When the music's over. Does music skill transfer to children's and young adolescents' cognitive and academic skills? A Meta-Analysis

Giovanni Sala, Fernand Gobet

PII: S1747-938X(16)30064-1
DOI: 10.1016/j.edurev.2016.11.005
Reference: EDUREV 209

To appear in: Educational Research Review

Received Date: 10 May 2016
Revised Date: 7 November 2016
Accepted Date: 25 November 2016

Please cite this article as: Sala, G., Gobet, F., 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 (2016), doi: 10.1016/j.edurev.2016.11.005.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
When the Music's Over. Does Music Skill Transfer to Children’s and Young Adolescents’ Cognitive and Academic Skills? A Meta-Analysis

Giovanni Sala and Fernand Gobet

University of Liverpool

Author Note

Giovanni Sala, Department of Psychological Sciences, University of Liverpool;
Fernand Gobet, Department of Psychological Sciences, University of Liverpool.

Correspondence concerning this article should be addressed to Giovanni Sala,
Department of Psychological Sciences, Bedford Street South, University of Liverpool,
Liverpool L69 7ZA, United Kingdom.
e-mail: Giovanni.sala@liv.ac.uk
1. Introduction

Recently, the question of whether music-related activities in school improve young people’s cognitive and academic skills has raised much interest among researchers, educators, and policy makers. Several studies have tried to establish the effectiveness of music training in enhancing children’s and young adolescents’ general intelligence (Rickard, Bambrick, & Gill, 2012), memory (Roden, Kreutz, & Bongard, 2012), spatial ability and mathematics (Mehr, Schachner, Katz, & Spelke, 2013), and literacy skills (Slater et al., 2014), among others (for a review, see Miendlarzewska & Trost, 2013). Music training comprises activities such as singing, playing instruments, clapping, and rhythm games beyond many others. Notably, several specific curricula have been designed to develop those cognitive skills involved in playing music (e.g., Kodály method; Houlahan & Tacka, 2015). The educational implications of this research are evident. If music training enhances children’s and young adolescents’ cognitive skills and school grades, then schools might consider implementing additional musical activities.

1.1 The Question of Transfer of Skills

Crucially, the importance of establishing whether music training provides any educational advantage is not limited to the field of education. In fact, this topic addresses the broader psychological question of transfer of skills. Transfer of learning takes place when skills learned in one particular area either generalize to new areas or increase general cognitive abilities. It is customary to distinguish between near- and far-transfer (Barnett & Ceci, 2002; Mestre, 2005). Whilst near-transfer takes place between areas that are tightly related (e.g., driving two different car models), far-transfer occurs where the relationship between source and target areas is weak (e.g., transfer from music to mathematics). Thus, postulating that music skill generalizes to other non-music-related cognitive and academic
abilities means assuming the occurrence of a far-transfer.

According to Thorndike and Woodworth’s (1901) common-element theory, transfer depends on the number of features that are shared between two areas; these features are hypothesized to engage common cognitive elements (Anderson, 1990). A direct consequence of this theory, well supported by empirical data in psychology and education, is that, while near-transfer should be frequent, far-transfer should be rare (Donovan, Bransford, & Pellegrino, 1999; Sala & Gobet, 2016).

1.2 Why Should Music Skill Transfer to non-Music Skills?

Music training has been claimed to enhance various cognitive and academic skills. Given the well-known difficulty of far-transfer to occur, it is possible that music training boosts context-independent cognitive mechanisms, which may, in turn, improve other non-music cognitive and academic skills. According to Schellenberg (2004, 2006), the most likely explanation for the alleged diverse benefits of music interventions is that such training enhances individuals’ general intelligence, which correlates with many cognitive and academic skills (Deary, Strand, Smith, & Fernandes, 2007; Rohde & Thompson, 2007).

Music training requires focused attention, learning complex visual patterns, memory, and fine motor skills. Thus, such a demanding activity may enhance children’s and young adolescents’ overall cognitive skill, which, in turn, would increase their academic performance. This hypothesis is corroborated by the fact that formal exposure to music in childhood appears to correlate with IQ scores and academic attainment (Schellenberg, 2006).

Another possible explanation relies on executive functions. Cognitive skills such as working memory, cognitive control, and cognitive flexibility are important predictors of academic achievement (e.g., Conway & Engle, 1996; Peng, Namkung, Barnes, & Sun, 2016). Learning to play an instrument engages executive functions (Bialystok & Depape, 2009;
George & Coch, 2011) and it is not impossible that such improvements generalize to non-music skills.

1.3 Does Music Skill Transfer to non-Music Skills? A Look at the Empirical Evidence

Several correlational studies have shown that music skill is associated with non-music-specific skills such as literacy (Anvari, Trainor, Woodside, & Levy, 2002; Forgeard et al., 2008), mathematics (Cheek & Smith, 1999), short-term and working memory (Lee, Lu, & Ko, 2007), and general intelligence (Lynn, Wilson, & Gault, 1989; Schellenberg, 2006; Schellenberg & Mankarious, 2012). Anvari et al. (2002) found that music perception skills predicted reading abilities in preschool children. Similarly, Forgeard et al. (2008) reported that music discrimination ability correlated with phonological processing skill in a sample of typically developing and dyslexic children. Concerning mathematical ability, Cheek and Smith (1999) showed that students who had received private lessons of music performed better in the mathematics portion of the Iowa Test of Basic Skills. Consistent with the latter two studies, Wetter, Koerner, and Schwaninger (2009) found a positive relationship between being engaged in music activities and overall academic achievement.

Music skill seems to be positively associated to cognitive ability too. In Lee et al.’s (2007) study, music-trained children and adults were compared to age-matched control groups in a series of digit span and spatial span tasks. The music-trained groups outperformed the controls in all the measures. Regarding overall cognitive ability, a convincing amount of evidence suggests that music skill and general intelligence are significantly related. Lynn et al. (1989) found a correlation between the scores on Raven’s Standard Progressive Matrices (Raven, 1960) and a series of music tests in a group of 9-11 year-old children. Moreover, Schellenberg (2006) reported a positive correlation between duration of the music training and IQ in children and undergraduate students. Crucially, this result remained even after
controlling for potentially confounding variables, such as parental income and education. Finally, this finding was confirmed in a more recent study involving 7- and 8-year-old children (Schellenberg & Mankarious, 2012).

