for causal language. Inferring causation from correlation was notably high in general, evident in 64% of the articles across disciplines, which was particularly notable because each instance represented a fundamental error in scientific reasoning. As expected, the odds of inferring causation were two to four times greater among neuroscientists than among psychologists, depending on the particular analysis.

Why do many of us believe, or want to believe, that music has transformative powers beyond the pleasure it gives the listener, the feeling of connectedness to others it provides, the happy and sad memories it evokes, and the sense of wonder and awe that listeners often experience? For one thing, scholars and arts advocates complain that music is not taken seriously as a school subject, and that when budget-tightening is the rule of the day, music courses suffer more than math, science, and English courses. The National Endowment for the Arts (NEA) notes, moreover, that music and visual-arts education in schools has declined since the 1970s, and that studying the arts in school predicts subsequent participation in the arts, such as going to the ballet, the opera, the theatre, or to concerts that feature classical or jazz music (Rabkin & Hedberg, 2011). One has to wonder, though, why the NEA did not consider concerts by popular contemporary artists (e.g., Beyoncé, Kanye West, Rihanna). One could also argue that music education and participation have not declined, but simply changed, as consumers now have access to an infinite amount of music via smart phones and the internet, and individuals with no formal training can make professional-sounding recordings in the comfort of their homes, with relatively inexpensive software, a personal computer, and no musical instruments.

In my view, government support for music and other art forms is always going to be lukewarm (or cold) when compared to its support for STEM subjects (science, technology, engineering, and mathematics) and core subjects in the humanities such as English and history. Reports from neuroscience and psychology, which claim that music has beneficial, nonmusical side effects, are unlikely to change things. Moreover, justifying support for music based on nonmusical benefits implies, with just a little slippage, that without such benefits, music is unimportant. In short, science and advocacy make strange bedfellows.

Another explanation of the radical environmentalism that has taken over this research area is that it is much easier to receive funding for research programs that seek to document positive findings, rather than null results, about links between music training and nonmusical abilities. The same argument applies to publications in peer-reviewed journals. Recent evidence suggests, however, that instead of making you smarter, playing music in groups facilitates social bonding and social behavior (Kirschner & Tomasello, 2010; Schellenberg, Corrigall, Dys, & Malti, 2015), a view that is consistent with some evolutionary accounts of music's universality (e.g., Dunbar, 2012; Tarr, Launay, & Dunbar, 2014). Moving in synchrony appears to play a central role in the effect. For example, toddlers who are bounced in synchrony to music with an experimenter are more likely to be helpful to the same experimenter (Cirelli, Einarson, & Trainor, 2014), or to her friend (Cirelli, Wan, & Trainor, 2016), than they are to a stranger, or to someone who bounced out-of-sync with the toddler. Similar but smaller effects of synchrony are evident even without the music (Cirelli, Wan, Spinelli, & Trainor, 2017). In other words, music with a beat is likely to promote synchronous movement, which in turn facilitates social bonding. More generally, music listening often makes us feel good, and making music often makes us feel good together. Isn't that enough?

Overview

What has neuroscience added to our understanding of the nonmusical consequences of music training, over and above psychology? The reviewed findings are largely negative. In my opinion, neuroscience has detracted from research because neuroscientists routinely impute causality from correlation, even more so than psychologists do. The lack of behavioral evidence for far-transfer effects—in the case of music training and otherwise—means that the search for neural mechanistic explanations is unwarranted.

What are the concrete implications of research in this area, and opportunities for translation to education? Music training improves *musical* skills, which should be enough.

How can the reader learn more about the topic? Overviews of the available literature are provided by a recent meta-analysis of music training in particular (Sala & Gobet, 2017b), and by an article that provides a concise review of the lack of evidence for far transfer in general (Sala & Gobet, 2017a).

Note

1. https://files.eric.ed.gov/fulltext/ED581247.pdf

References

Alain, C., Khatamian, Y., He, Y., Lee, Y., Moreno, S., Leung, A. W. S., & Bialystok, E. (2018). Different neural activities support auditory working memory in musician and bilinguals. *Annals of the New York Academy of Sciences*, *1423*, 435–446.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*, 366–377.

Benz, S., Sellaro, R., Hommel, B., & Colzato, L. S. (2016). Music makes the world go round: The impact of musical training on non-musical cognitive functions—a review. *Frontiers in Psychology*, 6. doi:10.3389/fpsyg.2015.02023

Bouchard, T. J. Jr., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology*, *54*, 4–45.

Butkovic, A., Ullén, F., & Mosing, M. A. (2015). Personality related traits as predictors of music practice: Underlying environmental and genetic influences. *Personality and Individual Differences*, *74*, 133–138.

Chobert, J., François, C., Velay, J.-L., & Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex*, *24*, 956–967.

Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental Science*, *17*, 1003–1011.

