Examining the Influence of Music Training on Cognitive and Perceptual Transference: A Quantitative Meta-Analysis

by

Katherine Jennifer Thompson

A thesis submitted in conformity with the requirements for the degree of Master of Arts (M.A.) Psychology

Department of Psychology
University of Toronto

© Copyright by Katherine J. Thompson (2015)
Examining the Influence of Music Training on Cognitive and Perceptual Transference: A Quantitative Meta-Analysis

Katherine Jennifer Thompson

Masters of Arts (M.A.) Psychology

Department of Psychology

University of Toronto

2015

Abstract

Two meta-analyses were conducted in order to delineate whether musicians and non-musicians differ on specific cognitive and music processing domains, and determine the magnitude of any differences. Moreover, potential covariates outlined by the literature were included as moderator variables. As predicted, musicians significantly differed from non-musicians on all music processing tasks (melody perception: $d=1.18$; pitch perception: $d=1.00$; temporal perception: $d=.834$). Furthermore, as predicted, musicians significantly differed on the cognitive domains related to musical ability (Motor: $d=1.08$, and Audition: $d=.837$). Significant differences were also obtained in General Intelligence ($d=.348$), Memory ($d=.421$), Orientation and Attention ($d=.283$), Perception ($d=.190$), and Verbal Functions and Language Skills ($d=.423$). Meta-regression and sub-group analyses of moderating variables indicated that years of music training, age group, and categories (breakdown of groups within domains) had an effect on summary effects, however this impact was not consistent across all domains. The effects of other potential moderating variables are discussed in the context of existing literature.
Acknowledgements

I would like to express my gratitude to my supervisor Dr. Mark Schmuckler for his guidance and knowledge of the field, and providing me the exciting opportunity to pursue this line of research. My sincere thanks also go to Dr. Konstantine Zakzanis for his immense knowledge of meta-analyses, and his advice and counsel. I would also like to thank Dr. Lee Bartel for serving on my committee, and Drs. Annabel Cohen, June Countryman, and Karem Simon, whose expertise and encouragement inspired this thesis.

Last but never least, I would like to thank Branden, Jim, Anne, Alex, Allie, Beth and Arni for their love, emotional support, editing and care packages: I would not have completed this thesis without you all.

This work was supported through the Canada Graduate Scholarship (CGS-M) provided by the Social Sciences and Humanities Research Council (SSHRC).
Chapter 3 Results ...........................................................................................................................................34

3.1 Music Processing Results ......................................................................................................................35
   Basic Analyses .........................................................................................................................................35
   Publication Bias .........................................................................................................................................35
   Tests of Heterogeneity ..............................................................................................................................37
Further Analyses (Sub-group analyses and Meta-Regression) .................................................................37
   Pitch Perception .......................................................................................................................................38
   Melody Perception ...................................................................................................................................40
   Temporal Perception ...............................................................................................................................42

3.2 Cognition Results ..................................................................................................................................43
   Basic Analyses .........................................................................................................................................43
   Publication Biases ....................................................................................................................................44
   Tests of Heterogeneity ..............................................................................................................................47
Further Analyses (Sub-group analyses and Meta-Regression) .................................................................47
   Audition ....................................................................................................................................................47
   Concept Formations and Reasoning (Lezak et al., 2012) ......................................................................49
   General Intelligence ..................................................................................................................................49
   Memory (Lezak et al., 2012) ......................................................................................................................50
   Orientation and Attention (Lezak et al., 2012) ......................................................................................51
   Perception (Lezak et al., 2012) ...............................................................................................................52
   Verbal Functions and Language Skills (Lezak et al., 2012) .....................................................................52

Chapter 4 Discussion ..................................................................................................................................53

4.1 Music Processing (Near-transfer effects) ..............................................................................................53
4.2 Cognition (Far-transfer effects) .............................................................................................................56
Implications ..................................................................................................................................................61
Limitations ........................................................................................................................................62
  Methodological limitations .............................................................................................................62
  Music training and ability: causal or correlational?......................................................................65
Future directions ..................................................................................................................................65
References ..........................................................................................................................................67
List of Tables

Table 1. Research questions and hypotheses
Table 2. Data extraction sheet: Information coded from each article included in the meta-analysis
Table 3. Music Processing Domain Categorization based on Tirovolas & Levitin (2011)
Table 4. Cognition Domain Categorization based on (Lezak et al., 2012)
Table 5. Effect Sizes Grouped by Music Processing Domains
Table 6. Effect Sizes Grouped by Cognitive Domains
Table 7. Publication bias: Fail-Safe N Test
Table 8. Summary of findings from cognition analysis
List of Figures

Figure 1. Funnel plots for pitch perception and melody perception in assessing publication bias. Fewer studies in the bottom left area suggest the presence of publication bias.

Figure 2. Each category in Pitch Perception was significant against the null hypothesis, except for loudness: Circles represent effect size, bold vertical line represents the regression line, surrounding vertical lines represent confidence interval.

Figure 3. Although it was hypothesized that there would be a linear relationship between years of training in musicians and summary effect sizes, the relationship was not significant in the pitch performance domain (seen above). Furthermore, the regression slope line is nearly vertical, suggesting that there is a more important difference between musicians and non-musicians, other than years of training. Note: Circles represent effect size, bold vertical line represents the regression line, and surrounding vertical lines represent confidence interval.

Figure 4. A visual representation of the categories in Melody Perception: each category within the domain tested significant against the null hypothesis. Furthermore, categories did not test significantly different from each other. Note. The circles represent effect sizes from primary studies, bold vertical line represents the regression line, and the surrounding vertical lines represent confidence interval.

Figure 5. A linear relationship between musicians’ years of training and performance on temporal/rhythm perception tasks was observed. These results suggest an incremental relationship, such that more training results in better rhythm ability.

Figure 6. General Intelligence funnel plot; that studies from smaller studies are located on both sides of the average indicates a lack of publication bias in this domain.

Figure 7. Funnel plot for memory. Publication bias typically manifest itself in a funnel plot with fewer studies on the bottom left, indicating a bias in not published smaller studies (fewer
participants) finding no effect. In this case, however, there is also a lack of publications on the bottom right, which suggests that there were no small studies in this analysis. That larger studies are widely distributed around the average suggests strong heterogeneity.

Figure 8. A significant effect of years of training in the musician group on the overall summary effect of audition

Figure 9. Significant effect of years of training in the musician group for the memory domain, however the slope is not in the expected direction. The down-sloping regression line indicates that increased years of training in musicians resulted in worse performance in the memory domain.
List of Appendices

Appendix A. Keyword Search
Appendix B. Article Bibliography
Appendix C. Categorization and breakdown of effect sizes
Appendix D. Forest Plots
Appendix E. Funnel Plots
Chapter 1
Introduction

1.1 Transfer effects of music training

Music training is a unique experience that is often accompanied by demanding cognitive and motor processing. Recent decades have seen a surge in the number of publications comparing musicians to non-musicians on a wide range of tasks, and investigating the impact of music training interventions (Levitin & Tirovolas, 2011). Although the goal of music training is typically to improve one’s musicality and technical ability, during its process there are many sensorimotor and cognitive skills being employed. For example, playing a musical instrument involves reading music notation and executing fine motor movement, while adjusting it based on auditory and tactile feedback (Schlaug et al., 2005). Playing music also involves memorization, improvisation, rhythmic entrainment, and emotional manipulation, amongst others. Many of the abilities developed are not unique to music performance, and have thus raised questions of transfer effects.

Transfer effects, or transference, refers to the enhancement of skills in one area that improves skills in another domain. A close association between a training domain and a transfer domain is known as near transfer, and a less obvious association between domains is far transfer (Hyde et al., 2009, p. 3019). Far transfer has also been referred to as cross-domain plasticity (Patel, 2014). Transfer effects have been substantiated by structural changes in brain organization through skill development. Neuroanatomical change has been found in expert musicians (Schlaug, Janacke, Huang, et al., 1995), as well as structural differences, such as increased grey matter (Zatorre, Fields, Johansen-Berg, 2012) and cortical thickness (Bermudez, Lerch, Evans & Zatorre, 2009) in the auditory cortices. That these structural changes may lead to improved functional ability in more general cognitive and perceptual abilities is the basis for investigations utilizing both standardized assessments and music perception tasks.

Studies examining the potential of music training to elicit far-transfer effects often compare the performance of musicians to non-musicians on more general tasks unrelated to musical ability, such as memory or language. The theory that music training can cause transfer effects in more general cognitive processes and executive functioning has resulted in many studies on far transfer, however this literature is conflicting. There is evidence of a causal link
between music training and spatial abilities (Bilhartz et al., 1999; Hetland, 2000; Jakobson et al., 2008), memory (George & Coch, 2011; Tierney et al., 2008), attention (Besson et al., 2011; Moreno et al., 2011) and language abilities (Forgeard et al., 2008; Ho et al., 2003). Findings from other investigations, however, did not support a causal relationship with music training in these same domains (visuospatial: Aleman et al., 2000; memory: Helmbold et al., 2005; attention: Costa-Giomi, 1999; Helmbold et al., 2005; language ability: Butzlaff, 2000). These conflicting results make it difficult to reliably determine far-transfer differences in specific cognitive abilities between musicians and non-musicians.

When it comes to near-transfer, there appears to be a lack of consensus regarding a definition and inclusion criterion. Traditionally, transfer occurs when an ability unrelated to music processing also improves. Although some studies consider skills directly related to music processing to be near-transfer (Corrigall & Trainor, 2011), the majority of studies define it as an ability with a close association to a musical skill that is outside the realm of music (ex. the typing speed of pianists) (Hyde et al, 2009). Not including music processing in near-transfer abilities, however, carries the assumption that music training leads to a global improvement in all music perception skills. Some researchers have posited that we are born with a certain level of musical ability (Ladinig et al., 2009), sometimes referred to as music aptitude (Schellenberg, 2006), suggesting that some non-musicians possess music processing abilities that may challenge the abilities of musicians on some domains. Additionally, others have suggested that, given the ubiquity of music in our society, non-musicians may be improving their music processing abilities through listening (Bigand, 2003) or exposure (Loui et al., 2009; Marin, 2009; Tillman et al., 2000). These suggestions are substantiated with findings in which non-musicians are able to perform tasks well above chance, such as on the perception of melody (Halpern et al., 1998), timbre (Jentschke & Koelsch, 2009), rhythm (Ladinig et al., 2009), and pitch (Ott et al., 2013; Smith et al., 1994). It remains unclear in the literature the extent to which music training improves or affects music processing, therefore, it is argued that near-transfer is inclusive of cognitive and perceptual abilities considered musical in nature that may be enhanced through training. As a result, near-transfer in the current study refers to music-processing abilities (ex. pitch perception, melody discrimination, rhythm reproduction, etc).
1.2 Music training and music processing

Within the field of music cognition, the cognitive strategies used to perceive and discriminate between musical elements are of great interest. As a result, there are many publications comparing the performance of musicians and non-musicians on a wide range of musical domains (e.g., tonal relationships, melodic structures, rhythm, etc). As one may expect, musicians tend to outperform non-musicians on general tasks of pitch perception (Tervaniemi et al., 2005), melody perception (Halpern, 1998; Koniari et al., 2001), and rhythm reproduction (Watanabe et al., 2007). Despite findings of superiority, many studies have either failed to replicate findings of differences, or have demonstrated sophisticated music processing in non-musicians (pitch: Ott et al., 2013; Smith et al., 1994; melody: Halpern, 1998; Hoover & Cullari, 1992; timbre: Beal, 1985; Jentschke & Koelsch, 2009; rhythm: Watanabe et al., 2007).

Non-musicians are typically characterized as individuals with no music training, however they are still able to appreciate musical structure, indicating some level of musical ability. These skills may have been developed by exposure to the Western tonal system (Bigand, 2003; Loui et al., 2009; Marin, 2009) or through everyday music listening and early exposure. The theory that skills are developed passively through exposure is supported by research showing that both musicians and non-musicians demonstrate lower accuracy for detecting mistuning in culturally unfamiliar patterns (Lynch et al., 1991). Although non-musicians are able to perform many music processing tasks, it appears that they use different cognitive strategies than musicians. One study found that musicians use areas of the cortex used devoted to short-term memory for pitch memory tasks, whereas non-musicians rely on auditory sensory areas (Strait et al., 2011).

If music processing skills exist in all individuals, regardless of training, it becomes possible that musicians may not significantly outperform non-musicians on all musical domains. Determining the tasks in which musicians consistently outperform non-musicians, and the tasks that perhaps these group do not differ, may help researchers further delineate perceptual skills related to musical skill development. The following section will summarize the different musical domains for which there are conflicting findings of difference between musicians and non-musicians.

**Pitch (frequency) perception**

The ability to recognize and discriminate between pitches is an essential ability in music. Frequency discrimination is developed in music training through the constant tuning of one’s
instrument, playing within an ensemble, and an emphasis on strict precision and clarity of note execution. Furthermore, music training that involves aural skill development emphasizes interval naming and recognition to foster relative pitch ability. Therefore, it seems obvious that musicians would have strong pitch perception skills. In an often-cited study on pitch change detection, Tervaniemi and colleagues (2005) compared active, expert musicians with experience on a wide variety of instruments to non-musicians. Participants heard two complex tones and were asked to determine which one was higher in pitch, with unlimited time. Results showed that musicians were both faster and more accurate than non-musicians (Tervaniemi et al., 2005). Another study found that non-musicians' mean pitch thresholds were six times larger than that of the musicians, and sought to determine how much training would be needed to lower their pitch thresholds closer to that of musicians. Even after a short period of training, the non-musicians’ pitch thresholds were still 2-3 times larger than findings from previous studies (Micheyl et al., 2006).

The effects of short term music training on pitch perception was examined by Besson and colleagues (2006), in which 8-year-old children who received only six months of a novel music training approach were able to detect small pitch changes better than a control group with paint training.

That musicians typically demonstrate smaller difference limens for frequency (DLFs), as compared to non-musicians, has also been supported elsewhere (Akin & Belgin, 2009; Schon, Magne & Besson et al., 2007). As an example, Kishon-Rabin and colleagues (2011) found that the DFLs in their musician group were nearly half the size to that of non-musicians, and that performance was associated with years of music training. An interesting finding from this study was that classical musicians demonstrated significantly lower pitch thresholds than musicians with contemporary (ex. pop or rock) experience.

Interestingly, non-musicians also demonstrate above-chance accuracy levels on a variety of pitch processing tasks. Non-musicians are able to detect “false notes” that do not belong to a sequence or group of pitches (Neuhaus et al., 2006) and pitches that are not a good musical-fit within a sequence, as determined by probe-tone tests (Oram & Cuddy, 1995). When tasked with deciding whether a pitch is higher or lower than the previous in a presentation, both musicians and non-musicians perform at ceiling level (Ott et al., 2013). Furthermore, when the stimuli used in a study are more ecologically valid (eg. real musical sounds rather than sine waves),
individuals without music training perform well on tasks involving the discrimination of intervals (Smith et al., 1994).

**Melody and Harmony perception**

Similar to that of pitch perception, melody perception is an important aspect of musicality that often improves through music and aural training. In discriminating between melodies, musicians are more accurate at detecting contour change than non-musicians (Cuddy & Cohen, 1975; Fujioka et al., 2004). Musicians also outperform non-musicians on discrimination tasks of contour and mode in unfamiliar melodies (Halpern et al., 1998). Finally, individuals with extensive training are able to recognize melodies faster with less melodic information than non-musicians (Bailes, 2010).