However, such correlational studies cannot ascertain any far-transfer of skill from music training to other areas, because no direction of causality can be inferred. For example, both music and non-music skills could stem from innate intellectual abilities. Stronger conclusions can be drawn from studies using an experimental design, where an experimental group without previous formal musical instruction receives musical training. However, the experimental studies on the benefits of music training have provided mixed results. For example, while some studies have reported positive results (Kaviani, Mirbaha, Pournaseh, & Sagan, 2014; Portowitz, Lichtenstein, Egorova, & Brand, 2009), others have showed modest evidence of music training on children’s performance on intelligence tests (Rickard, et al., 2012; Schellenberg, 2004). Analogously, studies investigating the effect of music training on cognitive ability such as spatial- and memory-related skills have provided no clear pattern of results. For example, in Bowels (2003), music training exerted a strong effect on children’s visuospatial ability. Analogously, in Degé, Wehrum, Stark, and Schwarz (2011) music training significantly enhanced the participants’ visual and auditory memory. However, Rickard et al. (2012) failed to find any effect in either of the above measure. With regard to academic achievement, previous meta-analyses suggest that music training slightly enhances students’ mathematical (Hetland & Winner, 2001; Vaughn, 2000) and literacy skills (Gordon, Fehd, & McCandliss, 2015). However, the overall effect sizes reported in these meta-analyses are modest, and the variability between studies is quite pronounced. Put simply, the effects of music training on skills such as spatial ability, memory, academic performance, and general intelligence are still controversial, and positive results have not always been replicated (Miendlarzewska & Trost, 2013).
Such variability in the results may be due to the design features of the studies, including (a) the age of the participants, (b) the random (or non-random) assignment to the treatment and control groups, and (c) the presence (or absence) of a group engaged in an alternative activity to control for non-music-specific effects, such as placebos. Age may affect the occurrence of transfer of skills in two ways. First, transfer effects may be a function of brain plasticity (Buschkuehl, Jaeggi, & Jonides, 2012), which, in turn, is a function of age. Second, as students grow up, the level of specificity of the activities they are engaged in increases (e.g., mathematics, literacy, etc.). Crucially, research on expertise has shown that the higher the level of a particular ability, the more specific the features of that ability will be, and consequently, the lower the likelihood that transfer will occur (Ericsson & Charness, 1994; Gobet, 2015).

Quality design-related features may be important moderators too. Without random allocation of the participants, it is not always possible to ensure the baseline equivalence between experimental and control groups, especially if the experimental group is self-selected. Controlling for placebo effects could be even more important. In fact, the experience of a new activity such as music training may cause, *ipso facto*, an enhancement in children’s and young adolescents’ cognitive and academic skills. Music-related activities are usually a novelty for young students and may induce a state of motivation and excitement, which, in turn, may be the real cause of the observed (and temporary) improvements. Comparing music training with other enrichment activities is thus essential to understand whether the observed benefits are specifically due to music, or just the consequence of non-specific placebo effects.

1.4 Aims of the Present Meta-Analysis

Because of the theoretical implications for theories of transfer, the possible
educational applications, and the current general interest in this topic, it is imperative to rigorously evaluate the putative benefits of music training for academic and cognitive skills. Similar claims have been made about the possibility of obtaining transferable benefits, both cognitive and academic, from playing video-games (Green, Li, & Bavelier, 2010; Green, Pouget, & Bavelier, 2010), working memory training (Melby-Lervag & Hulme, 2013, 2016), and playing chess in schools (Gobet & Campitelli, 2006; Sala & Gobet, 2016). However, research in these fields suggests that optimism about the positive effects of music training must be tempered by the possibility that the observed effects are due to confounding factors such as placebo effects (Boot, Blakely, & Simons, 2011; Gobet et al., 2014; Sala & Gobet, 2016) and lack of random assignment of the participants to the groups.

Our meta-analysis, then, examines the potential cognitive and academic benefits of music training for the general population of children and young adolescents (see 2.2. Inclusion/Exclusion Criteria). In a first stage, we estimated the overall size of the effects of music training on non-music cognitive and academic skills by comparing experimental groups to control groups. In a second phase, we assessed the potential role of several possible moderators on the effectiveness of music training. The analysis of these factors – along with the estimation of an overall effect size – aimed to test: (a) whether music training enhances students’ cognitive and academic skills, or whether far-transfer from music to other areas is null or negligible; (b) whether music training improves some specific skills more than others; (c) whether students’ age affects the benefits of music training; and (d) whether the methodological quality of the studies reviewed – i.e., random allocation of participants and comparisons with active (i.e., do-other) control groups to rule out placebo effects – influence the results.

Points a) and b) were tested by calculating a general overall effect size (see Section 3. Results) and the measure-specific overall effect sizes (see Sections 3.1 and 3.2), respectively.
Points c) and d) were addressed by performing a meta-regression analyses (see Section 3.1).

2. Method

2.1 Literature Search

In line with the PRISMA statement (Moher, Liberati, Tetzlaff, & Altman, 2009), a systematic search strategy was used to find the relevant studies. Using the following combination of the keywords “music” AND ("training" OR "instruction" OR "education" OR "intervention"), Google Scholar, ERIC, Psyc-Info, ProQuest Dissertation & Theses, and Scopus databases were searched to identify all the potentially relevant studies. Also, previous narrative reviews were examined, and we e-mailed researchers in the field (n = 11) asking for unpublished studies and inaccessible data.¹

2.2 Inclusion/Exclusion Criteria

The studies were included according to the following nine criteria:

1. The design of the study included music training; correlational and ex-post facto studies were excluded;
2. The independent variable (music training) was successfully isolated; the studies using integrated curricula (e.g., lessons of music and reading in the same intervention) were excluded;
3. The study presented a comparison between a music-treated group and, at least, one control group;
4. Music training was not merely environmental (e.g., background music, music videos);

¹ Unfortunately, no author replied to our e-mails.
5. During the study, a measure of academic and/or cognitive skill non-related to music was collected;
6. The participants of the study were pupils aged three to 16;
7. The participants of the study were pupils without any previous formal musical training (as stated by the authors of the included studies);
8. The participants of the study were pupils without any specific learning disability (e.g., developmental dyslexia) or clinical condition (e.g., autism);
9. The data presented in the study were sufficient to calculate an effect size.

