Montreal Evaluation of Musical Abilities on tablet for 4 to 6-year-old children

Documentation July 2021

Isabelle Peretz*[1], Jonathan Bolduc*[1,2], Manon Couvignou[3]

*Equally contributed to the work

[1]*International Laboratory of Brain, Music and Sound Research (BRAMS), University of Montreal, Montreal, Canada*

[2]*Faculty of Music, Canada Research Chair in Music and Learning, Université Laval, Quebec, Canada*

[3]*Unité de Recherche en Neurosciences Cognitives (Unescog), Center for Research in Cognition & Neurosciences (CRCN), Université Libre de Bruxelles (ULB), Brussels, Belgium*

Corresponding author:

Isabelle.Peretz@umontreal.ca

ABSTRACT

The present study introduces a novel tool -- The Montreal Battery of Evaluation of Musical Abilities (MBEMA) on tablet -- for assessing musical abilities in 4 to 6-year-old children. The battery contains three tests: melody discrimination, rhythm discrimination and memory recognition of unfamiliar tonal melodies. Each test comprises two examples followed by 12 trials and is presented in a different picturesque format for an average duration of 15 minutes. As of July 2021, the battery has been administered to 191 5-to-6-year-old children, of whom 181 were tested twice (89 three months apart; 92 eight months apart). The battery was sensitive to individual differences and music training, with the amount of music training predicting performance. Despite high interindividual variability, the three-month follow-up showed good test-retest reliability, with significant improvements from test to retest that may reflect a familiarization effect. Longitudinal follow-up after eight months also showed an increase in performance, which may reflect either a maturation effect or be related to literacy. Overall, the MBEMA on a tablet can serve as an objective, short and up-to-date test for assessing musical ability in early childhood.

Keywords: assessment tool; music perception; early childhood

The major goal of the present investigation was to provide a means of adapting the The Montreal Battery of Evaluation of Musical Abilities (MBEMA) in order to assess musical abilities before the age of six. It is an abridged version of the The Montreal Battery for Evaluation of Musical Abilities (MBEMA; Peretz et al., 2013) implemented on an Android tablet. The present document aims to describe the tool and provide norms for researchers interested in using it. For full access please write to the corresponding author.

Procedure: On the tablet, the child initiates each trial by pressing on a picture. Two practice trials precede each task, and once the children have understood the task they perform the task alone. On every trial, one or two melodies are played (depending on the task), and response buttons are shown for "same"/"different" or "heard"/"new". In the melody test, the child presses flower pictures to hear the melodies. Then, two button icons appear, one with two bees on the left of the screen for the "same" melodies one with one bee and a beetle for the "different" melodies (see Figure 1). The child selects one of the two icons to indicate their response. In the subsequent rhythm discrimination test, the child presses on the planet of their choice to listen to two successive melodies. Then, the response icons appear, and the child responds by selecting two rockets or one rocket and a plane for the "same" and "different" melodies, respectively (Figure 1). The third and final test assesses recognition. The child presses on the house of their choice to listen to a melody that has been presented in the previous tests or not. After each melody, a picture of a teddy bear appears for the "yes, I have heard that melody before" response and a frog for the "no" response (Figure 1).