Cirelli, L. K., Wan, S. J., Spinelli, C., & Trainor, L. J. (2017). Effects of interpersonal movement synchrony on helping behaviors: Is music necessary? *Music Perception*, *34*, 319–326.

Cirelli, L. K., Wan, S. J., & Trainor, L. J. (2016). Social effects of movement synchrony: Increased infant helpfulness only transfer to affiliates of synchronously moving partners. *Infancy*, *21*, 807–821.

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00222

Costa-Giomi, E. (2012). Music instruction and children's intellectual development: The educational context of music participation. In R. A. R. MacDonald, G. Kreutz, & L. Mitchell (Eds.), *Music, health & wellbeing* (pp. 339–355). Oxford: Oxford University Press.

Costa-Giomi, E. (2015). The long-term effects of childhood music instruction on intelligence and general cognitive abilities. *Update: Applications of Research in Music Education*, *33*, 20–26.

Deary, I. J., Yang, J., Davies, G., Harris, S. E., Tenesa, A., Liewald, D., . . . Visscher, P. M. (2012). Genetic contribution to stability and change in intelligence from childhood to old age. *Nature*, *482*, 212–215.

Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology*, *2*, 124. doi:10.3389/fpsyg.2011.00124(2011).

Degé, F., Wehrum, S., Stark, R., & Schwarzer, G. (2011). The influence of two years of school music training in secondary school on visual and auditory memory. *European Journal of Developmental Psychology*, 8, 608–623.

Doidge, N. (2007). *The brain that changes itself*. New York: Viking.

Dumont, E., Syurina, E. V., Feron, F. J. M., & van Hooren, S. (2017). Music interventions and child development: A critical review and further directions. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.01694

Dunbar, R. I. M. (2012). On the evolutionary function of song and dance. In N. Bannan & S. Mithen (Eds.), *Music, language and human evolution* (pp. 201–214). Oxford: Oxford University Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406.

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS One*, *10*(9), e0138715. doi:10.1371/journal.pone.0138715

Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., . . . Schön, D. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*, 392. doi:10.3389/fnhum.2014.00392

Foster, J. L., Harrison, T. L., Hicks, K. L., Draheim, C., Redick, T. S., & Engle, R. W. (2017). Do the effects of working memory training depend on baseline ability level? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1677–1689.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*, 2038–2043.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*, 3–10.

Goswami, U., Huss, M., Mead, N., Fosker, T., & Verney, J. P. (2013). Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex*, *49*, 1363–1376.

Gromko, J. E. (2005). The effect of music instruction on phonemic awareness in beginning readers. *Journal of Research in Music Education*, *53*, 199–209.

Habibi, A., Cahn, B. R., Damasio, A., & Damasio, H. (2016). Neural correlates of accelerated auditory processing in children engaged in music training. *Developmental Cognitive Neuroscience*, *21*, 1–14.

Habibi, A., Damasio, A., Ilari, B., Veiga, R., Joshi, A. A., Leahy, R. M., . . . Damasio, H. (2017). Childhood music training induces change in micro and macroscopic brain structure: Results from a longitudinal study. *Cerebral Cortex*. Advance online publication. doi:10.1093/cercor/bhx286

Habibi, A., Ilari, B., Crimi, K., Metke, M., Kaplan, J. T., Joshi, A. A., . . . Damasio, H. (2014). An equal start: Absence of group differences in cognitive, social, and neural measures prior to music or sports training in children. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00690

Hambrick, D. Z., & Tucker-Drob, E. M. (2015). The genetics of music accomplishment: Evidence for gene-environment correlation and interaction. *Psychonomic Bulletin & Review*, *22*, 112–120.

Haywood, S., Griggs, J., Lloyd, C., Morris, S., Kiss, Z., & Skipp, A. (2015). *Creative futures: Act, sing, play. Evaluation report and executive summary*. London: Educational Endowment Foundation.

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, *76*, 486–502.

Herrera, L., Lorenzo, O., Defior, S., Fernandez-Smith, G., & Costa-Giomi, E. (2011). Effects of phonological and musical training on the reading awareness of native and foreign-Spanish-speaking children. *Psychology of Music*, *39*, 68–81.

Howe, M. J. A., Davidson, J. W., & Sloboda, J. A. (1998). Innate talents: Reality or myth? *Behavioral & Brain Sciences*, *21*, 399–442.

Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, *47*, 674–689.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, *29*, 3019–3025.

Ilari, B. S., Keller, P., Damasio, H., & Habibi, A. (2016). The development of musical skills of underprivileged children over the course of 1 year: A study in the context of an El Sistema-inspired program. *Frontiers in Psychology*, *7*. doi:10.3389/fpsyg.2016.00062

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 6829–6833.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 10081–10086.