To identify studies meeting these criteria, we searched for relevant published and unpublished articles in the last 30 years (from January 1, 1986, through March 1, 2016), and scanned reference lists.

Among the studies screened (n = 166), we found 38 studies, conducted from 1986 to 2016, that met all the inclusion criteria. These studies included 40 independent samples and 118 effect sizes, and a total of 3085 participants.

2.3 Moderators

We selected four potential moderators. The first two, which we termed theoretical moderators, referred to features of the dependent variables and the participants of the studies, while the last two, which we termed methodological moderators, addressed more general methodological aspects:

1. Outcome measure (categorical variable): This variable includes literacy, mathematics, memory, intelligence, phonological processing, and spatial skills.\(^2\) Effect sizes that

---

\(^2\) These broad categories were built by aggregating different outcomes related to a particular cognitive or academic ability (e.g., reading and writing both under the category of literacy). For all the details about the dependent variables of the reviewed studies, see Table 1. See the Table S1 in the Supplemental material for
were not related to these categories (e.g., visual-auditory learning and visual attention) were labelled as *others*;

2. Age: The age of the participants in years (continuous variable);

3. Random allocation (dichotomous variable): Whether participants were fully randomly allocated to the groups;\(^3\)

4. Presence of active control group (dichotomous variable): Whether the music training group was compared to another activity.

Table 1

*Studies, dependent variables, and moderators of the 118 effect sizes included in the meta-analysis*

