Movement-based music in the classroom: Investigating the effects of music programs incorporating body movement in primary school children

Borbála Lukács[1], Kata Asztalos[2], Emese Maróti[3], Tamara Farnadi[4], Gabriella Deszpot[2], Borbála Szirányi[2], László Norbert Nemes[2], Ferenc Honbolygó[3,5]

[1] Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary
[2] Kodály Institute of the Liszt Ferenc Academy of Music, Budapest, Hungary
[3] Brain Imaging Centre, Research Centre for Natural Sciences, Budapest, Hungary
[4] János Richter Secondary School of Music, Győr, Hungary
[5] Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

Author Note

This work was supported by the Content Pedagogy Research Program of the Hungarian Academy of Sciences (SZ-009/2016) and the János Bolyai Research Fellowship of the Hungarian Academy of Sciences awarded to F.H. (BO/00273/17). We have no conflicts of interest to disclose. We are grateful to all children and parents for their participation in this study and the primary school teachers for their support. We wish to thank Edina Barabás, Sára Buda, and Katalin Körtvési for their help in developing and implementing the active music educational models. In addition, we thank Tamara Biró, Dorottya Denke, Bernadett Ferencz, Borbála German, Márta Lőrincz-Fazekas, Sarolt Pintér, and Petra Prjevara for their help in data collection.

Movement-based music in the classroom: Investigating the effects of music programs incorporating body movement in primary school children

Abstract

The present studies investigated the impacts of the different implementation of body movement into classroom-based comprehensive music programs on the development of music-related and non-musical abilities in Hungarian primary school children. In Study 1, science-focused classes received Kodály music lessons completed with teacher-directed movements or no movement activities. In Study 2, intensive music classes participated in Kodály music lessons combined either with teacher-directed or improvised movement elements. From the beginning of schooling, participants were measured three times over one and a half years for musical abilities, sensorimotor entrainment, phonemic awareness, rapid naming, reading, executive functions, and IQ. Results revealed distinct developmental trajectories for melody discrimination, phonemic awareness, and verbal IQ in the science classes; however, the classes' comparable performance at the first and last measurements indicated that their overall growth was similar. Moreover, performance of the intense music classes was comparable at the end of the second school year even though the class with the music curriculum using teacher-directed movements showed greater improvements in rapid naming of pictures and verbal IQ. These findings suggest that in the early school years, diverse movement-based music programs provided in classroom settings supported musical, sensorimotor entrainment, early literacy, and cognitive development similarly.

Keywords: classroom music lessons, body movement, musical abilities and entrainment, early literacy skills, cognitive development

Introduction

A large body of research indicates that active engagement with music during childhood can facilitate the development of a wide range of auditory and cognitive functions. Participation in music instruction during the early primary school years have been associated with improvements in music audition (Ilari et al., 2016; Roden, Könen, et al., 2014) and advantages in rhythmic synchronization skills (Ilari et al., 2016). Moreover, compared to children receiving non-musical training or no additional training, school-aged children taking music lessons often demonstrate greater growth even for some components of non-musical skills, such as reading (Hurwitz et al., 1975; Rautenberg, 2015), memory (Rickard et al., 2010; Roden et al., 2012), executive functions (Frischen et al., 2021; Jaschke et al., 2018; Roden, Grube, et al., 2014), and intelligence (Schellenberg, 2004). Most of the findings, however, have been obtained from studies investigating the effects of out-of-school music training with specialized music programs, indicating specific and highly inconsistent developmental gains (for recent meta-analyses, see Cooper, 2019; Sala & Gobet, 2020). Furthermore, whether comprehensive music educational programs delivered in general learning contexts could support broad skill development is underrepresented in the literature. Therefore, here we provide an extensive assessment conducted in Hungary regarding the influence of school-based music instruction programs lasting for one and a half years on the development of 6−7-year-old children’s musical abilities, rhythmic entrainment, phonological processing, reading skills, executive functions, and intelligence.

The relevance of studying the impacts of school music education originates in the context of learning itself. School music instruction covers structured group music programs that are implemented into the general school curriculum, providing children with music lessons for free during the school hours. With respect to school music instruction, Hungary has a specific position as the homeland of the music pedagogical concepts of Zoltán Kodály.

The approach of Kodály, music composer and educator, represents a comprehensive music educational concept which is centered on the sequential development of music culture and literacy from the early childhood years, placing primary emphasis on singing and folk song heritage in teaching the basics of music, training listening skills, and thereby acquiring music comprehension (Dobszay, 1972). Following the concept of Kodály, which aims at making music attainable to and appreciated by all children (Choksy, 1999), music education has been a part of the Hungarian National Core Curriculum in primary schools since 1995. The curriculum determines the framework of compulsory music education for the first to the tenth grades at public schools. Accordingly, all Hungarian students receive structured music instruction in their classrooms from school entry twice a week in lower grades, regardless of the residence and socioeconomic background of their families. In addition, several schools in Hungary (so-called music primary schools) offer classes with a specialized music curriculum including three or four music lessons per week with optional choir lessons (Hejja & Szalai, 2020). Instrumental education, however, is not incorporated in the curriculum but typically provided by art schools, requiring a tuition fee to be paid by families.

Considering the practical benefits of music learning in the classroom, a few studies have examined the impacts of various in-school music education programs in children starting primary school. In the pioneering study by Hurwitz et al. (1975), first-grade children receiving daily school music lessons for seven months based on the comprehensive Kodály curriculum outperformed children not receiving Kodály lessons on various tests measuring spatial-temporal abilities and reading achievement. A few Hungarian studies investigated the impacts of school music instruction in the 1970s (Barkóczi & Pléh, 1977; Kokas, 1972) and the 1980s (Laczó, 1985, 1987), revealing that students in intense music classes showed greater improvements in musical abilities and intelligence compared to students in classes with weekly music lessons. Other studies conducted with 7–8-year-old German children showed that unlike the groups that underwent enhanced natural science instruction or no additional lessons in the school, the group that participated in weekly school instrumental lessons for 18 months improved more in the measures of verbal memory (Roden et al., 2012), the tests of the phonological loop and central executive components of working memory (Roden, Grube, et al., 2014), and the rhythmic and tonal aspects of music audiation (Roden, Könen, et al., 2014). In a recent study by Jaschke et al. (2018), 6-year-old Dutch children engaging in classroom-based comprehensive music instruction with instrumental elements for 2.5 years demonstrated greater enhancements on the test of verbal intelligence and specifically on the measures of inhibition and planning than children in the visual arts group or children in the no-arts control group.

These studies revealed that classroom-based music lessons could improve specific but not all components of musical abilities and cognitive functions in the early primary school years. Despite these promising findings, we cannot draw an overall conclusion regarding the efficacy of music instruction due to their differences in the focus areas and length of research, the content of the curricula, and the size of groups applied. The lack of explicit differentiation between the effects of diverse types of music programs targeting the training of different domains in various contexts critically challenge the generalization of the results in the field (for a framework, see Holochwost et al., 2021).

Another issue that has not yet been clarified regarding the efficacy of comprehensive music education programs is whether the specific involvement of movement in musical activities could contribute to developmental advantages. Movement is often regarded as an integral part of musical experience. Bodily responses to music seem to occur instinctively (Zelechowska, 2020), being typically expressed in the spontaneous synchronization of movements with the rhythm of music (Repp & Su, 2013) and with those of other individuals (Knoblich et al., 2011). Research has indicated a shared neural network underlying auditory perception and the perception and production of movement (Gazzola et al., 2006) and the engagement of the vestibular system in the interaction between auditory and motor system, mediating rhythm perception (Todd, 1999). Moreover, there is empirical evidence demonstrating that motor and cognitive functions have some common neural bases and similarly protracted developmental trajectories (Diamond, 2000). Particularly during the early school years, performance on tasks requiring complex motor skills and higher order cognitive abilities show strong associations (van der Fels et al., 2015). Based on the links of the auditory, motor, and cognitive domains, it is possible that music educational programs involving body movement may facilitate further growth in both music-related abilities and non-musical cognitive functions.

Body movement forms have otherwise been frequently incorporated as supplementary components accompanying musical activities in comprehensive music education programs. Most commonly, various fine and gross body movements, such as clapping, stepping, body percussions, drumming, tapping, are deliberately applied in alignment with the rhythmic, dynamic, or pitch-related changes of music. Alternatively, improvised movement using the whole body is encouraged, which allows free movement experiments to the properties of music, often resulting in dance-like expressions. It seems that practice may improve both directed and improvised movement responses to music. Training the rhythmic coordination of perception and action is linked to enhanced timing precision of bodily movements (e.g., Janzen et al., 2014; Repp, 2010), whereas the exploration for original and adaptive movement variants under different task constraints promotes motor creativity (Orth et al., 2017). Furthermore, there is evidence suggesting that synchronized movements and motor creativity utilizes shared cognitive resources, both relying heavily on attentional processes and working memory (e.g., de Dreu et al., 2012; Moraru et al., 2016; Tierney & Kraus, 2013). It is possible, therefore, that integrating either directed or improvised movements in music education may have additional effects on the development of general cognitive abilities, such as executive functions.

Nevertheless, only a few studies have focused on investigating the effects of school music lessons that emphasized the use of movement in musical activities. For instance, Lewis (1988) studied how various movement activities embedded in classroom music instruction affected first graders' progress in music listening skills compared to the music curriculum including no movement elements. After 12 sessions of music lessons, performance on the dynamics perception subtest but not on the melodic, tempo, meter, and rhythm perception subtests was enhanced in the group with the combined music-movement curriculum. Recently, as a pilot for the current studies, Maróti et al. (2019) estimated over eight months how two movement-based curricula completing the Kodály curriculum either with fixed or free movement components impacted 6–7-year-old Hungarian children's musical, sensorimotor, literacy, and cognitive development in comparison to the solfeggio-based Kodály curriculum. Results revealed that both movement-based classes improved more than the Kodály class in pitch discrimination, phonological skills, and working memory. However, the Kodály music class exhibited larger enhancements in executive functions compared to the movement-based classes. Overall, these findings indicate that the inclusion and diverse application of movement in classroom musical activities might have specific developmental impacts in school-aged children.

In the current studies, we conducted an extensive evaluation on how comprehensive music instruction programs combined with movement elements influence 6- to 7-year-old Hungarian children's development in the school context. From the beginning of schooling, we measured participants' progress three times over one and a half years in musical abilities (perception, auditory-visual connection), rhythmic entrainment, phonological processing skills (phonemic awareness, rapid naming), reading, executive functions, and IQ, for which the beneficial effects of school music instruction had been reported previously. To estimate the role of movement in promoting improvements, two studies were designed with children from naturalistic school classes, all receiving music instruction as part of the school curriculum. In Study 1, we examined the additional effects of the inclusion of movement in music lessons by comparing children's developmental trajectories in the class receiving the traditional Kodály curriculum to the class receiving Kodály music lessons completed with teacher-determined movement elements. Based on findings of Maróti et al (2019), we predicted that the class with the movement-based music curriculum would exhibit more pronounced improvements specifically in pitch discrimination, phonemic awareness, and working memory capacity, and presumably in sensorimotor entrainment. In Study 2, we explored whether the different application of movement elements would result in distinct developmental courses of children in the classes following the Kodály curriculum combined either with teacher-directed or improvised movement elements. Since the repetitive use of fixed movement elements and choreographies might more explicitly reinforce perception-action couplings than spontaneous movements, we anticipated greater growth of entrainment skills in the class that engaged in the intense music curriculum applying teacher-directed movements. Moreover, we made the exploratory hypothesis that enhanced sensorimotor entrainment might be associated with higher levels of executive functions (Miendlarzewska & Trost, 2014).

General Method

Studies were conducted in accordance with the Declaration of Helsinki. Ethical approval for this study was obtained from the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary (approval number: 2016/062). Parents provided written informed consent and children provided verbal assent before enrolment. Children received small toys for their participation in the studies.

Participants

All participants started taking part in classroom music instruction at the beginning of schooling. Regardless of the music curricula applied, professionally trained music educators delivered music lessons to whole school classes, involving groups of 20–30 children. Participants received music lessons for 63 weeks (excluding holidays) until the third assessments were carried out at the end of the second school year.

Measures

Potential confounding variables, such as age, gender, socio-economic status (SES), and children's musical background were measured through a background questionnaire. To estimate children's development during the first and the second school year, three assessments were carried out using the same measures. To evaluate children's musical abilities, we applied discrimination and auditory-visual connection subtests. Sensorimotor entrainment skills were assessed using paced tapping and continuation tapping tasks. Early literacy skills were examined with reading, phonemic awareness, and rapid automatized naming tests. Verbal (working) memory and verbal fluency were measured as indicators of executive functions. We administered one verbal and one nonverbal subtest to assess children's IQ.

Background questionnaire

Parents completed a questionnaire asking about the families' socio-economic background and their children's age, gender, and formal musical experience at the beginning of the study. Socio-economic status (SES) was assessed by maternal educational achievement and family income. *Mother's highest level of education* was rated on a 9-point scale (1 = No graduation, 2 = Primary education, 3 = Technical school, 4 = Vocational high school, 5 = Matriculation standard, 6 = Post-secondary tertiary education, professional qualification, 7 = Bachelor degree, 8 = Master degree, 9 = Doctoral degree). *Family income* was recorded on a 5-point scale (1 = less than 100 000 HUF; 2 = 100 000–150 000 HUF; 3 = 150 000–200 000

HUF; 4 = 200 000–250 000 HUF; 5 = more than 250 000 HUF), indicating monthly earnings per capita. *Participants' musical background* was determined by the duration (measured in months) of formal music training children participated in before school instruction. Because the parents of 8 participants did not send back the questionnaire, we used data of only 32 children regarding musical background and SES in statistical analyses.

