Abstract

Studies present conflicting results as to the potential for music training to increase cognitive abilities in schoolchildren. While some studies show large effects, studies with active control groups often show smaller effect sizes or non-significant results. A random-effects meta-analysis was conducted to calculate the overall effect size of musical interventions on measures of cognitive ability in schoolchildren. Results showed small to medium overall effects (g = .28), however, significant moderators related to methodological quality rendered findings less strong (g = .08). Additional moderator analysis showed no clear advantage in one area of cognitive function (verbal, g = .28) compared to another (non-verbal, g = .28). Results did not differ by geographical locale or type of music intervention. When compared to active control groups, music training yielded more improvement on a range of cognitive measurements (g = .21). Overall, results suggest music training may be a positive cognitive intervention for schoolchildren.

Keywords: cognition, executive function, intelligence, interventions, music education, music training, verbal memory, working memory

Introduction

1.1 Historical Background

The utility of music training in schools has received much attention in the United States, with the pendulum of advocacy topics swinging back and forth between aesthetic and extramusical benefits as the key argument for music's place in the core curriculum. Though it has been suggested Lowell Mason's singing schools were primarily started for aesthetic reasons, to improve the singing in local churches, Mason himself proposed to the Boston school board that his vocal programs had extramusical benefits, namely that singing facilitated healthy lung development in young children (Mark & Gary, 2007). While some scholars remain tied to advocating for music training for autotelic purposes, others remain convinced there are extramusical benefits to music training that if found, may be used as mortar to reify music education's status in the core curriculum of American schools.

1.2 Theoretical, policy, and/or practical issues related to the question

It might be said that the now infamous "Mozart Effect" study (Rauscher, Shaw, & Ky, 1993) was the catalyst for the "music makes you smarter" argument advocated through much of the 1990s and early 2000s. Media outlets such as the *New York Times* and *Boston Globe* popularized results to the point where the governor of Georgia proposed giving every infant a CD or cassette of classical music at birth. A later meta-analysis (Pietschnig, Voracek, & Formann, 2010) and replication attempt (McCutcheon, 2000) revealed small to no effects, subjecting these early findings to scrutiny. It should be noted that Rauscher herself is on the record as stating her research team never made the claim that listening to Mozart increases intelligence ("Mozart Sonata's IQ Impact: Eine Kleine Oversold?", 1999).

Despite potential spurious findings early on, research on the cognitive benefits of music training has continued and for noble reasons; a proven cognitive intervention for schoolchildren in a domain as enjoyable and sociable as music has immense utility in academic settings. Early studies measuring the cognitive benefits of music training tended to lack an active control, potentially inflating the effects of music training compared to groups of schoolchildren who had received no cognitive intervention (e.g. Bilhartz, Bruhn, & Olson, 2000; Dege, Wehrum, Stark, & Schwarzer, 2011; Gromko & Poorman, 1998; Ho, Cheung, & Chan, 2003; Rauscher & Zupan, 2000). Findings from these studies became difficult to interpret as music interventions appeared to enhance visual and auditory memory in some studies (Dege, Wehrum, Stark, & Schwarzer, 2011), increase spatial-temporal reasoning in other studies (Bilhartz, Bruhn, & Olson, 2000; Gromko & Poorman, 1998; Rauscher, 2000), but not spatial-temporal memory in all studies (Ho, Chueng, & Chan, 2003).

A later wave of research elucidated the methodological issue of no active control group and addressed the concern that any cognitive training should be more beneficial to cognitive growth in studies compared to studies with no type of control group training. These later studies looked at the premise of music training as being superior to other types of cognitive interventions for schoolchildren, though their findings again gave mixed results (e.g. Bhide, Power, & Goswami, 2013; Bugos & DeMarie, 2017; Flaugnacco et al., 2015; Janus, Lee, Moreno, & Bialystok, 2015; Rickard, Bambrick, & Gill, 2012; Roden, Grube, Bongard, & Kreutz, 2014). One study found music training to be a superior intervention for all cognitive outcomes measured (Flaugnacco et al., 2015), while two studies found benefits in some but not all measures of cognitive ability (Bugos & DeMarie, 2017; Roden, Grube, Bongard, & Kreutz, 2014). The majority of findings from studies with active control groups, however, claimed no such superiority for music training as a superior cognitive intervention (Bhide, Power, & Goswami, 2013; Janus, Lee, Moreno, & Bialystok, 2015; Rickard, Bambrick, & Gill, 2012). As interest in this topic continues to accelerate music education advocacy discourse, conflicting findings and pushes for high-quality data make it appropriate to conduct a meta-analysis on the cognitive benefits of music training in schoolchildren at this time.

1.3 Rationale for the selection and coding of moderators

Due to conflicting results about which areas of cognitive growth may be enhanced with music training, the first moderator to be evaluated was the type of cognitive outcome measured. Because of the auditory nature of music, it logically followed that verbal cognition would stand to benefit most from music training, perhaps at a greater rate than spatial ability or other cognitive measures, in this study labeled “non-verbal.” When a study did not specify subsections of test batteries, “full scale” was used to label the outcome reported.

Studies on the cognitive benefits of music training have piqued global curiosity. Unfortunately, standards about the consent process, random allocation of participants, ethical research standards, and reporting of non-significant findings may vary from location to location. The author of this paper was interested to see if any particular region appeared to contribute different effect sizes to the literature. If so, further assessment could be applied to the methodological processes inherent in such studies to investigate causes for different findings.