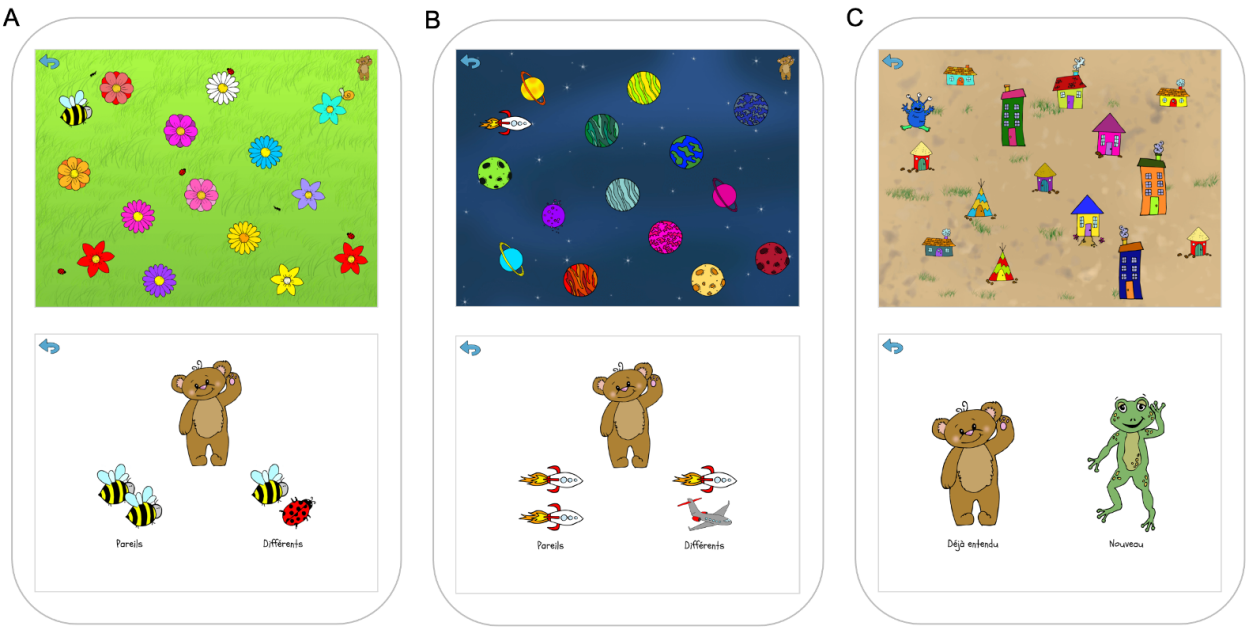

Figure 1. Illustration of a trial in the (A) *melody*, (B) *rhythm*, and (C) *memory* test. For each test, the upper panel shows the screen that allows the child to listen to the melodies and the lower panel shows the response choices.

Material

The *melody* and *rhythm* tests use the same set of melodies. Each trial is made of one standard melody followed by a comparison melody after a 1.5-sec silent interval. Each test comprises two practice trials and 12 test trials (6 same and 6 different). In any given trial, the stimuli are presented with the same timbre, whereas the timbre changes from one trial to the next. Ten different timbres (e.g., piano, marimba, guitar, flute) are used to make the tests as engaging as possible. The trials are presented in a fixed order with no more than three successive trials of either same or differing stimuli. The task requires a same-different response.

In the *melody* test, three types of changes are applied to six differing melodies. Two comparison melodies are different by one out-of-key note while retaining the original melodic contour (i.e., a scale violating change); two comparison melodies have one changed note that alters the pitch direction of the surrounding intervals while maintaining the original key (i.e., a contour-violating change); two comparison melodies have one changed note that alters two intervals while preserving the original contour and key (i.e., an interval-violating change; see Figure 2 for each type of change applied to one melody). The serial position of the modified note varies across melodies and never occurs in the first and last position. Average pitch interval changes are equivalent across the three types of pitch changes, with a mean of 1.5 (range: -1 to +2), 4 (-4 to +4) and 4 (-4 to +4) semitones from the original pitch for *scale*-violating, *contour*-violating, and *interval*-violating changes, respectively.

In the *rhythm* test, the manipulation consists of changing the durations of two adjacent tones so as to alter the rhythmic grouping of notes while retaining the number of notes and original meter. This is accomplished by changing two quarter notes to a dotted quarter and an eighth note (i.e., in 4 trials), or by reversing the order of two successive duration values (i.e., in two trials). The serial position of these changes varies across melodies.