---

\(^3\) The category of “non-random” encompasses both pre-post-test studies and only-post-test studies. Two studies reported only post-test results: Cardarelli (2003) and Geoghegan and Mitchelmore (1996).
<table>
<thead>
<tr>
<th>Study</th>
<th>Dependent variable</th>
<th>Outcome measure</th>
<th>Age (years)*</th>
<th>Randomization</th>
<th>Active control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhide, Power, and Goswami (2013) - M1</td>
<td>Working Memory (digit span)</td>
<td>Memory</td>
<td>6.8</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bhide, Power, and Goswami (2013) - M2</td>
<td>Phonological awareness</td>
<td>Phonological processing</td>
<td>6.8</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bhide, Power, and Goswami (2013) - M3</td>
<td>Spelling</td>
<td>Literacy</td>
<td>6.8</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bhide, Power, and Goswami (2013) - M4</td>
<td>Reading</td>
<td>Literacy</td>
<td>6.8</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bilhartz, Bruhn, and Olson (2000)</td>
<td>Intelligence (vocabulary)</td>
<td>Intelligence</td>
<td>4.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Bowels (2003) - M1</td>
<td>Spatial temporal ability</td>
<td>Spatial</td>
<td>6.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Bowels (2003) - M2</td>
<td>Reading</td>
<td>Reading</td>
<td>6.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Bowels (2003) - M3</td>
<td>Phonological awareness</td>
<td>Phonological processing</td>
<td>6.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Cardarelli (2003) - M1</td>
<td>Reading</td>
<td>Reading</td>
<td>9.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cardarelli (2003) - M2</td>
<td>Mathematics</td>
<td>Mathematics</td>
<td>9.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cogo-Moreira, de Avila, Ploubidis, and Mari (2013) - M1</td>
<td>Phonological awareness</td>
<td>Phonological processing</td>
<td>9.2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Cogo-Moreira, de Avila, Ploubidis, and Mari (2013) - M2</td>
<td>Reading</td>
<td>Literacy</td>
<td>9.2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Costa-Giomi (2004) - M1</td>
<td>Mathematics</td>
<td>Mathematics</td>
<td>9.0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Costa-Giomi (2004) - M2</td>
<td>Language</td>
<td>Literacy</td>
<td>9.0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Costa-Giomi (2004) - M3</td>
<td>Mathematics</td>
<td>Mathematics</td>
<td>9.0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Degé and Schwarzer (2011) - S1</td>
<td>Phonological awareness</td>
<td>Phonological processing</td>
<td>5.8</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Degé and Schwarzer (2011) - S2</td>
<td>Phonological awareness</td>
<td>Phonological processing</td>
<td>5.8</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Degé, Wehrum, Stark, and Schwarzer (2011) - M1</td>
<td>Visual memory</td>
<td>Memory</td>
<td>10.8</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Degé, Wehrum, Stark, and Schwarzer (2011) - M2</td>
<td>Memory (auditory)</td>
<td>Memory</td>
<td>10.8</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Geoghegan and Mitchelmore (1996)</td>
<td>Mathematics</td>
<td>Mathematics</td>
<td>4.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Gromko (2005)</td>
<td>Phonemic awareness</td>
<td>Phonological processing</td>
<td>5.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Study and Method</td>
<td>Domain</td>
<td>Subdomain</td>
<td>Score</td>
<td>Includes</td>
<td>Excludes</td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
<td>-----------</td>
<td>-------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Gromko and Poorman (1998)</td>
<td>Intelligence</td>
<td>Intelligence</td>
<td>3.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hanson (2001) - M1 - S1</td>
<td>Intelligence</td>
<td>Intelligence</td>
<td>5.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Hanson (2001) - M1 - S2</td>
<td>Intelligence</td>
<td>Intelligence</td>
<td>5.5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Hanson (2001) - M2 - S1</td>
<td>Spatial-temporal ability</td>
<td>Spatial</td>
<td>5.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Hanson (2001) - M2 - S2</td>
<td>Spatial-temporal ability</td>
<td>Spatial</td>
<td>5.5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Hanson (2001) - M3 - S1</td>
<td>Spatial recognition</td>
<td>Spatial</td>
<td>5.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Hanson (2001) - M3 - S2</td>
<td>Spatial recognition</td>
<td>Spatial</td>
<td>5.5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Herrera, Lorenzo, Defior, Fernandez-Smith, and Costa-Giomi (2011) - M1 - S1</td>
<td>Phonological awareness</td>
<td>Phonological processing</td>
<td>4.6</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Herrera, Lorenzo, Defior, Fernandez-Smith, and Costa-Giomi (2011) - M2 - S1</td>
<td>Naming speed</td>
<td>Phonological processing</td>
<td>4.6</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Herrera, Lorenzo, Defior, Fernandez-Smith, and Costa-Giomi (2011) - M1 - S2</td>
<td>Phonological awareness</td>
<td>Phonological processing</td>
<td>4.6</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Herrera, Lorenzo, Defior, Fernandez-Smith, and Costa-Giomi (2011) - M2 - S2</td>
<td>Naming speed</td>
<td>Phonological processing</td>
<td>4.6</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hole (2013)</td>
<td>Reading</td>
<td>Reading</td>
<td>8.9</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Hunt (2012)</td>
<td>Phonological discrimination</td>
<td>Phonological processing</td>
<td>3.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Janus, Lee, Moreno, and Bialystok (2016) - M1</td>
<td>Working Memory (verbal)</td>
<td>Memory</td>
<td>5.0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Janus, Lee, Moreno, and Bialystok (2016) - M2</td>
<td>Working Memory (spatial)</td>
<td>Memory</td>
<td>5.0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Janus, Lee, Moreno, and Bialystok (2016) - M3</td>
<td>Executive Control (verbal fluency)</td>
<td>Other</td>
<td>5.0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Janus, Lee, Moreno, and Bialystok (2016) - M4</td>
<td>Attention (sentence judgement)</td>
<td>Other</td>
<td>5.0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Janus, Lee, Moreno, and Bialystok (2016) - M5</td>
<td>Attention (visual search)</td>
<td>Other</td>
<td>5.0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Kaviani, Mirbaha, Pournaseh, and Sagan (2014) - M1</td>
<td>Intelligence (IQ)</td>
<td>Intelligence</td>
<td>5.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Kaviani, Mirbaha, Pournaseh, and Sagan (2014) - M2</td>
<td>Abstract reasoning</td>
<td>Intelligence</td>
<td>5.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Kaviani, Mirbaha, Pournaseh, and Sagan (2014) - M3</td>
<td>Verbal reasoning</td>
<td>Intelligence</td>
<td>5.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Kaviani, Mirbaha, Pournaseh, and Sagan (2014) - M4</td>
<td>Quantitative reasoning</td>
<td>Intelligence</td>
<td>5.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Kaviani, Mirbaha, Pournaseh, and Sagan (2014) - M5</td>
<td>Short-term memory</td>
<td>Memory</td>
<td>5.5</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Legette (1993) - M1</td>
<td>Mathematics</td>
<td>Mathematics</td>
<td>6.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Legette (1993) - M2</td>
<td>Reading</td>
<td>Reading</td>
<td>6.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Study</td>
<td>Domain</td>
<td>Task</td>
<td>Score</td>
<td>Include 1</td>
<td>Include 2</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>-------------------------</td>
<td>-------------------------------</td>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Lu (1986)</td>
<td>Reading</td>
<td>Reading</td>
<td>6.0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Mehr, Schachner, Katz, and Spelke (2013) - M1 - S1</td>
<td>Spatial navigation reasoning</td>
<td>Spatial</td>
<td>4.0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mehr, Schachner, Katz, and Spelke (2013) - M1 - S2</td>
<td>Spatial navigation reasoning</td>
<td>Spatial</td>
<td>4.0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Mehr, Schachner, Katz, and Spelke (2013) - M2 - S1</td>
<td>Visual form analysis</td>
<td>Spatial</td>
<td>4.0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mehr, Schachner, Katz, and Spelke (2013) - M2 - S2</td>
<td>Visual form analysis</td>
<td>Spatial</td>
<td>4.0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Mehr, Schachner, Katz, and Spelke (2013) - M3 - S1</td>
<td>Numerical discrimination</td>
<td>Mathematics</td>
<td>4.0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mehr, Schachner, Katz, and Spelke (2013) - M3 - S2</td>
<td>Numerical discrimination</td>
<td>Mathematics</td>
<td>4.0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Mehr, Schachner, Katz, and Spelke (2013) - M4 - S1</td>
<td>Receptive vocabulary</td>
<td>Literacy</td>
<td>4.0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mehr, Schachner, Katz, and Spelke (2013) - M4 - S2</td>
<td>Receptive vocabulary</td>
<td>Literacy</td>
<td>4.0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Moreno et al. (2009)</td>
<td>Reading</td>
<td>Literacy</td>
<td>8.3</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Moreno, Friesen, and Bialystok (2011) - M1</td>
<td>Rhyming</td>
<td>Phonological processing</td>
<td>5.3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Moreno, Friesen, and Bialystok (2011) - M2</td>
<td>Visual-auditory learning</td>
<td>Other</td>
<td>5.3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Moritz, Yampolsky, Papadelis, Thomson, and Wolf (2013) - M1</td>
<td>Rhyming</td>
<td>Phonological processing</td>
<td>5.6</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Moritz, Yampolsky, Papadelis, Thomson, and Wolf (2013) - M2</td>
<td>Isolation of phonemes</td>
<td>Phonological processing</td>
<td>5.6</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Myant, Armstrong, and Healy (2008) - M1</td>
<td>Alliteration</td>
<td>Phonological processing</td>
<td>4.3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Myant, Armstrong, and Healy (2008) - M2</td>
<td>Rhyming</td>
<td>Phonological processing</td>
<td>4.3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Portowitz, Lichtenstein, Egorova, and Brand (2009) - M1</td>
<td>Intelligence</td>
<td>Intelligence</td>
<td>8.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Portowitz, Lichtenstein, Egorova, and Brand (2009) - M2</td>
<td>Memory</td>
<td>Memory</td>
<td>8.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Portowitz, Peppler, and Downton (2014)</td>
<td>Working Memory (spatial)</td>
<td>Memory</td>
<td>9.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Rauscher and Zupan (2000) - M1</td>
<td>Working memory</td>
<td>Memory</td>
<td>5.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Rauscher and Zupan (2000) - M2</td>
<td>Spatial-temporal ability</td>
<td>Spatial</td>
<td>5.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Register (2004) - M1 - S1</td>
<td>Letter naming</td>
<td>Phonological processing</td>
<td>6.0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Register (2004) - M1 - S2</td>
<td>Letter naming</td>
<td>Phonological processing</td>
<td>6.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Register (2004) - M2 - S1</td>
<td>Sounds fluency</td>
<td>Phonological processing</td>
<td>6.0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Register (2004) - M2 - S2</td>
<td>Sounds fluency</td>
<td>Phonological processing</td>
<td>6.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Study</td>
<td>Task</td>
<td>Domain</td>
<td>Time</td>
<td>Training</td>
<td>Transfer</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>---------------------------</td>
<td>-----------------</td>
<td>-------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Register (2004) - M3 - S1</td>
<td>Reading Literacy</td>
<td>6.0</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Register (2004) - M3 - S2</td>
<td>Reading Literacy</td>
<td>6.0</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M1 - S1</td>
<td>Memory</td>
<td>Memory</td>
<td>12.7</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M1 - S2</td>
<td>Memory</td>
<td>Memory</td>
<td>12.7</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M2 - S1</td>
<td>Intelligence (IQ)</td>
<td>Intelligence</td>
<td>12.7</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M2 - S2</td>
<td>Intelligence (IQ)</td>
<td>Intelligence</td>
<td>12.7</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M3 - S3</td>
<td>Reading Literacy</td>
<td>Literacy</td>
<td>10.9</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M3 - S4</td>
<td>Reading Literacy</td>
<td>Literacy</td>
<td>10.9</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M4 - S3</td>
<td>Writing Literacy</td>
<td>Literacy</td>
<td>10.9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M4 - S4</td>
<td>Writing Literacy</td>
<td>Literacy</td>
<td>10.9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M5 - S3</td>
<td>Speaking Literacy</td>
<td>Literacy</td>
<td>10.9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M5 - S4</td>
<td>Speaking Literacy</td>
<td>Literacy</td>
<td>10.9</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M6 - S3</td>
<td>Spatial Memory</td>
<td>Spatial</td>
<td>10.9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M6 - S4</td>
<td>Spatial Memory</td>
<td>Spatial</td>
<td>10.9</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M7 - S3</td>
<td>Number Mathematics</td>
<td>Mathematics</td>
<td>10.9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M7 - S4</td>
<td>Number Mathematics</td>
<td>Mathematics</td>
<td>10.9</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M8 - S3</td>
<td>Structure Mathematics</td>
<td>Mathematics</td>
<td>10.9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M8 - S4</td>
<td>Structure Mathematics</td>
<td>Mathematics</td>
<td>10.9</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M9 - S3</td>
<td>Measurement Mathematics</td>
<td>Mathematics</td>
<td>10.9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M9 - S4</td>
<td>Measurement Mathematics</td>
<td>Mathematics</td>
<td>10.9</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M10 - S3</td>
<td>Mathematics Mathematics</td>
<td>Mathematics</td>
<td>10.9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rickard, Bambrick, and Gill (2012) - M10 - S4</td>
<td>Mathematics Mathematics</td>
<td>Mathematics</td>
<td>10.9</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Roden et al. (2014) - M1</td>
<td>Visual attention</td>
<td>Other</td>
<td>7.9</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Roden et al. (2014) - M2</td>
<td>Speed processing</td>
<td>Other</td>
<td>7.9</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Roden, Grube, Bongard, and Kreutz (2014) - M1</td>
<td>Working memory (visuospatial)</td>
<td>Memory</td>
<td>7.5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Roden, Grube, Bongard, and Kreutz (2014) - M2</td>
<td>Working memory (phonological)</td>
<td>Memory</td>
<td>7.5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Roden, Grube, Bongard, and Kreutz (2014) - M3</td>
<td>Working memory (CE)</td>
<td>Memory</td>
<td>7.5</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Roden, Kreutz, and Bongard (2012) - M1 - S1</td>
<td>Verbal memory</td>
<td>Memory</td>
<td>7.7</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Note. For studies with multiple samples, the result of each sample (S1, S2, etc.) is reported separately, and for studies with multiple outcome measures, the result of each measure (M1, M2, etc.) is reported separately.
When the mean age was not provided, the medium range was inserted in the model. Similarly, when the grade of the students was provided the medium range was considered (e.g., first graders, six-year-olds).
2.4 Effect Size