The ability to perceive and discriminate between melodies does not seem to be restricted to those with music training. Non-musicians can still discriminate and recognize similar melodies, even when transposed and manipulated rhythmically (Halpern, 1998). In a study with 10-11 year old children (Koniari et al., 2001), non-musicians were able to recognize musical materials above chance, despite the musicians showing greater accuracy overall (67.7% vs. 97.6%, respectively). Individuals without music training are also able to recognize regular and irregular endings (Koelsch et al., 2005) or abnormalities in melodies, caused by the manipulation of various musical elements (Seung et al., 2005). Finally, when participants are asked to categorize major and minor intervals without the use of musical terms, non-musicians use similar strategies to that of musicians (Smith et al., 1994).

Though perhaps outside the realm of perception related to frequency, distortions in loudness appear to produce equivalent performance between musicians and non-musicians. In a study on loudness, musicians and non-musicians were presented with 1-min melodic excerpts of various genres of music. Participants were instructed to match the loudness of the excerpts with a neutral stimulus. Non-musicians were more accurate than musicians, however both the effect ($d=0.24$) and sample ($n=25$) were small (Hoover & Cullari, 1992).

The ability to perceive and discriminate between harmonic structures, such as chords, is also a vital musical processing ability. As one might expect, formal music training has been associated with ability to process key and harmony (Corrigall & Trainor, 2009) and the ability to detect a mistuned harmonic (Zendel & Alain, 2009). However, both musicians and non-musicians are able to identify chords that are more appropriate in a certain tonal context (Bigand
et al., 2001), and are responsive to tonal hierarchies (Cuddy & Badertscher, 1987; Cuddy & Thompson, 1992); this suggests a sophisticated music processing ability in non-musicians, despite differences in training. Indeed, some studies have been unable to replicate findings of significant differences between musicians and non-musicians on harmonic perception (adults: Corrigall & Trainor, 2009; children: Marin & Barnes, 1985). In one such study, expert musicians and non-musicians were asked to judge the relatedness between chords; performance between groups did not significantly differ (Marin & Barnes, 1985).

**Timbre perception**

Although an important aspect to music, there is little research that compares musicians to non-musicians on timbre related tasks. Musicians, however, have demonstrate greater accuracy on tasks related to detecting specific instrumental timbres (Baumann et al., 2008), and determining whether two timbres came from the same source or not (Chartrand & Belin, 2006).

Despite a lack of training and exposure to timbres through performance, non-musician adults can detect timbre deviants within musical stimuli (Koelsch et al., 2013). In a study on timbre perception (Jentschke & Koelsch, 2009), children with and without music training were asked to discriminate between instruments and voices. Although the musically trained group was more accurate on the instrument timbre condition, the difference was not significant. Other studies have also failed to find significant differences between musicians and non-musicians on timbre discrimination (Beal, 1985). Accuracy in timbre discrimination abilities may depend on the timbres being used in a study (Crummer et al., 1994), as musicians demonstrate a greater neural response to the timbre of their primary instrument (Pantev et al., 1998). Therefore, it is possible that musicians are not as sensitive to the timbres of other instruments, perhaps to the point where their performance does not differ from that of non-musicians.

**Temporal (rhythmic) perception**

The perception of rhythm, though not necessarily unique to music, is often enhanced through music training. Musicians tend to exhibit more accurate rhythmic discrimination, and identify rhythm deviations better than non-musicians (Drake et al., 2000a; Franek et al., 1991; Habibi, Wirantana & Starr, 2014). Individuals with substantial music training also show greater sensitivity to meter than those without training (Geiser, et al., 2010), are more sensitive to small
duration changes (Musacchia, et al., 2007) and more accurately synchronize physical movement to a beat (Drake et al., 2000b).

Although rhythm is an essential musical element, it is not restricted to music. Rhythm is expressed through language, physical motion and even in our physiology. As a result, the ability to perceive temporal events may be innate (Baruch & Drake, 1997; Demany et al., 1977), although the ability to physically entrain to an external stimulus is not evident until around 4 years of age (Zentner & Eerola, 2010). Therefore, it is not surprising that rhythm perception is evident in those without music training (Grahn & Rowe, 2009). It has been suggested that hierarchical representations for rhythm are formed preattentively in the auditory system (Ladinig et al., 2009), giving a basis to how non-musicians perceive meter and rhythm.

Stemming from this possibility, many studies demonstrate sophisticated rhythmic processing in non-musicians. In a reproduction task where participants were asked to synchronize their tapping to a complex rhythmic motor stimulus, non-musicians did not differ from early- and late-trained musicians (Watanabe et al., 2007). Finally, untrained listeners are able to make consistent judgments about the appropriateness of the tempo of pieces based on various musical elements. (Quinn & Watt, 2006).

1.3 Music training and cognitive transfer

Along with differences in music processing skills between musicians and non-musicians, there has been a growing interest in the potential transfer of more general cognitive processing from music training. To illustrate, a search for the keywords “music training/education/participation” and “transfer/transference” on PsychINFO resulted in 4882 results published in the last two decades (1995-2015), but only 892 results published from 1970 to 1994. In an analysis of all studies published in Music Perception from 1983-2010, Tirovolas and Levitin (2011) found that there has been a significant increase in the use of standardized assessments for both musical and non-musical abilities; this finding they attribute to the growing interest of possible transfer effects from music training. The next section will review the literature comparing musicians to non-musicians, or administering music training to a control group, within the context of larger cognitive domains.
Auditory ability

Perhaps the most consistent finding between musicians and non-musicians is in auditory processing ability. This makes sense, as music is primarily auditory in nature. Musicians appear to have an advantage over non-musicians on tests of central auditory processing. Specifically, musicians tend to exhibit enhanced speech-in-noise (SIN) ability (Parbery-Clark et al., 2009), which is further enhanced if music training began early in life (Strait et al., 2014). Enhanced SIN ability has been found in children between 7-12 years with at least four years of training (Strait et al., 2012), and in children as young as three to five years of age with at least 12 consecutive months of music training (Strait et al., 2013). Only one study has failed to replicate these results with adults (Ruggles, Freyman & Oxenham, 2014). This enhanced ability in musicians has been owed to superior working memory capacity and auditory temporal acuity, which has been posited to form the basis for perception of SIN (Besser et al., 2013; Parbery-Clark et al., 2009, 2011). The advantage that musicians demonstrate on tests of central auditory processing appear to be independent of hearing sensitivity (Zendel & Alain, 2012), suggesting that musicians have superior listening skills, rather than hearing ability (Schellenberg, 2015).

In attempt to explain why music training enhances auditory processing more so than speech, Patel (2014) suggested the OPERA hypothesis. This hypothesis posits, “musical training drives adaptive plasticity in speech processing networks when 5 conditions are met”. These conditions include, a) overlap in anatomical brain structures, b) precision, that music places a higher demand on these structures than speech, c) emotion, the music activities elicit an emotional response, d) repetition, the targeted network is repeatedly activated by musical activity and, e) attention, focused attention is caused by the musical activity.

Language ability

The influence of music training on the development and processing of language ability is an important cognitive ability with far reaching implications. As a result, there is a large body of research examining differences between musicians and non-musicians from a behavioral, structural and neural standpoint. Studies investigating the effect of music training on language through standardized assessments point towards a possible causal link.

Auditory processing of language and music share similar brain mechanisms, however benefits of music training have also extended into other areas of language such as verbal fluency (Forgeard, Winner, Norton & Schlaug, 2008), verbal memory (Ho, Cheung & Chan, 2003),
language development (Shook et al., 2013) and phonological awareness (as reviewed by Tierney & Kraus, 2013). Moreover, music training has been associated with enhanced second language acquisition (Milovanov et al., 2008). In a study conducted by Moreno and colleagues (2011), children who received computerized music training demonstrated greater verbal intelligence after only 20 days, as compared to controls who received computerized training in visual arts.

This enhanced processing of language extends into speech perception, with musicians demonstrating greater sensitivity to speech intonation (Besson, Chobert, Marie, 2011) and speech encoding (Tierney et al., 2013), perhaps due to better pitch discrimination (Tervaniemi et al., 2005). In a recent study with children between the ages of 8 and 10 years, Chobert and colleagues (2014) found that children in a music training intervention were better in discriminating voice onset time and syllabic duration than children with painting training.

Visual spatial ability

There is evidence of a causal link between music training and visuospatial abilities (Brochard et al., 2004; Hetland, 2000), possibly due to an association between pitch (frequency) and spatial representations (Lidji et al., 2007; Platel et al., 1997; Rusconi, et al., 2006; Weiss, et al., 2014). Findings have even suggested that the processing of pitch in music requires the ability to process spatial representations in the non-musical modalities, such as in vision (Douglas & Bilkey, 2007).

In a study conducted by Bilhartz, Bruhn and Olson (1999), 71 children between the ages of 4-6 years were randomly assigned to either a music training or control group. The children were all given the Stanford-Binet Intelligence Scale (4th edition) before and after the intervention. Results showed that the musically trained children performed significantly better than controls on the Bead Memory task, in which children are shown an image of beads on a string, and then asked to reproduce it exactly. This task is associated with various measures of visuospatial ability. Other studies have also found enhanced performance on standardized tasks of visuospatial memory, such as on the Rey Visual Design Learning Test (Jakobson et al., 2008) and the line bisection task (Patson et al., 2006).

Despite findings that music training may cause transfer effects in visual-spatial abilities, some studies have been unable to support such claims. Aleman and colleagues (2000) randomly assigned participants to a music and control group in order to determine whether improved musical imagery ability transferred to visual imagery. Results indicated that there was no
difference between groups in visual imagery. Other studies have been unable to find differences between musicians and non-musicians on tasks of spatial memory (Helmbold et al., 2005), spatial-navigation (Mehr et al., 2013), visual memory (Ho, Cheun & Chan, 2003) and visual-spatial ability (Hansen, Wallentin & Vuust, 2012). These conflicting results are likely due to differences in methodological approaches (Weiss et al., 2014) and differences in musician and non-musician samples.

Memory

There is conflicting evidence for whether there are transfer effects from music training to memory processes. In a study conducted by Tierney, Bergeson-Dana and Pisoni (2008), experienced musicians were compared to three control groups (gymnasts, university students and video game players). Participants were tested on sequence memory and word familiarity tasks. It was found that musicians demonstrated greater memory spans and better reproduction of stimuli.

Musicians have also been found to demonstrate superior working memory over non-musicians, specifically on tasks of visual, phonological and executive memory (George & Coch, 2011). Tasks involving rhythmic stimuli have found an increased reliance on working memory in musicians (Saito & Ishio, 1998). Musicians typically exhibit greater verbal working memory (Cohen et al., 2011; Douglas & Bilkey, 2007; Franklin et al., 2008; Ho, Cheun & Chan, 2003), however not always (Helmbold et al., 2005), and it appears to be restricted to the auditory domain (Cohen et al., 2007). It has been suggested that memory enhancements in musicians may be restricted to the domain of music (Oura & Hatano, 1988).

Attention

Few studies have demonstrated transfer effects of music training on attention (Roden et al., 2014), however existing evidence suggests that a causal relationship may exist (Besson, Chobert & Marie, 2011; Moreno et al., 2011). In a longitudinal study with 142 primary school children (Rickard et al., 2010), children were randomly assigned to one of three groups: music training, juggling training or no training. Although the music group significantly outperformed the no training group on measures of attention, they did not significantly differ from the juggling group. In contrast, Roden and colleagues (2014) randomly assigned 345 children to either music or natural science training over a period of 18 months. Children were tested on various measures of attention and processing speed. While the musically trained children demonstrated superior
information processing speed and music ability, children trained in natural sciences (control group) actually outperformed the music group on visual attention tasks.

As can be expected, musicians often demonstrate superior auditory attention ability, in that they are more attuned to changes in melody, timbre and rhythm (Madsen & Geringer, 1990). In a study by Patson, Hogg and Tippett (2006), lateralization of attention was explored by comparing musicians to non-musicians. Although neurologically intact adults typically make mistakes in the left visual space, musicians in this study demonstrated a more balanced attentional capacity and enhanced visuomotor ability than non-musicians. This was exhibited with faster reaction times and accuracy rates. Music training advantages, however, are not always found in attention tasks (Costa-Giomi, 1999; Helmbold et al., 2005).

**Intelligence**

Intelligence is typically assessed by way of a standardized assessment, thus many researchers have compared musicians to non-musicians in attempt to further understand the effects of music training. Despite the important implications that intelligence scores between musicians and non-musicians may have, there does not exist one agreed-upon definition of intelligence. One theory is that intelligence is made up of a variety of skills, such as Gardner’s theory of multiple intelligences (1982) or Thurstone’s primary mental abilities (1938). Alternatively, intelligence can be conceptualized as a quantifiable, composite ability, such as Spearman’s g factor (Spearman, 1927).

Researchers have also expressed concern about the testing of intelligence, specifically whether they are culturally dependent or biased in other ways. Raven’s Progressive Matrices (Raven, 1941) and the Cattell’s Culture Fair Intelligence Test (Cattell, 1949) are two tests often used as tests of non-verbal intelligence used cross-culturally. The former is a test of reasoning, in which participants are presented with patterns in the form of matrices and asked to determine the missing element. The Cattell’s Culture Fair Intelligence Test estimates intelligence independent of socioculture and environment.

Although there exists various conceptions of intelligence, the present study defines it as scores between musicians and non-musicians from standardized assessments, such as the Weschler’s Adult Intelligence Scale (WAIS). Thus it becomes possible to investigate possible associations between performance on intelligence tests and musical ability. Musicians tend to demonstrate greater academic achievement and IQ scores than non-musicians (Schellenberg,
MUSIC TRAINING AND TRANSFER EFFECTS: A META-ANALYSIS

2011; Wetter, Koerner & Schwaninger, 2009). However, some have suggested that individuals with an increased interest in education may be more likely to begin, and continue, music education (Schellenberg, 2011). Moreover, music training tends to involve increased attention from teachers and extra time in an educational setting. This may develop a more positive attitude towards learning, thus refining focus in other subject areas (Elpus, 2013). Regardless, there does appear to be evidence that music training is associated with general intelligence (Babo, 2004; Butzlaff, 2000; Schellenberg, 2011).

The transfer of abilities measured by intelligence tests from music training has been supported in experimental studies with children. In a large study that extracted information from a survey database (SOEP), adolescents at age 17 who started private music training before age 8 had higher IQ and cognitive ability than both non-musicians and athletes with the same criteria as the musicians, even after controlling for demographics (Hille & Schupp, 2014). Moreno and colleagues (2011) found similar results after comparing the performance of children between the ages of 4 and 6 years on measures of verbal intelligence. The children were randomly assigned to computerized training in either music or visual arts. Results showed that children in the music group demonstrated enhanced performance on verbal intelligence, supporting the notion of transfer of high-level cognitive skills from music to non-music domains and enhanced executive functioning. These findings contrast a study with 272 elementary school boys (Hille et al., 2011), in which participants took the German adaptation of the Cattell’s Culture Fair Intelligence Test and a musical background questionnaire. Results showed that boys who played an instrument demonstrated a higher non-verbal IQ than those who didn’t, but there was no difference in IQ between boys who sang in a choir and those who didn’t. Although it appears that musicians outperform non-musicians in IQ scores, conflicting results make it difficult to interpret these findings against that of other studies.

1.4 Defining a musician

The way in which an individual study defines musician and non-musician groups has important ramifications when attempting to conduct a literature review. Existing research has identified numerous variables related to music training and study methodology that may explain between-study variance. The most common categorization of musician samples is based on years of music training, however other considerations, such as age of training onset, type of training
and instrumental focus, might provide a more comprehensive understanding of potential differences between musicians and non-musicians.