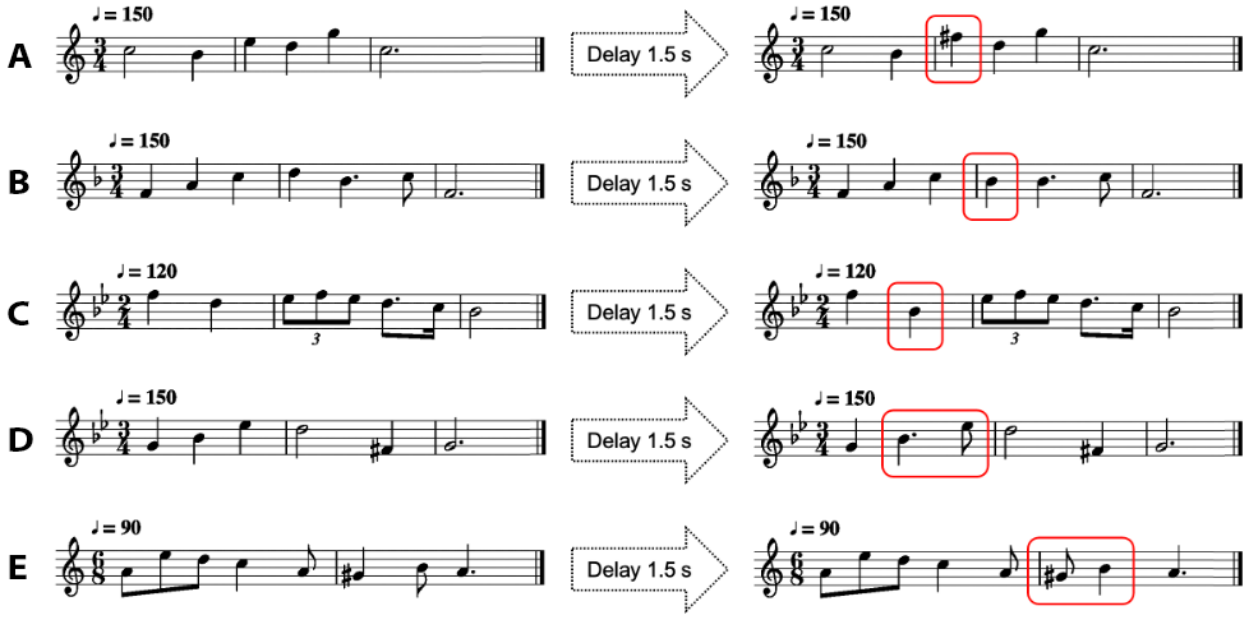

Figure 2. Example of the stimuli: The three types of changes in the *melody* test are (A) scale, (B) contour, and (C) interval; the two types of changes in the *rhythm* test are (D) duration interval and (E) duration order. Changes are framed in red. The example can be heard at www.peretzlab.ca/knowledge_transfer

From the initial set of 12 standard melodies, six are selected for the *memory* test. Each old melody has been presented at least twice in the same format in the *melody* or *rhythm* test. In addition to these old melodies, there are six foils or new melodies. The new melodies are constructed in the same manner as the old melodies but differ in their exact temporal and pitch patterns. The task here is to press the bear if the melody has been presented earlier and the frog otherwise. This test assesses incidental memory because the children are not informed that their memory for the melodies will be tested subsequently.

Participants

Data were acquired in two collaborating sites, in Quebec City, Canada, and Brussels, Belgium. The Canadian sample consisted of 89 children aged 5 to 6 attending public schools in Quebec City. The Belgian sample consisted of 102 children aged 5 to 6 recruited from public schools in Brussels; 33 were in a school with a pedagogical project focused on music from the age of 3 (Couvignou et al., in preparation). Exclusion criteria included brain injury, hearing impairments, and developmental disorders such as oral speech impairment, attention deficit hyperactivity disorder, or autism as reported by parents. The detailed characteristics of the two groups in terms of age, sex ratio, grade, and music training are presented in Table 1. All were native French speakers except two Canadian children who were English-speaking.

Ethical approval was granted by the local Ethics Committees (Université Laval in Canada, agreement number: 2016-194; Université Libre de Bruxelles in Belgium, agreement number: 034/2017). Written informed consent was obtained from the parents of each child, as well as oral agreement from each participant.