Already identified as a potential moderator was the conflicting findings from studies with active control groups and a passive control groups. In a previous meta-analysis on the relationship between music training and adolescents’ cognitive and academic skills, it was suggested that methodological quality played a large part in finding positive effects of musical training (Sala & Gobet, 2017). However, the claim of exaggerated effects did not consider the fact that positive effects were indeed found, but that they were simply not greater than that of the active control groups.

An additional level of methodological evaluation not present in the previous meta-analysis was the setting of the study, labeled in this study as natural or lab-based. Natural settings, such as classrooms and community centers, offer less mobile participants the opportunity to be part of a research study, which is a positive. However, it is not clear in many natural setting studies if participants were given a chance to opt-out, as all children from single classrooms would receive instruction with no attention to random assignment. Additionally, multiple settings in some studies (e.g. Bilhartz, Bruhn, & Olson, 2000; Portowitz, Lichtenstein, Egorova, & Brand, 2009) meant that the same instructor was not present between control and multiple experimental groups. The benefits of lab-based settings include instruction and testing done by the same teacher or set of researchers, more strict attention to attendance and reporting of attrition, and less distractions from typical school occurrences, classroom interruptions, and other uncontrollable variables. The presence of an active control group and the study being conducted in a lab setting were combined and evaluated as an additional moderator variable, reflecting the highest level of methodological rigor in the present collection of studies.

The final moderator tested was the type of musical training experienced by participants. Many studies used age-appropriate standardized musical training programs such as Kodaly or Kindermusik. However, a few studies had students studying in a private lesson setting or a traditional ensemble setting. The author of this paper was curious whether any type of training appeared to have cognitive advantages not afforded by other types of musical training.

1.4 Types of outcome measures used

In a previous meta-analysis by Sala & Gobet (2017), there was a choice to report cognitive measures with academic achievement. The present study looked only at cognitive measures for the following reasons. Conflating cognitive skills with academic outcomes such as standardized reading tests or math facts worksheets fails to acknowledge that cognitive ability and academic achievement are two discrete concepts. Achievement tests are more likely to encapsulate other variables such as motivation and knowledge of specific facts, while cognitive measures should be designed to look at ability separate of memorized facts.

1.5 Hypotheses

The purpose of this meta-analysis was to measure the overall mean effects of music training on cognitive measures in schoolchildren. Several moderators were identified to find a truer understanding of the overall mean effect. The following hypotheses were tested:

Main Hypothesis: Musical training will show small to medium effects on cognitive ability.

Moderator #1: The effects of music training will be larger for verbal than non-verbal (not verbal) measures of cognitive ability. This is due to the auditory processing component inherent in both domains.

Moderator #2: Studies published in North America will have larger effect sizes than studies published in Europe. The author proffers that non-significant results are less likely to be published in North America, where research journals tend to publish only significant findings.

Moderator #3-5: The methodological quality of studies will dramatically impact the found effect sizes. Studies in lab settings will have smaller effect sizes (#3), as will those with active control groups (#4). The presence of an active control group in a lab setting will yield the smallest effect sizes (#5).

Moderator #6: It is not expected that any one type of musical training will be superior to another.

Method

2.1 Study Criteria

The following eight criteria were established for studies to be included in the meta-analysis:

1. Participants had to be school-aged children.
2. Studies had to use an experimental or quasi-experimental design.
3. If random assignment was not used (quasi-experimental), justification and an attempt to match participants on relevant measures was required.
4. Outcomes (dependent variables) must clearly be cognitive measures and not measures of academic achievement, which can be biased and/or unreliable.
5. Effect sizes for musical aptitude or ability tests were not included, even if cognitive in nature. The aim of the study was to evaluate non-musical outcomes only.
6. Music training (independent variable) was described thoroughly enough to ensure students in the experimental group received a legitimate musical treatment.
7. Music training had to exceed two sessions. Studies that took pre-test measures, provided a brief musical intervention, and took post-test measures were not included. The aim of this study was not immediate change in cognition, but long-term benefits.
8. Studies had to be available in English, but could be conducted globally. The researcher is not proficient enough in other languages to include non-English studies.
9. Enough data were present to calculate effect sizes.

2.2 Moderator analyses

In the search for understanding the true effects of musical training on cognitive measures, several potential moderators were analyzed:

1. Type of outcome measure (categorical): This variable identified which type of cognitive measurement was reported. Some studies reported only a full-scale cognitive

score, while others identified specific verbal or non-verbal subscales or measurements. As verbal cognition measures were most frequently used, “non-verbal” in this meta-analysis could best be understood as “not verbal,” containing cognitive measurements related to spatial, temporal, “nonverbal”, or quantitative ability.

2. Region (categorical): This variable identified the continent on which a study was conducted.

3. Setting (dichotomous): This variable identified whether the study was conducted in a natural setting (e.g. a classroom or community center) or lab, meaning participants had to go to a university or other setting not part of their normal routine. This study assumed a lab setting is ideal, as researchers can control better for attendance, distractions, and other extraneous variables.

4. Active versus passive control group: (dichotomous): This variable identified whether the music training group was compared to an active control group. A “no” meant the music training group was compared to a passive group that received no type of intervention.

5. Active control and lab setting? (dichotomous): This variable identified if a study met the highest level of methodological quality. A “yes” meant a study had both an active control group and took place in a lab setting.

6. Musical training (categorical): This variable identified what type of musical training participants experienced. General would be a “general music” type program such as Kodaly or Kindermusik, where students are engaged in making music in a communal setting. Lessons indicates students received individual music lessons in a one-on-one setting. Ensemble refers to learning instruments together in small or large groups. One

study using computers was excluded from this moderator analysis. In another study, it was unclear if musical training happened individually or in groups, and thus was excluded from this moderator analysis as well. Both studies were included in the overall and previous moderator analyses.