_Years of music training_

Music training can be defined as the purposeful and focused act of developing a set of musical skills. But how many years of training does one require to become a musician? It is entirely possible that an individual with any amount of music training will demonstrate superior performance on some cognitive or music processing tasks; however, most studies examine expert or professional musicians, a criterion that is typically defined as more than 10 years of training (Tirovolas & Levitin, 2011). One could argue that these types of musicians represent a small subset of musically trained individuals within society and have achieved a level that is perhaps not obtainable at all stages of life. Conversely, some studies have defined a musician as an individual with as few as four years of training (Shook et al., 2013). It is unclear whether such samples would demonstrate a comparable performance to the expert musicians, or if they would fall between that of experts and non-musicians. Moreover, years of training is problematic, as it strips away potentially important information; for example, two individuals may have equal years of formal training behind them, but one may be completing a DMA, having focused on refining technical ability on one instrument the entire time, while the other may have switched between multiple instruments. It is unclear whether such samples would demonstrate a comparable performance to the expert musicians, or if they would fall between that of experts and non-musicians.

Few studies have examined the impact of short-term music training on cognition. One such study (Hyde et al., 2009) randomly assigned preschool children to either 15 months of weekly half-hour keyboard lessons or no training. The music-group demonstrated structural changes in the primary auditory and motor cortices, which correlated to improved motor skills and auditory melodic and rhythmic discrimination skills. That musically trained children demonstrated structural and functional enhancements after only 15 months supports the possibility that even a small amount of training can elicit change. But, are these short-term changes only possible in young children, or could adults benefit from similar enhancements?

In a study examining pitch discrimination in young adults Micheyl and colleagues (2006) found after only four to eight hours of auditory training, non-musicians demonstrated equivalent thresholds in pitch discrimination to the musician group. These findings demonstrate superior
ability in music processing abilities after very brief training. Furthermore, since pitch discrimination skills have been associated with speech perception (Besson et al., 2007), these findings also support cognitive transfer after short-term training. Importantly, the researchers used stimuli that consisted of both sine waves and complex tones in their pitch training, thus was not entirely musical in nature. However, pitch discrimination is a necessary ability in music, and is a basic skill improved through music training. Therefore, one could posit that short-term music training does improve pitch processing, which, in turn, may lead to transfer in the language domains. A similar study with children found that, after only eight weeks of training, participants in the music group showed enhanced pitch perception skills (Moreno & Besson, 2005).

Finally, in a study with 31 older adults over the age of 60 years (Bugos et al., 2007), participants were randomly assigned to either six months of piano instruction or no training. Participants in the music group received one half-hour lesson per week, with three additional hours of practice each week. These lessons had an emphasis on technical ability (scales and arpeggios), which is consistent with formal music instruction. After only six months, the musically trained group demonstrated significantly improved performance on Digit Span, which assesses visual attention and processing speed, and the Trail Making Test, which is used to assess working memory (Joy et al., 2004). The control group, however, had no improvement.

**Nature of the relationship between years of training and ability**

If musicians and non-musicians differ on some tasks of music processing and cognition, an important question is whether there is an incremental, or stepwise, relationship between years of training and performance. In other words, is the hypothetical association between music training and a musical or cognitive domain qualitative (based on categories defined by amount of training), or quantitative (linear/continuous)? If there is an association with a musical or cognitive domain that can be measured quantitatively, such as a standardized assessment, then it becomes possible to predict the musical or cognitive ability elicited by a specific number of years of training.

There is evidence for a linear relationship between music training and cognitive transfer in the literature. In a recent study conducted by Oeschslin and colleagues (2013), three levels of music ability (non-musicians, amateur, and expert musicians) were compared on measures of harmonic sensitivity. The fMRI data demonstrated evidence for "stepwise modulation of brain responses" (p. 2214) by strength of harmonic violation and level of musical expertise.
Behavioural results showed an incremental enhancement in visual working memory between each of the groups. A similar study conducted by Hanna-Pladdy and MacKay (2011) with older adults separated participants into three levels of musical experience: non-musicians, low-activity musicians (1-9 years of training), and high-activity musicians (more than 10 years of training). High-activity musicians significantly outperformed non-musicians on tasks of nonverbal memory recall, visuomotor speed and sequencing, and cognitive flexibility. Despite non-significance between the high- and low-activity musicians, the latter group’s scores fell between the high-activity and non-musicians, suggesting a linear relationship. Another important finding from this study revealed that active involvement in music was not necessary to demonstrate this incremental relationship. In other words, it provides support for the possibility that enhancing effects of music training persist into older age after training stopped, which was substantiated in a follow-up study (Hanna-Pladdy & Gajewski, 2012).

Other studies have also supported a linear relationship between years of training and transfer ability. In the music perception domain, an incremental relationship between years of training and perception of various musical elements was found (Seither-Preisler et al., 2007). Finally, Halpern and colleagues (1995) found a linear relationship with verbal intelligence in younger adults with low (less than 2 years), moderate (2-8 years), and high (more than 8 years) of formal music training.

Structural differences between musicians with varying levels of training have supported the possibility of an incremental relationship between music and functional ability. In a study comparing amateur and expert male pianists to non-musicians (Gaser & Schlaug, 2003), expert musicians showed larger gray matter density in motor, auditory, and visual-spatial brain regions than in amateur and non-musicians. Amateur musicians also had enhanced gray matter density in the same regions, but to a lesser extent than the expert musicians. These findings correlated to years of musical training to support behavioural results in other studies, providing more evidence for the possibility of an incremental relationship between music training and enhancing effects. Groussard and colleagues (2014) were also able to find a linear relationship between grey matter density and years of music training; however structural modifications in the left superior temporal, posterior cingulate and right supplementary motor areas only appeared after several years of practice, thus were not linear.
Differences in length of training in musician and non-musician samples between studies may help explain conflicting findings. Although many studies characterize musicians as having more than 10 years of training, studies using musicians with fewer years of training may help shed light on whether there is a linear relationship to transfer effects. Moreover, examining the criteria for a ‘non-musician’ may also assist in understanding conflicting findings; traditionally a non-musician is an individual with no training, though these individuals can be rare, given the ubiquity of in-school music classes. In some cases, studies classify non-musicians as having some training, so long as it is not extensive in length.

**Age of music training onset**

It has been suggested that there may be a critical or sensitive (Hanna-Pladdy & Gajewski, 2012; Musacchia et al., 2008) period of music acquisition around nine years of age due to a more “plastic” brain at a younger age. There is evidence from neuroimaging studies supporting a sensitive period for music (Bailey, Zatorre & Penhune, 2014; Steele et al., 2013), however no brain area seems to be specifically influenced by age of onset. Specifically, one study failed to find difference in gray matter density between musicians who started training before the age of 7 versus after (Groussard et al., 2014).

Behavioural evidence is also not very clear. There is support for an early-trained advantage in some near-transfer skills, such as motor ability (Bailey et al., 2014; Watanabe et al., 2007) and visual- and auditory-motor synchronization (Bailey & Penhune, 2012), even after controlling for years of experience. Far-transfer differences between early- and late-trained musicians have also been found in verbal working memory (Hanna-Pladdy & Gajewski, 2012), but working memory advantages in early-trained musicians has been met with conflicting results (Bailey & Penhune, 2012).

The fundamental principle of brain plasticity suggests a sensitive period arises when “maturational plasticity in a brain region associated with a specific skill is paired with intensive experience or practice of that skill” (Bailey et al., 2014, p. 756). This means that there should be structural or functional differences between individuals who begin training at different childhood ages, but interestingly, this does not always appear to be the case (Skoe & Kraus, 2012; Strait et al., 2014).
Type of training

Also inconsistent across studies is the type of music training that musician samples have received. There are many distinct kinds of training employed in music education, making it an important consideration. Traditional music training focuses on reading music notation while either playing an instrument or singing. There is often very little opportunity for improvisation. Training may also occur in jazz, where there is a heavy element of improvisation and very little music notation reading. The type of training one engages in has implications in terms of the cognitive skills being used. It is possible that classically trained musicians, who focus on notation reading, may exhibit stronger visual processing transfer than jazz musicians, who may have stronger auditory working memory abilities. Some have suggested that differences between classical and jazz training may be as different as auditory or visual training (Korenman & Peynircioğlu, 2007).

Cross-sectional studies that involve musician samples typically define musicians based on number of years of formal music training. As a result, most cross-sectional studies actually focus on the effect of private music lessons on cognitive ability. In contrast to this, experimental studies typically administer a music training intervention that manifests in the form of small, less formal group music lessons. One such study found that 2 years of informal training led to improved auditory and neural processing of speech sounds (Kraus et al., 2014), suggesting that informal training can elicit change. In contrast, another study found that musicians with private training have lower thresholds for detecting pitch mistuning (Lynch et al., 1991), perhaps due to more rigorous training.

Differences in instrumental focus, including voice, may also change the effect of music training on near- and far-transfer abilities. There is an obvious language element associated with vocal training, which may produce stronger transfer effects to speech processing and perception than in instrumentalists. However, the fine motor skills associated with playing an instrument likely contribute to superior performance on sensorimotor tasks than vocalists. As one might expect, drummers physically synchronize to a beat better than pianists, singers and non-musicians; but drummers do not outperform pianists at discriminating between rhythms (Krause et al., 2010), perhaps because piano playing also has a percussive element. Moreover, musicians who play an instrument in which fine-tuning is required, such as violin, show greater pitch discrimination abilities in chords than other musicians (Koelsch et al., 1999). Importantly, one
study comparing instrumentalists to vocalists did not find differences on auditory pitch discrimination (Nikjeh et al., 2008), suggesting that perhaps instrumental focus does not affect music processing.

Whether musicians are actively involved in music training or performance also appears to make a difference. In a study by Hanna-Pladdy and MacKay (2011), musicians were separated into low- and high-activity groups. The study showed that, when participants were matched on years of formal music training, the cognitive performance of low-activity musicians fell between that of non-musicians and high-activity musicians.

**Issues with design**

The vast majority of studies examining the effect of music training on cognition are cross-sectional, or quasi-experimental in design, meaning that they have treatment groups and outcome measures, but do not employ random-assignment (Zakzanis, 1998). One problem with this methodology is the inherent assumption, upon finding significance, that the only difference between musicians and non-musicians is music training. Such correlational studies often compare individuals with a varying amount of music training to individuals with very little to no music training on various standardized tasks. A problem with such studies is that it is unclear whether differences between musicians and non-musician groups are largely the result of music training, or pre-existing differences (Babo, 2004; Corrigal et al., 2013; as reviewed by Perez & Hyde, 2003; Schellenberg, 2006). Alternatively, studies with experimental designs randomly assign individuals to a treatment group, such as music training, and a control group. Cross-sectional, or correlational, studies provide researchers with insight into the differences between musicians and non-musicians, whereas experimental studies allow researchers to isolate the effect of music training on various cognitive tasks.

Only one study, to the author’s knowledge, has examined the effect of music training on cognitive transfer, specifically reading ability. Butzlaff (2000) conducted two separate meta-analyses on the effect of music instruction on reading ability; one for correlational studies (quasi-experimental) and one for studies that employed random-assignment (experimental). A more robust, albeit small, effect was found for the quasi-experimental studies analysis than for experimental studies. Since one cannot conclude causation from correlation, Butzlaff concluded that pre-existing differences must be the cause for findings of differences between musicians and
non-musicians on reading ability. As a result, study design is an important consideration when attempting to delineate the effect of music training on transfer.

1.5 Need for Study

Conflicting evidence and differences within musician samples have made it difficult to determine what differences, if any, exist in terms of music processing and cognitive transfer between musicians and non-musicians. However, conflicting results does not imply that there is no effect, or that something important has not already been discovered (Borenstein et al., 2009, p. 12). Although a substantial body of research has developed, to the author’s knowledge, this literature has not yet been reviewed quantitatively by way of meta-analytic methods. Considering the large number of studies examining differences between musicians and non-musicians, a meta-analysis is most appropriate. This type of analysis goes beyond a systematic review of music training, to yield average strengths of the relationship between musical experience and transfer abilities. One of the greatest limitations of a systematic review is its focus on statistical significance. Misunderstandings of statistical significance have led to many errors in interpretation. One of the major errors made has to do with the interpretation of the \( p \)-value itself. Often it is believed that a \( p \)-value of .05 is “less significant” than .01, or that finding significance implies a large effect. However, a \( p \)-level is simply an a priori measure that researchers determine, and only states “how improbable an event could be under the null hypothesis” (Bakan, 1966, p. 429). Moreover, the probability of rejecting the null hypothesis and finding significance is influenced by many decisions regarding study design that the researcher has control over such as sample size, one- versus two-tailed testing and alpha level. A common belief is that significance testing is an objective measure, however most often “psychologists convert intuitions and hypotheses into procedures that will yield significant results” (Bakan, 1966, p. 426).

Systematic reviews tend to dichotomize \( p \)-values from studies as either significant or not, which strips away meaningful data and leads to misinterpretation of findings (Zakzanis, 1998). Finding significance is often viewed as the most important finding, however some have argued that it should be “a tiny part of an inquiry concerned with the size and importance of relationships” (Ziliak & McCloskey, 2009). Moreover, finding a significant result only refers to the probability of two groups coming from the sample population, but nothing about the impact
of a treatment variable. Another issue is with the terminology itself; statistical significance is often confused as being scientifically significant, or of practical importance (Ziliak & McCloskey, 2009) An ancillary and often more meaningful measure is that of effect size. An effect size is a value that reflects the “magnitude of the treatment effect or strength of a relationship between two variables” (Borenstein et al., 2009, p. 3), and is independent of sample size. Effect sizes in individual studies can be combined and understood in a more objective manner than systematic reviews through a meta-analysis.

Thus, the current study will seek to undertake a meta-analysis of all published studies that compare musicians to non-musicians, or employ music training as an intervention. Through meta-analyses, this study will examine whether musicians and non-musicians differ on tasks related to both music processing and cognitive domains through meta-analyses. Possible confounds, as indicated by the literature, will be used as moderating variables in order to assess their impact on overall summary effects. The first analysis will examine the effect of music training on music processing domains (ex. pitch, melody and temporal perception), and a second analysis on cognitive domains (ex. orientation and attention, memory, and executive functions). Thus, the current study will attempt to answer six research questions, presented alongside of hypotheses in table 1.