Table 1. Characteristics of the children

AGE	5	6	**Total**
Canadian children			
SAMPLE SIZE (Gender)	83 (38M, 45F)	6 (3M, 3F)	**89**
SCHOOL GRADE			
Kindergarten	83	6	**89**
MUSIC TRAINING			
No Lesson	83	6	**89**
Lessons/mean duration in months	0	0	**0**
Belgian children			
SAMPLE SIZE (Gender)	57 (32M, 25F)	45 (23M, 22F)	**102**
SCHOOL GRADE			
Kindergarten	57	45	**102**
MUSIC TRAINING			
No Lesson	35	24	**59**
Lessons/mean duration (SD) in months	21.27 (5.15)	24.00 (11.85)	**22.60 (9.06)**
From Age 1	0	1	**1**
Age 2	0	1	**1**
Age 3	17	10	**27**
Age 4	5	5	**10**
Age 5	0	4	**4**

Each child was tested individually in a quiet room of their school twice: 3 months apart for the group in Canada (January 2017 - April 2017), and 8 months apart for the group in Belgium (via two recruitment waves: June 2018 - March 2019 and June 2019 - March 2020). Note that of the 102 Belgian children, only 92 participated in the second testing phase due to the COVID-19 pandemic and subsequent school closure in March 2020. The tests were presented through a *Samsung Galaxy A10* tablet and headphones set at a comfortable volume. Children were free to request breaks between tests, and each child received a "diploma" as a reward for their time and participation.

Results and comments

Sensitivity

Children's global score (across the three tests) at 5 and 6 years of age ranged from 13 to 33, with 18 representing chance performance and 36 a perfect score. Table 2 shows the mean scores and SD by age group. As can be seen in Figure 3A, the distribution of cumulative score was slightly skewed but did not violate normality [$W_{(140)} = .990$, $p = .426$; $W_{(51)} = .982$, $p = .628$, for 5- and 6-year-olds, respectively, by Shapiro-Wilk test]. However, scores on individual tests violated normality in most cases. Thus, the total score is more sensitive than individual test scores in distinguishing normal from abnormal performance.

Table 2. Mean (SD) score on each test and on all tests (global) by age group.

	n	Melody (/12)	Rhythm (/12)	Memory (/12)	Global (/36)
5 years	140	6.7 (1.58)	7.7 (2.04)	8.4 (1.70)	22.8 (3.75)
6 years	51	7.6 (1.76)	8.8 (1.77)	8.8 (1.90)	25.2 (3.90)