For the studies with an only-post-test design, the standardized means difference (Cohen’s $d$) was calculated with the following formula:

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

(1)

where $SD_{pooled}$ is the pooled standard deviation, and $M_e$ and $M_c$ are the means of the experimental group and the control group, respectively. For the studies with a repeated-measure design, the standardized means difference was calculated with the following formula:

$$d = (M_{g-e} - M_{g-c}) / SD_{pooled-pre}$$

(2)

where $SD_{pooled-pre}$ is the pooled standard deviation of the two pre-test standard deviations, and $M_{g-e}$ and $M_{g-c}$ are the gain of the experimental group and the control group, respectively (Schmidt & Hunter, 2015, p. 353).

Analogously to other recent meta-analyses (e.g., Macnamara, Hambrick, & Oswald, 2014), the effect sizes with $z$-scores greater than 3 ($n = 9$) or smaller than –3 ($n = 2$) were Winsorized to $z$-scores equal to 2.99 and –2.99, respectively. This procedure was adopted to reduce the weight of potential outliers in the analysis (Lipsey & Wilson, 2001; Schmidt & Hunter, 2015, pp. 235-236; Tukey, 1962). Finally, the Comprehensive Meta-Analysis (Version 3.0; Biostat, Englewood, NJ) software package was used for computing the effect sizes and conducting statistical analyses.