Table 1. Research questions and hypotheses

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do musicians and non-musicians differ on cognitive and music perception abilities, and if so, on what specific domains do they differ?</td>
<td>Music training will have a greater effect in tasks directly relevant to musical skills, such as the perception and discrimination of pitch, temporal events (rhythm), and timbre. It is predicted that music training will have the greatest effect on cognitive tasks similar to those used in music, such as motor speed and audition. Conversely, music training is predicted to have a small, but reliable, effect on other non-music related tasks.</td>
</tr>
<tr>
<td>2. What is the magnitude of any observed differences in performance on these tasks?</td>
<td></td>
</tr>
<tr>
<td>3. Does music training have an incremental relationship with performance on tasks in which musicians and non-musicians differ,</td>
<td>Due to some research pointing towards a stepwise relationship between level of music expertise and cognitive performance (Oeschslin et al, 2013), it is</td>
</tr>
<tr>
<td>Question</td>
<td>Answer</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1. Are there summary effects such that more training results in better performance?</td>
<td>also predicted that increased years of music training will have a large effect.</td>
</tr>
<tr>
<td>4. What effect does type of music training (e.g., formal versus informal, or solo versus group) have on cognitive or music processing performance?</td>
<td>It is predicted that type of training will have an impact, such that private music lessons are likely to have a greater effect than group lessons. This is substantiated by the pedagogical differences between one-on-one versus group lessons.</td>
</tr>
<tr>
<td>5. Does age of training onset, or age of the participants, have an effect on performance?</td>
<td>Age of training onset is predicted to influence the summary effects across domains. Furthermore, age of the participants is expected to have an effect on performance, such that older adults will exhibit lower effects than adults or children.</td>
</tr>
<tr>
<td>6. Is there a difference in summary effects between experimental and quasi-experimental studies?</td>
<td>There will be a difference between summary effects grouped by study design, such that summary effects in experimental designs will be smaller than in quasi-experimental designs, due to previous findings (Butzlaff, 2000).</td>
</tr>
</tbody>
</table>

Chapter 2

Methods

2.1 Meta-Analytic Procedures

A meta-analysis was conducted to answer the above research questions regarding music training’s effect on cognitive ability. A meta-analysis is a “quantitative summary of research domains that describe the typical strength of an effect, its variability, its statistical significance, and the nature of the moderator variables” (Rosenthal, 1995, p. 183). Meta-analysis was developed as a method of quantitatively synthesizing the results of a large number of similar studies together (O'Rourke, 2007). The term 'meta-analysis', coined in 1976, initially referred to “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” (Glass, 1976, p. 3). However it is important to note that many researchers had been long advocating the importance of effect size reporting alongside of inferential statistics (Bakan, 1966).
In the current investigation, effect sizes were calculated from primary studies based on the performance of musicians and non-musicians on various tasks related to music processing and cognition. Since summary effects are comprised of effect sizes from primary studies, it is important that a thorough search for all articles within a given domain is conducted.

2.1 Literature Search

Two separate database searches were conducted in order to extract all articles relevant to music processing and cognition. An initial search was conducted through the following online databases: PsychINFO, PubMed, Medline, Web of Science, and Scopus. Articles related to “music processing” (tasks evaluating music perception/processing skills) and musical experience were obtained with the following keywords: [music AND perception, memory, recognition, AND acoustics, atonal, audit*, contour, dynamics, emotion, entrain, harmon*, imagery, instrument*, interval, meaning, melod*, meter, metric*, music*, octave, performance, phrase, pitch, practice, rhythm, skill, sing*, structure, sychron*, temporal, tension, timbre, time, tonal, tones, tonality, vocal]. Articles related to “cognition” (tasks evaluating skills related to cognitive transfer), and musical experiences were obtained with the following keywords: [music*, music education, music expertise AND neuropsychology, memory, attention, executive function, language, visual spatial, motor speed, manual dexterity], (see Appendix A). After eliminating duplicate articles, there were 7527 articles in the music processing database, and 3162 articles in the cognition database. Articles obtained through the online database search were downloaded and examined. Any articles that were not available for download through the University of Toronto library database were obtained through either inter-library loan or retrieved manually.

In order to ensure the inclusion of all published articles related to music training and cognition or music processing, a manual hand search of select music cognition journals was also conducted. The following journals and volumes were hand searched: Music Perception (Vol. 1-16, all issues), Psychology of Music (Vol. 1-21, all issues), and Musicae Scientiae (Vol. 1-8, all issues). Finally, additional articles were retrieved through the snowball method, by reading included studies and ensuring that any cited articles of relevance were included.
2.3 Criteria for Inclusion

Articles were included in the final analysis if they met the following criteria: (1) empirical study that compared musicians to non-musicians, (2) healthy participants (no clinical populations), (3) published in English (due to time restraints), (4) peer-reviewed, (5) task must have a correct and incorrect answer. The lead author of the investigation assessed each article based on these criteria. Case studies were not included in the final analysis, due to the lack of control group, and dissertations were not included unless they were peer-reviewed and published. Clinical samples present a problem in a meta-analysis when combined with healthy samples, as the effect of music training may be more or less in such studies due to other influences or circumstances. Finally, some studies utilized tasks, particularly related to music processing, in which there was not a correct answer; these studies were often exploratory, thus were not included. In the event that an article met all of the inclusion criteria but did not provide enough information with which to calculate effect sizes, authors were contacted by email for more information. A total of 13 studies (5 music processing, 8 cognition) were excluded, as authors did not provide sufficient information in time for the analysis.

After evaluating articles based on the inclusion criteria, data was extracted from 82 studies (music processing database), and 73 studies (cognition database). A data extraction sheet was developed in Excel to organize information. Three types of variable coding were used: substantive, methodological and extrinsic (Sanchez-Meca & Marin-Martinez, 2010). Substantive variables include those related to the research questions, such as type of music training and mean age of participants. Methodological variables are those related to the study design, such as whether the study is experimental or independent. Finally, extrinsic variables are those not related to the question or design, but have an impact on the results, such as country where the study was carried out. In total, 22 codes were assigned to each study (see Table 1).

The $d$ effect-size statistic (Cohen, 1988), or standardized mean difference, was calculated for each comparison between musicians and non-musicians by the means and pooled standard deviation. When means and standard deviations were not reported, the effects were calculated based on the available statistics and calculated using the formulas provided by Zakzanis (2001). The variance for each $d$ was also calculated (its square root yielding standard deviation). In some cases, multiple effect sizes were calculated from the same study, when various tasks were administered to participants. Cohen’s $d$ is a more preferred measure of effect size over Hedge’s $g$.
MUSIC TRAINING AND TRANSFER EFFECTS: A META-ANALYSIS

(Zakzanis, 2001) and Pearson’s r (Borenstein et al., 2009) in psychological research, due inherent biases associated with each measure.

The interpretation of Cohen’s $d$ is often reduced to cut-off levels, such that a small effect is represented by $d=0.2$, and a medium and large effect by 0.5 and 0.8, respectively (Cohen, 1977). However, many researchers have cautioned that these cut-offs should be used as guidelines rather than strict rule (Baguley, 2009), as these values are often accompanied by great practical and clinical significance. For example, although 0.2 is considered by Cohen (1977) to be “small”, this value means that there is a higher statistical probability (56%) that an individual randomly chosen from the musician group will have a higher mean than a non-musician. Moreover, this same value means that for every 100 musicians, just over 6 individuals will demonstrate superior performance (Magnusson, 2014). Therefore, even a small Cohen’s $d$ has pragmatic meaning. In the current analysis, each summary effect size will be tested against the null hypothesis ($d=0$) through significance testing. Similarly, effects obtained through meta-regression and sub-group analysis were statistically compared the same way that means are tested in primary studies. Although significance testing is limited in the information it provides, pairing it with effect sizes provides rich and meaningful data.

Table 2. Data extraction sheet: Information coded from each article included in the meta-analysis

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome</strong></td>
<td>The specific neuropsychology or music processing task(s) employed in the study</td>
</tr>
<tr>
<td><strong>Authors</strong></td>
<td>Authors of the study</td>
</tr>
<tr>
<td><strong>Title</strong></td>
<td>Publication title</td>
</tr>
<tr>
<td><strong>Year</strong></td>
<td>Year of publication</td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td>Study design: Experimental/matched pairs or Quasi-Experimental/Cross-sectional</td>
</tr>
<tr>
<td><strong>Country</strong></td>
<td>Country where study took place</td>
</tr>
<tr>
<td><strong>SigLevel</strong></td>
<td>The significance levels obtained in the study for each task</td>
</tr>
<tr>
<td><strong>YrsTrainNM</strong></td>
<td>Average years of training the non-musician group</td>
</tr>
<tr>
<td><strong>YrsTrainM</strong></td>
<td>Average years of music training the music group</td>
</tr>
<tr>
<td><strong>TrainType</strong></td>
<td>Type of music training employed or received. Options: formal, informal, self-</td>
</tr>
</tbody>
</table>
taught or unspecified.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SoloGroup</td>
<td>Whether music training occurred privately or in a group.</td>
</tr>
<tr>
<td>AgeGroup</td>
<td>Age demographic of participants: Children [0-17yrs], Adults [18-60yrs] or Older adults [60+yrs]</td>
</tr>
<tr>
<td>AgeOnset</td>
<td>Average age of music training onset (age that musicians began training)</td>
</tr>
<tr>
<td>ConAge</td>
<td>Mean age of the control group</td>
</tr>
<tr>
<td>ExpAge</td>
<td>Mean age of the experimental/music group</td>
</tr>
<tr>
<td>N(Con)</td>
<td>Number of participants in the control group</td>
</tr>
<tr>
<td>N(Exp)</td>
<td>Number of participants in the experimental group</td>
</tr>
<tr>
<td>Total N</td>
<td>Total number of participants in study</td>
</tr>
<tr>
<td>Effect Size</td>
<td>Cohen’s d (standardized mean difference) and its standard deviation</td>
</tr>
<tr>
<td>Comment</td>
<td>Personal comments on study</td>
</tr>
</tbody>
</table>

### 2.4 Data Extraction

**Data Extraction and Categorization – Music Processing**

Effect sizes in the music processing database were calculated from various procedures spanning nearly five decades of research. Categorization of these tasks posed a unique challenge, as the tasks were often designed for a single investigation rather than the use of a standardized test. Therefore, effects were coded based on the outcome being measured (ex. frequency threshold) and placed into categories outlined by Tirovolas and Levitin (2011). In that study, every article that had been published in the journal *Music Perception* between 1983 and 2010 were categorized based on a number of measures (ex. outcome being investigated, type of participants, stimuli used, etc). Therefore, effects were grouped by *Domains*, and then further broken down into more specific *Categories* (as outlined by Tirovolas & Levitin, 2011) (see Table 3).

**Table 3. Music Processing Domain Categorization based on Tirovolas & Levitin (2011)**

<table>
<thead>
<tr>
<th>Domains</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch Perception</td>
<td>Chord Perception/Discrimination</td>
</tr>
<tr>
<td></td>
<td>Consonance/Dissonance Perception</td>
</tr>
<tr>
<td></td>
<td>Frequency Discrimination (Isolated Pitches)</td>
</tr>
<tr>
<td></td>
<td>Interval Discrimination (Isolated)</td>
</tr>
<tr>
<td></td>
<td>Loudness Perception</td>
</tr>
<tr>
<td></td>
<td>Pitch Processing</td>
</tr>
<tr>
<td>Melody Perception</td>
<td>Chord Perception/Discrimination</td>
</tr>
<tr>
<td></td>
<td>Contour Perception</td>
</tr>
<tr>
<td></td>
<td>Interval Discrimination (Sequential)</td>
</tr>
<tr>
<td></td>
<td>Melody Recognition</td>
</tr>
<tr>
<td></td>
<td>Scale Perception</td>
</tr>
<tr>
<td></td>
<td>Transposition</td>
</tr>
<tr>
<td>Timbre Perception</td>
<td>Timbre Perception/ Discrimination</td>
</tr>
<tr>
<td>Temporal Perception</td>
<td>Temporal/Rhythmic Discrimination</td>
</tr>
<tr>
<td></td>
<td>Tempo Perception</td>
</tr>
<tr>
<td></td>
<td>Reproduction</td>
</tr>
<tr>
<td>Musical Memory</td>
<td>Imagery</td>
</tr>
<tr>
<td></td>
<td>Melody Recognition/Discrimination</td>
</tr>
<tr>
<td></td>
<td>Memory for Pitch</td>
</tr>
</tbody>
</table>

Effect sizes from articles were categorized based on Tirovolas & Levitin’s (2011) analysis of all publications from Music Perception from 1983-2010.

**Data Extraction and Categorization – Cognition**

Each calculated effect size was organized into categories for analysis. Standardized assessments in the cognition database were organized according to Lezak et al. (2012) in *Neuropsychological Assessment*. Each standardized assessment, initially coded as *outcome*, was assigned to a specific *Domain* (organized by chapter), *Category* (within chapter) and *Task Categorization* (sub-headings within chapters). In total, the cognition analysis comprised of 7 domains, 15 categories, and 37 task categorizations (see Table 2). In some cases, particularly when an assessment designed for children was used in a study, tasks from primary studies were not present in Lezak’s (2012) textbook. Therefore, these tasks were placed in the most appropriate category according to its testing criteria and ability being measured. Importantly, the
three domains (represented in bold in table 2) were created for this analysis in order to evaluate their effect sizes with regards to musical experience.

**Table 4. Cognition Domain Categorization based on (Lezak et al., 2012)**

<table>
<thead>
<tr>
<th>Domains</th>
<th>Category</th>
<th>Task Categorization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Audition</strong></td>
<td>Auditory Perception</td>
<td>Speech-in-Noise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speech Perception</td>
</tr>
<tr>
<td>Concept Formation and Reasoning</td>
<td>Concept Formation</td>
<td>Sort and Shift</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concept and Formation Tests in Verbal Format</td>
</tr>
<tr>
<td></td>
<td>Reasoning</td>
<td>Verbal Reasoning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reasoning about visually presented materials</td>
</tr>
<tr>
<td></td>
<td>Mathematical Procedures</td>
<td>Calculations</td>
</tr>
<tr>
<td>Construction and Motor Performance</td>
<td>Assembly and Building</td>
<td>Two-Dimensional Construction</td>
</tr>
<tr>
<td></td>
<td>Drawing</td>
<td>Copying</td>
</tr>
<tr>
<td>Executive Functions</td>
<td>The Executive Functions</td>
<td>Self-Regulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Planning and Decision Making</td>
</tr>
<tr>
<td><strong>General Intelligence</strong></td>
<td>Full-Scale IQ</td>
<td>Performance IQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Verbal IQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-Verbal IQ</td>
</tr>
<tr>
<td>Memory</td>
<td>Visual Memory</td>
<td>Visual Recall</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Visual Reproduction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Visual Learning</td>
</tr>
<tr>
<td></td>
<td>Verbal Memory</td>
<td>Words</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Story Recall</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Supraspan</td>
</tr>
<tr>
<td><strong>Motor</strong></td>
<td>Fine Motor Skills</td>
<td>Motor</td>
</tr>
<tr>
<td>Orientation and Attention</td>
<td>Attention, Processing Speed, and</td>
<td>Attentional Capacity</td>
</tr>
<tr>
<td></td>
<td>Working Memory</td>
<td>Complex Attention Tests</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Working Memory/Mental Tracking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Divided Attention</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concentration/Focused Attention</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Processing Speed</td>
</tr>
</tbody>
</table>
Publication Bias

Publication trends pose a unique problem for meta-analysts, as articles that find significant results tend to be published more often than non-significant results, termed publication bias (Borenstein et al., 2009). This phenomenon has also been referred to as the “file-drawer” problem, as “unpublished results are imagined to be tucked away in researchers’ file cabinets” (Scargle, 2000, p. 91). The problem with publication bias is that it leads to the illusion that a variable has an effect on an outcome due to the greater number of statistically significant results. Additionally, it is a problem for meta-analyses, as results may show a greater effect than actually exists. In response to this problem, Fail-Safe File Drawer methods have been suggested to determine whether a proposed effect is likely to exist. One such method, Rosenthal’s Fail Safe N, calculates the number of non-significant studies needed to bring an effect to zero, which is then subjectively interpreted to validate findings (Scargle, 2000). However, there are problems with this calculation; Rosenthal’s Fail Safe N focuses on significance testing and assumes that a mean effect is zero, thus cannot be negative (Borenstein et al., 2009, p. 285).

To ascertain the validity of included studies, the author subjectively determined the placement of studies into each code of the data extraction sheet. To define the heterogeneity (variability) in primary effect sizes, it was hypothesized that the effect size may differ according to differences in methodology, such as study design (experimental vs. quasi-experimental).
2.5 Moderator variables

In a primary study, independent variables are tested against a dependent variable in order to determine their effect on the outcome. In a meta-analysis, the same conceptual analysis occurs, however the effect size is the dependent variable, and covariates are at the level of the study rather than the subject (Borenstein et al., 2009, p. 187). In the current investigation, demographic variables (age, design, year, country and publication) and seven targeted moderator variables were used based on coding (Table 2): years of music training (ControlCode and MusicCode), type of training (TrainType and SoloGroup), age of music training onset (AgeOnset), design of study (Design) and grouping of tasks from primary studies into large ability being tested (Category). We are able to compare standardized effect sizes between experimental and cross-sectional or correlational designs in the same meta-analysis because Cohen’s $d$ has the same meaning regardless of study design (Borenstein et al., 2009, p. 30).