2.3 Search strategies

A systematic search strategy was employed with databases made available through the University of South Florida library system. The terms “music” and “children” were searched along with “effects,” “cognition,” “cognitive,” “training,” “education,” “memory,” “executive function,” or “benefits.” There was not a restriction on the year in which a study was published. Indexes included ERIC, Google Scholar, JSTOR, Proquest Dissertation & Theses, Psyc-Info, and Scopus. Additionally, reference lists from collected studies and literature reviews were examined to identify studies not found using the search terms. A limitation in this study was that the researcher did not contact other researchers in the field to seek out unpublished or unreported data.

2.4 Coding procedures

The coding of moderator variables was done first by the author of the study, a published researcher in the field of music education. The second coder was a graduate student in Applied Behavioral Analysis with experience as a reading interventionist for students with special needs, who was experienced in administering a selection of the cognitive measurements used in the collected studies. The second coder was provided a list of moderators and moderator variables as well as a copy of each study and asked to make their ratings. Unsurprisingly, the coders were in full agreement over the type of musical training, country of study, presence of active control group, and type of experimental setting. The second coder disagreed with the first coder on four of the labels assigned to the type of cognitive measurement, yielding a *kappa* of .921. For these discrepancies, both coders read over the description of the measurements in the original study and discussed where the misunderstanding took place. This activity resulted in the first coder agreeing with second coder in three of the four cases and the second coder agreeing with the first coder on the final case.

To test for methodological quality, the presence of an active control group and the location of the study were considered. Studies that took place in natural settings, such as a classroom or community center, are not able to control for as many extraneous variables resulting from environmental factors. Studies with an active control group in a homogenous lab setting were considered the highest level of methodological quality. Due to researcher interest, these two variables were additionally analyzed separately.

2.5 Statistical methods

The raw data and subsequent analyses were evaluated by a full professor with extensive publishing credentials in psychology, specifically in the area of psychological meta-analysis methodology. The reviewing professor found an error in initial moderator interpretation which has since been corrected. Additionally, the reviewing professor suggested the author run the overall analysis with averaged effect sizes for each study rather than multiple individual effect sizes weighted by sample size to check for differences in means and confidence intervals between the two methods. The findings and explanation of this method are provided later in this section.

Effect sizes were collected using sample size, standardized mean differences, and standard deviations to calculate Cohen's *d*. Two studies reported Pearson correlations, one study reported log-odds, and one study reported eta-squared, which were used along with sample size to calculate *d*. Subscripts *c, e,* and *g* in the following equations refer to control group, experimental group, and pre- to post-test gain, respectively. Formulas were taken from Borenstein, Hedges, Higgins, & Rothstein (2009).

The formula used to calculate effect sizes that reported only post-test measures was:

$$d = (M_e - M_c) / SD_{pooled}$$

The formula used to calculate effect sizes that reported pre- and post-test measures was:

$$d = (M_{eg} - M_{cg}) / SD_{pooled\ (of\ pre\text{-}test)}$$

Each effect size and variance was transformed using Hedges' *g* to correct for bias:

$$V_g = [1 - 3/(4*df\text{-}1)]^2*V_d \qquad g = [1 - 3/(4*df\text{-}1)]*d$$

The formula used to calculate confidence intervals was:

$$95\%\ \mathrm{CI} = g \pm 1.96*sem$$

The formula used to calculate credibility intervals was:

$$95\%\ \mathrm{CR} = g \pm 1.96*\sqrt{REVC}$$

More than one effect size was able to be calculated for many of the studies. In this analysis, each effect size was treated as an independent outcome rather than nested data. It has been suggested violation of statistical independence has little to no effect on means, standard deviations, or confidence intervals for meta-analyses (Bijmolt & Pieters, 2001). Additionally, combining each study into a single effect size in the present study negated the uniqueness of each instrument to measure a different component or function of cognitive ability. At the behest of the reviewing professor, the initial findings were compared to a random-effects meta-analysis using a single weighted average of effect sizes per study. The differences between analyses showed a mean effect difference of g = .02 with slightly larger confidence intervals throughout. The small differences in this sample specifically were hardly noteworthy and corroborate the suggestion of Bijmolt & Pieters (2001).

Random effects was the chosen method of meta-analysis because the methods in each study varied significantly, and therefore, it was assumed that mean effects, variance, and distributions of scores would differ from sample to sample. A fixed-effects model would only be appropriate to assess the cognitive benefits of music training if each study was designed and implemented identically. The overall meta-analysis, tests for heterogeneity, moderators, sensitivity, and outliers were assessed using the 'metafor' package (v. 2.0-0) in R Studio (v. 3.4.3).

Results

3.1 Search Results

Using the criteria established in sections 2.1 and 2.3, the search yielded 52 potential studies for inclusion. Participant ages, types of outcome variables, lack of experimental or quasi-experimental research designs, and lack of enough data reported to calculate effect sizes reduced the final number of studies in the meta-analysis to 21 (*n*). From these 21 studies, 105 effect sizes (*k*) were able to be calculated with 5612 participants (*N*).

3.2 Sensitivity analysis

The final meta-analysis was conducted with the removal of 5 outliers containing studentized *z*-scores greater than |2.5| (Table 1). Trim-and-fill analysis suggested 13 missing studies, but to the right of the mean, so they were not included in the final analysis. Egger's regression test for asymmetry supported not including the trim-and-fill data, $z = -1.60$, $p = .11$.