Musical abilities

We measured musical abilities using an online battery based on the tests of Asztalos and Csapó (2017). The battery included six subtests: five subtests assessed musical discrimination and one assessed auditory-visual connection. Discrimination subtests evaluated children's ability to memorize and differentiate musical patterns and required children to decide whether the two successive musical stimuli were identical or different. In the *Melody Discrimination* subtest, the changed note was one second or third higher or lower than the original one, preserving the tonality of the original melody. Each melody contained two or four bars with simple rhythmic patterns (quarters, eighth notes, quarter rests). The *Pitch Discrimination* subtest presented two musical notes consecutively with a potential pitch difference of one semitone. The *Rhythm Discrimination* subtest comprised three- or four-bar-long rhythmic stimuli comprising simple (quarter, eighth notes, quarter rests) and complex rhythmic patterns (syncopation, sixteens, dotted quarters, triplets). In case of difference, the number of the notes remained unchanged. Items in *Harmony Discrimination* were chords of three notes. The successive triads could differ only in one note, and the interval difference between the initial and altered note was either a minor or a major second. In the *Tempo Discrimination* subtest, two successive melodies were played in the same tempo or the second one was slower or faster than the first one. The tempo values were between 80 and 130 bpm (beats per minute), the smallest difference was 10 bpm and the highest was 30 bpm between two consecutive musical samples.

The auditory-visual connection subtest measured the ability to associate musical auditory (melodic and rhythmic) patterns with their visual representations. In the case of *Melody Connection*, short melodies were played, and the task was to choose one out of the three pictures that represented the contour of the melody heard. For *Rhythm Connection*, short rhythmic patterns were played. In the pictures, quarter notes were depicted by big drums, whereas eighth notes were depicted by small drums. The task was to choose one out of the three pictures of drums corresponding to the rhythmic pattern heard.

Data collection was carried out via the eDia (Electronic Diagnostic Assessment; Csapó & Molnár, 2019) system providing an easy-to-use online platform for children to complete the test on the computer at their own pace. Instructions, short auditory explanations for musical terms, and test stimuli were presented via headphones. For discrimination subtests, children had to click on the green check mark in case the stimuli were identical or the red cross in case the stimuli were different. Each discrimination subtest included 15 items, and the auditory-visual connection subtest comprised 10 items related to melody and five items related to rhythm, resulting in a total of 90 test items. We calculated an accuracy score for each subtest separately based on the number of correct answers. As the two connection tasks measured different aspects of auditory-visual integration, Melody Connection and Rhythm Connection were included as separate variables in statistical analyses.

Rhythmic entrainment

Rhythmic synchronization skills were assessed using two tapping tasks described by Maróti et al. (2019) and designed after Tierney and Kraus (2013). Children had to tap along with the tempo given by a metronome (*Paced Tapping*) and had to continue tapping at the same tempo after the metronome stopped (*Continuation Tapping*). The metronome was presented as a click sound via headphones. The test started with three practice trials with different tempi to ensure that children understood the instructions given by the experimenter.

Participants were asked to start tapping on the spacebar using the computer keyboard as soon as they heard the metronome. Trials started with 20 metronome clicks to habituate the child to the tempo, and another 20 clicks were presented immediately as test stimuli. Trials ended with the continuation phase which terminated when the child produced 20 taps in the absence of the metronome. Each tapping task ran for three trials with 2 Hz, 2.5 Hz, and 1.5 Hz tempi, respectively. Data collection was carried out using MATLAB (R2015a; The MathWorks Inc., 2015). Tapping accuracy was measured in terms of tapping variability: we calculated the standard deviation of the inter-tap intervals for each tempo, which were averaged for the paced and continuation trials separately.

Early literacy skills

Reading, Phoneme Deletion, and Rapid Automatized Naming (RAN) subtests from the Hungarian version of the 3DM-H (Dyslexia Differential Diagnosis Maastricht; Blomert & Vaessen, 2009; Hungarian version; Tóth et al., 2014) were recorded to measure early literacy skills. The *Reading* subtest presented children blocks of frequent words and asked them to read aloud accurately as many words as possible in 30 seconds. Instructions were presented on the screen and heard via headphones simultaneously. To be able to assess early reading skills at the beginning of schooling, we revised and shortened the original subtest. At baseline (T0) and at the end of the first school year (T1), five practice words were presented to investigate whether the child was already familiar with letters. We administered the subtest only if the participant was able to read the one-syllable words correctly. After the practice phase, two blocks of 15 items were presented with CVC (C = consonant, V = vowel) words in the first block and CVCC/CCVC words in the second block. At the end of the second school year (T2), all participants were tested using the original subtest from the 3DM-H battery, measuring high-frequent word reading. The structure of the first two blocks was identical with the structure of the shortened test. Additionally, the third block of CVCVC words, the fourth block of complex 2-syllable words, and the fifth block of 3-syllable words were included in the original test. For each measurement point, a reading fluency score was calculated based on the number of letters read correctly in a second.

Phoneme Deletion subtest was applied to measure phonemic awareness. Children were asked to delete a specified speech sound from one-syllable pseudo-words, then to pronounce the remaining sound sequence. Like in the reading task, a revised and shortened form of the original subtest was used at T0 and T1. First, we presented two practice items to ensure that the child understood the instruction properly. We administered the subtest if the child performed the practice trials correctly. At T0 and T1, test items were simple CVC pseudo-words (e.g., ‘cák’ without ‘k’ [=cá]) and participants had to delete a phoneme from the beginning or the end of each pseudo-word. During the testing, a total of 4 items were presented. At T2, three blocks of trials were presented with items varying in complexity. In the first and second blocks, children were required to delete the initial or the last phoneme from CVC and CCVC/CVCC pseudo-words, whereas they had to delete an inner phoneme from more complex one-syllable pseudo-words in the third block. The subtest comprised 27 items. Participants heard the instructions and the test items via headphones. For each measurement point, we calculated the accuracy score reflecting the proportion of correct answers, which were transformed and weighted by the parameters of hierarchic IRT models; thus, the mean of the corrected accuracy scores was 0 and the standard deviation was 1.

RAN tasks were used to measure the effectiveness of visual-verbal integration mechanisms and phonological lexical retrieval. We registered two RAN tasks, requiring children to sequentially name digits (e.g., 1, 4, 5, 6, 8) and pictures of simple objects (e.g., fish, chair, pear, scissors, dog) as quickly and accurately as possible. During the practice phase, five items were displayed to ensure that children used the correct name of the items. Stimuli were arranged in 3×5 matrices and presented in four blocks (two blocks for each item type) within which the items occurred in a pseudo-random order. Instructions were shown on the screen and heard via headphones simultaneously. A speed score was defined for the two subtests separately by averaging the number of items named correctly in a second.

Executive functions

Three tests were registered to measure executive functions and the executive aspects of verbal abilities. The test of *Verbal Fluency* (Mészáros et al., 2011) assessed a set of cognitive skills: organized search for task-relevant items, generation and use of strategies, inhibition of task-irrelevant items, updating information, flexible shifting between conditions, as well as verbal abilities, such as lexical access, verbal concept formation, and expressive language ability. The test required children to generate as many items within a specific category as possible in 60 seconds. Phonemic fluency was measured in three tasks asking children to list words starting with a specified phoneme (*k*, *t*, and *s*, respectively). In two tasks, semantic fluency was tested, and children had to name first as many animals, then as many fruits as they could. We calculated fluency scores based on the total number of correct words listed for each condition, excluding repetitions and out-of-category items from the calculation. By summing the scores from the five tasks, a composite fluency score was created.

We applied two tests to assess children's working memory (WM). *Digit Span* subtest of the WISC-IV (Hungarian adaptation; Nagyné Réz, Lányiné Engelmayer, Kuncz, Mészáros, & Mlinkó, 2008; Wechsler, 2003) was used to measure verbal short-term and working memory. Children were required to repeat sequences of digits in forward and backward order. We calculated a span score by summing the longest correctly repeated sequence in the forward and the backward condition and used the raw scores in analyses to follow the developmental tendencies within the same age group. *Counting Span* test (Case et al., 1982) was used as a complex WM task to measure information processing, storage, and rehearsal concurrently. Pictures were presented as stimulus sets on a computer screen containing blue circles as target stimuli as well as yellow circles and blue squares as distractor stimuli. The number of blue circles depicted in a picture ranged between two and eight. Three blocks were presented during the test, each containing five sequences of pictures. The first sequence in a block comprised two pictures and the number of pictures increased with one in the subsequent trials. The last trial in each block was the fifth sequence containing six pictures. Children were asked to count the blue circles aloud, then to repeat the final count. At the end of a picture sequence, children had to recall the final counts related to the pictures in the order of appearance. In case the child was not able to recall the counts in the correct order, we switched to the next block. The counting span score was calculated for each block, reflecting the number of final counts the child recalled in the correct order. Scores were averaged across the three blocks as an indicator of counting span capacity.

Intelligence

We administered two subtests from the Hungarian standardized version of WISC-IV (Nagyné Réz et al., 2008; Wechsler, 2003) to assess verbal and nonverbal IQ. Verbal IQ and verbal knowledge were assessed with the *Vocabulary* subtest. Children had to define the meaning of the word that was presented verbally by the experimenter and scores were given according to the sophistication of definition. *Block Design* was recorded to test nonverbal IQ and visuo-spatial skills. Children were presented colored designs and asked to construct the model using red-and-white blocks. Scores were given for accurately designed items completed within the time limit. For both IQ measures, we used raw scores in analyses to follow the developmental tendencies within the same age group.

Procedure

Children were tested individually in two 45-minute sessions in their schools. The paper-pencil tests (measures of IQ, verbal fluency tasks) were administered in one session, and the computer-based tests (language-related subtests, counting span task, tapping tests) were run in another session. The order of the tests within the sessions was the same at all testing points for all children, and the order of the two sessions were counterbalanced across participants. In an additional 45-minute session with groups of 10 to 15 children, the online tests measuring musical abilities were completed by the participants using the computer on their own in the school lab. Data collection was run and monitored by the authors and trained assistants.

Baseline measurements were carried out at the beginning of the first school year (T0). We retested children six months later at the end of the first school year (T1), and the third measurement was conducted a year later, at the end of the second school year (T2). Data collection was part of a 4-year longitudinal project involving additional tests that measured children's creativity (Pásztor et al., 2015) and empathy (Bryant, 1982).

Data analysis

As several test results violated the assumption of normality and there were only a limited number of participants in the groups, we carried out nonparametric analysis of longitudinal data using the nparLD package in R (Noguchi et al., 2015) to investigate improvements over one and a half years (T0−T2) in each measure. We used the F1-LD-F1 statistical design, where LD indicates "longitudinal data" and F1 stands for the number of whole-plot (between-subjects) and sub-plot (within-subjects) factors used, respectively. We ran ANOVA-type analyses with *Group* as the between-subjects factor and *Time* (T0, T1, T2) as the within-subjects factor. Analyses provided relative treatment effects (RTEs) for each behavioral test in each group as a measure of effect size. The RTE suggests the probability that a randomly chosen observation from the whole sample obtains a smaller value than a randomly chosen observation from a certain subgroup. Varying between 0 and 1, an RTE value below 0.50 indicates a tendency that a person in a subgroup scores lower than a random person from the whole sample, whereas a value above 0.50 indicates a tendency that a person in a subgroup scores at least as high as a randomly chosen person from the whole sample. A value of 0.50 indicates no effect (Noguchi et al., 2012).

Post hoc analyses were performed using the nparcomp package in R (Konietschke et al., 2019) that provided rank-based methods for comparisons between the groups. We tested group differences at each testing point using the npar.t.test function and examined the differences between the relative time effects in each group using the nparcomp function. We report RTEs for between-subject and within-subject comparisons as an indicator of effect size. All main effects, interactions, and post hoc group differences were regarded as significant if the *p*-value was below .050.

Study 1

Methods

Participants

Children starting their first year at primary school were recruited from three classes in Budapest, Hungary. The allocation of participants to the classes was based on the parents' and children's interests as well as the selection process of the institute. Children in the participating classes received enhanced education in natural sciences, providing students twice-a-week 45-minute lessons regarding natural environment and health maintenance. Also, as part of the Hungarian school curriculum, all children engaged in 45-minute classroom music lessons twice a week according to the pedagogical concepts of Kodály.

The initial sample consisted of 49 participants. Twenty-two children from two classes constituted the science class receiving traditional Kodály music lessons which did not incorporate movement elements into musical activities (*No-Movement – NM-science class*). In addition, 27 children attended the science class with the movement-based music curriculum providing Kodály music lessons with teacher-directed movement elements (*Directed*

Movement – DM-science class). Two children from the NM-science class and two children from the DM-science class did not complete the study. Additional five children from the NM-science class were excluded from analyses due to having missing data on more than two measures at either measurement point.

The final sample (82% of the initial sample) consisted of 40 first-grade children (14 girls, M_{age} = 6.95 years, SD = 0.31 years at the first measurements) of which 15 attended the NM-science class and 25 attended the DM-science class. All participants were native speakers of Hungarian. The sample comprised four bilingual children with Hungarian as their first language and with Spanish (n = 2) or Russian (n = 2) as their second language.

