Individual Differences in Musical Ability are Stable Over Time in Childhood

Haley E. Kragness*[1,2], Swathi Swaminathan*[1,3], Laura K. Cirelli[2], and E. Glenn Schellenberg[1,4]

*indicates shared first-authorship

[1]Department of Psychology, University of Toronto Mississauga

[2]Department of Psychology, University of Toronto Scarborough

[3]Brain and Mind Institute, University of Western Ontario

[4]Centro de Investigação e Intervenção Social (CIS-IUL), Instituto Universitário de Lisboa

Author Contributions: All authors contributed to the conceptualization and design of the study. SS and HEK led data collection. HEK and EGS analyzed the data and wrote the manuscript, with contributions from SS and LKC.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

Data Availability Statement: The data from this study are available upon request to the corresponding author. The data are not openly available due to restrictions in our ethics protocol.

Acknowledgements: Funded by the Natural Sciences and Engineering Research Council of Canada and the Fundação para a Ciência e a Tecnologia (FCT) em Portugal. Our thanks to the families who volunteered their time to participate, and to Shreya Jha, Sarah Dehbi, Sarah Selvadurai, Lily Ho, Salma Faress, and Erika Wharton-Shukster, for assistance with recruiting, testing, and data entry.

Abstract

The development of human abilities stems from a complex interplay between genetic predispositions and environmental factors. Numerous studies have compared musicians with non-musicians on measures of musical and non-musical ability, frequently attributing musicians' superior performance to their training. By ignoring preexisting differences, however, this view assumes that taking music lessons is akin to random assignment. In the present longitudinal study, the musical ability of 5- to 10-year-olds was measured at Time 1 with a test of music perception and cognition. Five years later, at Time 2, the children took the same test and a second test designed for older listeners. The test-retest correlation for aggregate scores was remarkably high, $r \approx .7$, and remained strong when confounding variables (age, cognitive abilities, personality) were held constant. At both time points, music training was associated with musical ability, but the association at Time 2 became nonsignificant when musical ability at Time 1 was held constant. Time 1 musical ability also predicted duration of *subsequent* music training. These data are consistent with results from genetic studies, which implicate genes in all aspects of musical behavior and achievement, and with meta-analyses, which indicate that transfer effects from music training are weak. In short, early musical abilities significantly predicted later abilities, demonstrating that individual differences are stable over time. We found no evidence, however, to suggest that music training predicted musical ability after accounting for prior ability. The results underscore the importance of considering preexisting abilities in any type of learning.

Keywords: music, aptitude, expertise, training, melody, rhythm

Research Highlights

- Individual differences in musical ability were longitudinally stable ($r \approx .7$) over a 5-year period in childhood
- Music training did not significantly predict musical ability at Time 2 after controlling for musical ability 5 years prior
- Early musical ability predicted duration of music training over the subsequent 5 years
- Early musical abilities appear to influence which children seek out, stick with, and benefit from music training

Individual Differences in Musical Ability are Stable Over Time in Childhood

The prevalence of music is consistent with the idea that musical ability is part of human nature (Honing, 2018). Although accomplished musical *performance* is limited to a certain few, musical *competence* appears to be universal. Through simple exposure to music, such competence allows individuals with no music lessons to acquire fine-grained implicit knowledge of their native musical systems and to respond emotionally but consistently to short musical excerpts (Bigand & Poulin-Charronnat, 2006). Musically untrained individuals can also recognize an almost unlimited number of melodies (Halpern & Bartlett, 2010), even without words, and even when the key, tempo, and timbre are unfamiliar (Schellenberg & Habashi, 2015).

As with any ability, musical ability varies across individuals. In the present investigation, we sought to determine whether individual differences in musical ability are stable over time. Ericsson et al. (1993) documented that highly accomplished musicianship is the consequence of years of deliberate practice. Ericsson's emphasis on learning and practice as the dominant origin of musical ability has been challenged in recent years, however, by evidence documenting genetic contributions to numerous aspects of musical abilities and behaviors in the general population (e.g., Tan et al., 2014; Mosing et al. 2017).