3.2 Tables and graphic summaries

Table 1
List of studies and moderator variables

Author	Type of Instruction	Type of Measurement	Country of Study^	Type of Setting	Active/ Control?
Bhide, Power, & Goswami (2013)	Lessons	Verbal	UK	L	Yes
Bhide, Power, & Goswami (2013)	Lessons	Verbal	UK	L	Yes
Bhide, Power, & Goswami (2013)	Lessons	Verbal	UK	L	Yes
Bhide, Power, & Goswami (2013)	Lessons	Verbal	UK	L	Yes
Bhide, Power, & Goswami (2013)	Lessons	Verbal	UK	L	Yes
Bhide, Power, & Goswami (2013)	Lessons	Verbal	UK	L	Yes
Bhide, Power, & Goswami (2013)	Lessons	Verbal	UK	L	Yes
Bilhartz, Bruhn, & Olson (2000)	General	Verbal	USA	N	No
Bilhartz, Bruhn, & Olson (2000)	General	Verbal	USA	N	No
Bilhartz, Bruhn, & Olson (2000)	General	Verbal	USA	N	No
Bilhartz, Bruhn, & Olson (2000)	General	Verbal	USA	N	No
Bilhartz, Bruhn, & Olson (2000)	General	Verbal	USA	N	No
Bilhartz, Bruhn, & Olson (2000)	General	Verbal	USA	N	No
Bugos & DeMarie (2017)	General	Verbal	USA	L	Yes
Bugos & DeMarie (2017)	General	Verbal	USA	L	Yes
Bugos & DeMarie (2017)	General	Verbal	USA	L	Yes
Bugos & DeMarie (2017)	General	Verbal	USA	L	Yes
Dege, Wehrum, Stark, & Schwarzer (2011)	EL	Full Scale	Germany	N	No
Dege, Wehrum, Stark, & Schwarzer (2011)	EL	Non-Verbal	Germany	N	No
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Flaugnacco, et al. (2015)	General	Verbal	USA & Italy	N	Yes
Gromko & Poorman (1998)	General	Non-Verbal	USA	N	No
Gromko & Poorman (1998)	General	Non-Verbal	USA	N	No
Gromko (2005)	General	Verbal	USA	N	No
Gromko (2005)	General	Verbal	USA	N	No
Gromko (2005)	General	Verbal	USA	N	No
Hanson (2003)	General	Non-Verbal	USA	N	No
Hanson (2003)	General	Non-Verbal	USA	N	No
Hanson (2003)	General	Non-Verbal	USA	N	Yes
Hanson (2003)	General	Verbal	USA	N	No
Hanson (2003)	General	Non-Verbal	USA	N	Yes
Hanson (2003)	General	Verbal	USA	N	Yes
Ho, Cheung, Chan (2003)	EL	Non-Verbal	Hong Kong	L	No
Ho, Cheung, Chan (2003)	EL	Full Scale	Hong Kong	L	No
*Ho, Cheung, Chan (2003)	EL	Verbal	Hong Kong	L	No
Janus, Lee, Moreno, & Bialystok (2015)	General	Non-Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Non-Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Non-Verbal	Canada	L	Yes

Table 1, *continued*

Janus, Lee, Moreno, & Bialystok (2015)	General	Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Non-Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Non-Verbal	Canada	L	Yes
Janus, Lee, Moreno, & Bialystok (2015)	General	Verbal	Canada	L	Yes
Kaviani, Mirbaha, Pournaseh, & Sagan (2013)	General	Non-Verbal	Iran	L	No
Kaviani, Mirbaha, Pournaseh, & Sagan (2013)	General	Non-Verbal	Iran	L	No
Kaviani, Mirbaha, Pournaseh, & Sagan (2013)	General	Full Scale	Iran	L	No
Kaviani, Mirbaha, Pournaseh, & Sagan (2013)	General	Verbal	Iran	L	No
Mehr, Schachner, Katz, & Spelke (2013)	General	Non-Verbal	USA	L	Yes
Mehr, Schachner, Katz, & Spelke (2013)	General	Verbal	USA	L	No
Mehr, Schachner, Katz, & Spelke (2013)	General	Non-Verbal	USA	L	Yes
Mehr, Schachner, Katz, & Spelke (2013)	General	Verbal	USA	L	Yes
Mehr, Schachner, Katz, & Spelke (2013)	General	Non-Verbal	USA	L	No
Mehr, Schachner, Katz, & Spelke (2013)	General	Non-Verbal	USA	L	No
Mehr, Schachner, Katz, & Spelke (2013)	General	Non-Verbal	USA	L	No
Mehr, Schachner, Katz, & Spelke (2013)	General	Non-Verbal	USA	L	Yes
Moreno, Driesen, Bialoystok (2011)	Unknown	Verbal	Canada	L	Yes
Portowitz, Lichtenstein, Egorova, & Brand (2009)	General	Non-Verbal	Israel	N	No
Portowitz, Lichtenstein, Egorova, & Brand (2009)	General	Non-Verbal	Israel	N	No
Portowitz, Lichtenstein, Egorova, & Brand (2009)	General	Non-Verbal	Israel	N	No
Portowitz, Peppler, Downton (2014)	Computer	Non-Verbal	Israel	L	No
Portowitz, Peppler, Downton (2014)	Computer	Non-Verbal	Israel	L	No
Rauscher & Zupan (2000)	Lessons	Non-Verbal	USA	N	No
Rauscher & Zupan (2000)	Lessons	Non-Verbal	USA	N	No
*Rauscher & Zupan (2000)	Lessons	Non-Verbal	USA	N	No
Register (2004)	Unknown	Verbal	USA	N	No
Register (2004)	Unknown	Verbal	USA	N	No
Rickard, Bambrick, & Gill (2012)	General	Verbal	Australia	N	Yes
Rickard, Bambrick, & Gill (2012)	General	Verbal	Australia	N	Yes
Rickard, Bambrick, & Gill (2012)	General	Verbal	Australia	N	Yes
Rickard, Bambrick, & Gill (2012)	General	Verbal	Australia	N	Yes
Rickard, Bambrick, & Gill (2012)	General	Verbal	Australia	N	Yes
Rickard, Bambrick, & Gill (2012)	General	Verbal	Australia	N	Yes
Rickard, Bambrick, & Gill (2012)	General	Verbal	Australia	N	Yes
Rickard, Bambrick, & Gill (2012)	General	Verbal	Australia	N	Yes
Rickard, Bambrick, & Gill (2012)	General	Non-Verbal	Australia	N	Yes
Rickard, Bambrick, & Gill (2012)	General	Non-Verbal	Australia	N	Yes
*Roden, et al. (2014a)	EL	Non-Verbal	Germany	L	Yes
Roden, et al. (2014a)	EL	Non-Verbal	Germany	L	Yes
Roden, Grube, Bongard, & Kreutz (2014b)	EL	Non-Verbal	Germany	N	Yes
Roden, Grube, Bongard, & Kreutz (2014b)	EL	Non-Verbal	Germany	N	Yes
Roden, Grube, Bongard, & Kreutz (2014b)	EL	Non-Verbal	Germany	N	Yes
Roden, Grube, Bongard, & Kreutz (2014b)	EL	Non-Verbal	Germany	N	Yes
Roden, Grube, Bongard, & Kreutz (2014b)	EL	Verbal	Germany	N	Yes
*Roden, Grube, Bongard, & Kreutz (2014b)	EL	Non-Verbal	Germany	N	Yes
*Roden, Grube, Bongard, & Kreutz (2014b)	EL	Verbal	Germany	N	Yes
Schellenberg (2004)	Lessons	Full Scale	USA	L	Yes
Schellenberg (2004)	Lessons	Full Scale	USA	L	Yes
Schellenberg (2004)	Lessons	Full Scale	USA	L	No
Schellenberg (2004)	Lessons	Full Scale	USA	L	No