School music curricula

Music curricula were based on the pedagogical approach of Kodály in both classes. Music lessons included collective singing and listening activities to train vocal and basic auditory skills as well as music comprehension. Clapping and tapping were generally employed as supplementary movement elements in rhythmic activities. Additionally, children learned the fundamental theoretical concepts of music, relative solmization, and musical reading/writing.

Participants in the NM-science class obtained music lessons following the traditional Kodály music curriculum. Participants in the DM-science class, however, received an improved Kodály-based music curriculum in which musical activities incorporated movement elements directed by music educators. This movement-based music curriculum was based on the educational model of *Creative Singing-Movement Games*, which enabled children to experience music acoustically, visually, and kinesthetically during musical activities. In singing-movement games, predefined movement forms (e.g., clapping, snapping, touching different body parts, walking, stamping, jumping) presented by the teacher were linked to the rhythmic, melodic, or formal characteristics of music deliberately. Moreover, a ‘body-rhythm’ system has been developed for rhythmic activities in which the traditional rhythmic duration syllables were associated with specific movement forms. The model was adapted to classroom music instruction by music educators Borbála Szirányi and Edina Barabás (Kodály Institute of the Liszt Ferenc Academy of Music). More details about this model of movement-based music education are presented in **Supplementary Material 1.**

Results and Discussion

Baseline results

Data were analyzed using JASP (Version 0.13; JASP Team, 2020). Independent samples *t*-tests or Mann-Whitney tests were conducted to examine whether there was any baseline difference between the classes at the beginning of the study. Analyses revealed no significant differences between the groups in age, $t(38) = -0.57$, $p = .571$, $\eta_p^2 = -0.19$, gender distribution, $\chi^2(1) = 1.44$, $p = .231$, $V = 0.19$, family income, $W = 100$, $p = .897$, $r = -.03$, mother's educational achievement, $W = 143.5$, $p = .054$, $r = .39$, or participants' musical background, $W = 158.5$, $p = .229$, $r = -.16$. Because of the lack of significant differences, we excluded these factors from further analyses.

Descriptive statistics of children's performance for all measures and measurement points (T0−T2) are presented in **Table 1**. At the beginning of the study, only a subset of participants was able to perform the Reading test (DM-science class: $n = 13$, NM-science class: $n = 7$) and the Phoneme Deletion test (DM-science class: $n = 17$, NM-science class: $n = 5$). Thus, medians regarding Reading Fluency and Phoneme Deletion at T0 and T1 are shown only for this subsample.

--- TABLE 1 NEAR HERE ---

Results from the analyses examining baseline group differences are shown in **Supplementary Material 2, Table S1**. Analyses revealed that baseline performance of the two groups differed significantly in Tempo Discrimination, $t(38) = 2.62$, $p = .013$, $d = 0.86$, and Paced Tapping variability, $t(38) = 2.19$, $p = .035$, $d = 0.72$. At the beginning of schooling, the DM-science class scored higher than the NM-science class in the Tempo Discrimination subtest, whereas the NM-science class demonstrated lower variability in the paced tapping task as compared to the DM-science class. No additional between-group difference reached significance (all *p*s ≥ .052).

Longitudinal analyses

Results concerning improvements over one and a half years (T0–T2) in the DM-science and the NM-science classes are shown in **Supplementary Material 2, Table S2**. We report only the significant results in detail below.

Musical abilities

Concerning musical discrimination, longitudinal analyses showed significant main effect of Time for Pitch Discrimination, $F(1.90, \infty) = 3.78$, $p = .025$, $RTE_{T0} = 0.43$, $RTE_{T1} = 0.46$, $RTE_{T2} =0.56$, and Tempo Discrimination, $F(1.80, \infty) = 3.35$, $p = .040$, $RTE_{T0} = 0.45$, $RTE_{T1} = 0.43$, $RTE_{T2} = 0.54$, which implies that both groups showed increases over the two school years in pitch and tempo perception. For Melody, Rhythm, and Harmony Discrimination, no significant main effect of Time was found (*p*s ≥ .061). The main effect of Group was significant in Rhythm Discrimination, $F(1, \infty) = 10.03$, $p = .002$, $RTE_{DM\text{-}science} = 0.57$, $RTE_{NM\text{-}science} = 0.38$, Harmony Discrimination, $F(1, \infty) = 5.61$, $p = .018$, $RTE_{DM\text{-}science} = 0.55$, $RTE_{NM\text{-}science} = 0.41$, and Tempo Discrimination, $F(1, \infty) = 10.14$, $p = .002$, $RTE_{DM\text{-}science} = 0.58$, $RTE_{NM\text{-}science} = 0.36$. Based on the RTEs, the DM-science class outperformed the NM-science class in these subtests, suggesting the overall superior abilities of the DM-science class in these specific components of music perception. For Melody and Pitch Discrimination, the main effect of Group was not significant (*p*s ≥ .075).

With respect to Melody Discrimination, a significant Group × Time interaction emerged (see **Figure 1A**), $F(1.95, \infty) = 5.06$, $p = .007$; DM-science: $RTE_{T0} = 0.49$, $RTE_{T1} =$

0.57, RTE_{T2} = 0.55; NM-science: RTE_{T0} = 0.53, RTE_{T1} = 0.31, RTE_{T2} = 0.49. Post hoc analyses showed no significant difference between the groups at T0 ($p = .730$) and T2 ($p = .526$), whereas the NM-science class scored significantly lower than the DM-science class at T1, $T = -3.53$, $p = .001$, RTE = 0.24. The analysis of time effects revealed that scores in the DM-science class did not change significantly neither from T0 to T1 ($p = .606$) nor from T1 to T2 ($p = .956$). Performance scores in the NM-science class declined significantly from T0 to T1, $T = -2.74$, $p = .030$, RTE = 0.25, but did not change significantly from T1 to T2 ($p = .168$). Even though the pattern of changes was different in the classes, the lack of group difference at T2 indicated no significant developmental advantages for the DM-science class over 18 months in melody perception.

Regarding the auditory-visual connection tasks, analyses revealed a significant main effect of Time for both Melody Connection, $F(1.71, \infty) = 17.55$, $p = 1.843 \times 10^{-7}$, RTE_{T0} = 0.33, RTE_{T1} = 0.51, RTE_{T2} = 0.61, and Rhythm Connection, $F(1.88, \infty) = 9.30$, $p = .0001$, RTE_{T0} = 0.39, RTE_{T1} = 0.49, RTE_{T2} = 0.58, indicating that both groups improved significantly over the two school years. The main effect of Group and Group × Time interactions were not significant for any auditory-visual connection subtest (ps $\geq .091$).

--- FIGURE 1 NEAR HERE ---

Rhythmic entrainment

The main effect of Time was significant for Paced Tapping, $F(1.93, \infty) = 3.78$, $p = .024$, RTE_{T0} = 0.50, RTE_{T1} = 0.55, RTE_{T2} = 0.41, but not for Continuation Tapping ($p = .281$), implying that in both groups, children’s tapping performance improved specifically in tapping along to the metronome. The main effect of Group and Group × Time interactions did not reach significance for any tapping measure (ps $\geq .072$).

Early literacy skills

As only a subset of participants performed the Reading and the Phoneme Deletion tests at baseline, we conducted the longitudinal data analyses based on T0 and T1 performance using data from this subsample. The main effect of Time was significant for both Reading Fluency, $F(1, \infty) = 108.92$, $p = 1.689 \times 10^{-25}$, $RTE_{T0} = 0.31$, $RTE_{T1} = 0.66$, and Phoneme Deletion, $F(1, \infty) = 32.63$, $p = 1.113 \times 10^{-8}$, $RTE_{T0} = 0.32$, $RTE_{T1} = 0.64$, suggesting significant enhancements regarding reading and phonemic awareness in both groups. No significant main effect of Group was found (*p*s ≥ .273). A significant Group × Time interaction emerged for Phoneme Deletion (see **Figure 1B**), $F(1, \infty) = 6.54$, $p = .011$; DM-science class: $RTE_{T0} = 0.43$, $RTE_{T1} = 0.61$; NM-science class: $RTE_{T0} = 0.21$, $RTE_{T1} = 0.67$, but not for Reading Fluency ($p = .189$). Post hoc analyses revealed that the performance of the groups on the Phoneme Deletion test did not differ significantly at any measurement point (*p*s ≥ .135). When examining the performance from T0 to T1, greater improvement was observed in the NM-science class, $T = 9.39$, $p = <.001$, RTE = 0.96, compared to the DM-science class, $T = 2.35$, $p = .029$, RTE = 0.67. However, the classes' comparable performance at both measurement points shows that the NM-science class and the DM-science class did not have considerably different developmental trajectory in phonemic awareness.

To estimate group performances on more complex measures of reading and phonemic awareness at T2, independent samples *t*-tests and Mann-Whitney tests were conducted using data of the whole sample. We found no significant difference between the groups neither in Reading Fluency, $t(38) = 1.69$, $p = .100$, $d = 0.55$, nor in Phoneme Deletion, $W = 257$, $p = .053$, $r = .37$, indicating that the two groups performed similarly at the end of the second school year in both tests.

Regarding RAN, nonparametric longitudinal data analyses revealed a significant main effect of Time for both RAN Digits, $F(1.93, \infty) = 126.24$, $p = 9.413 \times 10^{-54}$, $RTE_{T0} = 0.24$,

$\mathrm{RTE}_{\mathrm{T1}} = 0.52$, $\mathrm{RTE}_{\mathrm{T2}} = 0.73$, and RAN Pictures, $F(2, \infty) = 60.77$, $p = 4.748 \times 10^{-27}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.32$, $\mathrm{RTE}_{\mathrm{T1}} = 0.49$, $\mathrm{RTE}_{\mathrm{T2}} = 0.68$, which suggests an overall improvement over 18 months in rapid naming skills. No significant main effect of Group or Group × Time interaction were found in any RAN measures (*p*s ≥ .115).

Executive functions

Longitudinal data analysis showed a significant main effect of Time for Digit Span, $F(1.96, \infty) = 22.67$, $p = 2.154 \times 10^{-10}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.37$, $\mathrm{RTE}_{\mathrm{T1}} = 0.53$, $\mathrm{RTE}_{\mathrm{T2}} = 0.62$, Counting Span, $F(1.80, \infty) = 16.28$, $p = 3.083 \times 10^{-7}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.37$, $\mathrm{RTE}_{\mathrm{T1}} = 0.50$, $\mathrm{RTE}_{\mathrm{T2}} = 0.63$, and Verbal Fluency, $F(1.84, \infty) = 74.97$, $p = 8.467 \times 10^{-31}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.31$, $\mathrm{RTE}_{\mathrm{T1}} = 0.53$, $\mathrm{RTE}_{\mathrm{T2}} = 0.70$, implying that both groups improved significantly over the two school years in these tests. We found no significant main effect of Group or Group × Time interaction regarding executive functions (*p*s ≥ .230).

Intelligence

The nonparametric ANOVA-type analysis revealed a significant main effect of Time both in Verbal IQ, $F(1.90, \infty) = 85.77$, $p = 2.831 \times 10^{-36}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.30$, $\mathrm{RTE}_{\mathrm{T1}} = 0.50$, $\mathrm{RTE}_{\mathrm{T2}} = 0.75$, and Nonverbal IQ, $F(1.76, \infty) = 36.56$, $p = 6.440 \times 10^{-15}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.35$, $\mathrm{RTE}_{\mathrm{T1}} = 0.45$, $\mathrm{RTE}_{\mathrm{T2}} = 0.65$, indicating that both groups improved significantly over one and a half years in intellectual abilities. In case of Nonverbal IQ, the main effect of Group and Group × Time interaction did not yield significance (*p*s ≥ .061). For Verbal IQ, the main effect of Group yielded significance, $F(1, \infty) = 4.03$, $p = .045$, $\mathrm{RTE}_{\text{DM-science}} = 0.46$ and $\mathrm{RTE}_{\text{NM-science}} = 0.57$, with the RTEs suggesting the better performance of the NM-science class compared to the DM-science class. We found a significant Group × Time interaction in Verbal IQ (see **Figure 1C**), $F(1.90, \infty) = 3.52$, $p = .032$, DM-science: $\mathrm{RTE}_{\mathrm{T0}} = 0.25$, $\mathrm{RTE}_{\mathrm{T1}} = 0.39$, $\mathrm{RTE}_{\mathrm{T2}} = 0.73$; NM-science: $\mathrm{RTE}_{\mathrm{T0}} = 0.35$, $\mathrm{RTE}_{\mathrm{T1}} = 0.60$, $\mathrm{RTE}_{\mathrm{T2}} = 0.76$. Post hoc contrast analysis showed that the groups did not differ at T0 and T2 (*p*s ≥ .052), while the NM-science class scored significantly higher than the DM-science class at T1, $T = 3.09$, $p = .005$, RTE = 0.76. Comparisons for within-subject effects revealed that both groups improved significantly from T0 to T1 (DM-science: $T = 2.47$, $p = .049$, RTE = 0.69; NM-science: $T = 3.47$, $p = .009$, RTE = 0.80) and from T1 to T2 (DM-science: $T = 6.86$, $p = 8.308 \times 10^{-8}$, RTE = 0.86; NM-science: $T = 2.64$, $p = .048$, RTE = 0.75), with the RTEs suggesting greater improvements in the NM-science class from T0 to T1 and larger increases in the DM-science class from T1 to T2. As there were no group differences at baseline and the last measurements, these developmental patterns imply that the rate of improvements but not the overall growth differed between the DM-science and the NM-science classes in verbal IQ.