Previous studies have been inconsistent in defining musician groups in terms of years of training; therefore, a meta-analysis provides an opportunity to determine the potential impact years of training has on music and cognitive ability. For example, whether there is an incremental relationship. Moreover, the type of training musician groups will receive or have received, such as private versus group lessons, is a potential covariate that can be examined as a moderator variable. Previous research has suggested that age of training onset may be key to understanding the influence of music training on cognitive transfer (Hanna-Pladdy & Gajewski, 2012; Watanabe et al., 2007), therefore this variable was also included. Information related to these moderator variables was extracted from each article; information that was not reported in primary studies resulted in a blank cell for that moderator.

Considering age as a moderator variable is important when examining the effect of music training on cognition. It has been posited that music training may have a greater effect on children’s cognition, as their brains may be more “plastic” than the brains of adults (Jancke, 2009). This literature also poses some interesting questions regarding the brains of older adults, often defined as over the age of 60 years, and whether music training can elicit transfer effects. Therefore, the three age code categories (AgeGroup, ConAge and ExpAge) were also employed as covariates.
2.6 Data Analysis

Data was analyzed with the Comprehensive Meta-Analysis (V.3) software (Borenstein, Hedges, Higgins, & Ruthstein, 2009). Excel was used to organize the coding information extracted from primary studies. The dataset in a meta-analysis is composed of a matrix where the rows are the studies and the columns are the moderator variables, effect sizes and variance. Analysis of the dataset has the following objectives: calculate average effect sizes and their standard deviations and CIs, assess the heterogeneity of the effect sizes, and assess how the moderator variables can explain heterogeneity (Sutton & Higgins, 2008).

A random-effects model was used in all of the analyses, and each category was analyzed using meta-regression and sub-group analyses to examine moderator variables. Meta-regression employs the same techniques as multiple regression, however with studies as individual data points rather than participants, and effect size as the dependent variable (Borenstein et al., 2009, p. 203). Since there is no one true effect size when discussing the influence of music training on cognition, the current study will use a random-effects model. The goal of this model is to estimate the “mean of a distribution of effects” (Borenstein et al., 2009, p. 79), or the weighted average effect size, across studies. To illustrate, the null hypothesis in the fixed-effects model is that the effect size is zero in every study, however in the random-effects model it is that the mean or average effect size is zero (Borenstein et al., 2009, p. 83). The random-effects model makes the assumption that the studies included in the meta-analysis are a random sample of a larger population; however, the fixed-effects model questions whether a treatment was beneficial only for the population included in the analysis (Russo, 2007). This means that, despite having less power than the fixed-effects model, the random-effects model allows for generalization past the study to the same population (Rosenthal, 1995). In other words, a random-effects model for the current study means that the results can be used to further understand general differences between musicians and non-musicians. Since that is the intention of the current study, a random-effects model is most appropriate.

In the current study, the effect sizes for individual studies were bound by a confidence interval, reflecting “the precision with which the effect has been estimated in the study” (Borenstein et al., 2009, p. 5); smaller intervals indicate estimates that are more precise. Thus, individual studies with more precise estimates are assigned more weight towards the final analysis than studies with less precision. Ultimately, the goal of a meta-analysis is to calculate the
summary, or average, effect size, made of individual effects from primary studies. It then
determines how much that effect varies across studies; large variability suggests very different
implications than a robust effect (Borenstein et al., 2009, p. 9).

The confidence interval for the summary effect size was calculated by \( d \pm t \times \text{critical value} \),
where \( t \) is the critical value (significance cut-off based on \( k \)) multiplied by the standard error
(standard deviation divided by the square root of \( k \)). The \( k \) value represents the number of studies used to calculate
the effect size (Howell, 2009). The collections of effect sizes are visually represented with a
forest plot, pictorially showing the distribution of each individual effect in the analysis

**Tests of heterogeneity**

Variability within a single study is often described by the standard deviation of scores and
range. In a meta-analysis, however, heterogeneity refers to the variation in the true effect sizes,
which includes both true heterogeneity and within-study error (Borenstein et al., 2009, p. 107).
Heterogeneity shows the amount of consistency in results across studies. It is assessed by
comparing observations with what is expected if all of the studies shared the same effect size.
The resulting excess is then used to assess various measures of heterogeneity (Borenstein et al.,

Heterogeneity in the current study was tested using the Q statistic (Cochran, 1954), \( I^2 \)
statistic (Higgins & Thompson, 2002), and \( T^2 \) statistic. The Q statistic is considered to be the
confidence interval in a meta-analysis (Borenstein et al., 2009). A limitation of the Q statistic is
that it is influenced by the number of studies in the analysis, leading to poor power when the
number is low, and over-estimates heterogeneity when the sample is high (Huedo-Medina et al.,
2006). As a result, domains in the current study with few studies may result in smaller Q values
than is actually the case. Overall, the Q-statistic is only able to indicate whether effect sizes from
individual studies within the same domain significantly differ from each other. Significant Q
values may indicate an issue with how tasks were grouped into domains, or it may be the
consequence of different procedures used between tasks to measure an outcome.

The \( \tau^2 \) statistic, \( T^2 \), reflects the total amount of heterogeneity on the same scale used in
the meta-analysis, in this case the standardized mean difference (Cohen’s \( d \)) (Borenstein et al.,
2009). It reflects the amount that the true population effect sizes, from the primary studies in the
In a meta-analysis, the T² statistic equals the variance of a summary effect; therefore, its square root (T) is the standard deviation of the summary effect.

Finally, the I² statistic measures the percentage of total variation across studies due to heterogeneity, in order to assess the effects of music training; it is calculated by subtracting Q by its degrees of freedom (k-1), and dividing it by Q multiplied by 100. This results in a percentage of the amount of variability between effect sizes due to between-study heterogeneity (Huedo-Medina et al., 2006, p.194). As an example, I²=75 means that 75% of the variability between primary effect sizes is true variance, rather than sampling error, and thus could be explained by a moderating variable, such as years of music training (Borenstein et al., 2009). Through meta-regression, moderator variables result in R² statistics; this statistic is a percentage of the amount of true variance accounted for by a moderator. For example, if “years of music training” results in R² = 25%, it means that this variable accounts for 25% of the true variance in the I² statistic.

A classification suggested by Higgins and Thompson (2002) is that I²=25 corresponds to low heterogeneity, whereas I²=50 and I²=75 would correspond to a medium and high heterogeneity, respectively (Huedo-Medina et al., 2006). This measure is advantageous in that it is not affected by the number of studies and is accompanied by its own confidence interval (Briel et al., 2004). Like the Q statistic, I² is affected by a low number of studies, and statisticians recommend that both statistics be reported. Both I² and T² are positively correlated, but are not on the same scale.

In sum, Q tests the null hypothesis, that there is no dispersion across effects, while I² quantifies that dispersion. The I² statistic is represented as a percentage of the observed variance between studies that is due to real differences in the effect sizes (rather than random error). While standard deviation is conventionally used in primary studies to quantify dispersion between scores, Tau (T) is used in a meta-analysis to express between study (effect size) dispersion (Borenstein et al., 2009).

Funnel Plots

Forest and funnel plots are also used to determine whether there is publication bias in a meta-analysis. Both of these graphs demonstrate the relationship between study size and effect size. Since studies with larger than normal effect sizes tend to be published, these graphs provide visual representation of the overall dispersion of effects (Borenstein et al., 2009). In the forest
plot, study names go down the left column and effect size, bounded by a confidence interval (CI), are lined in the next column. A vertical line indicating no effect intersects all of the effects and corresponding CIs. In a funnel plot, studies are plotted on an axis by effect size and its standard error. Since smaller studies tend to have more standard error in effect sizes (Borenstein et al., 2009, p. 283) they often have a wide range of variation in effect size than larger studies. The result is a plot that resembles a funnel, and symmetry is assessed for indications of publication bias (Higgins & Green, 2011). Typically publication bias is manifested in a funnel plot with few studies in the bottom left area, as this represents small studies with non-significant findings; such results are usually attributed to a lack of power. This type of graph would require additional analyses to determine the robustness of the summary effect in that domain. These analyses were conducted in order to assess the potential impact of publication bias in all analyses.

Assumption of Independence

Independence between data points is an important assumption in many statistical procedures. It asserts that each observation, or data point, in an analysis is independent from each other. The consequences for violating this assumption are biased p-values and confidence intervals, and an over- or under-estimated summary effect size in a meta-analysis. In a meta-analysis, the data points are individual effect sizes from primary studies.

In many cases, multiple effect sizes were obtained from the same study. As a result, many of the domains in both the music processing and cognition databases comprised of multiple effects from the same study, which posed a significant problem related to the assumption of independence. When multiple effects from the same study are treated as independent from each other, the overall summary effect is biased and may be over-estimated. Therefore, multiple effects obtained from one study were grouped by study in each analysis (resulting in one averaged effect) rather than individual effects. Although this decreased power (by lowering the number of data points), it resulted in data that does not violate assumptions of independence. A pilot analysis was conducted with each task, individual effect sizes, at the unit of analysis to compare summary effects. Indeed, summary effect sizes calculated when each task was the unit of analysis was higher than when the study was at the unit of analysis.
Chapter 3

Results

There were a total of 5 music-processing domains and 10 cognition domains that each underwent a three part analysis; (1) basic statistics were conducted, in which a summary effect and its confidence interval was calculated, heterogeneity was assessed ($I^2$, $T^2$, Q and forest plots) and publication bias was assessed (funnel plots and File-Drawer Fail-Safe N); (2) sub-group analyses were conducted in order to assess the impact of categorical moderator variables (e.g. design, age group, type of training); (3) meta-regression analyses were conducted in order to examine potential interactions and assess the impact of continuous moderator variables (e.g. years of training, age of onset) against the summary effect. For a list of all studies used in the analysis, see Appendix B (bibliography). A breakdown of individual effect sizes for each test can be seen in Appendix C.

The standardized mean difference (Cohen’s $d$) is the metric that was used for calculating effect sizes in primary studies and the summary effects. Cohen’s $d$ can be negative or positive depending on whether the desired effect in primary studies was obtained. In this context, $d$ indicates a difference between musicians and non-musicians, with a positive value representing a better performance by musicians, a negative value representing superior performance by the non-musicians, and zero indicating no difference. Importantly, the larger the effect size, irrespective of direction, the greater the difference between musicians and non-musicians. It is worth noting that the direction of the effect is not based solely on higher or lower scores from primary studies, but rather if the scores were in the predicted direction. For example, lower scores in reaction time tasks indicate better performance whereas higher scores on other tasks represent superior performance.

Despite the large number of studies fitting the inclusion criteria, some domains and categories did not have sufficient enough information with which to carry out more advanced analyses. It has been suggested that 10 studies is the minimum size required in order to conduct a meta-analysis (Borenstein et al., 2009). Therefore, some domains were ultimately excluded from additional analyses (meta-regression and sub-group analysis).

The following section will begin with the results from the music processing analyses, followed by the cognition results. Within each section, basic statistics (including heterogeneity
and publication bias) will be reported, followed by results from sub-group analyses and the meta-regression analyses.

3.1 Music Processing Results

Basic Analyses

A total of five music-processing domains, however musical memory and timbre perception were excluded from further sub-group and meta-regression analyses due to a low number of studies. Therefore only three domains, pitch perception, melody perception and temporal perception, underwent more refined analyses. Overall summary effects and confidence intervals for each domain, including timbre perception and musical memory, are presented in Table 5. However, one should be cautioned against generalizing results of summary effects from domains in which less than 10 studies were obtained. Forest plots for each domain can be seen in Appendix D.

Tests of the null hypothesis in each domain were run to determine whether the observed summary effects significantly differ from $d=0$. Resulted indicated significance against the null in every music processing domain, $p<.001$.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Md</th>
<th>SD (T)</th>
<th>CI</th>
<th>N</th>
<th>TestNull</th>
<th>Min.d</th>
<th>Max.d</th>
<th>MusN</th>
<th>NMusN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch Perception</td>
<td>1.00</td>
<td>.316</td>
<td>95%(.84,1.2)</td>
<td>28</td>
<td>$p&lt;.001$</td>
<td>.180</td>
<td>2.77</td>
<td>822</td>
<td>803</td>
</tr>
<tr>
<td>Melody Perception</td>
<td>1.18</td>
<td>.391</td>
<td>95%(1.0,1.4)</td>
<td>33</td>
<td>$p&lt;.001$</td>
<td>.250</td>
<td>2.23</td>
<td>946</td>
<td>878</td>
</tr>
<tr>
<td>Timbre Perception</td>
<td>.632</td>
<td>.404</td>
<td>95%(.16,.1)</td>
<td>3</td>
<td>$p&lt;.001$</td>
<td>.396</td>
<td>.402</td>
<td>56</td>
<td>75</td>
</tr>
<tr>
<td>Temporal Perception</td>
<td>.834</td>
<td>.288</td>
<td>95%(.68,99)</td>
<td>28</td>
<td>$p&lt;.001$</td>
<td>.201</td>
<td>2.00</td>
<td>1350</td>
<td>1397</td>
</tr>
<tr>
<td>Musical Memory</td>
<td>1.37</td>
<td>.738</td>
<td>95%(.89,1.9)</td>
<td>8</td>
<td>$p&lt;.001$</td>
<td>.187</td>
<td>2.05</td>
<td>247</td>
<td>244</td>
</tr>
</tbody>
</table>

Note. Md is the summary effect size; SD (T), standard deviation, or tau, of the summary effect; N, number of effect sizes; TestNull, significance test of the null hypothesis; Min.d is the smallest effect size; Max.d is the largest effect size; MusN is the total number of musicians; NMusN is the total number of non-musicians.

Publication Bias

Tests of publication bias were conducted in order to assess the robustness of the summary effects in each cognitive domain, and determine the number of unpublished studies necessary to
bring a test of the null to $p > .05$. Subjective examination of funnel plots indicated the presence of some publication bias in every music-processing domain (see Appendix E), particularly in melody and pitch perception (see Figure 1). Classic fail-safe N analyses were conducted in order to determine the number of studies that would be required to bring the significance value of effect sizes to $p > .05$. This analysis showed that the following number of unpublished studies would be needed: 1985 (pitch perception); 3650 (melody perception); 2124 (temporal perception); and 280 (musical memory).

**Pitch Perception**

*Funnel Plot of Standard Error by Std diff in means*
Figure 1. Funnel plots for pitch perception and melody perception in assessing publication bias. Fewer studies in the bottom left area suggest the presence of publication bias.

Tests of Heterogeneity

Strong evidence of heterogeneity was observed in Melody Perception ($Q=83.046$, $df=32$, $p<.0001$, $I^2=61.5\%$), Pitch Perception ($Q=56.8$, $df=27$, $p<.0001$, $I^2=52.5\%$), and Temporal Perception ($Q=71.7$, $df=27$, $p<.0001$, $I^2=62.3\%$). In attempt to explain the true variance estimated by the $I^2$ statistic, sub-group and meta-regression analyses were run.