Because genetic and environmental influences are inseparable, the relative influence of environmental factors and predispositions is not clear cut for most individuals in the general population, who are neither professional musicians nor tone-deaf. Phenotypes are shaped by gene-environment feedback loops in which individuals are drawn to certain environmental factors, such as music training, by virtue of having a particular genetic profile, such as good listening skills (Scarr & McCartney, 1983). Accordingly, beat- or tone-deaf children are unlikely to become professional musicians no matter how many years of training are administered. Equally unlikely candidates would be children with good natural abilities but no opportunity for lessons. In our view, phenotypic musical ability results from a genetic profile that includes certain predispositions (e.g., listening skills) working in combination with environmental factors (e.g., music lessons, socio-economic status, supportive social networks). Indeed, both musical ability and the propensity to practice are genetically influenced (Hambrick & Tucker-Drob, 2015; Mosing et al., 2014).

Historically, objective tests of musical ability were designed for children, in order to determine whether their natural ability (i.e., musical *aptitude*) made them appropriate candidates for music training (for a review see Schellenberg & Weiss, 2013). The term “aptitude” has since fallen out of favor, with less deterministic terms such as *competence* (Wallentin et al. 2010), *ability* (Law & Zentner, 2012; Peretz et al., 2013), or *musicality* (Müllensiefen et al., 2014; Ullén et al., 2014) being used in its place. Regardless of nomenclature or publication date, most objective tests measure how well listeners perceive and remember auditory sequences comprising notes and beats that vary in pitch, time, timbre, and so on. They typically use same-different (AX) tasks, with subtests indexing melody (pitch) and rhythm (time) discrimination abilities (e.g., Law & Zentner, 2012; Peretz et al., 2013; Ullén et al., 2014; Wallentin et al., 2010). Although positive correlations between test scores and music training are used to provide evidence for test validity, the causal direction of this association is unclear. Music training may enhance performance, but listeners who perform well on such tests may also be more likely than other individuals to take music lessons.

The present 5-year longitudinal study is the first to examine whether individual differences in musical ability are stable over time in childhood. Stability here refers to differences among individuals—whether high (or low) performers at Time 1 remain high (or low) performers at Time 2. As with IQ and other psychological traits that are known to be stable (Mackintosh, 2011), absolute levels of ability would improve throughout childhood.

Method

The study protocol was approved by the Research Ethics Board at the University of Toronto.

Participants

Participants were recruited from a larger sample of children who came to the laboratory previously (Time 1; Swaminathan & Schellenberg, 2019). Of those 91 children, 44 (21 girls) returned for the present study (Time 2). On average, children were almost exactly 5 years older at Time 2 (M = 13.05, SD = 1.17, range = 10.29-14.80) compared to Time 1 (M = 8.11, SD = 1.22, range = 5.23-9.89).

We attempted to recruit as many children as possible from the earlier sample. Accordingly, sensitivity rather than power analysis was conducted with G*Power 3.1 (Faul et al., 2009). It tested the independent contribution of a single predictor variable in a multiple-regression model with six other predictors. With N = 44, we were 80% certain of detecting an effect size of at least $f^2 = .18$ (partial correlation ≅ .40), conventionally considered to be slightly larger than *medium* in size (two-tailed α= .05; Cohen, 1988).

Comparisons of children in the present sample with those who did not return from the earlier study (Swaminathan & Schellenberg, 2019) revealed no differences in age, gender, mother's education, father's education, family income, working memory, IQ, duration of music training, or performance on three subtests of musical ability, all ps > .1. The one exception was auditory short-term memory (Digit Span-Forward), p = .022, with the present sample having higher scores. With 11 tests conducted simultaneously, however, there was a high probability of a Type I error. In any event, variance did not differ between groups on any measure, all *p*s > .1 (Levene's test).