Taken together, results revealed that the application of movement in classroom-based Kodály music lessons did not have additional developmental benefits in children starting primary school. The classes participating in Kodály music lessons either with or without teacher-directed movement forms demonstrated comparable improvements over one and a half years. Interestingly, we did not observe significant enhancements for any of the classes in melody, rhythm, and harmony discrimination as well as for continuation tapping performance. However, the DM-science class showed superior abilities for rhythm, harmony, and tempo perception. Although the classes had distinct developmental trajectories in specific skills (melody perception, phonemic awareness, verbal IQ), the differences did not indicate that any of the classroom music education programs (the traditional or movement-based music curricula) could support larger improvements during the first school years.

In Study 2, to explore whether the different application of movement in classroom music education would induce distinct longitudinal changes over one and a half years, we compared the developmental impacts of movement-based music lessons using either teacher-directed or improvised movement activities.

Study 2

Methods

Participants

Two first-grade classes were recruited to engage in intense classroom music instruction according to one of the movement-based music curricula which provided children with Kodály music lessons combined either with directed or improvised movement elements. Originally, the sample comprised 53 children. Thirty-two children from a primary school in Budapest, Hungary, received Kodály music lessons complemented with teacher-directed body movements (*Directed Movement – DM-music class*). Twenty-one children from a primary school in Győr (county seat in the north-western region of Hungary) had Kodály music lessons involving improvisational movement elements (*Improvised Movement – IM-music class*). In both schools, the allocation of participants to the class with the intense music curriculum was dependent on the parents' and children's interests and the selection process of the institute. Four participants from the DM-music class and one participant from the IM-music class did not finish the study. Moreover, six children from the DM-music class and two children from the IM-music class had missing data on more than two measures in either measurement point, therefore were excluded from data analyses.

Thus, 40 first-grade children (22 girls, M_{age} = 7.02 years, SD = 0.34 years at the first measurements) were included in the final sample (76% of the initial sample) with the DM-music class comprising 22 children and the IM-music class comprising 18 children. Participants were native speakers of Hungarian, including one bilingual child who had English as the second language.

School music curricula

In both classes, intense music instruction included four music lessons per week. Both movement-based music curricula followed the pedagogical goals of Kodály; however, movement forms were used differently during musical activities in the DM-music class and the IM-music class.

In the DM-music class, children participated in movement-based music lessons following the pedagogical model of Creative Singing-Movement Games which integrated teacher-defined movement elements into musical activities. In contrast, children in the IM-music class received movement-based music lessons according to the *Dynamic Music Learning model* that used free movement improvisations in music education. The model followed the complex art education program of Klára Kokas, which was adapted to Kodály-based classroom music education by Tamara Farnadi (János Richter Secondary School of Music) and Gabriella Deszpot (Kodály Institute of the Liszt Ferenc Academy of Music). Movement was employed as the mean of reception and expression. During listening or singing, children were encouraged to use creative movement combinations freely and spontaneously to respond to the dynamically changing qualities of musical pieces. Neither the form of motion nor the role of partners was determined in the movement choreographies. During the lessons, the receptive and expressive phases varied imperceptibly with reflective phases, which enabled children to share their musical experiences with the group members through movement demonstrations, verbal expressions, or visual artwork creations. More details on the movement-based music educational models are provided in **Supplementary Material 1**.

Results and Discussion

Baseline results

At baseline, no significant differences were observed between the groups in age, $t(38) = -1.89$, $p = .066$, $\eta_p^2 = -0.60$, gender distribution, $\chi^2(1) = 0.33$, $p = .565$, $V = 0.09$, family income, $W = 254$, $p = .063$, $r = .34$, or children’s musical background, $W = 319$, $p = .536$, $r = .11$. Analysis revealed a significant difference between the DM-music class and the IM-music class concerning mother's educational achievement, *W* = 260.5, *p* = .028, *r* = .38. As a great proportion of mothers graduated from university (89.8%), the initial responses were converted into a dichotomous format (0 = "no university graduation" and 1 = "graduated from university"). When we re-analyzed data using this recoded variable, the difference between the two groups in maternal education level was no longer significant, *W* = 211.5, *p* = .238, *r* = .12. Hence, none of the background variables was included in further analyses.

Table 2 reports the medians of the two groups for all measures and measurement points (T0−T2). Only a subset of participants performed the Reading test (DM-music class: *n* = 9, IM-music class: *n* = 14) and the Phoneme Deletion test (DM-music class: *n* = 13, IM-music class: *n* = 18) at the beginning of schooling; therefore, medians regarding Reading Fluency and Phoneme Deletion at T0 and T1 are shown only for this subsample.

--- TABLE 2 NEAR HERE ---

Baseline group differences are shown in **Supplementary Material 2, Table S3** of which we report only the significant results in detail below. At baseline, the two classes differed significantly regarding Verbal IQ, *t*(38) = −2.63, *p* = .012, *d* = −0.84, RAN Pictures, *W* = 77, *p* = <.001, *r* = −0.61, Harmony Discrimination, *W* = 308.50, *p* = .002, *r* = .56, and Tempo Discrimination, *t*(38) = 2.37, *p* = .023, *d* = 0.75. Children in the IM-music class performed better than children in the DM-music class in Verbal IQ and RAN Pictures. In contrast, children in the DM-music class earned higher Harmony and Tempo Discrimination scores than children in the IM-music class. Baseline group differences did not reach significance in any other measures (*p*s ≥ .083).

Longitudinal analyses

Supplementary Material 2, Table S4 shows the results of the longitudinal data analyses regarding the development of the DM-music class and the IM class over the two school years (T0–T2). Only the significant results are reported in detail below.

Musical abilities

Nonparametric longitudinal data analyses showed significant main effect of Time for all musical discrimination subtests, Melody: $F(1.63, \infty) = 7.92$, $p = .001$, $\text{RTE}_{\text{T0}} = 0.40$, $\text{RTE}_{\text{T1}} = 0.47$, $\text{RTE}_{\text{T2}} = 0.62$; Pitch: $F(1.85, \infty) = 39.10$, $p = 1.429 \times 10^{-16}$, $\text{RTE}_{\text{T0}} = 0.35$, $\text{RTE}_{\text{T1}} = 0.45$, $\text{RTE}_{\text{T2}} = 0.69$; Harmony: $F(1.88, \infty) = 13.95$, $p = 1.676 \times 10^{-6}$, $\text{RTE}_{\text{T0}} = 0.37$, $\text{RTE}_{\text{T1}} = 0.48$, $\text{RTE}_{\text{T2}} = 0.63$; Tempo: $F(1.99, \infty) = 9.06$, $p = .0001$, $\text{RTE}_{\text{T0}} = 0.41$, $\text{RTE}_{\text{T1}} = 0.45$, $\text{RTE}_{\text{T2}} = 0.62$, except for Rhythm Discrimination ($p = .481$). These results indicate an overall improvement during the two school years in pitch-related and tempo perception. We found a significant main effect of Group in Rhythm, $F(1, \infty) = 4.25$, $p = .039$, $\text{RTE}_{\text{DM-music}} = 0.56$, $\text{RTE}_{\text{IM-music}} = 0.43$, Harmony, $F(1, \infty) = 6.08$, $p = .014$, $\text{RTE}_{\text{DM-music}} = 0.57$, $\text{RTE}_{\text{IM-music}} = 0.42$, and Tempo Discrimination, $F(1, \infty) = 5.87$, $p = .015$, $\text{RTE}_{\text{DM-music}} = 0.57$, $\text{RTE}_{\text{IM-music}} = 0.42$. The RTEs suggest that the DM-music class performed better than the IM-music class in these subtests. For Melody and Pitch Discrimination, no significant main effect of Group was observed (ps $\geq .130$). Group × Time interactions did not yield significance in any musical discrimination subtest (ps $\geq .057$).

In case of the auditory-visual connection tasks, we observed a significant main effect of Time for both Melody Connection, $F(1.79, \infty) = 43.16$, $p = 1.066 \times 10^{-17}$, $\text{RTE}_{\text{T0}} = 0.32$, $\text{RTE}_{\text{T1}} = 0.49$, $\text{RTE}_{\text{T2}} = 0.68$, and Rhythm Connection, $F(1.89, \infty) = 15.94$, $p = 2.414 \times 10^{-7}$, $\text{RTE}_{\text{T0}} = 0.39$, $\text{RTE}_{\text{T1}} =0.44$, $\text{RTE}_{\text{T2}} = 0.66$, implying that performance of children in both classes increased significantly over the two school years. The main effect of Group and Group × Time interactions did not reach significance in any auditory-visual connection subtest (ps $\geq .165$).

Rhythmic entrainment

The main effect of Time reached significance for both Paced Tapping, $F(1.99, \infty) = 3.24$, $p = .040$, $\text{RTE}_{\text{T0}} = 0.56$, $\text{RTE}_{\text{T1}} = 0.50$, $\text{RTE}_{\text{T2}} = 0.43$ and Continuation Tapping, $F(1.95,$

∞) = 10.83, $p = 2.451 \times 10^{-5}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.60$, $\mathrm{RTE}_{\mathrm{T1}} = 0.50$, $\mathrm{RTE}_{\mathrm{T2}} = 0.39$. Children in both groups became more accurate and tapped with less variability by the end of the second school year. The main effect of Group and Group × Time interaction were not significant in any entrainment task (*p*s ≥ .301).

Early literacy skills

Regarding Reading and Phoneme Deletion tests, the longitudinal data analysis was based on T0 and T1 performance of participants who performed the tests at the baseline measurement. The main effect of Time reached significance in both Reading Fluency, $F(1, \infty) = 27.53$, $p = 1.549 \times 10^{-7}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.36$, $\mathrm{RTE}_{\mathrm{T1}} = 0.62$ and Phoneme Deletion, $F(1, \infty) = 80.05$, $p = 3.654 \times 10^{-19}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.32$, $\mathrm{RTE}_{\mathrm{T1}} = 0.69$, suggesting that children in both classes improved significantly in reading and phonemic awareness. No significant main effect of Group or Group × Time interaction was found (*p*s ≥ .464).

Additional Mann-Whitney tests were conducted using data of the whole sample to examine between-group differences at T2 in more complex measures of reading and phonemic awareness tasks. We observed a significant difference between the groups in Reading Fluency, $W = 120$, $p = .034$, $r = -.39$, but not in Phoneme Deletion, $W = 204$, $p = .882$, $r = .03$. The IM-music class ($Mdn = 5.74$) significantly outperformed the DM-music class ($Mdn = 3.99$) at the end of the second school year specifically in the reading test.

With respect to RAN, nonparametric longitudinal data analyses revealed a significant main effect of Time for both RAN Digits, $F(1.82, \infty) = 170.58$, $p = 1.671 \times 10^{-68}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.24$, $\mathrm{RTE}_{\mathrm{T1}} = 0.53$, $\mathrm{RTE}_{\mathrm{T2}} = 0.75$, and RAN Pictures, $F(1.83, \infty) = 29.16$, $p = 1.831 \times 10^{-12}$, $\mathrm{RTE}_{\mathrm{T0}} = 0.35$, $\mathrm{RTE}_{\mathrm{T1}} = 0.53$, $\mathrm{RTE}_{\mathrm{T2}} = 0.65$. The main effect of Group also yielded significance in both RAN Digits, $F(1, \infty) = 5.22$, $p = .022$, $\mathrm{RTE}_{\mathrm{DM\text{-}music}} = 0.45$, $\mathrm{RTE}_{\mathrm{IM\text{-}music}} = 0.57$, and RAN Pictures, $F(1, \infty) = 8.36$, $p = .004$, $\mathrm{RTE}_{\mathrm{DM\text{-}music}} = 0.42$, $\mathrm{RTE}_{\mathrm{IM\text{-}music}} = 0.60$, with the RTEs suggesting that in both RAN tasks the IM-music class performed better than the DM-music class. Additionally, a significant Group × Time interaction emerged in RAN Pictures (see **Figure 2A**), $F(1.83, \infty) = 4.51$, $p = .013$; DM-music: $\text{RTE}_{\text{T0}} = 0.20$, $\text{RTE}_{\text{T1}} = 0.43$, $\text{RTE}_{\text{T2}} = 0.62$; IM-music: $\text{RTE}_{\text{T0}} = 0.50$, $\text{RTE}_{\text{T1}} = 0.62$, $\text{RTE}_{\text{T2}} = 0.69$, but not in RAN Digits ($p = .416$). For RAN Pictures, post hoc contrast analyses indicated that the IM class outperformed the DM-music class at T0, $T = 4.18$, $p = <.001$, RTE = 0.81, and T1, $T = 2.62$, $p = .014$, RTE = 0.73, but no significant difference emerged across the groups at T2 ($p = .324$). For within-subject effects, we found that the DM-music class improved both from T0 to T1, $T = 4.39$, $p = .0002$, RTE = 0.80, and from T1 to T2, $T = 3.01$, $p = .012$, RTE = 0.73, whereas no significant change was observed in the IM-music class (ps ≥ .154). The pattern of performance changes suggests greater benefits for the DM-music class in rapid naming of pictures; however, the initial developmental lag might have provided much room for growth in the DM-music class, enabling to reach the superior performance level of the IM-music class by the end of the second school year.