Further Analyses (Sub-group analyses and Meta-Regression)

Only three music processing domains had a sufficient sample size with which to run sub-group analyses and meta-regression: Melody Perception, Pitch Perception, and Temporal Perception. A total of three categorical moderator variables were assessed through the sub-group analysis: design (quasi-experimental vs. experimental), age group of participants (children, adults or older adults) and type of training (formal vs. informal). Three moderator variables were assessed through meta-regression: years of training (musicians and non-musicians), age of
training onset, and categories (specific outcomes measured, categories broken down from domain).

*Pitch Perception*

Significant differences in summary effect sizes were found when age-group was analyzed in the *pitch perception* domain (Q=5.72, df=2, p=.05). Post hoc analyses demonstrated that this difference laid between adults (d=1.08, sd=.324) and older adults 60+ years of age (d=.511, sd=.570), p<.05 (see Figure 2). Every study included in the *pitch perception* domain was quasi-experimental in design, and there was not sufficient information reported in order to run type of training as a moderator variable.

In a meta-regression, a significant difference was obtained for categories in *pitch perception*, Q=8.52, df=3, p=.036, (Figure 2). Moreover, sub-group analysis confirmed that every category was significant against the null (Chord discrimination: d=1.70, sd=.354, p<.001; Consonance/Dissonance: d=.993, sd=.286, p<.001; Frequency discrimination: d=.909, sd=.105, p<.001), except for perception of loudness (d=.239, sd=.439, p=.587).

No significant differences, however, were found for years of music training (see figure 3) or age of training onset (R²=12%). In total, 19 studies reported information for years of training and 7 studies for age of training onset, out of a total of 28 studies.
Figure 2. Each category in Pitch Perception was significant against the null hypothesis, except for loudness: Circles represent effect size, bold vertical line represents the regression line, surrounding vertical lines represent confidence interval.
Figure 3. Although it was hypothesized that there would be a linear relationship between years of training in musicians and summary effect sizes, the relationship was not significant in the Pitch Performance domain (seen above). Furthermore, the regression slope line is nearly vertical, suggesting that there is a more important difference between musicians and non-musicians, other than years of training. Note: Circles represent effect size, bold vertical line represents the regression line, and surrounding vertical lines represent confidence interval.

Melody Perception

Sub-group analyses according to study design (experimental $d=1.04$, $sd=.493$, quasi-experimental $d=1.19$, $sd=.493$) and age group were not significant ($R^2=0\%$). Insufficient information was provided by primary studies with which to run type of training as a moderator variable.

Multivariate meta-regression showed no differences in Melody Perception effect after allowing for the potential moderator variables of years of music training (musicians and non-
musicians), and age of training onset ($R^2=0\%$). In total, 17 studies reported years of training, and only 11 reported information for age of training onset (out of a total of 33 studies).

No significant difference was obtained when categories were entered into the analysis, $Q=0.53$, df=5, $p=.99$ (Figure 4). However, sub-group analysis was able to show that each category was significant against the null (chord perception: $d=1.19$, $p<.001$; contour perception: $d=1.06$, $p=.006$; interval discrimination (sequential): $d=1.08$, $p=.006$; melody recognition/discrimination: $d=1.23$, $p<.001$; summary: $d=1.19$, SD=.028). Note: The $R^2$ statistic is estimated by taking the difference of the total variance (model with no covariates), and the residual variance (variance not explained by the model when covariates are added). In theory, the residual variance should be smaller than the total variance, however the CMA (3rd ed) software estimates these values separately, and both are influenced by random error. When moderator variables account for very little true variance, and the residual variable is estimated to be higher than the true variance, the program estimates $R^2$ to be zero.

![Regression of Std diff in means on Category](image)

Figure 4. A visual representation of the categories in Melody Perception: each category within the domain tested significant against the null hypothesis. Furthermore, categories did not test significantly different
from each other. Note. The circles represent effect sizes from primary studies, bold vertical line represents the regression line, and the surrounding vertical lines represent confidence interval.

Temporal Perception

A near significant difference in age group between adults ($d=.852$, $sd=.288$) and children ($d=.508$, $sd=.446$) was found in temporal perception, $p=.056$, $R^2=0\%$. Although type of training was not significant ($p=.20$), it did account for some of the true variance between effects ($R^2=4\%$). Only one study design was experimental, and thus there was not sufficient information to run design as a sub-group.

A significant effect of years of music training in musicians was found in temporal perception in the 16 studies (out of 28) that reported the necessary information, $Q=3.98$, $df=1$, $p=.04$, $R^2=.41\%$ (see Figure 5). Thus years of training in musicians accounted for 41% of the true variance in the summary effect size. No significant differences were found for years of training in non-musicians ($p=.97$), age of training onset ($p=.45$), or categories ($p=.58$), $R^2=0\%$ for each moderator. Once again, however, sub-group analyses confirmed that summary effects for each category were significant against the null (Rhythm reproduction: $d=.104$, $sd=.330$, $p<.001$; Tempo perception: $d=.660$, $sd=.134$, $p<.001$; Temporal discrimination: $d=.847$, $sd=.10$, $p<.001$).
Figure 5. A linear relationship between musicians’ years of training and performance on temporal/rhythm perception tasks was observed. These results suggest an incremental relationship, such that more training results in better rhythm ability.

3.2 Cognition Results

Basic Analyses

A total of 10 cognition domains were assessed, 7 of which were categorized according to Lezak et al (2012), and 3 (Audition, General Intelligence, and Motor) that were created for the current investigation. The following three domains were excluded from further sub-group and meta-regression analyses, due to fewer than 10 studies: Construction and Motor Performance, Executive Functions, and Motor. Summary effects and their confidence intervals, along with
significant testing of the null hypothesis are shown in Table 6. Forest plots for each domain can be seen in Appendix D.

Table 6. Effect Sizes Grouped by Cognitive Domains

<table>
<thead>
<tr>
<th>Domain</th>
<th>Md</th>
<th>SD (T)</th>
<th>CI</th>
<th>N</th>
<th>TestNull</th>
<th>Min.d</th>
<th>Max.d</th>
<th>MusN</th>
<th>NMusN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audition</td>
<td>.837</td>
<td>.296</td>
<td>95% (.57, 1.1)</td>
<td>11</td>
<td>p &lt; .001</td>
<td>.271</td>
<td>1.89</td>
<td>307</td>
<td>335</td>
</tr>
<tr>
<td>Concept formation &amp; Reasoning</td>
<td>-.059</td>
<td>.391</td>
<td>95% (.33, .21)</td>
<td>11</td>
<td>p = .84</td>
<td>-.742</td>
<td>.535</td>
<td>914</td>
<td>879</td>
</tr>
<tr>
<td>Construction &amp; Motor Perform.</td>
<td>.229</td>
<td>.325</td>
<td>95% (-.05, .51)</td>
<td>8</td>
<td>p = .11</td>
<td>.133</td>
<td>1.00</td>
<td>354</td>
<td>295</td>
</tr>
<tr>
<td>Executive Functions</td>
<td>.100</td>
<td>.290</td>
<td>95% (-.21, .41)</td>
<td>5</td>
<td>p = .46</td>
<td>-.368</td>
<td>.837</td>
<td>293</td>
<td>258</td>
</tr>
<tr>
<td>General Intelligence</td>
<td>.348</td>
<td>.185</td>
<td>95% (.23, 4.7)</td>
<td>28</td>
<td>p &lt; .001</td>
<td>-.329</td>
<td>1.33</td>
<td>1181</td>
<td>1119</td>
</tr>
<tr>
<td>Memory</td>
<td>.421</td>
<td>.342</td>
<td>95% (.24, .61)</td>
<td>19</td>
<td>p &lt; .001</td>
<td>-.581</td>
<td>1.26</td>
<td>1827</td>
<td>1250</td>
</tr>
<tr>
<td>Motor</td>
<td>1.08</td>
<td>.041</td>
<td>95% (.83, 1.3)</td>
<td>4</td>
<td>p &lt; .001</td>
<td>.702</td>
<td>2.10</td>
<td>186</td>
<td>154</td>
</tr>
<tr>
<td>Orientation and Attention</td>
<td>.283</td>
<td>.344</td>
<td>95% (.15, .42)</td>
<td>35</td>
<td>p &lt; .001</td>
<td>-.719</td>
<td>1.51</td>
<td>2754</td>
<td>2708</td>
</tr>
<tr>
<td>Perception</td>
<td>.190</td>
<td>.163</td>
<td>95% (.02, .36)</td>
<td>11</td>
<td>p = .025</td>
<td>-.726</td>
<td>.665</td>
<td>524</td>
<td>511</td>
</tr>
<tr>
<td>Verbal Functions &amp; Lang. Skills</td>
<td>.423</td>
<td>.422</td>
<td>95% (.20, .65)</td>
<td>18</td>
<td>p &lt; .001</td>
<td>-.227</td>
<td>1.52</td>
<td>756</td>
<td>822</td>
</tr>
</tbody>
</table>

Note. Md is the summary effect size; SD (T), standard deviation, or tau, of the summary effect; N, number of effect sizes; TestNull, significance test of the null hypothesis; Min.d is the smallest effect size; Max.d is the largest effect size; MusN is the total number of musicians; NMusN is the total number of non-musicians.

Publication Biases

Tests of publication bias were conducted in order to assess the robustness of the summary effects in each cognitive domain, and determine the number of unpublished studies necessary to bring a test of the null to p > .05. This analysis was unnecessary in three cognitive domains, in which non-significance results were already obtained (concept formation and reasoning, construction and motor performance, and executive functions). Publication bias appeared to be evident in most of the cognitive domains (see Appendix E), except for general intelligence (Figure 6). A funnel plot is designed so that the size of the study is on the y-axis, and effect sizes from primary studies are on the x-axis. The funnel plot for memory (Figure 7) is abnormal in that larger studies should be closer to the average, however the larger studies are distributed the way
that smaller studies (which were not retrieved, according to this graph) normally would be; this indicates an issue with heterogeneity.

Figure 6. *General Intelligence* funnel plot; that studies from smaller studies are located on both sides of the average indicates a lack of publication bias in this domain.
Figure 7. Funnel plot for Memory. Publication bias typically manifest itself in a funnel plot with fewer studies on the bottom left, indicating a bias in not published smaller studies (fewer participants) finding no effect. In this case, however, there is also a lack of publications on the bottom right, which suggests that there were no small studies in this analysis. That larger studies are widely distributed around the average suggests strong heterogeneity.

Classic Fail-Safe N analyses were conducted in order to determine the number of studies that would be required to bring the significance value of effect sizes to $p > .05$. The number of unpublished studies needed for each domain to bring the summary effect to zero is presented in Table 7.

Table 7. Publication bias: Fail-Safe N Test

<table>
<thead>
<tr>
<th>Domain</th>
<th>Number of studies observed in analysis</th>
<th>Number of studies to achieve $p &gt; .05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor</td>
<td>4</td>
<td>75</td>
</tr>
<tr>
<td>General Intelligence</td>
<td>28</td>
<td>385</td>
</tr>
</tbody>
</table>
Memory 19 368
Orientation and Attention 35 615
Perception 11 28
Verbal Functions and Language Skills 18 357
Audition 11 173

Tests of Heterogeneity

Strong evidence of heterogeneity was observed in Concept Formation and Reasoning (Q=57.06, df=10, p<.001, I²=82.48%), memory (Q=76.31, df=18, p<.001, I²=76.41%), Orientation and Attention (Q=178.39, df=34, p<.001, I²=80.94%), and Audition (Q=32.01, df=11, p<.001, I²=65.64%). Moderate evidence of heterogeneity was observed in General Intelligence (Q=43.37, df=26, p=.02, I²=40.01%). Finally, no heterogeneity was observed in Perception, Q=15.24, df=10, p=.12, I²=34.36. In attempt to explain the true variance estimated by the I² statistic, sub-group and meta-regression analyses were run.

Further Analyses (Sub-group analyses and Meta-Regression)

Seven cognitive domains underwent sub-group analyses and meta-regression: Concept Formations and Reasoning, General Intelligence, Memory, Orientation and Attention, Perception, Verbal Functions and Language Skills, and Audition. A total of three categorical moderator variables were assessed through the sub-group analysis: design (quasi-experimental vs. experimental), age group of participants (children, adults or older adults) and type of training (formal vs. informal). Three moderator variables were assessed through meta-regression: years of training (musicians and non-musicians), age of training onset, and categories (specific outcomes measured, categories broken down from domain). For a list of tasks in each domain, and the breakdown of their categories, see Appendix C.

Audition

Sub-group analysis showed a significant difference in type of training received by musicians, such that formal training (d=.983, sd=.366, n=8) had the highest effect, followed by informal (d=.949, sd=.672, 1) and a combination of both formal and informal (d=.295, sd=.469,
n=2). No significant difference against audition was found for study design ($p=.301$) or age group ($p=.365$).

Results from a meta-regression on audition showed that there was a significant effect of years of music training in musicians ($Q=7.86, \text{df}=1, p=.005, R^2=6\%$), (see figure 8). A significant effect of years of training in non-musicians was also found, and accounted for a greater amount of true variance in the audition summary effect ($Q=8.77, \text{df}=1, p=.003, R^2=72\%$). The significant effect of musicians’ years of training is pictorially shown in Figure 8. There was no significant effect of age of training onset ($p=.24$) and it accounted for only a small amount of the true variance ($R^2=5\%$).

Figure 8. A significant effect of years of training in the musician group on the overall summary effect of Audition
No significant difference in effect sizes were observed for study design ($p=.08$), although the summary effect for quasi-experimental studies ($d=-.16$, $sd=.40$, $n=8$) was in the opposite direction as in experimental studies ($d=.24$, $sd=.41$, $n=3$). Similarly, no significant differences were found for age group ($p=.554$), and insufficient information was provided by primary studies with which to analyze the effect of type of training.

Multivariate meta-regression showed no significant difference in *concept formation and reasoning* after allowing for the potential moderator variables of years of music training (musicians, $p=.57$, and non-musicians, $p=.67$), and categories ($p=.47$); none of these variables accounted for any estimated true variance ($R^2=0\%$). In total, 9 studies reported years of training in musicians, and 11 reported years of training in non-musicians (out of a total of 11 studies). Not enough studies reported information for age of training onset, thus this moderator was not analyzed.

No significant differences were obtained when categories were entered into the analysis, $Q=1.5$, $df=2$, $p=.47$. Moreover, sub-group analysis showed that none of the categories were significant against the null (concept formation: $d=-.24$, $p=.54$; mathematical procedures: $d=.23$, $p=.06$; reasoning: $d=.30$, $p=.36$), substantiating a non-significant overall summary effect.

### General Intelligence

No significant differences were obtained when either study design ($p=.15$), age-group ($p=.82$), or type of training ($p=.533$) were run as sub-groups. However, effect size for informal training ($d=+.514$, $sd=.456$) was slightly higher than that of formal training ($d=.328$, $sd=.263$).

No significant difference in *general intelligence* was found after allowing for the variables of years of training (musicians, $p=.83$, and non-musicians, $p=.52$), age of training onset ($p=.77$), or categories ($p=.94$); further, none of these variables were able to account for an estimate portion in true variance ($R^2=0\%$). In total, 20 studies reported years of musicians’ training, 27 reported years of non-musicians’ training, and 9 reported age of training onset (out of total of 28).

Sub-group analysis of the categories within *General Intelligence* showed that performance IQ ($d=.574$, $sd=.621$, $p=.14$) and full-scale IQ ($d=.469$, $sd=.508$, $p=.071$) were not significant against the null. The other two categories were significant against the null, (verbal IQ: $d=.441$, $sd=.41$, $p=.01$; non-verbal: $d=.386$, $sd=.266$, $p<.001$ (Figure 8).
Memory (Lezak et al., 2012)

No significant differences were obtained when either study design ($p=.14$) or age group ($p=.57$) were run as sub-groups. Although type of training yielded a significant result ($Q=19.99$, df=3, $p<.001$), with informal ($d=.509$, sd=.428) having a higher summary effect than formal training ($d=.360$, sd=.332), there were only 2 effects in the informal group compared to 15 in the formal group.