Measures

At both time points, a background questionnaire asked parents about basic demographics and their child's history of music training. At Time 2, we also included the Goldsmith's Musical Sophistication Index v1.0 (Gold-MSI; Müllensiefen et al., 2014), which provided a Music Training subtest score (hereafter, *GMSI-Music Training*). Measured demographics included annual family income and both parents' highest level of education. For the statistical analyses, we extracted the principal component representing socio-economic status (SES) in order to reduce collinearity and measurement-specific error.

At both timepoints, participants were administered the three-subtest version of the Montreal Battery for Evaluation of Musical Abilities (MBEMA; Peretz et al., 2013). Although the test was designed for typically developing children, we increased its user-friendliness by adding animations (Swaminathan & Schellenberg, 2019). On the Melody and Rhythm subtests (administered in that order), two melodies were played on each trial. Children determined if they were the same or different. On different (10 of 20) trials, one of the notes was displaced in pitch on the Melody subtest; durations of two adjacent tones were swapped on the Rhythm subtest. In a final Memory-for-Music subtest, children heard a single melody on each of 20 trials and judged whether they heard it previously during the Melody or Rhythm subtest. Half of the melodies were old.

The Musical Ear Test (MET; Wallentin et al., 2010) was added at Time 2 because we were concerned that the MBEMA might be too easy. The MET is a 20-min, computer-administered test of music-discrimination ability that has two subtests: Melody and Rhythm (in that order). Both have 52 trials (50% same). Although the test was designed for adults, the children in our sample, who were 10- to 14-year-olds at Time 2, performed well above chance levels on both subtests, *p*s < .001 (*M*s ≈ 68%, chance = 50%). On each trial, they heard two auditory sequences and judged whether they were the same or different. In the Melody and Rhythm subtest, respectively, the stimuli comprised piano melodies and wood-block rhythms. On different trials, one sound was displaced in pitch in the Melody subtest. In the Rhythm subtest, one or more sounds were added or displaced in time.

We operationally defined musical ability as the principal component extracted from three subtests of ability at Time 1, and five subtests at Time 2. Although the subtests were objective indexes of melody and rhythm discrimination, and of musical memory, they did *not* measure *all* aspects of musical ability, or any aspect of performance ability. Nevertheless, they measured ability in a manner that has become the norm historically (e.g., Gordon, 1965, 1982; Seashore, 1915; for review see Schellenberg & Weiss, 2013) and contemporaneously (Law & Zentner, 2012; Peretz et al., 2013; Ullén et al., 2014; Wallentin et al., 2010). Moreover, the principal component (shared variance) was particularly likely to have construct validity and to be free from measurement-specific error. At Time 1, the principal component (hereafter, *musical ability at Time 1*) accounted for 67% of the variance in the three measured subtests (all loadings ≥ .72). At Time 2, the principal component (hereafter, *musical ability at Time 2*) accounted for 58% of the variance in the five measured variables (all loadings ≥ .71).

For detailed information about music training, see Supplementary Table 1. For statistical analyses, we summed duration of training (in months) across instruments (or voice), then square-root transformed the sum because of positive skew, as in previous research (Swaminathan et al., 2018; Swaminathan & Schellenberg, 2017, 2018). The GMSI-Music Training subtest (completed at Time 2 only) considered—in addition to years of formal training—current practice, peak-level practice, music theory, number of instruments played, and hours of attentive listening to music. GMSI-Music Training was correlated with our stand-alone measure (i.e., square-root total duration) at Time 2, $r = .698$, $N = 44$, $p < .001$ (age held constant). In absolute terms, however, GMSI-Music Training had higher correlations with four of five measures of musical ability at Time 2. (The fifth measure, MET-Rhythm scores, was not associated with GMSI-Music Training or duration-of-training scores.) Because GMSI-Music Training indexed multiple facets of training and maximized statistical power, we used this score as our measure of music training at Time 2 throughout the analyses. The pattern of findings remained unchanged when alternate codings of music training were used (Supplementary Table 2).