--- FIGURE 2 NEAR HERE ---

Executive functions

Nonparametric longitudinal data analysis revealed a significant main effect of Time for Digit Span, $F(1.74, \infty) = 49.65$, $p = 1.030 \times 10^{-19}$, $\text{RTE}_{\text{T0}} = 0.33$, $\text{RTE}_{\text{T1}} = 0.51$, $\text{RTE}_{\text{T2}} = 0.67$, Counting Span, $F(1.94, \infty) = 13.44$, $p = 2.028 \times 10^{-6}$, $\text{RTE}_{\text{T0}} = 0.36$, $\text{RTE}_{\text{T1}} = 0.58$, $\text{RTE}_{\text{T2}} = 0.56$, and Verbal Fluency, $F(1.93, \infty) = 60.59$, $p = 3.416 \times 10^{-26}$, $\text{RTE}_{\text{T0}} = 0.29$, $\text{RTE}_{\text{T1}} = 0.56$, $\text{RTE}_{\text{T2}} = 0.67$, indicating that both groups developed significantly over the two school years in all measures of executive functions. The main effect of Group was significant in Verbal Fluency, $F(1, \infty) = 3.97$, $p =.046$, $\text{RTE}_{\text{DM-music}} = 0.44$, $\text{RTE}_{\text{IM-music}} = 0.57$, but not in Digit Span ($p = .969$) and Counting Span ($p = .784$), which implies that the IM-music class performed better than the DM-music class specifically in the fluency tasks. Group × Time interaction did not yield significance in any measures of executive functions (ps ≥ .269).

Intelligence

Longitudinal analysis showed a significant main effect of Time for Nonverbal IQ, $F(1.84, \infty) = 31.60$, $p = 1.735 \times 10^{-13}$, $\text{RTE}_{\text{T0}} = 0.38$, $\text{RTE}_{\text{T1}} = 0.47$, $\text{RTE}_{\text{T2}} = 0.65$, and Verbal IQ, $F(1.80, \infty) = 58.81$, $p = 6.253 \times 10^{-24}$, $\text{RTE}_{\text{T0}} = 0.31$, $\text{RTE}_{\text{T1}} = 0.49$, $\text{RTE}_{\text{T2}} = 0.72$, which suggests that verbal and nonverbal IQ improved significantly over one and a half years in both groups. The main effect of Group did not reach significance in any IQ measures (*p*s ≥ .115). Group × Time interaction was significant for Verbal IQ (see **Figure 2B**), $F(1.80, \infty) = 3.83$, $p = .026$; DM-music: $\text{RTE}_{\text{T0}} = 0.22$, $\text{RTE}_{\text{T1}} = 0.43$, $\text{RTE}_{\text{T2}} = 0.73$; IM-music: $\text{RTE}_{\text{T0}} = 0.40$, $\text{RTE}_{\text{T1}} = 0.56$, $\text{RTE}_{\text{T2}} = 0.70$, but not for Nonverbal IQ (*p*s ≥ .052). For Verbal IQ, post hoc contrast analysis revealed that the IM-music class scored significantly higher than the DM-music class at T0, $T = 2.46$, $p = .020$, RTE = 0.71, but there was no significant difference between the groups at T1 and T2 (*p*s ≥ .155). Comparisons for within-subject effects showed that the DM-music class significantly improved from T0 to T1, $T = 3.22$, $p = .009$, RTE = 0.74, and from T1 to T2, $T = 5.07$, $p = .0001$, RTE = 0.84, while the IM class showed significant increases only from T0 to T2, $T = 4.30$, $p = .0003$, RTE =0.81, but not between T0–T1 and T1–T2 (*p*s > .168). Based on the group difference at baseline, it is possible that the greater growth of the DM-music class might have originated from its lower performance at the beginning of schooling. Nonetheless, by the end of the first school year, the DM-music class had already reached but could not surpass the performance level of the IM-music class in verbal IQ.

Study 2 demonstrated that the diverse use of movement in intense classroom music education did not have differential impacts on the development of school-aged children’s musical abilities, rhythmic entrainment, phonological processing skills, executive functions, and intelligence. The initial developmental differences between the classes in rapid naming of pictures and verbal IQ disappeared as the skills of the DM-music class improved considerably while the skills of the IM-music class stayed constant. Regarding reading, the IM-music class outperformed the DM-music class at the end of the second school year despite the classes' similar enhancements during the first school year. Furthermore, the IM-music class exhibited superior verbal fluency skills, whereas the DM-music class showed superior temporal and harmony perception. Neither of the classes progressed significantly in rhythm perception.

General Discussion

The present studies were designed to examine the impacts of the application of movement in comprehensive music instruction programs on the development of musical abilities, rhythmic entrainment, phonological processing skills, reading, executive functions, and IQ in children starting primary school. We followed whole classes receiving classroom music lessons as part of the Hungarian school curriculum for one and a half years and assessed participants' improvements at three measurement points. We hypothesized that children would demonstrate more pronounced improvements in pitch discrimination, rhythmic entrainment, phonemic awareness, and working memory when taking Kodály music lessons with but not without teacher-determined movement elements (Study 1). Moreover, we expected to find greater growth in rhythmic entrainment skills, potentially improving executive functions in children receiving intense Kodály music instruction complemented with teacher-directed but not with improvised movement elements (Study 2). The results, however, did not confirm our hypotheses.

Musical abilities

Our findings showed that the various incorporation of movement components in school music education did not benefit children's musical development differently. Concerning perceptual abilities, children in both the science class receiving traditional Kodály music lessons (NM-science class) and the science class receiving Kodály music lessons complemented with directed movement (DM-science class) showed marked improvements specifically in pitch and tempo discrimination but not in melody, harmony, and rhythm discrimination. Pitch discrimination ability, however, was not more increased in the DM-science class as we predicted based on the findings of Maróti et al. (2019). On the other hand, these findings are consistent with the results of Lewis (1988), showing that children who engaged in twice-a-week music lessons with or without movement activities demonstrated comparable performance on the tests of melody, tempo, and rhythm perception.

Nonetheless, children attending the music classes receiving intense Kodály music lessons with directed movements (DM-music class) or improvised movements (IM-music class) demonstrated significant progress for all perceptual abilities, excluding rhythm perception. Considering the specific improvement of pitch perception in the science classes, the general improvements of pitch-related music perception abilities in the music classes might indicate that the more intense application of the movement-based music curricula supported the development of higher-order pitch processing abilities. In the pilot of the current studies, Maróti et al. (2019) detected similar developmental courses for pitch-related music discrimination after 8 months in the classes taking intense school music instruction according to the movement-based music curricula, while children’s rhythm discrimination ability appeared to stay constant. As neither the science classes nor the music classes demonstrated enhancements for rhythm discrimination in the present studies, it indicates that irrespective of the intensity and application of movement, classroom-based Kodály music programs applied in our studies could not facilitate the development of rhythm perception even over a longer 18-month period in the early school years. Moreover, the results might suggest that pitch-related and rhythm discrimination abilities vary in their developmental trajectories.

Regarding auditory-visual connection skills, we detected general enhancements in both the melodic and rhythmic tasks. The similarity of developmental courses among the science classes and the intense music classes possibly reflects school-aged children's emerging capability of associating musical sounds with corresponding symbols. It might have been acquired through the learning of music notation which was included in all classes' music curricula.

Rhythmic entrainment

For performance in tapping to the metronome, the science classes showed similar advancements, while none of them improved in continuation tapping. These results are in accordance with the increasing evidence indicating that the ability to maintain a constant tapping tempo is developed earlier than synchronization to an external beat (Kertész & Honbolygó, 2021; Provasi & Bobin-Bègue, 2003). On the other hand, it suggests that even when including body movement activities, twice-a-week classroom music lessons were not efficient in supporting further improvements in keeping the steady beat as an index of increased timing precision of perception-action coordination. By contrast, children in the intense music classes improved significantly over the two school years in both synchronized and continuation tapping performance. These findings are not in line with the results reported by Maróti et al. (2019), indicating no significant growth for first-grade children in any of the tapping tasks after implementing the movement-based music educational programs for eight months. It is therefore conceivable that advancements for rhythmic synchronization skills from classroom music instruction using movement activities may require intense practice and take a longer time to emerge in tapping performance.

Nonetheless, the lack of group differences in both studies was unexpected given that synchronization skills were explicitly trained by practicing the coordination of motions and music perception in the classes with the directed-movement music curriculum (i.e., the DM-science and DM-music classes). The absence of the classes' developmental advantages may be explained by the educational setting that more likely promoted the practice of social entrainment (Phillips-Silver et al., 2010). When synchronizing motor actions to music in a social context, children could monitor their partners' movements, often making the prediction of others' movements and the temporal matching of movements to the pacesetter easier (Overy & Molnar-Szakacs, 2009). However, our tapping tasks measuring rhythmic entrainment required children to align their movements with the click sounds in an individual setting, which provided no partners as additional references. Thus, entraining to the metronome beat might have acted as a relatively novel and challenging task for every child regardless of the music program they received in their classrooms. It also suggests that the directed-movement music curriculum could not considerably enhance attentional processes underlying entrainment through training movement adjustments in the present of other individuals. On the other hand, all children practiced entraining to complex musical stimuli when listening to musical pieces or singing songs. It has been shown that first-grade children may be more successful in tapping tasks employing musical stimuli instead of metronome clicks (Kertész & Honbolygó, 2021). Our findings, therefore, shed light on the need for considering the role of context and musical stimulus in future studies when measuring the effects of movement-based music lessons on sensorimotor synchronization skills.

Early literacy skills

The present results did not confirm our predictions about the benefits of providing movement-based music lessons with directed movement activities by indicating general increases over one and a half years in phonemic awareness, rapid naming skills, and reading in both the science classes and the music classes. Regarding the science classes, the trend towards superior phonemic awareness at the last measurements in the DM-science class suggests that over a longer period we might have observed the advantages of the twice-a-week movement-based music curriculum on the development of phonemic awareness, and possibly on reading acquisition. Concerning children in the intense music classes, the comparable enhancements in phonemic awareness are concordant with the findings of the study by Maróti et al. (2019), revealing no developmental differences when comparing the classes with the movement-based music curricula to the class with the traditional Kodály curriculum. Nonetheless, the IM-music class exhibited higher reading skills than the DM-music class at the end of the second school year. However, as reading performance of the whole sample was assessed only at the last measurements, we cannot draw any valid conclusions about the causal associations between the involvement of movement in music education and the specific advancements emerged for reading but not for phonological processing skills.

Executive functions and intelligence

Based on the findings of Maróti et al (2019), we predicted greater growth in the classes receiving music lessons combined with teacher-directed movement activities, which placed higher demands on a wide range of executive functions than Kodály music lessons employing free improvised movements or no movements. However, group differences for improvements in executive functions and intellectual abilities were not evident after 18 months of school instruction neither in the science classes nor in the intense music classes. In the current studies, the similarity of the developmental trajectories might reflect the general improvement of executive functions and intelligence during the first school years, possibly due to the commencement of schooling (Ceci & Williams, 1997). According to the arguments that executive functions mediate the effects from music instruction to non-musical cognitive abilities, such as reading (Moreno & Farzan, 2015) and IQ (Degé et al., 2011), the present findings regarding the classes' similar growth in the sub-components of executive functions might be responsible for the absence of distinct developmental trajectories for phonological processing, reading, and intelligence.

Considerations regarding the incorporation of movement in school music education

Taken together, our findings suggest that during the first school years, the diverse implementation of body movement into classroom music education may not influence children's developmental trajectories in musical abilities, rhythmic entrainment, early literacy skills, executive functions, and intellectual abilities distinctively. Plausible explanations for the absence of movement-related benefits primarily relate to the *setting of music education*. First, music instruction in a large-group classroom setting presumably could not ensure the sufficient learning conditions for considerable improvements to occur in multiple domains. In contrast to previous studies in which school music instruction typically took place in groups of 5 to 15 children, music lessons were provided to whole classes including 20 to 30 children in our studies. However, learning in large groups may present difficulties for children in focusing their attention and provide less support for individual development. We argue that in the early grades, music education in the school environment would be more efficient in promoting improvements when occurring in smaller classes.

On the other hand, it is conceivable that the *length and intensity of implementation* determined the benefits of the movement-based music educational programs. Both movement-based music curricula used body movements as kinesthetic reinforcement to learn to attend to, analyze, perceive, and comprehend the qualities of music. Moreover, it has been shown that practice plays an important role in the development of synchronized movements and motor creativity (Janzen et al., 2014; Orth et al., 2017; Repp, 2010). We speculate that due to the complex auditory, sensorimotor, and cognitive demands of accompanying music with motion during listening and music making activities, extended and intense practice might be needed for the strengthening of music-movement associations, leading to music instruction-related advantages over a protracted period.