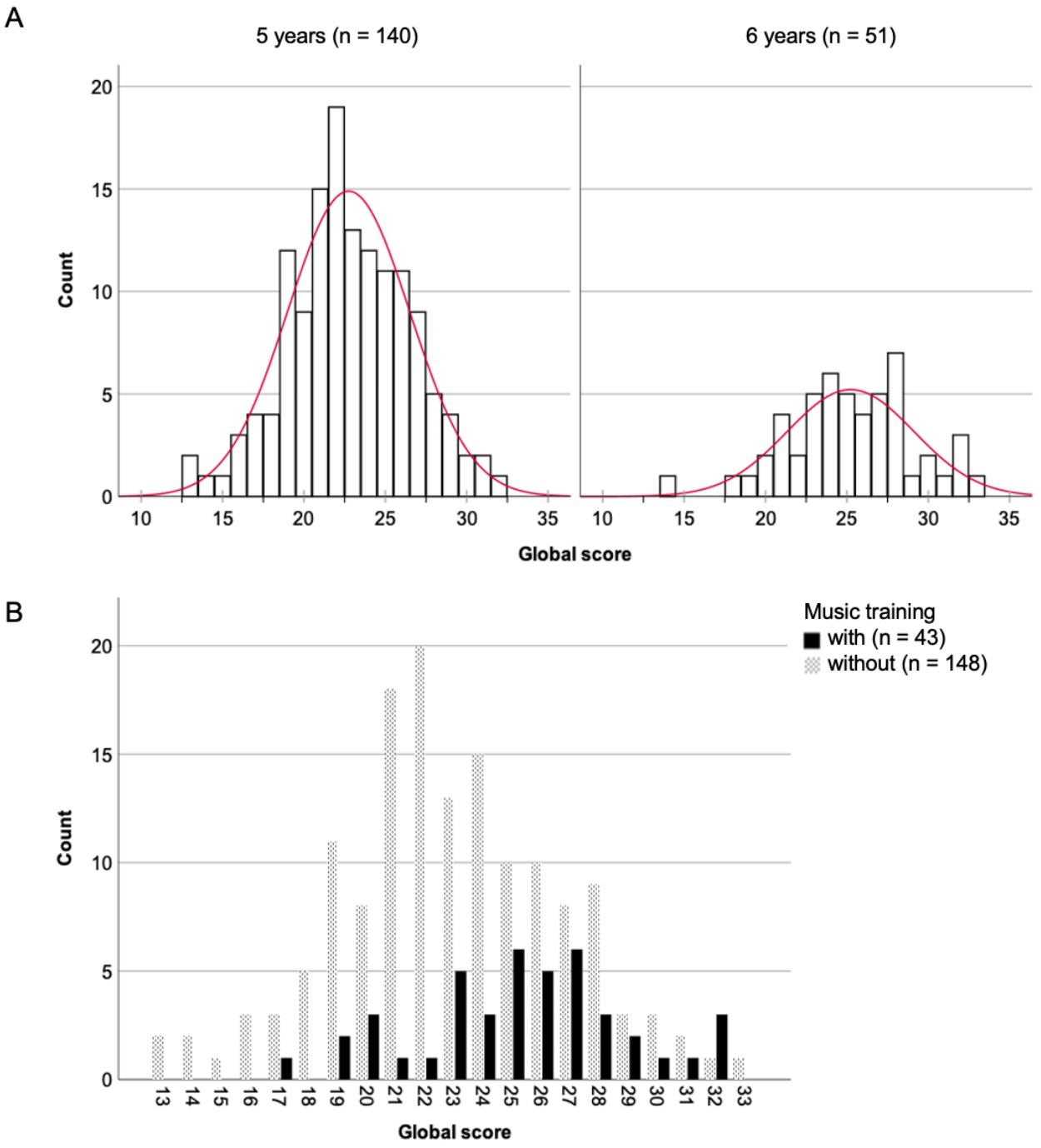

Figure 3. Distribution of the global scores as a function of (A) age and (B) music training.

Age

We measured age in months and correlated it with the global score (Figure 4) and the scores obtained in each test. Age predicted global, *melody*, and *rhythm* performance [$r_{(191)}$ = .197, p = .006 ; $r_{(191)}$ = .249, p < .001 ; $r_{(191)}$ = .149, p = .040, respectively], but not *memory* [$r_{(191)}$ = .031, p = .674].