At Time 1, we also measured auditory short-term memory with Digit-Span Forward, working memory with Digit-Span Backward, and general cognitive ability with the four subtests from the Wechsler Abbreviated Scale of Intelligence, 2nd edition (i.e., Block Design, Matrix Reasoning, Similarities, and Vocabulary; WASI-II, Weschsler, 2011). Principal components analysis of the six variables (raw scores) revealed a two-factor structure, with Digit-Span Forward loading highly onto one factor ($r = .943$), and the other five variables loading onto a second factor ($.667 \leq r\text{s} \leq .819$). A principal component representing *general cognitive ability* was therefore extracted from the four WASI-II subtests and Digit-Span Backward. At Time 1, a parent also completed the Big Five Inventory on the child's behalf, but only scores for openness-to-experience (hereafter, *openness*) were considered because it is the only personality dimension that predicts musical behaviors reliably (Swaminathan & Schellenberg, 2019).

Missing data included eight families who did not report annual income at either time point. At Time 2, a single datum was also missing for one child for MBEMA-Melody and MBEMA-Rhythm scores (equipment failure), for a second child for MET-Rhythm scores (experimenter error), and for a third child for Music Training as measured by the Gold-MSI (participant error). Missing data were replaced with the mean. Thus, the same 44 children were included in all of the statistical analyses.

Procedure

At both time points, children were tested individually and a parent completed the background questionnaire. At Time 1, children completed the MBEMA in a single session (< 30 min). All children had been to the laboratory before (*M* = 43 days earlier), when they were administered tests of language ability (speech perception, receptive grammar), auditory short-term memory (Digit-Span Forward), and general cognitive ability (Digit-Span Backward and the WASI-II). The language data were reported previously (Swaminathan & Schellenberg, 2019). At Time 2, children completed both the MBEMA *and* the MET, with the MET administered before the MBEMA in a single session (75-90 min). A parent assisted the child in completing the Gold-MSI.

Results

Preliminary analyses examined simple associations among variables. The principal analysis involved a linear multiple-regression model predicting musical ability at Time 2 as a function of music training, musical ability at Time 1, general cognitive abilities, and personality. Because children's ages varied widely at both timepoints (i.e., range of approximately 5 years), we included age as a covariate in all analyses. When an analysis included measures from both

Time 1 and Time 2, age at Time 1 *and* the increase in age from Time 1 to Time 2 were both included.

Gender was not associated with musical ability or music training at either time point, *p*s > .07, and not considered further. Likewise, no significant associations were observed with SES, *p*s > .1, so it was not considered further. Musical ability improved from Time 1 to Time 2 on the Melody, $t(43) = 4.95$, Rhythm, $t(43) = 5.64$, and Memory, $t(43) = 5.06$, subtests from the MBEMA, *p*s < .001. Effect sizes (Cohen's *d*) were large, ranging from .75 (Melody) to .85 (Rhythm).

Pairwise correlations between musical-ability measures administered at Time 1 and Time 2 are provided in Table 1. We also asked whether stability over time varied according to the particular subtest. As shown in Table 1, all but three of the correlations were significant. The exceptions included the correlations between Time 1 MBEMA-Melody and Time 2 MBEMA-Melody scores, Time 1 MBEMA-Memory and Time 2 MBEMA-Melody scores, and Time 1-Memory and Time 2 MET-Melody scores. Unexpectedly, the six associations with performance at Time 2 were higher for Rhythm at Time 1 (Table 1, column 2) than they were for Melody at Time 1 (column 1) or Memory at Time 1 (column 3), *p*s = .031 (sign test, two-tailed). Perhaps compared to the other subtests, scores on the Rhythm subtest at Time 1 tapped more into children's working-memory ability, which then contributed to performance at Time 2. When we tested this possibility by controlling for working memory, response patterns did not change (Supplementary Table 3).