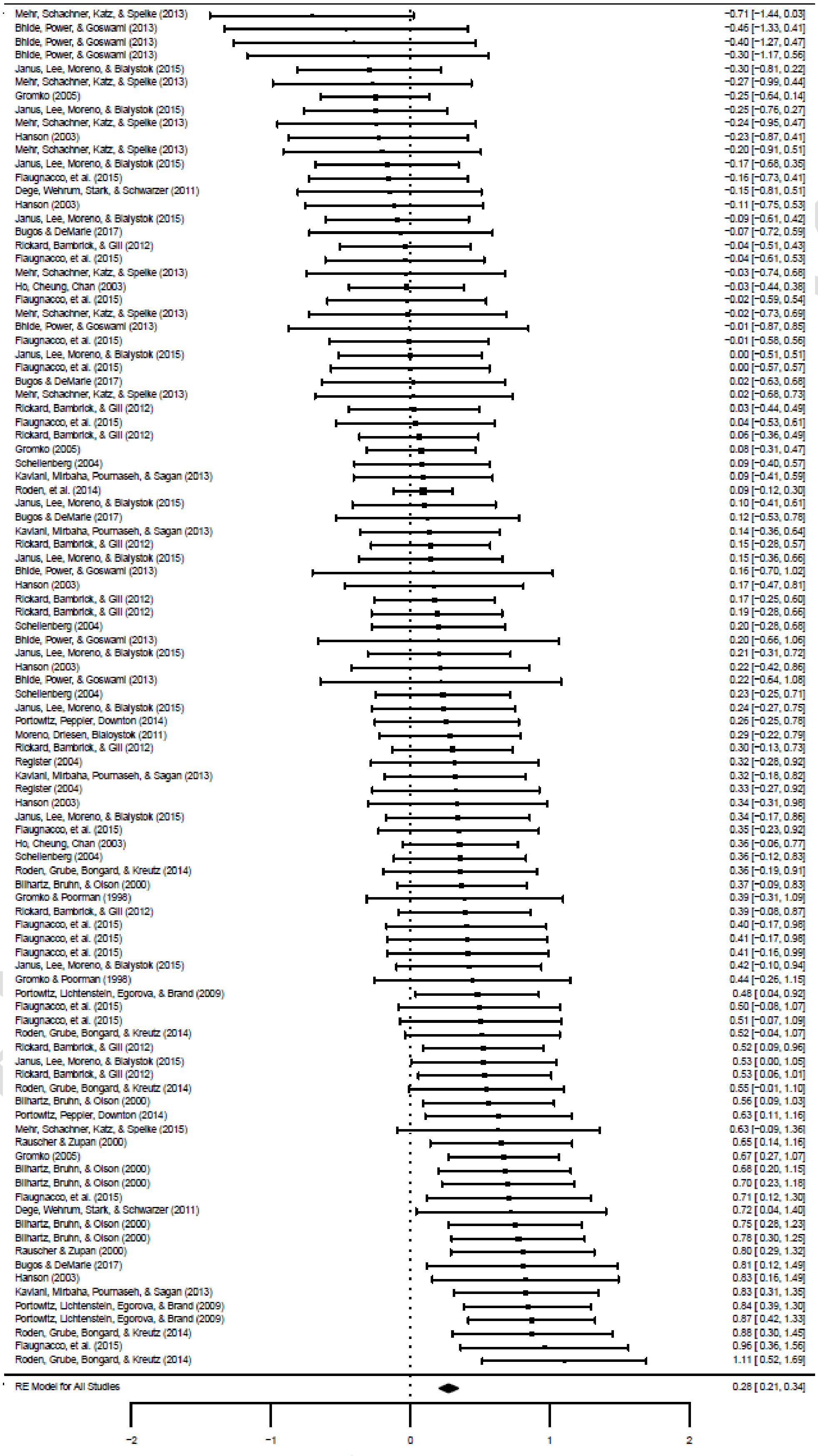
$n_{\text{studies}} = 21$, $k_{\text{ES}} = 100$ (*five ES removed as outliers); ^Moderator analyzed by region (North America, Europe, Other); **Type of Instruction** refers to music treatment; **EL** = Ensemble and Lessons; **General** = "General Music" program; **Lessons** = individual instruction; **L** = Lab setting, **N** = Natural setting (school or community center); **Active/Control?** refers to presence of an active control group.