---

4 All the formulas we used were taken from Schmidt and Hunter (2015).

5 If the $t$ statistic was provided, we used the regular formula $d = t \times \sqrt{(n_1 + n_2)/(n_1 \times n_2)}$.

6 We also performed additional analyses without Winsorizing the 11 effect sizes. No significant difference was found in the overall results (for the details, see Section S1 and Table S3 in the Supplemental material available online).
2.5 Statistical Dependence of the Samples

The effect sizes were calculated for each dependent variable reported in the studies (Schmidt & Hunter, 2015). Moreover, when the study presented a comparison between one experimental group and two control groups (do-nothing and active), two effect sizes were calculated (one for each comparison with experimental and control groups; see Table 1). As this procedure violates the principle of statistical independence, the method designed by Cheung and Chan (2004) was applied to both the main and the additional models (see Sections 3.1 and 3.2). This method reduces the weight in the analysis of dependent samples by calculating an adjusted (i.e., smaller) N. Since Cheung and Chan’s (2004) method cannot be used for partially dependent samples, we ran our analyses as if the comparisons between experimental samples and two different control groups were statistically independent. However, it must be noticed that the violation of statistical independence has little or no effect on means, standard deviations, and confidence intervals (Bijmolt & Pieters, 2001; Tracz, Elmore, & Pohlmann, 1992). Thus, the entire procedure is a reliable way to deal with the statistical dependence of part of the samples. For the list of the studies and the adjusted Ns, see Table S2 in the Supplemental material available online.

3. Results

The random-effects meta-analytic overall effect size was $\bar{d} = 0.16$, CI [0.09; 0.22], $k = 118$, $p < .001$. The forest plot is shown in Figure 1. The degree of heterogeneity (Borenstein, Hedges, Higgins, & Rothstein, 2009) between effect sizes was $I^2 = 46.94$, suggesting that some moderators had a potential effect.\(^7\)

\(\text{}\)

\textit{Insert Figure 1 around here}

\(\text{}\)

\(^7\) A degree of heterogeneity ($I^2$) around 50.00 is considered moderate, around 25.00 low, and around 75.00 high.
3.1 Meta-Regression Analysis

A meta-regression model including all the four moderators was run. The model fitted the data significantly, $Q(9) = 49.06$, $R^2 = .65$, $p < .001$. Age was not a significant moderator, $p = .944$. The statistically significant moderators were Outcome measure, $Q(6) = 21.78$, $p = .001$, Random allocation, $b = –0.16$, $p = .010$, and Presence of active control group, $b = –0.25$, $p < .001$. The last two moderators show that studies with random allocation of participants and studies comparing music treatment to another activity (active control group) tended to have weaker effect sizes. The overall effect sizes in randomized and non-randomized samples were $\bar{d} = 0.09$, CI $[–0.01; 0.18]$, $k = 57$, $p = .084$, and $\bar{d} = 0.23$, CI $[0.14; 0.31]$, $k = 61$, $p < .001$, respectively. The overall effect sizes when music training was compared to active control and do-nothing control groups were $\bar{d} = 0.03$, CI $[–0.07; 0.12]$, $k = 54$, $p = .562$, and $\bar{d} = 0.25$, CI $[0.17; 0.34]$, $k = 64$, $p < .001$, respectively. Finally, the overall effect size in randomized samples with active control groups was $\bar{d} = –0.12$, CI $[–0.27; 0.03]$, $k = 22$, $p = .113$, while the overall effect size in the non-randomized samples without active control group was $\bar{d} = 0.33$, CI $[0.23; 0.44]$, $k = 29$, $p < .001$.

3.2 Additional Meta-Analytic Models

Since Outcome measure was a significant moderator, we calculated the random-effects meta-analytic overall effect size of each of the seven measures, in order to investigate whether any measure showed an overall effect size appreciably larger (or smaller) than the others. The overall effect sizes are summarized in Table 2.

Table 2

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Effect size ($\bar{d}$)</th>
<th>95% CI</th>
<th>$k$</th>
<th>$p$-value</th>
</tr>
</thead>
</table>

Overall effect sizes, confidence intervals, ks, and $p$-values in each outcome measure.
The meta-regression analysis showed that only memory- and intelligence-related overall effect sizes were significantly different compared to the other measures ($b = 0.26, p = .041$, and $b = 0.30, p = .029$, respectively).

### 3.3 Publication Bias Analysis

Begg and Mazumdar’s (1994) rank correlation test showed no evidence of publication bias ($p = .433$, one-tailed). In addition, to test the robustness of results (Kepes & McDaniel, 2015), we ran a $p$-curve analysis for the detection of publication bias (Simonsohn, Nelson, & Simmons, 2014). We selected the $p$s according the following two rules: (a) only positive results (i.e., $z > 0$) were considered; and (b) to avoid redundancy, only one $p < .01$ per study was inserted into the analysis. The results had evidential value (i.e., no evidence of publication bias) because we found more low $p$-values ($p < .01$) than high $p$-values ($0.01 < p < .05$), $z(14) = -4.24, p < .001$ (Figure 2).

*Insert Figure 2 around here*

Finally, Duval and Tweedie’s (2000) method found no publication bias in any of the seven models (i.e., no studies trimmed left of the mean).
3.4 Sensitivity Analysis

Since Rickard et al.’s (2012) study reported a large number of effect sizes \((k = 20)\), we conducted a sensitivity analysis by excluding those effect sizes from all the models. The random-effects meta-analytic overall effect size was still modest, \(\bar{d} = 0.20, \text{CI } [0.14; 0.27], k = 98, p < .001\). The degree of heterogeneity between effect sizes was \(I^2 = 39.31\), suggesting that some moderators had a potential effect. For the list of the studies and the adjusted \(N_s\), see Table S4 in the Supplemental material available online.