The discrepancies between the findings of the present and previous studies might also indicate that the outcomes of classroom music lessons in the first school years may critically depend on the *focus of music educational programs* employed. For instance, both the studies by Roden et al. (Roden et al., 2012; Roden, Grube, et al., 2014; Roden, Könen, et al., 2014) and Jaschke et al. (2018) implemented school music programs that comprised focused instrumental training beyond the general elements like listening and singing activities. Since music making and instrumental activities heavily rely on executive functions (Okada & Slevc, 2018), their results showing larger improvements for the music group in verbal memory (Roden et al., 2012), working memory (Roden, Grube, et al., 2014), planning, and inhibition (Jaschke et al., 2018) might have emerged from the quality of the applied instruction. Even though the music curricula using body movements demand executive functions to respond dynamically to musical changes with motions, both teacher-directed and improvised movement activities were included in listening and singing/rhythmic games, which might have provided more playful contexts for training executive functions. It is conceivable, therefore, that in the early school years, instrumental music programs provided in classroom settings may promote larger improvements for specific sub-components of executive functions than comprehensive music programs combined with movement activities.

The abovementioned arguments also raise the fundamental question whether it is valid to expect improvements in areas that had not been trained directly. It is more probable that the benefits of learning may transfer to other domains if target skills are trained explicitly (Salomon & Perkins, 1989; Winner et al., 2013). Our movement-based music programs were primarily designed to improve music-specific skills and knowledge but not domain-general cognitive abilities. These programs, therefore, did not necessarily include curricular elements that could directly train non-musical areas. However, it is also conceivable that other movement-based or music-related activities in which participants took part outside of school over the first school years indeed had a considerable influence on their development. At the beginning of the present studies, data were collected with respect to the participants' out-of-school activities, which indicated that most of the children were engaged in at least one extra-curricular lesson, including either movement-based (e.g., sports, dance), artistic (e.g., visual arts, drama lessons), linguistic (e.g., English lessons), and additional musical (e.g., instrumental education and solfeggio) activities. As we did not follow how participation in these activities changed over the course of this study, we cannot rule out their effects on children's cognitive development. These experiences might have played a critical role in promoting considerable growth for children attending the NM-science class, who showed similar achievements as children in the DM-science class even without participating in movement-based music lessons. Therefore, in further studies estimating the influence of movement-based music programs, it would be essential to frequently register participants' extra-curricular experiences in order to control for potential curriculum-relevant confounding factors.

Nevertheless, the outcomes of formal instruction are affected not only by learning but the interaction between the conditions of learning and maturation (Galván, 2010). The gains that can be achieved through long-term music instruction during childhood have proven to be independent of age *per se* (Sala & Gobet, 2020); rather, the benefits might be influenced by the *child's current level of functioning*. The developmental course of a domain determines the extent to which domain-specific skills can be shaped over the course of music education (Holochwost et al., 2021). Our data showed great inter-individual variability within the participating classes in all skill areas; however, we did not examine individual developmental trajectories which might have also varied substantially across children. Future research could evaluate the advantages of school music instruction based on individual improvements to unravel whether individual characteristics determine the gains that can be obtained through classroom music lessons in the early school years.

In addition, individual differences in the *experience of children in a specific music program* might mediate the success of music education in supporting considerable improvements. Perceiving the musical activity as enjoyable and rewarding makes children more motivated to be involved in that activity, which may also determine its potential gains (Ericsson et al., 2009). To support an optimal motivational basis for engagement in music education, it seems crucial to provide opportunities for practicing and developing music-relevant skills as well as fostering the feeling of agency and independence while interacting with others in a social context (Lamont, 2020). The rationale behind our pedagogical improvements was to pique children's curiosity and promote their motivation through bodily experiences, bringing them joy and the feeling of responsibility in the social context of their classes. Although assessing children's involvement and satisfaction with the programs was not in the focus of our current research, it is possible that in the first school years the diverse inclusion of movement in our music programs have specifically affected musical engagement and socioemotional skills but not musical and general cognitive abilities.

It is important to note, however, that the complex interplay between the abovementioned factors might modulate the effects of movement-based music programs on children's development. Therefore, it seems likely that each music program incorporating body movement might have its own potential to support improvements in particular domains, depending on its contextual characteristics, which challenge the comparability of different studies and the generalization of their results. An answer to these issues would be the unequivocal formulation of hypotheses specifying the form and context of education and the domains under investigation (Holochwost et al., 2021). This differentiation might enable the testing of whether in-school and out-of-school music education could result in similar benefits when provided under same conditions for the training of the same domain.

Limitations

Although the use of randomized controlled trials would be desirable to control for potential preexisting differences in follow-up studies, we did not intend to manipulate children’s natural school environment when examining the impacts of school music instruction to preserve the ecological validity of the present studies. However, since we used real-life group allocations, it was not feasible to change the basic pedagogical specifications of the school curricula in the participating classes. Thus, we could not include a comparison class with a science-focused curriculum that received music lessons combined with improvised movement activities in Study 1 and a comparison class participating in intense music lessons without movement components in Study 2. Moreover, this design did not allow us to increase the class sizes, which resulted in relatively small sub-samples by the end of our measurements, certainly compromising the statistical power of the results. It clearly indicates that in further studies, it would be inevitable to recruit multiple classes for each music program to reach sufficient sample size and resolve power issues, which are more stressed when applying non-parametric methods for hypothesis testing.

Another methodological issue involved the measures applied. Since the Reading and Phoneme Deletion tasks selected for first-year measurements were completed by only a sub-sample of participants, we could not evaluate the developmental courses of phonemic awareness and reading skills for the entire examination period in all children. It emphasizes the need for the development of computerized adaptive testing methods for beginning readers which adjust test items to the participant’s ability level. Furthermore, future research could attempt to use multiple tasks when evaluating children’s musical abilities, sensorimotor entrainment, reading-related skills, and cognitive abilities to control for possible task-dependent effects.

Conclusion

In summary, the present results demonstrated that in classroom educational settings, music lessons varying in the incorporation of body movements did not have distinctive impacts on children’s musical, sensorimotor entrainment, literacy, and general cognitive development in the early primary school years. Our findings raise the possibility that extra-curricular movement-based or music-related activities promoted similar development across the domains measured. On the other part, movement-based music lessons provided in the school context might have supported improvements in areas not examined. Future research should be dedicated to studying the direct influences of school music education employing body movements on participants’ motivation, musical engagement, socioemotional skills, and creativity, which are primarily targeted areas but fairly understudied in terms of the efficiency of classroom music instruction in the early school years.

References

Asztalos, K., & Csapó, B. (2017). Development of musical abilities: Cross-sectional computer-based assessments in educational contexts. *Psychology of Music*, *45*(5), 682–698. https://doi.org/10.1177/0305735616678055

Barkóczi, I., & Pléh, C. (1977). *An examination of the efficiency of Kodály's music educational method [Kodály zenei nevelési módszereinek hatásvizsgálata (Hungarian)]*. Kodály Institute.

Blomert, L., & Vaessen, A. (2009). *Differentiaal Diagnostiek van Dyslexie: Cognitieve analyse van lezen en spellen [Dyslexia Differential Diagnosis: Cognitive analysis of reading and spelling]*. Boom Test.

Bryant, B. K. (1982). An Index of Empathy for Children and Adolescents. *Child Development*, *53*(2), 413–425. https://www.jstor.org/stable/1128984

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal Experimental Child Psychology, 33*, 386–404. https://doi.org/10.1016/0022-0965(82)90054-6

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist*, *52*(10), 1051–1058. https://doi.org/10.1037/0003-066X.52.10.1051

Choksy, L. (1999). *The Kodaly Method I: Comprehensive Music Education (Third Edition)*. Prentice Hall.

Cooper, P. K. (2019). It's all in your head: A meta-analysis on the effects of music training on cognitive measures in schoolchildren. *International Journal of Music Education*, *38*(3), 321–336. https://doi.org/10.1177/0255761419881495

Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology, 10:1522*. https://doi.org/10.3389/fpsyg.2019.01522 de Dreu, C. K. W., Nijstad, B. A., Baas, M., Wolsink, I., & Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin*, *38*(5), 656–669. https://doi.org/10.1177/0146167211435795

Degé, F., Kubicek, C., & Schwarzer, G. (2011). Music lessons and intelligence: a relation mediated by executive functions. *Music Perception: An Interdisciplinary Journal*, *29*(2), 195–201. https://doi.org/10.1525/mp.2011.29.2.195

Diamond, A. (2000). Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child Development*, *71*(1), 44–56. https://doi.org/10.1111/1467-8624.00117

Dobszay, L. (1972). The Kodály Method and Its Musical Basis. *Studia Musicologica Academiae Scientiarum Hungaricae*, *14*(1/4), 15–33. https://doi.org/10.1556/ComEc.

Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2009). Toward a science of exceptional achievement: Attaining superior performance through deliberate practice. *Annals of the New York Academy of Sciences*, *1172*, 199–217. https://doi.org/10.1196/annals.1393.001

Frischen, U., Schwarzer, G., & Degé, F. (2021). Music lessons enhance executive functions in 6- to 7-year-old children. *Learning and Instruction*, *74*(101442). https://doi.org/10.1016/j.learninstruc.2021.101442

Galván, A. (2010). Neural plasticity of development and learning. *Human Brain Mapping*, *31*(6), 879–890. https://doi.org/10.1002/hbm.21029

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the Somatotopic Auditory Mirror System in Humans. *Current Biology*, *16*(18), 1824–1829. https://doi.org/10.1016/j.cub.2006.07.072

Hejja, B., & Szalai, T. (2020). Investigating attitudes and musical taste related to music lessons of primary school students in Debrecen. *Hungarian Educational Research*

Journal, *9*(2), 318–331. https://doi.org/10.1556/063.9.2019.1.27

Holochwost, S. J., Goldstein, T. R., & Wolf, D. P. (2021). Delineating the Benefits of Arts Education for Children’s Socioemotional Development. *Frontiers in Psychology, 12:624712*. https://doi.org/10.3389/fpsyg.2021.624712

Hurwitz, I., Wolff, P. H., Bortnick, B. D., & Kokas, K. (1975). Nonmusical Effects of the Kodaly Music Curriculum in Primary Grade Children. *Journal of Learning Disabilities*, *8*(3), 167–174. https://doi.org/10.1177/002221947500800310

Ilari, B. S., Keller, P., Damasio, H., & Habibi, A. (2016). The Development of Musical Skills of Underprivileged Children Over the Course of 1 Year: A Study in the Context of an El Sistema-Inspired Program. *Frontiers in Psychology, 7:62*. https://doi.org/10.3389/fpsyg.2016.00062

Janzen, T. B., Thompson, W. F., & Ranvaud, R. (2014). A developmental study of the effect of music training on timed movements. *Frontiers in Human Neuroscience, 8:801*. https://doi.org/10.3389/fnhum.2014.00801

Jaschke, A. C., Honing, H., & Scherder, E. J. A. (2018). Longitudinal Analysis of Music Education on Executive Functions in Primary School Children. *Frontiers in Neuroscience, 12:103*. https://doi.org/10.3389/fnins.2018.00103

JASP Team. (2020). *JASP (Version 0.13) [Computer software]*.

Kertész, C., & Honbolygó, F. (2021). Tapping to Music Predicts Literacy Skills of First-Grade Children. *Frontiers in Psychology, 12:741540*. https://doi.org/10.3389/fpsyg.2021.741540

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: theory and data. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 59–101). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-385527-5.00003-6

Kokas, K. (1972). *Képességfejlesztés zenei neveléssel [Fostering abilities by music education (Hungarian)]*. Zeneműkiadó.

Konietschke, F., Noguchi, K., & Rubarth, K. (2019). *Nparcomp: Multiple Comparisons and Simultaneous Confidence Intervals. R package version 3.0*. https://cran.r-project.org/web/packages/nparcomp/index.html

Laczó, Z. (1985). The Nonmusical Outcomes of Music Education: Influence on Intelligence. *Bulletin of the Council for Research in Music Education, 85*, 109–118. https://doi.org/10.1177/0022429414562783

Laczó, Z. (1987). The First Measurement of the Effectiveness of the Kodaly Concept in Hungary Using the Seashore-Test. *Bulletin of the Council for Research in Music Education, 91,* 87–96. https://www.jstor.org/stable/40318067

Lamont, A. (2020). Moving towards music: Viewing early years musical engagement through the lenses of movement, interaction, motivation, agency, identity, and context. *International Journal of Music in Early Childhood*, *15*(1), 25–38. https://doi.org/10.1386/ijmec_00010_1

Lewis, B. E. (1988). The Effect of Movement-Based Instruction on First- and Third-Graders' Achievement in Selected Music Listening Skills. *Psychology of Music, 16*, 128–142. https://doi.org/10.1177/0305735688162003

Maróti, E., Barabás, E., Deszpot, G., Farnadi, T., Nemes, L. N., Szirányi, B., & Honbolygó, F. (2019). Does moving to the music make you smarter? The relation of sensorimotor entrainment to cognitive, linguistic, musical, and social skills. *Psychology of Music, 47*(5), 663–679. https://doi.org/10.1177/0305735618778765

Mészáros, A., Kónya, A., & Kas, B. (2011). A verbális fluenciatesztek felvételének és értékelésének módszertana [Methods in the administration and evaluation of verbal fluency tests (Hungarian)]. *Alkalmazott Pszichológia, 2,* 53–76. http://ap.elte.hu/wpcontent/uploads/2012/09/APA_2011-2.pdf

Miendlarzewska, E. A., & Trost, W. J. (2014). How musical training affects cognitive development: Rhythm, reward and other modulating variables. *Frontiers in Neuroscience, 7:279*. https://doi.org/10.3389/fnins.2014.00279

Moraru, A., Memmert, D., & van der Kamp, J. (2016). Motor creativity: the roles of attention breadth and working memory in a divergent doing task. *Journal of Cognitive Psychology, 28*(7), 856–867. https://doi.org/10.1080/20445911.2016.1201084

Moreno, S., & Farzan, F. (2015). Music training and inhibitory control: A multidimensional model. *Annals of the New York Academy of Sciences, 1337*, 147–152. https://doi.org/10.1111/nyas.12674

Nagyné Réz, I., Lányiné Engelmayer, Á., Kuncz, E., Mészáros, A., & Mlinkó, R. (2008). *Wechsler Intelligence Scale for Children – fourth edition. Magyar adaptáció [Hungarian adaptation]*. OS-Hungary Ltd.

Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). NparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software, 50*(12), 1–23. https://doi.org/10.18637/jss.v050.i12

Noguchi, K., Latif, M., Thangavelu, K., Konietschke, F., Gel, Y. R., & Brunner, E. (2015). *NparLD: Nonparametric Analysis of Longitudinal Data in Factorial Experiments. R package version 2.1*. https://cran.r-project.org/web/packages/nparLD/nparLD.pdf

Okada, B. M., & Slevc, L. R. (2018). Individual differences in musical training and executive functions: A latent variable approach. *Memory and Cognition, 46*(7), 1076–1092. https://doi.org/10.3758/s13421-018-0822-8

Orth, D., van der Kamp, G. J. P., Memmert, D., & Savelsbergh, G. J. P. (2017). Creative motor actions as emerging from movement variability. *Frontiers in Psychology, 8:1903*. https://doi.org/10.3389/fpsyg.2017.01903

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception, 26*(5), 489–504. https://doi.org/10.1525/mp.2009.26.5.489

Pásztor, A., Molnár, G., & Csapó, B. (2015). Technology-based assessment of creativity in educational context: The case of divergent thinking and its relation to mathematical achievement. *Thinking Skills and Creativity*, *18*, 32–42. https://doi.org/10.1016/j.tsc.2015.05.004

Phillips-Silver, J., Aktipis, C. A., & Bryant, G. A. (2010). The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Perception*, *28*(1), 3–14. https://doi.org/10.1525/mp.2010.28.1.3

Provasi, J., & Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 2 1/2- and 4-year-old children. *International Journal of Behavioral Development*, *27*(3), 220–231. https://doi.org/10.1080/01650250244000290

Rautenberg, I. (2015). The effects of musical training on the decoding skills of German-speaking primary school children. *Journal of Research in Reading*, *38*(1), 1–17. https://doi.org/10.1111/jrir.12010

Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Human Movement Science*, *29*(2), 200–213. https://doi.org/10.1016/j.humov.2009.08.002

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic Bulletin and Review*, *20*(3), 403–452. https://doi.org/10.3758/s13423-012-0371-2

Rickard, N. S., Vasquez, J. T., Murphy, F., Gill, A., & Toukhsati, S. R. (2010). Benefits of a classroom based instrumental music program on verbal memory of primary school children: a longitudinal study. *Australian Journal of Music Education*, *1*, 36–47.

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

Roden, I., Könen, T., Bongard, S., Frankenberg, E., Friedrich, E. K., & Kreutz, G. (2014). Effects of music training on attention, processing speed and cognitive music abilities - findings from a longitudinal study. *Applied Cognitive Psychology*, *28*(4), 545–557. https://doi.org/10.1002/acp.3034

Roden, I., Kreutz, G., & Bongard, S. (2012). Effects of a school-based instrumental music program on verbal and visual memory in primary school children: A longitudinal study. *Frontiers in Psychology*, *3:572*. https://doi.org/10.3389/fpsyg.2012.00572

Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, *48*, 1429–1441. https://doi.org/10.31234/osf.io/7s8wr

Salomon, G., & Perkins, D. N. (1989). Rocky Roads to Transfer: Rethinking Mechanisms of a Neglected Phenomenon. *Educational Psychologist*, *24*(2), 113–142. https://doi.org/10.1207/s15326985ep2402_1

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*(8), 511–514. https://doi.org/10.1111/j.0956-7976.2004.00711.x

The MathWorks Inc. (2015). *MATLAB* (No. R2015a).

Tierney, A. T., & Kraus, N. (2013). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain and Language*, *124*(3), 225–231. https://doi.org/10.1016/j.bandl.2012.12.014

Todd, N. P. M. (1999). Motion in Music: A Neurobiological Perspective. *Music Perception: An Interdisciplinary Journal*, *17*(1), 115–126. https://doi.org/10.2307/40285814

Tóth, D., Csépe, V., Vaessen, A., & Blomert, L. (2014). *A diszlexia differenciáldiagnózisa. Az* *olvasás és helyesírás kognitív elemzése. Technikai kézikönyv. [3DM-H: Dyslexia Differential Diagnosis. Cognitive analysis of reading and spelling (Hungarian)]*. Kogentum.

van der Fels, I. M. J., te Wierike, S. C. M., Hartman, E., Elferink-Gemser, M. T., Smith, J., & Visscher, C. (2015). The relationship between motor skills and cognitive skills in 4-16 year old typically developing children: A systematic review. *Journal of Science and Medicine in Sport*, *18*(6), 697–703. https://doi.org/10.1016/j.jsams.2014.09.007

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children, fourth edition (WISC-IV)*. The Psychological Corporation.

Winner, E., Goldstein, T. R., & Vincent-Lancrin, S. (2013). *Art for Art's Sake?: The Impact of Arts Education*. Educational Research and Innovation, OECD Publishing. https://doi.org/10.1787/9789264180789-en

Zelechowska, A. (2020). *Irresistible Movement : The Role of Musical Sound, Individual Differences and Listening Context in Movement Responses to Music* [Doctoral dissertation, Faculty of Humanities, University of Oslo]. https://www.duo.uio.no/bitstream/handle/10852/82884/2/Agata_Zelechowska_PhD.pdf

Tables

Table 1

Medians for all tests at the three measurement points (T0−T2) in the science classes

Measures	DM-science class (*n* = 25)			NM-science class (*n* = 15)		
	T0	T1	T2	T0	T1	T2
Melody Discrimination	7	7	7	7	6	7
Pitch Discrimination	7	8	9	6	5	7
Rhythm Discrimination	7	9	9	7	5	7
Harmony Discrimination	7	9	9	7	7	8
Tempo Discrimination	9	9	10	7	7	8
Melody Connection	5	7	8	2.5	6	7
Rhythm Connection	2	3	4	2	2	3
Paced Tapping	0.10	0.11	0.08	0.08	0.09	0.08
Continuation Tapping	0.10	0.10	0.09	0.11	0.10	0.10
Reading Fluency [a]	1.20	4.06	4.38	1	1.98	3.95
Phoneme Deletion [a]	−0.81	−0.08	0.38	−1.12	0.13	−0.37
RAN Digits	1.16	1.52	1.84	1.13	1.54	1.82
RAN Pictures	0.93	1.06	1.23	0.88	1	1.25
Digit Span	12	13	14	12	14	14
Counting Span	2.5	2.67	2.67	2.33	2.67	3
Verbal Fluency	28	35	45	26	40	56
Verbal IQ	23	28	37	26	34	39
Nonverbal IQ	28	30	38	22	26	32

[a] Medians for T0 and T1 show data from the subset of participants who performed the test at baseline.

Table 2

Medians for all tests at the three measurement points (T0−T2) in the music classes

Measures	DM-music class (*n* = 22)			IM-music class (*n* = 18)		
	T0	T1	T2	T0	T1	T2
Melody Discrimination	7	7.50	9	6	6.50	9
Pitch Discrimination	8.50	12	14	8.50	8.50	14
Rhythm Discrimination	8.50	9	8	8	7	7.50
Harmony Discrimination	9	9.50	10.50	7	7	11
Tempo Discrimination	9	9	10	7	7	9.50
Melody Connection	5.50	7.50	10	4.50	7.50	8.50
Rhythm Connection	2.50	3	4	2	2	4
Paced Tapping	0.11	0.08	0.08	0.08	0.09	0.08
Continuation Tapping	0.10	0.08	0.08	0.09	0.08	0.07
Reading Fluency [a]	1.50	2.45	3.99	0.90	2.92	5.74
Phoneme Deletion [a]	−1.60	−1.79	0.24	0.16	0.02	0.11
RAN Digits	1.11	1.47	1.82	1.30	1.58	1.95
RAN Pictures	0.86	1.06	1.20	1.07	1.17	1.24
Digit Span	11	12	14	11.50	13.50	14
Counting Span	2.33	3	2.67	2.33	2.67	2.67
Verbal Fluency	25.50	35	41.50	31	42	48
Verbal IQ	23	28	35	26.50	31.50	35
Nonverbal IQ	26	26.50	32	22	22	33

[a] Medians for T0 and T1 show data from the subset of participants who performed the test at baseline.

Figure captions

Figure 1. Performance on the tests of Melody Discrimination (A), Phoneme Deletion (B), and Verbal IQ (C) for all measurement points in the science classes.

Note. Asterisks indicate significant differences between the groups. ***p* < .01.

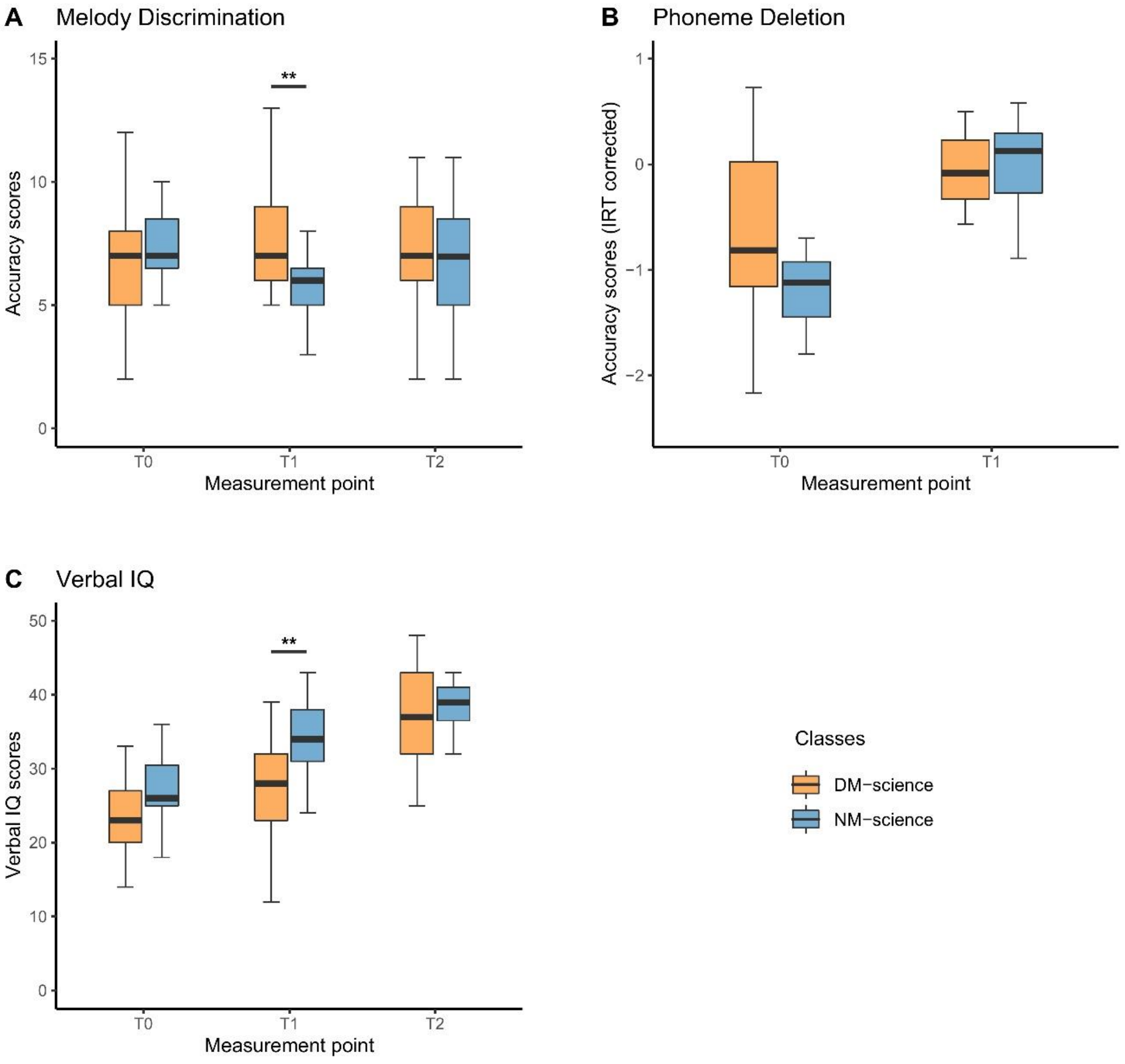
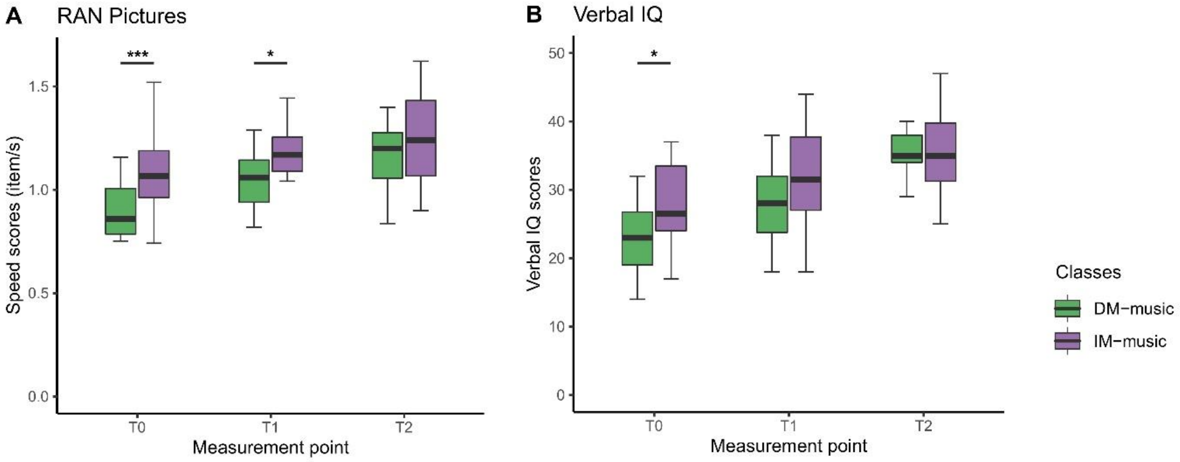


Figure 2. Performance on the tests of RAN Pictures (A) and Verbal IQ (B) for all measurement points in the intensive music classes.