A meta-regression demonstrated a significant effect of years of training in the musician groups in memory, $Q=4.35$, df=1, $p=.037$, $R^2=13\%$, n=18, but this was in the opposite direction than predicted (see Figure 9). No significant difference, however, was found for the variables of years of training in the non-musician groups ($p=.42$), age of training onset ($p=.73$), or categories ($p=.77$). Moreover, none of these variables were able to account for an estimate portion in true variance ($R^2=0\%$). In total, 17 reported years of non-musicians’ training, and 6 reported age of training onset (out of total of 20). Analysis of the categories showed that while verbal memory tested significant against the null ($d=.386$, sd=.134, $p=.005$), visual memory did not ($d=.322$, sd=.207, $p=.12$).
Figure 9. Significant effect of years of training in the musician group for the Memory domain, however the slope is not in the expected direction. The down-sloping regression line indicates that increased years of training in musicians resulted in worse performance in the memory domain.

Orientation and Attention (Lezak et al., 2012)

A significant effect of age group was found in Orientation and Attention, $p<.001$. Post-hoc analyses showed that older adults had the highest effect ($d=.458$, $sd=.321$), followed by adults ($d=.331$, $sd=.303$) and children ($d=.123$, $sd=.397$).

The study design (quasi-experimental or experimental) was not significant, $p=.605$.

Finally, the majority of the effects came from studies with formal training, thus type of training was not run as a moderator.

Meta-regression showed no significant difference in orientation and attention after allowing for the potential moderator variables of years of music training (musicians, $p=.48$, and non-musicians, $p=.30$), age of training onset ($p=.65$), and categories ($p=.14$); only categories
accounted for some of the true variance ($R^2=3\%$). In total, 27 studies reported years of training in musicians, 32 reported years of training in non-musicians, and 11 reported age of training onset (out of a total of 35 studies).

As previously stated, the overall summary effect was significantly different from the null hypothesis, $p<.001$. Sub-group analyses, however, showed that the only category within this domain to significantly differ from the null was processing speed, $d=.850$, $sd=.383$, $p<.001$. The rest of the categories were not significant (attentional capacity: $d=.301$, $sd=.439$, $p=.12$; concentration: $d=.203$, $sd=.752$, $p=.72$; divided attention: $d=.007$, $sd=.513$, $p=.98$; space: $d=.068$, $sd=.491$, $p=.78$).

**Perception (Lezak et al., 2012)**

A sub-group analysis for the Perception domain yielded non-significant results for study design ($p=.99$), age-group ($p=.62$), and type of training ($p=.62$).

Meta-regression showed no significant differences in the perception domain after allowing for years of music training in musicians ($p=.45$) and non-musicians ($p=.74$), or categories ($p=.28$). Insufficient information was reported with which to analyze the age of training onset in musicians ($n=3$). In total, 10 studies reported years of music training in musicians and non-musicians out of a possible 11 studies.

Overall, the summary effect for Perception ($d=.19$, $sd=.16$) tested significantly different against the null, $p=.025$. However, a sub-group analysis of categories showed that only visual organization was significantly different from the null hypothesis, $d=.479$, $sd=.472$, $p=.03$. Neither visual inattention ($d=.062$, $sd=.442$, $p=.75$) or visual recognition ($d=.103$, $sd=.366$, $p=.29$) were significant.

**Verbal Functions and Language Skills (Lezak et al., 2012)**

Sub-group analysis showed that there was no significant effect of study design ($p=.76$) or type of training ($p=.41$) on the summary effect. Sub-group null hypothesis analysis of age groups showed that adults ($d=.282$, $sd=.205$) did not significantly differ from the null, $p=.168$. Effect sizes for children ($d=.503$, $sd=.449$) and older adults ($d=.545$, $sd=.359$), however, did significantly differ from the null.

Meta-regression showed no significant difference in Verbal Functions and Language Skills after allowing for the potential moderator variables of years of music training (musicians,
Years of training in musicians and non-musicians accounted for 4% and 7%, respectively, of the true variance. Age of onset did not account for any of the true variance, but categories accounted for 9%. In total, 12 studies reported years of training in musicians, 18 reported years of training in non-musicians, and 5 reported age of training onset (out of a total of 18 studies).

The overall summary effect for verbal functions and language skills ($d=.423$, sd=.42) tested significantly different against the null hypothesis. However, sub-group meta-regression analysis showed that verbal comprehension was not significantly different from the null, $d=-.095$, sd=.373, $p=.50$. The other three categories were significant against the null: verbal academic skills: $d=.498$, sd=.401, $p=.002$; verbal expression: $d=.472$, sd=.472, $p=.034$; verbal memory: $d=.944$, sd=.346, $p<.001$.

Chapter 4
Discussion

A meta-analysis was conducted to delineate any differences in music processing and cognitive ability between musicians and non-musicians, and the magnitude of these differences. The impact of potential moderating variables on summary effects were carried out through further analyses, specifically that of years of music training, age of training onset, study design, and type of training. It was hypothesized that music training would have a large effect in tasks directly related to musical skills, such as the perception and discrimination of pitch, temporal events (rhythm), and timbre. Further, it was predicted that music training would have the greatest effect on cognitive tasks similar to those used in music, such as motor and audition. Conversely, music training was predicted to have a small, but reliable, effect on other non-music related tasks (ex. memory, and executive functions). Finally, regarding moderating variables, it was predicted that music expertise (years of training) would be incrementally related to summary effect size, and that formal training would have a significant influence.

4.1 Music Processing (Near-transfer effects)

As predicted, musicians significantly outperformed non-musicians on every music-processing domain. Summary effects over $d=1$ were obtained for Melody Perception ($d=1.18$), Musical Memory ($d=1.37$) and Pitch Perception ($d=1.00$). Practically, this means that over 80%
of musicians were over the mean of the control group in each of these domains (Magnusson, 2014). Temporal Perception ($d=.834$), while showing a relatively smaller effect that the first three music domains, was still a large effect according to Cohen (1977). A smaller effect may have been seen in Temporal Perception over the domains related to pitch because rhythm is more familiar to non-musicians, due to its ubiquity across modalities. Timbre Perception resulted in the smallest summary effect ($d=.632$), however, due to the small sample size of only three studies, it is difficult to know whether this effect would have increased or decreased with the presence of more data. Further, the summary effect of the three studies that comprised Timbre Perception resulted in a large standard deviation (see Appendix B). Importantly, sub-group analysis showed that every category within each music domain was also significant against the null. These results demonstrate that even though one could argue that many diverse music abilities were grouped together into major domains, musicians still outperformed non-musicians within each refined category (see Appendix B for a complete list of these categories).

A number of moderating variables were predicted to have an impact on findings of difference between musicians and non-musician. Two music domains (Timbre Perception and Musical Memory) were excluded from analyses involving moderator variables, due to small sample sizes. Years of music training, in particular, was expected to account for true variance in summary effect sizes, such that more years of training would result in better music processing performance. However, the predicted linear relationship was only found in the Temporal Perception summary effect, where it accounted for 41% of true variance. This suggests an incremental relationship, such that the more training in music one receives, the stronger one’s rhythm perception abilities. Since this domain included listening and reproduction abilities, this domain is inclusive of most rhythmic processing abilities. These results are consistent with findings from the literature outside the scope of studies used in the current study. In a study employing DJ training (manipulating existing musical tracks in a temporally-appropriate manner), non-musicians demonstrated significantly improved rhythm processing ability after only one week (Butler & Trainor, 2015).

Interestingly, this predicted linear relationship was not observed in the music domains related to frequency perception, specifically, Pitch and Melody Perception. The Pitch Perception domain included musician samples with hardly any training, to those with over 20 years of training, however the regression line is almost completely vertical across years of training. These
results suggest that there is something else, other than years of training, which distinguishes musicians from non-musicians on pitch and melody perception. Within the literature, there are suggestions that musicians are individuals who may be born with superior auditory perception skills, thus creating an internal motivation to pursue music lessons and stick with it (Ladinig et al., 2009; Schellenberg, 2006). It would appear that results from the frequency-related domains support a theory of a pre-disposition, however results from the temporal domain do not. This suggests that perhaps music processing abilities are not simply described by either nature (pre-disposition) or nurture (music training). Instead, it is possible that some perceptual abilities may be better described by a more innate ability, such as those involving pitch perception, and others may be more influenced by experience, such as rhythm perception. In other words, music perception itself is likely a more complex interaction of cognitive and perceptual abilities that is influenced by music and non-music domains, in addition to pre-disposed ability.

Age of training onset was another variable expected to have an influence on the summary effects. While this variable accounted for 12% of the true variance in Pitch Perception, it did not account for any variance in either Melody or Temporal Perception. This was an unexpected finding, as the age at which musicians begin training has been shown in the literature to result in findings of difference between groups (Hanna-Pladdy & Gajewski, 2012; Watanabe et al., 2007). Furthermore, that this variable accounted for variance in Pitch Perception but not in the Melody Perception domain may suggest that the proposed sensitive period for music acquisition is limited to the ability to discriminate between isolated pitches.

Overall, the moderating variables used in the music processing analysis were not able to explain most of the true variance in summary effects. This means that although summary effects and significance tests indicate that musicians reliably outperform non-musicians on all music processing tasks, it remains unclear what specific variables contribute to these differences. It is likely that a large portion of the true variance between effects in each of these domains is directly related to differences in methodology, specifically that of procedures, musical stimuli and other extenuating variables due to lack of standardization. Tasks in the music processing studies were often created specifically to measure a musical ability, thus resulting in a variety of procedures between-studies measuring the same outcome. Effect sizes were grouped in this analysis based on the musical ability being measured, which ultimately formed into categories and then domains. These differences were not quantified or categorized as covariates by the current investigation;
therefore it is unknown whether a single moderator that was not used in the current analysis could have explained more of the true variance.

Since significant summary effects were obtained, it was prudent that Fail-Safe N tests were run to evaluate publication bias. It is likely that there are studies that found no difference between musicians and non-musicians and were either not submitted for publishing, or were rejected. Determining the number of these studies to make the observed effect non-significant can be compared to the number of observed studies to subjectively determine whether the effect is robust. Results from Fail-Safe N tests showed that substantially more studies (than retrieved for the current investigation) finding no effect would be needed to bring the significance level above alpha in each domain. For example, 33 studies were used in the Melody Perception domain, and 3650 studies with no effect would be needed to bring the test of significance to $p > .05$.

Therefore, although there is evidence of publication bias from the funnel plots, we can be confident that there are likely not enough in existence to results in non-significance.

### 4.2 Cognition (Far-transfer effects)

Findings across the cognitive domains were not as consistent as in the music processing domains. For a summary of findings as they relate to the research questions, please refer to table 8. As predicted, the domains most related to musical ability, audition and motor, had the highest summary effects. Although the motor domain resulted in the largest effect, it was based on only four studies. That music training often involves demanding fine motor ability suggests that this effect would be found in additional studies, however it is also likely that this effect is an over-estimation of the difference between groups.

<p>| Table 8. Summary of findings from cognition analysis – Domains by Research Questions |
|---------------------------------|---------------------------------|----------------|----------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>Does performance of M and NM significantly differ?</th>
<th>What is the magnitude of observed differences?</th>
<th>Is there linearity b/w training and performance in musicians?</th>
<th>What is the effect of type of training on performance?</th>
<th>What is the effect of age of onset?</th>
<th>Is there a difference b/w effects grouped by design?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audition</td>
<td>Yes</td>
<td>$d=.84$, large</td>
<td>Yes</td>
<td>Not significant</td>
<td>Not significant</td>
<td>No</td>
</tr>
<tr>
<td>Concept</td>
<td>No</td>
<td>$d=-.06$, no</td>
<td>No</td>
<td>Not enough</td>
<td>Not enough</td>
<td>No</td>
</tr>
</tbody>
</table>
Three cognitive domains were not significant against the null hypothesis, and resulted in trivial summary effects: Concept Formation and Reasoning, Construction and Motor Performance, and Executive Functions. These domains encompass such abilities as mental flexibility and reasoning, visuoperceptual ability, and planning, decision-making and performance execution, respectively (Lezak et al., 2012). Interestingly, the negative summary found in Concept Formation and Reasoning indicates that non-musicians were superior to non-musicians (not significant), but that this effect was almost zero. A closer examination of categories within the Concept Formation and Reasoning domain resulted in non-significance for each of the three categories (concept formation, mathematical procedures and reasoning). However, mathematical procedures was near significance with a larger effect size than the other two categories (d=.23, p=.06); too few studies reported information within this category (4) to definitively determine whether musicians and non-musicians actually differ on this ability.
The Construction and Motor Performance and Executive Functions domains were excluded from more advanced analyses, due fewer than 10 studies in each domain. It is unclear whether a larger sample of data would have changes the summary effect in either direction. The Motor domain differed from the Construction and Motor Performance domain in that the former was created with tasks each measuring motor ability, but differing in procedures, and the latter is comprised of standardized assessments. Although the Construction and Motor Performance domain was not significant against the null, the summary effect did lay between that of a small and medium effect (Cohen, 1977), based on only 8 studies. It is usually recommended that meta-analyses have at least 10 studies to calculate a summary effect (Borenstein et al., 2009). As a result, one cannot conclude that musicians and non-musicians do not differ on these cognitive domains, as there is simply not sufficient enough information. Pragmatically this means that more research is needed on these three domains in order to further understand the influence of music training on these cognitive domains.

The remaining five cognitive domains (Perception, Orientation and Attention, General Intelligence, Memory, Verbal Functions and Language Skills) tested significant against the null hypothesis, with summary effects resulting in small to medium strengths (Cohen, 1977). This means that musicians outperformed non-musicians on each of these domains, encompassing skills such as attention, spatial ability, linguistic processing, and working memory (Lezak et al., 2012). Overall, musicians outperformed non-musicians the most in the Verbal Functions and Language Skills ($d=.422$) and Memory ($d=.421$) domains, resulting in small-medium effects. Practically this means that approximately 14 individuals out of 100 that take music training will demonstrate superior performance to those who do not take music. A small-medium effect in General Intelligence ($d=.35$) has similar practical significance, followed by findings in Orientation and Attention ($d=.28$), meaning that 10 individuals out of 100 who take music training will improve ability in this domain. The Perception ($d=.19$) domain yielded the smallest significant summary effect, and is considered to be marginal.

Sub-group analyses within domains were run in order to ensure that each category tested significant against the null hypothesis. Interestingly, while some domains were significant overall, not all of the categories within them were. For example, in the Memory domain, visual memory was not significantly different from zero. Similarly, two categories within the Perception domain (visual recognition and visual inattention) were also not significant and both
fall under the visual modality. Some researchers have suggested that memory advantages for musicians may not extend to the visual modality (Cohen et al., 2007; Oura & Hatano, 1988), perhaps explaining the non-significant results obtained here.

The robustness of each domain’s summary effect was examined through tests of publication bias. The only cognitive domain to raise concerns of bias was Perception, as the number of observed studies was 11, and the analysis indicated that only 28 studies would be needed to bring the effect below alpha. The Perception domain included tasks mostly related to visual perception, and previous research has suggested that there may be an implicit link between pitch and visual processing, due to the importance of spatial representations within both modalities (Lidji et al., 2007; Platel et al., 1997; Rusconi, et al., 2006; Weiss, et al., 2014). Therefore, although there is evidence of publication bias in this domain, it is unlikely that the addition of unpublished findings on the topic would significantly change the observed effect.