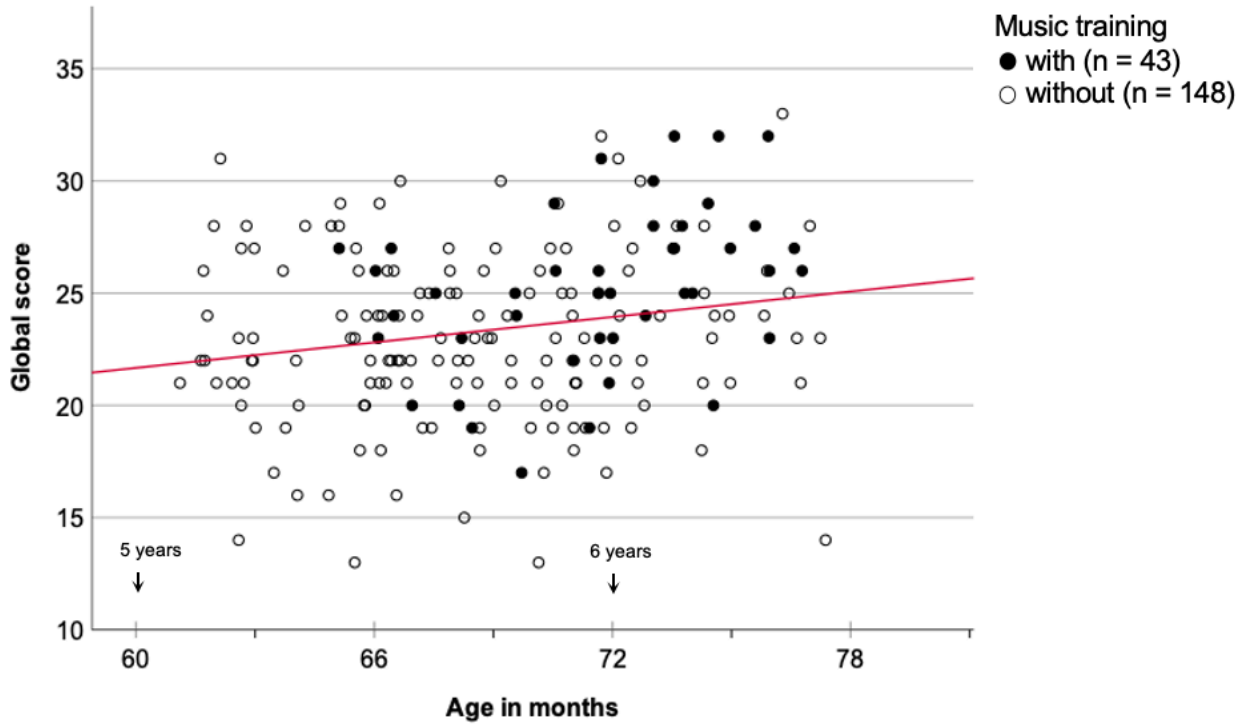

Figure 4. Distribution of individual global scores as a function of age and music training.

Gender

Gender did not influence performance, with girls performing similarly to boys in all tests [all ts < 1].

Music training

We examined the role of music training as a dichotomous variable, by comparing performance between the 43 children with music lessons and the 148 children with no lessons (Figure 3B). To do so, we examined the effect of music training on performance by considering months of age as a covariate. Indeed, opportunities for music training increase with age. The 2 x 3 ANCOVA considering music training (with, without) as a between-subjects factor and test (*melody*, *rhythm*, *memory*) as a within-subject factor revealed a main effect of music training [F(1,188) = 8.16, p = .005, η^2_p = .042] and a significant interaction between test and music training [F(2,376) = 3.13, p = .045, η^2_p = .016], but no other significant main effect or interaction [test: F(2,376) = 2.16, p = .117].

Regarding the interaction between test and music training, post-hoc comparisons using Tukey corrections indicated that, for *melody*, children with music training (M = 7.74, SD = 1.42) did not differ significantly from children without music training (M = 6.71, SD = 1.68) [$t(189) = 2.43$, $p = .148$]. In contrast, for *rhythm*, children with music training (M = 8.91, SD = 2.05) performed significantly better than children without music training (M = 7.70, SD = 1.96) [$t(189) = 3.31$, $p = .013$, $d = .239$]. For *memory*, children with music training (M = 8.61, SD = 1.71) did not differ from children without music training (M = 8.47, SD = 1.78) [$t(189) = .33$, $p = .999$].

Test-retest reliability after three months (n = 89)

Test-retest reliability was examined in the Canadian children who were retested three months later. After three months, children's global score ranged from 17 to 33 (M = 26.09, SD = 2.93), with no children obtaining a perfect score. Again, the distribution of performance was skewed but did not violate normality [$W_{(89)} = .979$, $p = .171$, by Shapiro-Wilk test]. Thus, the battery remained sensitive to individual difference three months later despite prior familiarization with the task.

Performance at Test (T1) predicted performance at retest (T2) both for global performance [$r_{(89)} = .767$, $p < .001$] and for each individual test [$r_{(89)} = .664$, $p < .001$, $r_{(89)} = .794$, $p < .001$, $r_{(89)} = .719$, $p < .001$, for *melody*, *rhythm*, and *memory*, respectively], hence showing good reliability.