Table 2
Summary of Overall Effect Size and Moderator Effect Sizes

Factor	Effect Size *(g)*	*se*	*z*-value	*k*	*N*	*p*-value	95% CI	95% CR
Overall ES	.276	.033	8.142	100	5612	< .001	[.21, .34]	[-.11, .66]
			Heterogeneity: $Q(df = 99) = 150.50$, $p = .0007$, $I^2 = 34.12\%$, *REVC* = .193; Plot Asymmetry: $z = -1.60$, $p = .11$.					
Scale Type								
Full Scale	.222	.121	1.834	7	448	.07	[-.02, .46]	[-.17, .61]
Verbal	.279	.045	6.193	59	3113	< .0001	[.19, .37]	[-.11, .67]
Non-Verbal	.282	.058	4.828	34	2051	< .0001	[.17, .40]	[-.11, .67]
			Heterogeneity: $QM(df = 2) = .21$, $p = .90$, $I^2 = 35.29\%$, *REVC* = .199					
Region								
Europe	.198	.079	2.515	22	1118	.01	[.04, .35]	[-.19, .59]
N. America	.275	.046	6.032	57	2917	< .0001	[.19, .36]	[-.12, .67]
Other	.333	.067	4.991	21	1577	< .0001	[.20, .46]	[-.06, .72]
			Heterogeneity: $QM(df = 2) = 1.72$, $p = .42$, $I^2 = 34.16\%$, *REVC* = .195					
Setting								
Lab	.205	.048	4.301	51	2824	< .0001	[.11, .30]	[-.16, .57]
Natural	.346	.047	7.306	49	2788	< .0001	[.25, .44]	[-.02, 71]
			Heterogeneity: $QM(df = 1) = 14.92$, $p = .0001$, $I^2 = 25.86\%$, *REVC* = .159					
Active Control?								
Yes	.214	.052	5.075	63	3396	< .0001	[.13, 30]	[-.13, .56]
No	.373	.042	7.117	37	2216	< .0001	[.27, 48]	[.03, .72]
			Heterogeneity: $QM(df = 1) = 5.60$, $p = .02$, $I^2 = 29.93\%$, *REVC* = .176					
Lab Setting and Active Control?								
Yes	.079	.060	1.309	31	1604	.19	[-.04, .20]	[-.23, .39]
No	.353	.037	9.438	69	4008	< .0001	[.28, .43]	[.05, .66]
			Heterogeneity: $QM(df = 1) = 14.98$, $p = .0001$, $I^2 = 25.33\%$, *REVC* = .157					
Musical Training								
Lessons	.221	.041	2.038	13	521	.04	[.01, .43]	[-. 20, .64]
General	.260	.101	6.337	72	3980	< .0001	[.18, .34]	[-.16, .68]
EL	.383	.105	3.653	10	1111	< .001	[.18, .59]	[-.03, .80]
			Heterogeneity: $QM(df = 2) = 1.41$, $p = .49$, $I^2 = 38.29\%$, *REVC* = .213					

Table 3

Forest plot of studies sorted by *g*



3.3 Results of main effect size

A random effects model (REML) of 100 effect sizes (*k*) with 5,612 participants (*N*) resulted in a main effect size of $g = .276$, $p <. 0001$, 95% CI [.21, .34], 95% CR [-.11, .66], *REVC* = .193 (Table 2). The test for heterogeneity showed the potential presence of moderator variables, $Q(df = 99) = 150.50$, $p = .0007$, $I^2 = 34.12\%$.

3.4 Results of moderators

Tests for moderator significance revealed significant differences by setting, $QM(df = 1) = 14.92$, $p = .0001$, $I^2 = 25.86\%$; active versus passive control group, $QM(df = 1) = 5.60$, $p = .02$, $I^2 = 29.93$; and the combined active control+lab setting variable; $QM(df = 1) = 14.98$, $p = .0001$, $I^2 = 25.33\%$. Tests for moderator significance revealed no significant differences by scale type, $QM(df = 2) = .21$, $p = .90$, $I^2 = 35.29\%$; region, $QM(df = 2) = 1.72$, $p =.42$, $I^2 = 34.16\%$, or type of musical training, $QM(df = 2) = 1.41$, $p = .49$, $I^2 = 38.29\%$. The mean effect sizes for each moderator variable with confidence and credibility intervals are available in Table 2.

Discussion

4.1 Statement of major findings

The mean effect size by itself shows a small to medium effect for musical training on cognitive measures. However, considering the large quantity of confidence intervals containing 0 (Table 3) and non-significant prediction intervals (Table 2), it would be brash to say there is any type of causal link. When a high degree of methodological quality is present, the evidence for musical training as superior to other forms of cognitive interventions looks questionable. These results taken together would suggest that music training is positive for cognitive growth, but so are a variety of other interventions.