Note. Asterisks indicate significant differences between the groups. **p* < .05. ****p* < .001.



Supplementary Material 1

Description of the movement-based music education models

The concept of movement-based music education

Movement-based music education refers to an experience-centered music pedagogical approach that emphasizes the integration of body movement in music instruction. The concept originates in the educational goals of Kodály and centers on the sequential development of music culture and literacy. Therefore, the music curriculum provides structured lessons including vocal and ear training as well as the learning of folk songs, relative solmization, hand signs, rhythmic duration syllables, fundamentals of music theory, and music notation. Movement-based music lessons complement these musical activities with body motions which respond to the temporal, tonal, and formal changes in music.

These movement-based music educational methods were specifically designed for children in their early school years and implemented into general classroom instruction. Despite their identical pedagogical goals, movement-based music educational models vary in how body motions are applied: the model of *Creative Singing-Movement Games* uses movement forms determined and directed by the music teacher, whereas the model of *Dynamic Music Learning* encourages children to improvise and use spontaneous movements during musical activities.

Model 1: Creative Singing-Movement Games

The pedagogical concept of *Creative Singing-Movement Games* integrates the principles of Kodály-based music instruction and some elements of the Dalcroze approach. The model was developed and adapted to classroom music instruction by music educators Borbála Szirányi and Edina Barabás (Kodály Institute of the Liszt Ferenc Academy of Music). The method implements body movement into musical activities deliberately by linking the experience of sound to specific movements.

During singing-movement games, movement forms are predefined and freely paired with rhythmic values or pitch variations. The emergent movement choreographies express and strictly follow the formal, rhythmic, or melodic characteristics of music. At an early stage, the teacher presents the movements to children, who imitate the movements first individually, then, depending on the choreography, in pairs or groups. Later, students are encouraged to improvise movement choreographies cooperating with their peers. Movement choreographies implemented in listening and vocal activities are repeated systematically to promote children's self-expression and improve their phrasing and singing technique.

In addition, a 'body-rhythm' system has been developed in which the traditional rhythmic duration syllables are associated with specific movement forms (e.g., clapping, snapping, touching different body parts, walking, stamping, and jumping). In rhythmic games, each movement form from the 'body-rhythm' set is paired with a specific rhythmic value, which enables children to experience rhythm acoustically, visually, and kinesthetically at a time. The rhythm-motion pairs are strengthened by systematic practice, aiming to support the decoding and interpretation of rhythmic patterns in music.

Model 2: Dynamic Music Learning

The *Dynamic Music Learning model* follows the educational concept and music pedagogical methods of Klára Kokas, which was adapted to the traditional classroom music education by Tamara Farnadi (János Richter Secondary School of Music) and Gabriella Deszpot (Kodály Institute of the Liszt Ferenc Academy of Music). The main aim of this complex art education program is to support the understanding of music by integrating music, movement, and imagination into a unified experience. The model is based on Hungarian folk songs, movement games for children, and classical music masterpieces, and employs movement as the mean of reception and expression.

Music lessons include collective singing phases as a framing for each session. Listening and reflection phases are embedded between the singing phases. During the music listening phase, children are encouraged to use creative movement combinations spontaneously, responding to the dynamically changing qualities of the presented musical piece. The form of motion and the role of partners are not determined: movement improvisations are based on children's creativity and can lead to free individual choreographies or complex group performances. The collective reflection phase enables children to share their musical experiences with the group members through movement demonstrations, verbal expressions, or visual artwork creations. The receptive, reflective, and expressive phases vary imperceptibly while presenting the same musical piece to the group several times in succession. Besides the acquisition of basic music knowledge, music appreciation using bodily movements supports the expression of emotions and attempts to indirectly facilitate the development of children's social skills.

Supplementary Material 2

Table S1

Baseline differences between the DM-science and the NM-science classes in all behavioral measures

Measurement	*t / W (df)*	*p*	*d / r* [a]
Melody Discrimination	0.15 (38)	.881	0.05
Pitch Discrimination	-0.43 (38)	.672	-0.14
Rhythm Discrimination	1.25 (38)	.220	0.41
Harmony Discrimination	0.96 (38)	.343	0.31
Tempo Discrimination	2.62 (38)	**.013**	0.86
Melody Connection	218.50	.203	0.25
Rhythm Connection	167.50	.834	-0.04
Paced Tapping	2.19 (38)	**.035**	0.72
Continuation Tapping	189	.978	0.01
Reading Fluency [b]	0.74 (18)	.469	0.35
Phoneme Deletion [b]	59	.218	0.39
RAN Digits	1.02 (38)	.314	0.33
RAN Pictures	0.49 (38)	.626	0.16
Digit Span	183.50	.921	-0.02
Counting Span	192.50	.724	0.07
Verbal Fluency	-0.86 (38)	.397	-0.28
Verbal IQ	-1.65 (38)	.108	-0.54
Nonverbal IQ	257	.052	0.37

Note. Significant *p*-values below .05 are written in bold.

[a] Effect size indicators: *d* = Cohen’s *d*; *r* = rank biserial correlation. [b] Results are based on data from the subset of participants who performed the test at baseline.

Table S2

Results of the longitudinal data analysis for the DM-science and the NM-science classes

Measurement	Main effect of Group		Main effect of Time		Group × Time interaction	
	F (df)	*p*	*F* (df)	*p*	*F* (df)	*p*
Melody Discrimination	1.83 (1)	.176	1.50 (1.95)	.224	5.06 (1.95)	**.007**
Pitch Discrimination	3.17 (1)	.075	3.78 (1.90)	**.025**	2.29 (1.90)	.105
Rhythm Discrimination	10.03 (1)	**.002**	0.87 (1.62)	.401	1.51 (1.62)	.224
Harmony Discrimination	5.61 (1)	**.018**	2.81 (1.97)	.061	1.27 (1.97)	.280
Tempo Discrimination	10.14 (1)	**.002**	3.35 (1.80)	**.040**	0.64 (1.80)	.514
Melody Connection	2.86 (1)	.091	17.55 (1.71)	**1.843×10^{-7}**	0.16 (1.71)	.815
Rhythm Connection	1.48 (1)	.224	9.30 (1.88)	**.0001**	2.20 (1.88)	.114
Paced Tapping	3.24 (1)	.072	3.78 (1.93)	**.024**	1.68 (1.93)	.187
Continuation Tapping	0.39 (1)	.530	1.25 (1.62)	.281	1.11 (1.62)	.321
Reading Fluency [a]	1.20 (1)	.273	108.92 (1)	**1.689×10^{-25}**	1.73 (1)	.189
Phoneme Deletion [a]	0.68 (1)	.410	32.63 (1)	**1.113×10^{-8}**	6.54 (1)	**.011**
RAN Digits	0.40 (1)	.529	126.24 (1.93)	**9.413×10^{-54}**	1.38 (1.93)	.251
RAN Pictures	0.44 (1)	.508	60.77 (2)	**4.748×10^{-27}**	2.16 (2)	.115
Digit Span	0.16 (1)	.688	22.67 (1.96)	**2.154×10^{-10}**	0.23 (1.96)	.787
Counting Span	0.0002 (1)	.990	16.28 (1.80)	**3.083×10^{-7}**	0.96 (1.80)	.374
Verbal Fluency	1.44 (1)	.230	74.97 (1.84)	**8.467×10^{-31}**	1.25 (1.84)	.284
Verbal IQ	4.03 (1)	**.045**	85.77 (1.90)	**2.831×10^{-36}**	3.52 (1.90)	**.032**
Nonverbal IQ	3.52 (1)	.061	36.56 (1.76)	**6.440×10^{-15}**	0.72 (1.76)	.471

Note. The denominator degrees of freedom are set to infinity for each nparLD analysis. Significant *p*-values below .05 are written in bold.

[a] Analysis was conducted using data from T0 and T1 and based on the performance scores of the subset of participants who performed the test at baseline.

Table S3

Baseline differences between the DM-music and the IM-music classes in all behavioral measures

Measurement	*t / W (df)*	*p*	*d / r* [a]
Melody Discrimination	216.50	.620	0.09
Pitch Discrimination	1.12 (38)	.271	0.36
Rhythm Discrimination	0.82 (38)	.418	0.26
Harmony Discrimination	308.50	**.002**	0.56
Tempo Discrimination	2.37 (38)	**.023**	0.75
Melody Connection	1.62 (38)	.114	0.51
Rhythm Connection	235.50	.302	0.19
Paced Tapping	0.67 (38)	.506	0.21
Continuation Tapping	-0.32 (38)	.753	-0.10
Reading Fluency [b]	59	.825	-0.06
Phoneme Deletion [b]	0.85 (29)	.401	0.31
RAN Digits	-1.78(38)	.083	-0.57
RAN Pictures	77	**< .001**	-0.61
Digit Span	0.04 (38)	.970	0.01
Counting Span	202.50	.911	0.02
Verbal Fluency	135.50	.091	-0.32
Verbal IQ	-2.63 (38)	**.012**	-0.84
Nonverbal IQ	245.50	.196	0.24

Note. Significant *p*-values below .05 are written in bold.

[a] Effect size indicators: *d* = Cohen’s *d*; *r* = rank biserial correlation. [b] Results are based on data from the subset of participants who performed the test at baseline.

Table S4

Results of the longitudinal data analysis for the DM-music and the IM-music classes

Measurement	Main effect of Group		Main effect of Time		Group × Time interaction	
	F (df)	*p*	*F* (df)	*p*	*F* (df)	*p*
Melody Discrimination	2.29 (1)	.130	7.92 (1.63)	**.001**	0.43 (1.63)	.607
Pitch Discrimination	1.56 (1)	.212	39.10 (1.85)	**1.429×10^{-16}**	1.22 (1.85)	.295
Rhythm Discrimination	4.25 (1)	**.039**	0.71 (1.82)	.481	1.11 (1.82)	.327
Harmony Discrimination	6.08 (1)	**.014**	13.95 (1.88)	**1.676×10^{-6}**	2.93 (1.88)	.057
Tempo Discrimination	5.87 (1)	**.015**	9.06 (1.99)	**.0001**	0.45 (1.99)	.637
Melody Connection	1.90 (1)	.168	43.16 (1.79)	**1.066×10^{-17}**	1.25 (1.79)	.285
Rhythm Connection	1.93 (1)	.165	15.94 (1.89)	**2.414×10^{-7}**	0.65 (1.89)	.513
Paced Tapping	0.63 (1)	.429	3.24 (1.99)	**.040**	1.08 (1.99)	.339
Continuation Tapping	1.07 (1)	.301	10.83 (1.95)	**2.451×10^{-5}**	1.05 (1.95)	.349
Reading Fluency [a]	0.54 (1)	.464	27.53 (1)	**1.549×10^{-7}**	0.44 (1)	.505
Phoneme Deletion [a]	0.53 (1)	.468	80.05 (1)	**3.654×10^{-19}**	0.06 (1)	.807
RAN Digits	5.22 (1)	**.022**	170.58 (1.82)	**1.671×10^{-68}**	0.86 (1.82)	.416
RAN Pictures	8.36 (1)	**.004**	29.16 (1.83)	**1.831×10^{-12}**	4.51 (1.83)	**.013**
Digit Span	0.002 (1)	.969	49.65 (1.74)	**1.030×10^{-19}**	0.36 (1.74)	.667
Counting Span	0.08 (1)	.784	13.44 (1.94)	**2.028×10^{-6}**	1.31 (1.94)	.269
Verbal Fluency	3.97 (1)	**.046**	60.59 (1.93)	**3.416×10^{-26}**	0.31 (1.93)	.729
Verbal IQ	2.48 (1)	.115	58.81 (1.80)	**6.253×10^{-24}**	3.83 (1.80)	**.026**
Nonverbal IQ	0.03 (1)	.853	31.60 (1.84)	**1.735×10^{-13}**	3.05 (1.84)	.052

Note. The denominator degrees of freedom are set to infinity for each nparLD analysis. Significant *p*-values below .05 are written in bold.

[a] Analysis was conducted using data from T0 and T1 and based on the performance scores of the subset of participants who performed the test at baseline.