The evolution of performance between T1 and T2 was examined by performing a 3 x 2 repeated measure ANOVA considering test (*melody*, *rhythm*, *memory*) and testing phase (T1, T2) as within-subject factors. The ANOVA revealed main effects of test [$F(2,176) = 36.53$, $p < .001$, $\eta^2_p = .293$] and testing phase [$F(1,88) = 178.33$, $p < .001$, $\eta^2_p = .670$], and a significant interaction between test and testing phase [$F(2,176) = 12.40$, $p < .001$, $\eta^2_p = .124$]. Post-hoc comparisons using Tukey corrections indicated significant improvements from test to retest for all tests: *melody* performance at T2 (M = 7.92, SD = 1.57) was significantly better than at T1 (M = 6.60, SD = 1.63) [$t(88) = 10.54$, $p < .001$, $d = 1.118$], *rhythm* performance at T2 (M = 8.98, SD = 1.42) was significantly better than at T1 (M = 7.69, SD = 1.83) [$t(88) = 10.28$, $p < .001$, $d = 1.089$], and *memory* performance at T2 (M = 9.19, SD = 1.51) was significantly better than at T1 (M = 8.61, SD = 1.49) [$t(88) = 4.65$, $p < .001$, $d = .492$]. The difference was larger for the rhythm test, explaining the interaction reported above (Figure 5A).

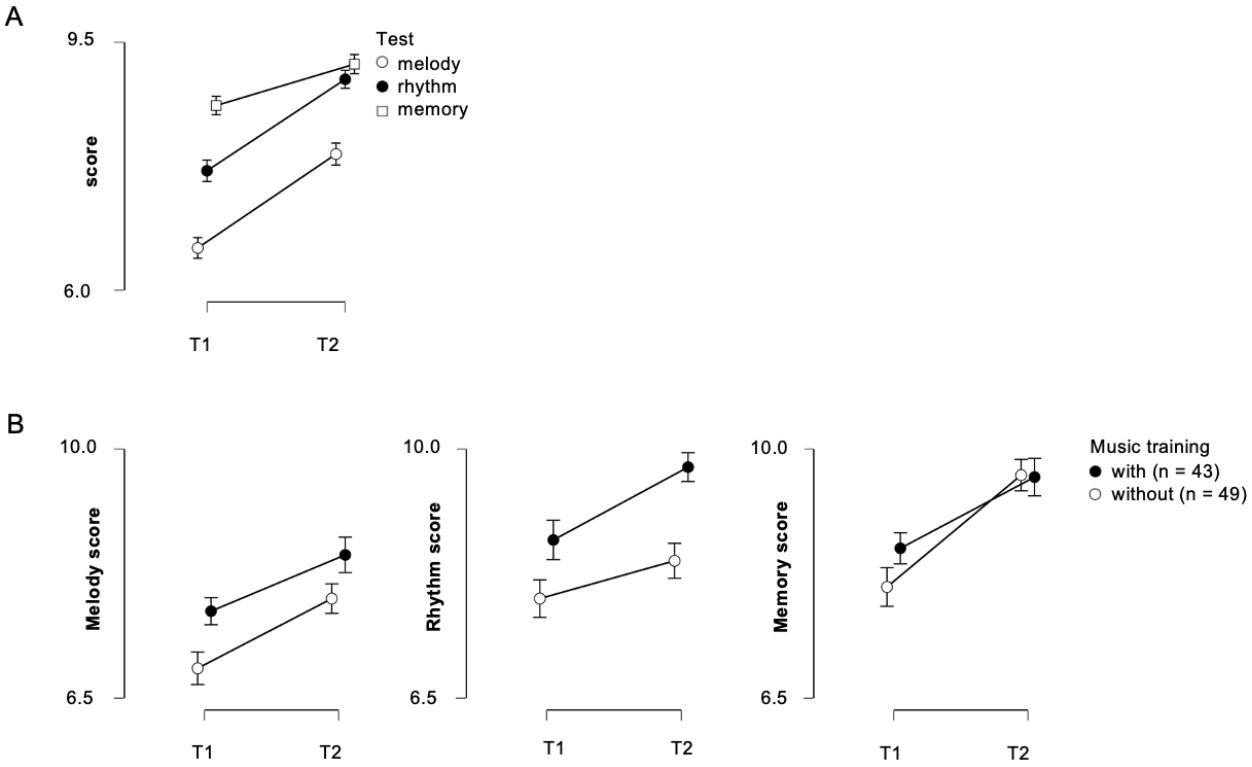

Figure 5. Average scores at test and retest for (A) Canadian children tested 3 months apart and (B) Belgian children tested 8 months apart, as a function of test and/or music training. Error bars represent standard errors.