The scatterplot in Figure 1 illustrates the strong association ($r = .668$) between musical ability at Time 1 and musical ability at Time 2. We also calculated Spearman's rank-order correlation, using the residuals after regressing musical ability (aggregate scores) on age. The rank-order association between musical ability at Time 1 and musical ability at Time 2 was strong, $r_s = .621$, $p < .001$, although on average, each child's rank changed by 8.5 places (*SD* = 6.6).

The data point marked with an arrow in Figure 1 represents a child who was an outlier, with a score at Time 2 that was more than 2 *SD* above the regression line. When this child was excluded from the sample, the association between musical ability at Time 1 and musical ability at Time 2 increased in magnitude: $r = .739$ and $r_s = .691$, *p*s < .001. This child was nevertheless included in subsequent analyses.

As expected, musical ability was correlated with duration of training at Time 1, $r = .365$, $p = .016$, and with GMSI-Music Training at Time 2, $r = .508$, $p < .001$. Musical ability at Time 1 also predicted months of music lessons (square-root transformed) that the children took between Time 1 and Time 2, $r = .374$, $p = .015$. In this instance, the additional lessons could not have caused *earlier* levels of musical ability. By contrast, Time 1 ability could have influenced the likelihood of taking music lessons between Times 1 and 2.

We then asked whether musical ability at Time 2 was better explained by musical ability at Time 1, or by GMSI-Music Training at Time 2. The association between musical ability at Time 1 and musical ability at Time 2 remained significant when GMSI-Music Training was held constant, $r = .559$, $p < .001$. By contrast, the correlation between musical ability at Time 2 and GMSI-Music Training was not significant when musical ability at Time 1 was held constant, $r = .292$, $p = .064$. Bayesian analyses (conducted with JASP and default priors; JASP Team, 2019) revealed that with GMSI-Music Training held constant, the observed data were 195 times more likely with a model that included musical ability at Time 1. With musical ability at Time 1 held constant, the observed data were equally likely with a model that included or excluded GMSI-

Music Training, such that the Bayes factor (BF_{10}) was very close to 1 (i.e., 1.37). According to common heuristics (Jarosz & Wiley, 2014; Jeffreys, 1961), the observed data provided *decisive* evidence for the partial association with musical ability at Time 1, but only *anecdotal* evidence for a partial association with GMSI-Music Training, and then only because the Bayes factor was slightly greater rather than less than 1. We also asked whether the GMSI-Music Training variable was associated with *change* in ability scores over time. It was not, whether we examined change in aggregate scores, $p > .7$, or change in MBEMA total scores, $p > .4$.

We also considered other measures that were likely to co-vary with musical ability (both timepoints) or with GMSI-Music Training, specifically auditory short-term memory, general cognitive ability, and openness, which were measured only at Time 1. Correlations are provided in Table 2. As auditory short-term memory improved, so did musical ability at Time 1 and Time 2. General cognitive ability was associated positively with GMSI-Music Training. Finally, higher levels of openness were accompanied by higher levels of musical ability at Time 1 and at Time 2, and higher levels of GMSI-Music Training.

In the final analysis, multiple linear regression was used to predict musical ability at Time 2 from musical ability at Time 1, GMSI-Music Training (Time 2), auditory short-term memory (Time 1), general cognitive ability (Time 1), and openness (Time 1). Age at Time 1 and the increase in age from Time 1 to Time 2 were included as covariates. The results are summarized in Table 3 and Figure 2. Tolerance values for all predictors were greater than 0.5 (Table 3), which confirmed that multicollinearity was not a major problem. The multiple-regression model accounted for almost two-thirds (64.0%) of the variance in musical ability at Time 2 (multiple $R = .800$), but only musical ability at Time 1 and auditory short-term memory made significant independent contributions to the model. The standardized slope indicated that a difference of one *SD* in musical ability at Time 1 was predictive of a difference of 43.5% of one *SD* in musical ability at Time 2, even with all other predictor variables held constant. GMSI-Music Training did not make a significant independent contribution to the model, and Bayesian analyses confirmed that the observed data were equally likely with a model that either included or excluded GMSI-Music Training (BF_{10} = 1.32). By contrast, the observed data were 22.9 times more likely with a model that included musical ability at Time 1.