4.2 Consideration of alternative explanations for observed results

One potential alternative for cognitive growth during this time is the natural maturation of schoolchildren. It is fair to question if observed positive effects of music training are spurious to the natural cognitive growth of children. In a small random sample, it is possible some children experienced faster natural rates of maturation as cognitive growth between children is not fixed. However, there were positive small effects when all studies were combined, even under the most rigorous of conditions, likely ruling out small sample sizes combined with unequal maturation as a potential explanation.

4.3 Generalizability of conclusions

The examination of studies from multiple countries and multiple types of musical trainings increases the generalizability of these findings to broader samples. Though there were significant moderators, they did not carry nearly the weight of going from medium effects (g = .35) to below small effects (g = .08) as was seen with methodological quality.

The researcher collected age ranges but did not analyze them as a moderator. These conclusions are based off of schoolchildren that are overwhelmingly from ages 4 to 10, a time of rapid cognitive development in healthy children. It would be inappropriate to generalize the findings of this study to older adults experiencing healthy or unnatural cognitive decline, those with brain injuries, those with cognitive delays, or those for which music is being prescribed as a therapy.

4.4 General limitations

The amount of highest-quality studies available will limit any meta-analysis. While there were enough studies available to conduct a worthwhile study, it should be noted that the highest-quality studies typically had smaller sample sizes as a result of more controlled allocation and matching of groups. As more high-quality studies become available, it would be worthwhile to conduct a future meta-analysis revisiting the issue of methodological quality found in this study.

With regards to the amount of present studies, it was mentioned earlier that the author of this paper could have been more exhaustive in searching for unpublished data. It is plausible there are other studies that meet the inclusion criteria, although the researcher suspects methodological quality would not be high in most cases.

A final limitation of this study is the lack of standardization present in the outcome variables from study to study. Most studies were explicit in the cognitive domain they were measuring, however, each measure was not always used for the same purpose or even defined consistently. For example, Bugos & DeMarie (2017) chose to use the *Matching Familiar Figures Test* (Egeland & Weinberg, 1976) as a continuous measure of time and errors rather than the intended use of categorizing children as reflective or impulsive. In the case of the *Raven* test (Raven, Court, & Raven, 1996), it was explicitly defined in one study as “designed to measure a person’s ability to form perceptual relations and to reason by analogy independent of language and formal schooling” (Portowitz, Lichtenstein, Egorova, & Brand, 2009) but generalized as simply “nonverbal cognitive functioning” (Janus, Lee, Moreno, & Bialystok, 2016) in another. For these reasons, it should be considered if cognitive measures from study to study were chosen carefully enough, collected appropriately, and interpreted accurately. In fairness, the same scrutiny could be applied to the coding of the measurement moderators in this study, as the two coders together still did not have experience with or expert knowledge of every measurement included in the meta-analysis.

4.5 Implications and interpretation

The key takeaway from this meta-analysis is that many interventions aimed to promote cognitive growth during early childhood were found to be beneficial and should therefore be structured into schools and community centers. However, even though other interventions throughout these studies demonstrated the potential for cognitive growth, the potential of music instruction as a sociable and enjoyable intervention should be taken into consideration when evaluating the efficacy and novelty of such programs, especially for children. Additionally, the fact that music training can be made progressively more challenging is a unique strength to consider when advocating for music training as a cognitive growth intervention.

4.6 Guidelines for future research

There were a few exemplar cases of methodological rigor applied to studying the cognitive benefits of music instruction (e.g. Mehr et al., 2013; Rickard, Bambrick, & Gill, 2012). Studies such as these should be used as guidelines for designing future studies. Key to these designs were the presence of an active control group. The author of this paper recommends future studies make every effort to have an active control group for not only obvious methodological reasons but for ethical reasons as well. As ethical researchers, denying school-aged participants potentially beneficial interventions is something that must be avoided. In the event that a passive control group is strongly desired, the control group should be offered music training at the end of the study. Additionally, if a researcher's hypothesis is that music training will be a superior cognitive intervention to an active control intervention, then it stands that the ethical practice would be to offer the active control group music training after data collection as well. The latter consideration in particular was not described in any single study.

One area of curiosity that remains unclear after this meta-analysis is the potential for cognitive growth in very specific areas as a result of music training, such as working memory or auditory discrimination,. Of the one hundred and five effect sizes initially analyzed in this study, nearly every effect size came from a different cognitive outcome measure. While these tests are similar in nature, they do not capture exactly the same cognitive function. As more studies are published, it would be wise to re-visit these data to begin parsing out a category like "non-verbal" into more specific categories such as spatial, temporal, and quantitative ability. The present body of literature does not contain enough effect sizes to justify such discrete compartmentalization during analysis at this time.

References

An asterisk (*) denotes that a citation was used in the final meta-analysis. Table 1 in the body of the report shows how many effect sizes were pulled from each individual study.

*Bhide, A., Power, A., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education, 7*, 113-123. http://dx.doi.org/10.1111/mbe.12016.

Bijmolt, T., & Pieters, R. (2001). Meta-analysis in marketing when studies contain multiple measurements. *Marketing Letters*, *12*, 157-169. http://dx.doi.org/10.1023/A:1011117103381.

*Bilhartz, T., Bruhn, R., & Olson, J. (2000). The effect of early music training on child cognitive development. *Journal of Applied Developmental Psychology, 20*(4), 615-636. http://dx.doi.org/10.1016/S0193-3973(99)00033-7.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis.* Chichester, UK: Wiley.