A meta-regression model including all the four moderators was run. The model fitted the data significantly, \(Q(9) = 36.94, R^2 = .74, p < .001\). The only two statistically significant moderators were Outcome measure, \(Q(6) = 20.16, p = .003\) and Presence of active control group, \(b = -0.17, p = .0014\). The overall effect sizes when music training was compared to do-nothing control and active control groups were \(\bar{d} = 0.28, \text{CI } [0.19; 0.36], k = 56, p < .001\), and \(\bar{d} = 0.08, \text{CI } [-0.03; 0.19], k = 42, p = .139\), respectively. Compared to Table 2, no significant difference was found in six of the seven Outcome measure-related overall effect sizes. The only exception was Mathematics \((\bar{d} = 0.35 \text{ vs } \bar{d} = 0.17; \text{Table } 3)\).

Table 3

Overall effect sizes, confidence intervals, ks, and \(p\)-values in each outcome measure.

<table>
<thead>
<tr>
<th>Outcome measure</th>
<th>Effect size ((\bar{d}))</th>
<th>95% CI</th>
<th>(k)</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literacy</td>
<td>0.07</td>
<td>[-0.07; 0.21]</td>
<td>16</td>
<td>.307</td>
</tr>
<tr>
<td>Mathematics</td>
<td>0.35</td>
<td>[0.16; 0.54]</td>
<td>7</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Memory</td>
<td>0.39</td>
<td>[0.25; 0.54]</td>
<td>16</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Intelligence</td>
<td>0.37</td>
<td>[0.21; 0.53]</td>
<td>11</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
4. Discussion

The present meta-analysis aimed to test the hypothesis that music training improves children’s and young adolescents’ cognitive and academic skills, and to evaluate the potential role of moderating variables. Along with a small overall effect size ($d = 0.16$, CI [0.09; 0.22]), which indicates that far-transfer from music to non-music skills was limited, the results showed a slightly greater positive effect of music training on some of the cognitive skills (i.e., intelligence and memory) and a non-significant effect on all the academic skills. Moreover, the design quality of the studies significantly affected the magnitude of the effects. A similar pattern of results was obtained in the sensitivity analysis model.

We did not correct for attenuation due to measurement error because only about half of the studies provided reliability coefficients. However, correcting for measurement error would not significantly affect the effect sizes. For example, if we assume that the reliability coefficients are between .80 and .90, then the corrected estimate of the overall effect size of the main model (i.e., $d = 0.16$) would be between 0.17 and 0.18, a difference of only 0.01 or 0.02 standard deviations.

4.1 Substantive Results

The outcomes of the present meta-analysis allow us to draw some important conclusions. First, the small overall effect size upholds Thorndike and Woodworth’s (1901) common-element theory. In line with previous research (Donovan et al., 1999; Sala & Gobet,
far-transfer from music to other cognitive or academic abilities seems to be small or null. Second, music training appears to moderately foster intelligence- and memory-related outcomes. However, no significant effect on academic skills was found (literacy, $d = -0.07$, CI $[-0.23; 0.09]$, $p = .386$; mathematics, $d = 0.17$, CI $[-0.02; 0.36]$, $p = .085$). This outcome suggests that improvements in memory and intelligence do not generalize to academic skills. Alternatively, and more likely, the observed positive effects of music training in intelligence- and memory-related outcomes are due to confounding variables (we will take up this point below). Either way, the hypothesis according to which the multiple benefits of music training, including academic benefits, stem from an improvement in general intelligence (or overall cognitive skill) is not corroborated. Third, the age of the participants is not a statistically significant moderator. Fourth, the meta-regression model accounts for a large proportion of the variance ($R^2 = .65$) between the effect sizes. The latter result implies that the statistically significant moderators explain, to a large extent, why the research on the effects of music training on children’s and young adolescents’ skills has produced mixed results up to now.

4.2 Methodological Results

The meta-regression analysis shows that both methodological moderators (i.e., random allocation of participants to the treatment groups and comparison to an active control group) affected the effect sizes. In other words, the better the design quality, the smaller the effect sizes. This outcome lends further support to the idea that the observed positive effects, when any, of music training on non-music-related outcomes, are probably due to confounding variables, such as placebo effects and lack of random allocation of participants.

Unfortunately, this conclusion seems to apply to memory- and intelligence-related effect sizes too. In fact, despite the greater overall effect sizes in these two outcome measures ($d = 0.34$, CI $[0.20; 0.48]$ and $d = 0.35$, CI $[0.21; 0.49]$, respectively), the reliability of these
positive results seems questionable. Only one study (Schellenberg, 2004) tested the effect of music training on children’s intelligence using a rigorous experimental design (i.e., random allocation of participants and active control group), and the effect was found to be modest ($d = 0.16$). Concerning the memory-related outcomes, none of the reviewed studies adopted such a design. Furthermore, as pointed out above, a genuine – i.e., not due to confounding variables – improvement in such critical cognitive skills should leave a trace in students’ academic skills, at least to some degree.

The sensitivity analysis (Section 3.4) showed that when Rickard et al.’s (2012) study and all its effect sizes were excluded, the overall effect size in mathematics became significantly positive. However, the only study comparing a music training group to an active control group and with random allocation of the participants to the groups – i.e., Mehr et al. (2013) – found a negative effect size ($d = -0.25$). These considerations uphold the conclusion that music training does not substantially enhance any non-music-related cognitive skill.

**4.3 Conclusions and Recommendations for Future Research**

The results of this meta-analysis fail to support the hypothesis that music skill transfers to cognitive or academic skills in the general population of children and young adolescents. Together with previous findings in psychology and education, these results suggest a sobering conclusion: when the potential occurrence of far-transfer is tested rigorously, the results are often, if not always, disappointing. Thus, this study lends further support to the hypothesis according to which far-transfer rarely occurs. Even when music training appears to foster some of the participants’ cognitive skills (intelligence and memory), the reliability of the results is doubtful. In fact, only one study investigated, with a proper design, the effects exerted by music training on the participants’ intelligence- and memory-related skills.
Due to the lack of well-designed studies, the question of whether music training enhances children’s and young adolescents’ intelligence- and memory-related skills is still unanswered. For this reason, future studies should strive for proper designs that include both random allocation of the participants and an active control group. Furthermore, future investigations should evaluate the effects of music training on both cognitive (especially intelligence and memory) and academic skills. Such a design makes it possible to empirically assess whether the potential benefits of music training on youngsters’ cognitive skills generalize to academic performance. Nonetheless, considering the previous unsatisfactory outcomes and the scarcity of far-transfer in the literature, it is our opinion that future experiments will show results in line with those presented in this meta-analysis.