We also formed a variable representing the interaction between musical ability at Time 1 and GMSI-Music Training at Time 2, and added it to the multiple-regression model tested above. The interaction term was not a significant predictor of musical ability at Time 2, $p > .5$ (see Supplementary Table 3). Hence, there was no evidence that music training moderated the stability of musical ability over time.

Discussion

Our major findings were twofold. First, although musical ability improved dramatically over time, it was remarkably stable in terms of who performed well or poorly. Second, we found no evidence that music training influenced stability or change in musical ability.

Consider the test-retest correlation for musical ability (≈ .7). In a study of children similar in age to ours, the 5-year test-retest correlation for general intelligence, sometimes considered to be the most stable psychological construct (Caspi et al., 2005), was .75 (Schneider et al., 2014). Personality is perhaps the next most stable psychological construct (Caspi et al., 2005), yet trait consistency is considerably lower than what we observed for musical ability, around .43 for 6- to 12-year-olds (Roberts & DelVecchio, 2000). In short, musical ability is very stable.

We also observed substantial change over time. For example, when children were ranked at both timepoints, the average child shifted more than 8 positions over 5 years. The source of such change remains a topic for future research. The outlier in Figure 1 was 1.7 *SD* below the mean at Time 1, yet 0.8 *SD* above the mean at Time 2. Perhaps inattention at Time 1, when the child was approximately 7.5 years of age, played a role in the low score. Other trait and state variables would also affect the rank order of individual scores.

Music training did not significantly predict Time 2 musical ability above Time 1 ability, or change in ability over time, and Bayesian statistics revealed that the observed data were more or less equally likely under the null and alternative hypotheses. Thus, we did not find compelling evidence for an association or no association with music training. Rather, if such an association exists, it is almost certain to be weak, and much weaker than the stability of musical ability over time. These results raise doubts about causal claims of training or practice effects on musical ability (Schellenberg, 2016). Training is clearly important for advanced performance skills (Ericsson et al., 1993), but in our sample, early musical ability predicted both later musical ability *and* duration of subsequent music training, underscoring genetic *and* gene-environment contributions to musical ability.

Our findings also challenge claims about far-transfer effects of music training (Kraus & Chandrasekaran, 2010; Patel, 2011), of which the vast majority are made from correlational data (Schellenberg, 2019a, 2019b). If music training does not improve musical ability substantially, it would be even less likely to improve skills in other domains. Indeed, other findings indicate that the link between music training and musical ability is surprisingly modest (Swaminathan & Schellenberg, 2018), and that language abilities in childhood and adulthood are better predicted by musical ability than they are by music training (Swaminathan & Schellenberg, 2017, 2019;

Swaminathan et al., 2018). A recent meta-analysis confirms, moreover, that the effect of music training on other domains is negligible (Sala & Gobet, 2020). For special populations, however, such as children with dyslexia or hearing loss, targeted music-training interventions, particularly those focusing on rhythm training, may indeed have beneficial non-musical effects (Flaugnacco et al., 2015; Hidalgo et al., 2017, 2019).

A surprising result emerged when we examined associations among the different musical-ability subtests. Specifically, performance on the MBEMA-Rhythm subtest at Time 1 had the strongest associations with *all* measures at Time 2. In terms of absolute magnitude, the association with MBEMA-Rhythm was even larger than test-retest correlations for the same subtests (i.e., MBEMA-Melody, MBEMA-Memory), which were administered identically at Times 1 and 2. This finding raises the possibility that precociousness in rhythm processing may be especially useful for identifying early musical ability.