*Bugos, J., & DeMarie, D. (2017). The effects of a short-term music program on preschool children’s executive functions. *Psychology of Music*, *45*(6), 855-867. doi:10.1177/0305735617692666

*Dege, F., Wehrum, S., Stark, R., & Schwarzer, G. (2011). The influence of two years of school music training in secondary school on visual and auditory memory. *European Journal of Developmental Psychology*, *8*(5), 608-623. doi:10.1080/17405629.2011.590668.

Egeland, B., & Weinberg, R. A. (1976). The Matching Familiar Figures Test: A look as its psychometric credibility. *Child Development, 47*(2), 483–491. http://dx.doi.org/10.2307/1128805.

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

*Gromko, J. (2005). The effect of music instruction on phonemic awareness in beginning readers. *Journal of Research in Music Education*, *53*(3), 199-209. http://dx.doi.org/10.1177/002242940505300302.

*Gromko, J., & Poorman, A. (1998). The effect of music training on preschools' spatial-temporal task performance. *Journal of Research in Music Education*, *46*(2), 173-181. http://dx.doi.org/10.2307/3345621.

*Hanson, M. (2003). Effects of sequenced Kodaly literacy-based music instructions on the spatial reasons skills of kindergarten students. *Research & Issues in Music Education, 1*(1), Article 4. Available at: http://ir.stthomas.edu/rime/vol1/iss1/4.

*Ho, Y., Cheung, M., & Chan, A. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology, 17*(3), 439-450. doi:10.1037/0894-4105.17.3.439.

*Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *Journal of Experimental Child Psychology, 144*, 84-97. http://dx.doi.org/10.1016/j.jecp.2015.11.009.

*Kaviani, H., Mirbaha, H., Pournaseh, M., & Sagan O. (2014). Can music lessons increase the performance of preschool children in IQ tests? *Cognitive processing, 15*(1), 77-84. doi:10.1007/s10339-013-0574-0.

Mark, M., & Gary, C. (2007). *A history of American music education* (3rd Ed.). Lanham, MD: Rowman & Littlefield Education.

McCutcheon, L. (2000). Another failure to generalize the Mozart effect. *Psychological Reports, 87*(1), 325-330. doi:10.2466/pr0.2000.87.1.325.

*Mehr, S. A., Schachner, A., Katz, R. C., & Spelke, E. S. (2013). Two randomized trials provide no consistent evidence for nonmusical cognitive benefits of brief preschool music enrichment. *PLoS ONE, 8*(12): e82007. https://doi.org/10.1371/journal.pone.0082007.

Pietschnig, J., Voracek, M., & Formann, A. (2010). Mozart effect-Shmozart effect: A meta-analysis. *Intelligence*, *38*(3), 314-323. https://doi.org/10.1016/j.intell.2010.03.001.

*Portowitz, A., Lichtenstein, O., Egorova, L., & Brand E. (2009). Underlying mechanisms linking music education and cognitive modifiability. *Research Studies in Music Education, 31*(2), 107-128. doi:10.1177/1321103X09344378

*Portowitz, A., Peppler, K., & Downton, M. (2014). In harmony: A technology-based music education model to enhance musical understanding and general learning skills. *International Journal of Music Education*, *32*(2), 242-260. doi:10.1177/0255761413517056.

Rauscher, F., Shaw, G., & Ky, K. (1993). Music and spatial task performance. *Nature, 365*(611). doi:10.1038/365611a0.

*Rauscher, F., & Zupan, M. (2000). Classroom keyboard instruction improves kindergarten children’s spatial-temporal performance: A field experiment. *Early Childhood Research Quarterly*, *15*(2), 215-228. http://dx.doi.org/10.1016/S0885-2006(00)00050-8.

Raven, J., Court, J., & Raven, J. (1996). *Coloured progressive matrices*. London: H. K. Lewis.

*Register, D. (2004). The effects of live music groups versus an educational children's program on the emergent literacy of young children, *XLI*(1), 2-27. http://dx.doi.org/10.1093/jmt/41.1.2.

*Rickard, N., Bambrick C., & Gill, A. (2012). Absence of widespread psychosocial and cognitive effects of school-based music instruction in 10-13 year-old-students. *International Journal of Music Education, 30*(1), 57-78. doi:10.1177/0255761411431399.

*Roden, I., Konen, T., Bongard, S., Frankenberg, E., Friedrich, E., & Kreutz, G. (2014a). Effects of music training on attention, processing speed and cognitive music abilities – findings from a longitudinal study. *Applied Cognitive Psychology, 28,* 545-557. doi:10.1002/acp.3034.

*Roden, I., Grube, D., Bongard, S., & Kreutz, G. (2014b). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychology of Music, 42*(2), 284-298. doi:10.1177/0305735612471239.

Sala, G., & Gobet, F. (2017). When the music's over. Does music skill transfer to children's and young adolescents' cognitive and academic skills? A meta-analysis. *Educational Research Review, 20*, 55-67. http://dx.doi.org/10.1016/j.edurev.2016.11.005.

*Schellenberg, E. (2004). Music lessons enhance IQ. *Psychological Science, 15,* 511-514. http://dx.doi.org/10.1111/j.0956-7976.2004.00711.x.

Weiss, R. (1999, August 30). *Mozart Sonata's IQ Impact: Eine Kleine Oversold?* Retrieved from https://www.washingtonpost.com/archive/politics/1999/08/30/mozart-sonatas-iq-impact-eine-kleine-oversold/d0a36eeb-3132-4cd9-87c7-858e70ebd216/?utm_term=.14566c966651.