5. References

References marked with an asterisk indicate studies included in the meta-analysis.


and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455-463.
doi:10.1111/j.0006-341X.2000.00455.x


doi:10.1525/MP.2008.25.4.383


doi:10.3389/fpsyg.2014.01337

doi:10.3389/fpsyg.2015.01777


Education Inspired by the Kodály Concept. New York, NY: Oxford University Press.


doi:10.1525/mp.2011.29.2.165


*Yazejian, N., & Peisner-Feinberg, E. S. Effects of a preschool music and movement curriculum on children's language skills. *NHSA Dialog, 12*, 327-341.

doi:10.1080/15240750903075255
Figures Captions

*Figure 1.* Overall effect size ($\bar{d}$) for music training groups compared to control groups. Cohen’s $d$s (circles) and 95% CIs (lines) are shown for all the effects entered into the meta-analysis. The diamond at the bottom indicates the meta-analytically weighted mean $\bar{d}$. When studies had multiple samples, the table reports the result of each sample (S1, S2, etc.) separately. Similarly, when studies used multiple outcome measures, the table reports the result of each measure (M1, M2, etc.) separately. Asterisks indicate adjusted (Winsorized) outliers.

*Figure 2.* $p$-curve analysis of the studies reporting significantly positive results. The blue line shows that most of the significant $p$-values are smaller than .01.
Study name  Std diff in means and 95% CI
Rickard, Bambrick, and Gill (2012) - M9 - S3
Rickard, Bambrick, and Gill (2012) - M10 - S3
Rickard, Bambrick, and Gill (2012) - M5 - S4
Rickard, Bambrick, and Gill (2012) - M8 - S4
Rickard, Bambrick, and Gill (2012) - M3 - S4
Rickard, Bambrick, and Gill (2012) - M9 - S4
Mehr, Schachner, Katz, and Spelke (2013) - M4 - S3
Mehr, Schachner, Katz, and Spelke (2013) - M4 - S1
Yazejian and Peisner-Feinberg (2009) - M4
Bhide, Power, and Goswami (2013) - M4
Hanson (2001) - M1 - S2
Mehr, Schachner, Katz, and Spelke (2013) - M2 - S2
Herrera, Lorenzo, Defior, Fernandez-Smith, and Costa-Giomi (2011) - M1 - S1
Register (2004) - M3 - S2
Myant, Armstrong, and Healy (2008) - M2
Hanson (2001) - M2 - S2
Roden, Kreutz, and Bongard (2012) - M2 - S1
Roden, Kreutz, and Bongard (2012) - M1 - S2
Roden, Kreutz, and Bongard (2012) - M2 - S2
Roden et al. (2014) - M2
Roden, Kreutz, and Bongard (2012) - M1 - S1
Roden, Kreutz, and Bongard (2012) - M2 - S2
Roden, Kreutz, and Bongard (2012) - M1 - S2
Roden, Kreutz, and Bongard (2012) - M1 - S1
Hunt (2012)
Mehr, Schachner, Katz, and Spelke (2013) - M2 - S1
Portowitz, Lichtenstein, Egorova, and Brand (2009) - M2
Roden, Kreutz, and Bongard (2012) - M3
Roden, Kreutz, and Bongard (2012) - M1 - S1
Roden, Grube, Bongard, and Kreutz (2014) - M2
Roden, Grube, Bongard, and Kreutz (2014) - M1
Roden, Grube, Bongard, and Kreutz (2014) - M3
Schellenberg (2004) - S2
Degé and Schwarzer (2011) - S2
Degé, Wehrum, Stark, and Schwarzer (2011) - M2
Kaviani, Mirbaha, Pournaseh, and Sagan (2014) - M5
Cardarelli (2003) - M2
Cardarelli (2003) - M1
Geoghegan and Mitchelmore (1996)
Degé, Wehrum, Stark, and Schwarzer (2011) - M2
Portowitz, Lichtenstein, Egorova, and Brand (2009) - M1
Roden, Grube, Bongard, and Kreutz (2014) - M1
Roden, Grube, Bongard, and Kreutz (2014) - M3
Cardarelli (2003) - M2
Rauscher and Zupan (2000) - M2
Cardarelli (1998)
Kaviani, Mirbaha, Pournaseh, and Sagan (2014) - M1
Roden, Grube, Bongard, and Kreutz (2014) - M1
Roden, Grube, Bongard, and Kreutz (2014) - M3
Cardarelli (2003) - M2
Gromko and Poorman (1998)
Costa-Giomi (2004) - M1
Portowitz, Lichtenstein, Egorova, and Brand (2009) - M2
Rauscher and Zupan (2000) - M1
Rauscher and Zupan (2000) - M2
Bilhartz, Bruhn, and Olson (1999)
Register (2004) - M2 - S1
Register (2004) - M1 - S2
Register (2004) - M2 - S2
Register (2004) - M1 - S1
Hanson (2001) - M1 - S2
Hanson (2001) - M1 - S1
Hanson (2001) - M2 - S2
Hanson (2001) - M3 - S1
Hanson (2001) - M3 - S2
Hanson (2001) - M2 - S1
Costa-Giomi (2004) - M1
Hanson (2001) - M1 - S1
Note: The observed p-curve includes 14 statistically significant (p < .05) results, of which 13 are p < .025. There were no non-significant results entered.
Highlights

- Music training is thought to improve youngsters’ cognitive and academic skills.
- Results show a small overall effect size ($\bar{d} = 0.16, K = 118$).
- Music training seems to moderately enhance youngsters’ intelligence and memory.
- The design quality of the studies is negatively related to the size of the effects.
- Future studies should include random assignment and active control groups.