Test-retest after eight months (n = 92)

As shown in Figure 5B, the 92 Belgian children (43 with music training, 49 without music training), who were examined before and in the middle of the first Grade, also improved performance. Note that six of them started music lessons at the beginning of the first Grade; they were included in the group "with music training". For children without music training, the gain after eight months of schooling (M = 3.08, SD = 5.13) was comparable to the gain after three months (M = 3.20, SD = 2.26), as indicated by an independent samples *t*-test conducted on differences in global scores T2 minus T1 [$t(136) = -.19$, $p = .849$]. Children's global score ranged from 16 to 34 (M = 26.85, SD = 3.80), with no child obtaining a perfect score. The distribution of performance was skewed and violated normality [$W_{(92)} = .972$, $p = .048$, by Shapiro-Wilk test]. Thus, the battery loses sensitivity by the middle of the first Grade.

The evolution of performance between T1 and T2 was examined by performing a 2 x 3 x 2 mixed-design ANOVA considering music training (with, without) as a between-subjects factor, and test (*melody*, *rhythm*, *memory*) and testing phase (T1, T2) as within-subject factors. The ANOVA revealed main effects of music training [$F(1,90) = 8.67$, $p = .004$, $\eta^2_p = .088$], test [$F(2,180) = 27.53$, $p < .001$, $\eta^2_p = .234$], and testing phase [$F(1,90) =$

43.47, $p < .001$, $\eta^2_p = .326$], but no significant interaction. Regarding the main effect of music training, post-hoc comparisons using Tukey corrections indicated that children with music training (M = 9.29, SD = 1.84) performed significantly better than children without music training (M = 8.65, SD = 1.87) [$t(90) = 2.94$, $p = .004$, $d = .307$]. Regarding the main effect of test, post-hoc comparisons using Holm corrections indicated that *melody* performance (M = 8.18, SD = 1.83) was significantly worse than both *rhythm* (M = 9.04, SD = 1.93) [$t(91) = -5.88$, $p < .001$, $d = -.613$] and *memory* performance (M = 9.62, SD = 1.60) [$t(91) = -6.81$, $p < .001$, $d = -.710$], and *rhythm* performance was similar to *memory* performance [$t(91) = -1.70$, $p = .094$]. Regarding the main effect of testing phase, post-hoc comparisons using Holm corrections indicated that T2 performance (M = 8.95, SD = 1.88) was significantly better than T1 performance (M = 7.96, SD = 2.01) [$t(91) = 6.66$, $p < .001$, $d = .694$]. As can be seen in Figure 5B, performance increased comparably between children who received music training and those who did not over time on all three tests.

In conclusion, the tablet version of the MBEMA appears adequate to test 5-to-6-year-old children, and provides a welcome addition to the original MBEMA which was not adequate for children who were younger than six years. It remains to be established if the MBEMA on tablet is adequate for children younger than 5. Preliminary results obtained in Canada are encouraging for the 4-year-olds.

Acknowledgements

MC is supported by a PhD Grant from The Belgian Kids' Fund for Pediatric Research ("Bourse en mémoire du Professeur André Kahn"), by the David and Alice Van Buuren Fund, and by the Jaumotte-Demoulin Foundation and was supported by the ULB Neuroscience Institute (UNI). We thank all the children, schools and families who collaborated to this study and thank Chloé Bertrand, Félicie Cazaubon, Martin Colleu, Mélanie Evrard, Marine Lefebvre, and Camille Lefèvre for testing. JB is grateful to the *Centre de services scolaire des Découvreurs* in Quebec.

Competing interests

The authors declare no competing interests.

ORCID

Isabelle Peretz http://orcid.org/0000-0003-3572-0262
Jonathan Bolduc http://orcid.org/0000-0001-9737-6175
Manon Couvignou https://orcid.org/0000-0002-6652-5286

References

Couvignou, M., Ramus, F., Kolinsky, R. (in preparation). Literacy, phonological, and musical skills in children with and without familial risk for dyslexia: a follow-up study from kindergarten to Grade 2.

Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., & Béland, R. (2013). A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in systems neuroscience*, *7*, 30.