Jäncke, L. (2009). The plastic human brain. *Restorative Neurology and Neuroscience*, *27*, 521–538. doi:10.3233/RNN-2009-519

Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short term music and second language training on executive control. *Journal of Experimental Child Psychology*, *144*, 84–97.

Jaschke, A. C., Honing, H., & Scherder, E. J. A. (2018). Longitudinal analysis of music education on executive functions in primary school children. *Frontiers in Neuroscience*, *12*. doi:10.3389/fnins.2018.00103

Kaviani, H., Mirbaha, H., Pournaseh, M., & Sagan, O. (2014). Can music lessons increase the performance of preschool children in IQ tests? *Cognitive Processing*, *15*, 77–84.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution & Human Behavior*, *31*, 354–364.

Kraus, N., Hornickel, J., Strait, D. L., Slater, J., & Thompson, E. (2014). Engagement in community music classes sparks neuroplasticity and language development in children from disadvantaged backgrounds. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.01403

Leong, V., & Goswami, U. (2014). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00096

Linnavalli, T., Putkinen, V., Lipsanen, J., Huotilainen, M., & Tervaniemi, M. (2018). Music playschool enhances children's linguistic skills. *Scientific Reports*, 8, 8767. doi:10.1038/s41598-018-27126-5

Love, J. M., Chazan-Cohen, R., Raikes, H., & Brookes-Gunn, J. (2013). What makes a difference: Early head start evaluation findings in a developmental context. *Monographs of the Society for Research in Child Development*, *78*(1).

Mackintosh, N. J. (2011). *IQ and human intelligence* (2nd ed.). Oxford: Oxford University Press.

Macnamara, B., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science*, *25*, 1608–1618.

McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Petrill, S. A., & Plomin, R. (1997). Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science*, *276*(5318), 1560–1563.

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(12), e82007. doi:10.1371/journal.pone.0082007

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*, 512–534.

Mongelli, V., Dehaene, S., Vinckier, F., Peretz, I., Bartolomeo, P., & Cohen, L. (2017). Music and words in the visual cortex: The impact of musical experience. *Cortex*, *86*, 260–274.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011a). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, *22*, 1425–1433.

Moreno, S., Friesen, D., & Bialystok, E. (2011b). Effect of music training on promoting preliteracy skills: Preliminary causal evidence. *Music Perception*, *29*, 165–172.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence of brain plasticity. *Cerebral Cortex*, *19*, 712–723.

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: No causal effect of musical practice on musical ability. *Psychological Science*, *25*, 1795–1803.

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.

Mosing, M. A., Pedersen, N. L., Madison, G., & Ullén, F. (2014). Genetic pleiotropy explains associations between musical auditory discrimination and intelligence. *PLoS One*, 9(11), e113874. doi:10.1371/journal.pone.0113874

Mosing, M. A., & Ullén, F. (2018). Genetic influence on musical specialization: A twin study on choice of instrument and music genre. *Annals of the New York Academy of Sciences*, *1423*, 427–434.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*, 473–478.

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.

Patscheke, H., Degé, F., & Schwarzer, G. (2016). The effects of training in music and phonological skills on phonological awareness in 4- to 6-year-old children of immigrant families. *Frontiers in Psychology*, *7*. doi:10.3389/fpsyg.2016.01647

Plomin, R., & DeFries, J. C. (1985). *Origins of individual differences in infancy: The Colorado adoption project*. Orlando, FL: Academic Press.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives in Psychological Science*, *11*, 3–23.

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, *19*, 148–159.

Portowitz, A., Lichtenstein, O., Egorova, L., & Brand, E. (2009). Underlying mechanisms linking music education and cognitive modifiability. *Research Studies in Music Education*, *31*, 107–128.

Power, R. A., & Pluess, M. (2015). Heritability estimates of the big five personality traits based on common genetic variants. *Translational Psychiatry*, *5*, e604. doi:10.1038/tp.2015.96

Rabkin, N., & Hedberg, E. C. (2011). *Arts education in America: What the declines mean for arts participation*. Washington, DC: National Endowment for the Arts.

Rauscher, F. H., & Zupan, M. A. (2000). Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: A field experiment. *Early Childhood Research Quarterly*, *15*, 215–228.

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.

Roden, I., Könen, T., Bongard, S., Frankenberg, E., Friedrich, E. K., & Kreutz, G. (2014). Effects of music training on attention, processing speed and cognitive music abilities—findings from a longitudinal study. *Applied Cognitive Psychology*, *28*, 545–557.

Roden, I., Kreutz, G., & Bongard, S. (2012). Effects of a school-based instrumental music program on verbal and visual memory in primary school children: A longitudinal study. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00572

Sala, G., & Gobet, F. (2017a). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science*, *26*, 515–520.

Sala, G., & Gobet, F. (2017b). When the music's over. Does musical skill transfer to children's and young adolescents' cognitive and academic skills? A meta-analysis. *Educational Research Review*, *20*, 55–67.