Given the robust summary effects in many of the cognitive domains, it was expected that the moderator variables included in the analysis would explain the true variance estimated between musicians and non-musicians. Interestingly, years of music training in musicians was not significant in most of the cognitive domains, despite a range in the way musicians were categorized in primary studies. This might support a theory that musicians and non-musicians differ on an unknown variable that is more important than years of training, such as performance ability or type of instrument. The only significant effect of musicians’ years of training in the predicted direction appeared in the Audition domain. The effect of years of training in musicians was significant in the Memory domain, however the slope of this relationship is not in the expected direction. One possible explanation for this unexpected finding may have to do with the interaction of active and passive training with years of music training. There is some suggestion within the literature that far-transfer ability may require active participation in music (Hanna-Pladdy & MacKay, 2011), meaning that some transfer skills elicited by music training may begin to deteriorate once participation in music has ceased. It is possible that perhaps the Memory domain is one such cognitive area that requires active participation in music in order to retain advantages over non-musicians. There was also a significant effect in the Audition domain, such that years of music training accounted for an estimated 6% of the true variance between effect sizes. A large portion of this domain was comprised of speech-in-noise tests, in which participants are required to repeat sentences spoken by single speakers within a noisy
environment of increasing decibels. The ability to hear speech-in-noise is believed to arise through advantages in the central auditory system, and rely on working memory ability (Besser et al., 2013; Parbery-Clark et al., 2009, 2011). Therefore, it makes sense that differences in years of training in musicians would be significant for both the Memory and Audition domains. Further investigation into the interaction of other variables is needed in order to determine the influence of music training on abilities related to memory.

Age of training onset is often discussed as another potential covariate within musician samples, and has been used to describe differences in brain structures (Bailey, Zatorre & Penhune, 2014; Steele et al., 2013) and functional ability (Hanna-Pladdy & Gajewski, 2012; Watanabe et al., 2007). Despite the relative importance of it in the literature, too few studies reported this information for musician groups. As a result, non-significant results were found for this moderating variable across cognitive domains, likely due to the lack of data reported by primary studies. Similarly, type of training received by musician groups was not examined in every cognitive domain due to lack of information reported in primary studies. The only domain in which type of training significantly differed was Memory, where informal training \((d = .509, \text{sd} = .428)\) had a higher summary effect than formal training \((d = .360, \text{sd} = .332)\). These results might suggest that informal training creates a more beneficial environment to cause far-transfer effects to cognitive domains. However, more data might have provided a different picture, or there could be other variables other than type of training that could better explain these differences, (ex. informal lessons often occur in group situations, whereas private lessons are often one-on-one).

Primary studies in each domain included individuals from across the life span, therefore it was no surprise that age groups significantly differed in some of the domains. In Orientation and Attention, older adults (60+ years of age) had a higher summary effect over adults (18-60 years). However, in Verbal Functions and Language Skills, children and older adults had higher effects than adults. One potential explanation for these differences is that children and older adults are often the targeted groups for experimental designs, where they are actively engaged in music training at the time of the investigations. Although some quasi-experimental studies specify that musicians are active in training, most do not. Therefore, these differences in age group could be due to age, or simply the result of being actively engaged in music training at the time of the experiment.
It was predicted that there would be a difference between summary effect sizes based on the study design, specifically between quasi-experimental and experimental studies. Correlational studies are often criticized for the lack of random-assignment, leading some researchers to suggest that groups differ on more than musical experience, such as IQ (Schellenberg, 2011) or personality (Corrigall, Schellenberg & Misura, 2013). However, sub-group analysis did not find significant differences between the effects gleaned from different experimental designs. That there were no significant differences between experimental and quasi-experimental studies’ summary effects suggests that one can use music training as a variable to explain differences between musicians and non-musicians.

Interestingly, the summary effect in the Concept Formation and Reasoning domain for experimental studies \((d=.24, sd=.41, n=3)\) was greater than in the quasi-experimental studies \((d=-.16, sd=.40, n=8)\); this difference, however, was not significant. This trend was also found in the Construction and Motor Performance domain, the only domain in which there was an equal number of studies from each design \((n=4)\), and again, the summary effect was higher in the experimental studies \((d=.48, sd=.19)\) than in the quasi-experimental \((d=-.03, sd=.12), ns.\) Again, these findings are not consistent with previous research examining summary effects between design type (Butzlaff, 2000). It is unclear whether these results would have been different with more experimental studies in these domains.

**Implications**

Although it may seem obvious that musicians have superior music processing skills, the current analysis was able to provide novel insight into the variables that differentiate musicians from non-musicians. Specifically, years of training is considered to be a defining feature of a musician; however, this variable did not account for any of the variance between musicians and non-musicians in either pitch or melody perception. That this variable did account for differences in the temporal perception domain suggests that perhaps rhythm processing is more influenced by musical experience. Since rhythm is not limited to musical domains, rather is present in our language and physical movements, it is possible that music training may help individuals with processing issues involving rhythm. Furthermore, results from the music processing domain may help explain why music training appears to elicit cognitive change in some domains (ex. verbal functions and language skills) and not others (ex. construction and motor processing).
Delineating the music processing abilities that are more influenced by years of music training has many implications for fields such as music therapy. In music therapy, techniques similar to those used in music training are employed in order to improve wellbeing or rehabilitate abilities outside the realm of music. Older adults in particular may benefit from research examining transfer effects from music training. For example, performing in an ensemble requires the ability to isolate various auditory streams from a dense body of sound. This is very similar to speech-in-noise perception, an ability that often deteriorates with age (Parbery-Clark et al., 2009). Thus older adults find it more difficult to hear speech in loud environments, which may lead to avoidance in the form of social isolation; the reclusion of older adults from social activities is believed to be a predictor of cognitive decline, specifically dementia (Cacioppo & Hawkley, 2009). Therefore there is pragmatism in delineating cognitive and perceptual abilities that can be enhanced through music training.

That music training may improve non-music skills in other domains also has implications for fields such as music education and neuroscience. Research regarding the effects of music training on non-music domains has a great influence on music education policy. Across the continent, music educators are asked to justify the inclusion of music in the public school curriculum. Although music education has shown to have a significant impact on social and emotional development (Hargreaves et al., 2003), policy is more often influenced by findings of transfer from music to non-music domains. This line of research might also contribute to neuroscientific research, a field that is currently finding anatomical and structural differences between musicians and non-musicians (Schlaug, Janacke, Huang, et al., 1995; Zatorre, Fields, Johansen-Berg, 2012).

Limitations

Methodological limitations

One of most important limitations of the current investigation was the lack of moderator information reported by primary studies. Inconsistent reporting of participant demographic information led to the exclusion of moderator variables from analyses in some cognitive and music processing domains. This resulted in a sub-set of data with few studies in each domain reporting necessary information, and caused three potential limitations: (1) summary effect sizes of categorical moderator variables in sub-group analyses may be over- or under-estimated, (2)
there may have been a lack of statistical power to find significant differences between sub-
groups, and (3) summary effects may be inherently biased, as it is unclear why some authors
chose to report information such as years of music training or age of training onset and some did
not.

Meta-analyses have often been criticized for combining studies that are very different
(Bailar, 1995; Feinstein, 1995; Meinert, 1989). This was avoided by having careful inclusion
criteria, such as only including studies with healthy adult samples (not clinical populations) and
excluding case studies. Arguments have also arisen regarding the emphasis of quantity over
quality, such that one “bad” study may weaken the overall results; therefore, only studies
published in a peer-reviewed journal were included. Familiarity bias also leads to issues in meta-
analyses, when only articles familiar to the researcher are used. This was countered by searching
for articles in non-music cognition journals, such as large medical and psychological databases.

Effect sizes in the cognitive domains were calculated from standardized assessments
between individuals with and without music training. That standardized tasks were used implies
that there was a common procedure utilizing the same stimuli across studies. However, due to the
diversity in the actual tasks employed in primary studies, there was not enough data to calculate
summary effects for each standardized task. Therefore tasks were organized into smaller
categories that formed larger domains according to Lezak et al (2012). This is an important
limitation, as one could argue that some tasks may not have belonged in certain domains, or
could have been placed in other domains. In these situations, tasks were placed in the domain that
more appropriately matched the outcome being measured in primary studies. For example,
Raven’s Progressive Matrices was placed in the General Intelligence domain, however one could
argue that this test could be more appropriately placed in the Concept Formation and Reasoning
domain. Primary studies utilized this test as a measure of non-verbal intelligence; therefore, since
the outcome being measured was intelligence, it was placed in that domain. If the tasks were not
categorized, or grouped appropriately, one would expect there to have been a significant
difference between effect sizes between categories within a domain. However, sub-group analysis
did not show any significant differences between categories in any of the cognitive domains.
Therefore although the way tasks were organized may have been a limitation, it likely would not
account for a substantial amount in variance within summary effects.
Unlike the cognitive tasks used in the current meta-analysis, the music perception tasks were not standardized. Even the stimuli used in the music processing studies varied, particularly due to changes in technology over the past five decades. The only way that music processing studies could be combined is by the specific outcome being measured in each study. Effect sizes were calculated between musicians and non-musicians, and tasks were labeled based on their outcomes. When a task could arguably fit in multiple domains, the outcome specifically attached to the hypothesis was examined. Another challenge to categorizing these tasks is that it is often difficult to separate the perception of one musical characteristic from another. For example, melody perception involves the simultaneous perception of pitch, rhythm, timbre, loudness, etc. Accordingly, when participants were asked to discriminate between sequences containing these elements, the task was placed into “melody discrimination” rather than “pitch/frequency discrimination”. Therefore, although the domains used were based on Tirovolas & Levitin (2011), there may have been a more precise way to organize the music processing tasks.

Another criticism of meta-analyses is that generalizations made from results may not be appropriate; for example, the limitations attached to the initial studies are sometimes forgotten through the meta-analysis process (Bailar, 1995; Feinstein, 1995). In the current study, sub-group analyses were run in order to examine whether there were differences in summary effect sizes between correlational/cross-sectional studies and experimental/matched pairs studies. The implications of each of these study designs are very different. Since quasi-experimental studies compare individuals who are already musicians to non-musicians, one cannot conclude that it is only music training that has an effect, since there are likely other differences between these groups. However, experimental designs with true-random assignment isolates the effect of music training, such that its effects can be generalized. Although most of the studies used in the currently investigation were quasi-experimental in design, there were no significant differences between designs in either the music processing or cognition domains. This suggests that perhaps there is something about the music-training variable that has an influence on differences in cognition or music processing. However, the current study was not successful in isolating what aspect of the training variable is most important due to lack of information reported in many primary studies. Years of music training are typically viewed as the most important aspect of music training, but this moderator variable was not significantly related to the summary effect in
all of the domains. These results suggest that perhaps the differences in training between musicians are more complex than previously believed.

One frequently acknowledged issue with meta-analyses is the inability to collect all possible articles on a topic. To ensure that as many studies as possible were included in the current meta-analysis, a wide range of journals were thoroughly searched in large, multi-disciplinary databases. Moreover, to ensure that publications were not missed, peer-reviewed music cognition journals were hand-searched.

Music training and ability: causal or correlational?

The majority of studies that examine the effect of music training on music processing and perceptual abilities compare the performance of musicians to non-musicians. This raises problems regarding correlation and causation, as one cannot conclude that music training is the sole difference between these groups. It is possible that differences in cognitive ability found between musicians and non-musicians may be the result of pre-existing differences, such as IQ (Babo, 2004), academic achievement (Schellenberg & Weiss, 2013), or personality (Corrigall, Schellenberg & Misura, 2013). It is possible that individuals who possess a natural curiosity or enjoyment of learning are more internally motivated to pursue education, such as music training. Moreover, one cannot conclude that music training is the reason why musicians demonstrate superior pitch or timbre discrimination abilities, or any other enhanced music perception skills over non-musicians. It is equally plausible that some individuals are born with stronger auditory processing abilities, and as a result, naturally gravitate towards, and stick with, music education (Trainor & Hannon, 2012, p.454). Despite criticism of research comparing musicians to non-musicians, evidence does exist to suggest that music training has an effect on cognitive transfer (Hyde et al., 2009; Pantev et al., 1998).

Future directions

The current study attempted to delineate the cognitive tasks in which musicians differ from non-musicians, and examine the impact of various moderator variables on summary effects. Although some of the moderators could be examined, very rarely did every primary study report the necessary information with which to compare effect sizes. For example, although musical experience (years of training) is considered to be an important characteristic of musician samples, not every study reported this information. Further, although age of training onset is believed to be
an important aspect of training (Hanna-Pladdy & Gajewski, 2012) only around half of the studies included in the current meta-analysis reported this information.

There are other moderator variables other than the ones used in the current analysis that could not be examined due to a lack of information provided by primary studies. Initially, whether musicians were instrumentalists or vocalists was recorded from studies, however far too few studies compared vocalists to non-musicians and instrumentalists. It is possible that there are differences between vocalists and instrumentalists on some music processing or cognitive domains, due to differences in training and practice techniques. Furthermore, more specific information regarding type of music training may have accounted for more of the true variance within domains. For example, there are many differences between classically-trained musicians and those trained in jazz, such as the emphasis on score reading in classical music, and of “playing by ear” in jazz music. Expert musicians from these two genres may differ on measures of audition or visual processing, or other cognitive domains.

Therefore, in order for researchers to conduct adequate and comprehensive reviews of literature, the full reporting of participant musical experience is necessary. Specifically, it is recommended that the following variables be considered when collecting information from musicians for an experiment: (1) years of training, (2) years of experience (total, including years spent training), (3) age of training onset, (4) type of training received (ex. private vs. group lessons, and classical vs. pop or jazz), (5) intensity of training (goal of music lessons, ex. university degree, RCM certificate level, or informal), and (6) primary and secondary instruments (instrumentalist vs. vocalist).

Results from the current study suggest that there may be some music processing skills that are innate (pitch and melody perception), and others that may be developed through music training (rhythm perception). In order to further understand the ways in which pre-disposed abilities and environmental influences impact near- and far-transfer abilities is through random-assignment. This type of procedure is more common in studies on cognitive transfer, however might provide novel insight in music processing ability.

Finally, a lack of data in some music processing and cognitive domains resulted in the exclusion from more fine-grained analyses. Specifically, the author was only able to find three studies comparing musicians and non-musicians on Timbre Perception. This raises the question of whether this is a musical domain in which either: musicians and non-musicians do not
significantly differ, or if it is more influenced by innate perception abilities (as pitch and melody perception appear to be) or music training (rhythm perception). More studies on the perception of timbre would provide a clearer picture of the overall effect of music training on timbre perception. In the cognitive domain, there were not enough studies to run more detailed analyses on domains of Construction and Motor Performance or Executive Functions. Again, further investigations of these cognitive areas may provide a better insight into the differences between musicians and non-musicians, and the role of music training.

To the author’s knowledge, this is the first quantitative synthesis, by way of meta-analysis, which compared musicians to non-musicians on both near- and far-transfer abilities. The current study attempted to provide researchers working within the field with a meaningful summary of studies in both the music processing and cognitive-transfer literature. Furthermore, it attempted to isolate and detail variables related to music training that might be acting as confounds in primary studies. Findings from the field of music perception and cognition transcend into many other fields, and have widespread implications throughout academia and in society. Thorough and standardized reporting of participant information may be the key to further understanding differences between musicians and non-musicians.

References


Mehr, S., Schachner, A., Katz, R. & Spelke, E. (2013). Two randomized trials provide no consistent evidence for nonmusical cognitive benefits of brief preschool music enrichment. *PLoSOne, 8*(12), e82007.