Sala, G., & Gobet, F. (2017c). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology*, *53*, 671–685.

Scarr, S., & McCartney. K. (1983). How people make their own environments: A theory of genotype → environmental effects. *Child Development*, *54*, 424–435.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*, 511–514.

Schellenberg, E. G. (2005–2006). Music lessons enhance IQ: A reply to Black (2005) and Steele (2005). *Scientific Review of Mental Health Practice*, *4*(2), 10–13.

Schellenberg, E. G. (2016). Music training and nonmusical abilities. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 415–429). Oxford: Oxford University Press.

Schellenberg, E. G. (2018). *Correlation = causation: The case of music training, psychology, and neuroscience*. Manuscript submitted for publication.

Schellenberg, E. G., Corrigall, K. A., Dys, S. P., & Malti, T. (2015). Group music training and children's prosocial skills. *PLoS One*, *10*(10), e0141449. doi:10.1371/journal.pone.0141449

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 499–550). Amsterdam: Elsevier Academic Press.

Schlaug, G. (2001). The brain of musicians: A model for functional and structural adaptation. *Annals of the New York Academy of Sciences*, *930*, 281–299.

Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the analysis of covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, 6. doi:10.3389/fpsyg.2015.00474

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*, 103–186.

Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural Brain Research*, *291*, 244–252.

Slater, J., Strait, D. L., Skoe, E., O'Connell, S., Thompson, E., & Kraus, N. (2014). Longitudinal effects of group music instruction on literacy skills in low-income children. *PLoS One*, 9(11), e113383. doi:10.1371/journal.pone.0113383

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*, *24*, 1077–1096.

Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: Musicians as a model of auditory learning. *Hearing Research*, *308*, 109–121.

Swaminathan, S., & Schellenberg, E. G. (2016). Music training. In T. Strobach & J. Karbach (Eds.), *Cognitive training: An overview of features and applications* (pp. 137–144). New York: Springer.

Swaminathan, S., & Schellenberg, E. G. (2018a). Musical competence is predicted by music training, cognitive abilities, and personality. *Scientific Reports*. doi:10.1038/s41598-018-27571-2

Swaminathan, S., & Schellenberg, E. G. (2018b). Music training and cognitive abilities: Associations, causes, and consequences. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford handbook of music and neuroscience*. Advance online publication. doi:10.1093/oxfordhb/9780198804123.013.26

Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? *Intelligence*, *62*, 119–124.

Swaminathan, S., Schellenberg, E. G., & Venkatesan, K. (2018). Explaining the association between music training and reading in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi:10.1037/xlm0000493

Tarr, B., Launay, J., & Dunbar, R. I. (2014). Music and social bonding: "Self-other" merging and neurohormonal mechanisms. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.01096

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*, 46–64.

Thomson, J. M., Leong, V., & Goswami, U. (2013). Auditory processing interventions and developmental dyslexia: A comparison of phonemic and rhythmic approaches. *Reading and Writing*, *26*, 139–161.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions (I). *Psychological Review*, *8*, 247–261.

Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 10062–10067.

Tierney, A., Krizman, J., Skoe, E., Johnston, K., & Kraus, N. (2013). High school music classes enhance the neural processing of speech. *Frontiers in Psychology*, *4*, 855. doi:10.3389/fpsyg.2013.00855

Ullén, F., Hambrick, D. Z., & Mosing, M. A. (2016). Rethinking expertise: A multifactorial gene-environment interaction model of expert performance. *Psychological Bulletin*, *142*, 426–446.

Ullén, F., Mosing, M. A., Holm, L., Eriksson, H., & Madison, G. (2014). Psychometric properties and heritability of a new online test for musicality, the Swedish musical discrimination test. *Personality and Individual Differences*, *63*, 87–93.

U.S. Department of Health and Human Services & Administration for Children and Families. (2010). *Head start impact study: Final report*. Washington, DC. Retrieved from www.acf.hhs.gov/opre/resource/head-start-impact-study-final-report-executive-summary

Vernon, P. A., Martin, R. A., Schermer, J. A., & Mackie, A. (2008). A behavioral genetic investigation of humor styles and their correlations with the big-5 personality dimensions. *Personality and Individual Differences*, *44*, 1116–1125.

Wai, J., Worrell, F. C., & Chabris, C. F. (2017). The consistent influence of general cognitive ability in college, career, and lifetime achievement. In K. L. McClarty, K. D. Mattern, & M. N. Gaertner (Eds.), *Preparing students for college and careers: Theory, measurement, and educational practice* (pp. 46–56). New York: Routledge.

Yang, H., Ma, W., Gong, D., Hu, J., & Yao, D. (2014). A longitudinal study on children's music training experience and academic development. *Scientific Reports*, *4*. doi:10.1038/srep05854