As in previous research (Corrigall et al., 2013; Schellenberg, 2006, 2011a, 2011b; Swaminathan & Schellenberg, 2017, 2018, 2019; Swaminathan et al., 2017, 2018), musical ability and music training covaried with non-musical individual differences in auditory short-term memory, general cognitive ability, and openness. Only auditory short-term memory, however, made a significant independent contribution to musical ability at Time 2, along with musical ability at Time 1. The contribution of auditory short-term memory can be explained in a straightforward manner based on the format of auditory same-different tasks, which require participants to compare two auditory sequences heard in succession. Because most contemporary tests of musical ability use same-different tasks (Law & Zentner, 2012; Peretz et al., 2013; Ullén et al., 2014; Wallentin et al., 2010), auditory short-term memory is likely to contribute to performance on all such tests (e.g., Hansen et al., 2013; Swaminathan et al., in press; Wallentin et al., 2010). By contrast, performance on tests of other musical abilities, such as determining whether a rhythm is a waltz or a march (Peretz et al., 2003), or whether the vocal track from a familiar recording is mistuned (Larrouy-Maestri et al., 2019), could be independent of auditory short-term memory.

In sum, there is no doubt that music lessons and practice train the fine motor skills and procedural knowledge required to play specific instruments, and the declarative information and stylistic nuances that support expert performance (Ericsson et al., 1993). Nevertheless, the superior listening skills exhibited by musicians over non-musicians appear to reveal a classic interaction between genes and the environment. Trait-like predispositions increase or decrease the likelihood that individuals choose to take music lessons.

References

Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100–130. https://doi.org/10.1016/j.cognition.2005.11.007

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, *56*(1), 453–484. https://doi.org/10.1146/annurev.psych.55.090902.141913

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology, 4:* 222. https://doi.org/10.3389/fpsyg.2013.00222

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100, 363–406.* http://doi.org/10.1037/0033-295X.100.3.363

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

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

Gordon, E. (1965). *Musical aptitude profile*: *Manual*. Houghton Mifflin.

Gordon, E. E. (1982). *Intermediate measures of music audiation* (Grade 1-6). GIA Publications.

Halpern, A. R., & Bartlett, J. C. (2010). Memory for melodies. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Springer handbook of auditory research: Vol. 36. Music perception* (pp. 233-258). Springer. https://doi.org/10.1007/978-1-4419-6114-3_8

Hambrick, D. Z., & Tucker-Drob, E. M. (2015). The genetics of music accomplishment: Evidence for gene–environment correlation and interaction. *Psychonomic Bulletin & Review*, *22*(1), 112–120. https://doi.org/10.3758/s13423-014-0671-9

Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. *Psychology of Music, 41*(6), 779-793. https://doi.org/10.1177/0305735612452186

Hidalgo, C., Falk, S., & Schön, D. (2017). Speak on time! Effects of a musical rhythmic training on children with hearing loss. *Hearing Research, 351*, 11-18. http://dx.doi.org/10.1016/j.heares.2017.05.006

Hidalgo, C., Pesnot-Lerousseau, J., Marquis, P., Roman, S., & Schön, D. (2019). Rhythmic training improves temporal anticipation and adaptation abilities in children with hearing loss during verbal interaction. *Journal of Speech, Language, and Hearing Research, 62*(9), 3234-3247. https://doi.org/10.1044/2019_JSLHR-S-18-0349

Honing, H. (Ed.) (2018). *The origins of musicality*. MIT Press.

Jarosz, A., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving, 7*(1), 2–9. http://dx.doi.org/10.7771/1932-6246.1167

JASP Team (2019). JASP (Version 0.10.2) [Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11*, 599-605. https://doi.org/10.1038/nrn2882

Larrouy-Maestri, P., Harrison, P. M. C., & Müllensiefen, D. (2019). The mistuning perception test: A new measurement instrument. *Behavior Research Methods, 51*(2), 663-675. https://doi.org/10.3758/s13428-019-01225-1

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PLoS ONE*, *7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

Mackintosh, N. (2011). *IQ and human intelligence* (2nd ed.). Oxford University Press.

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: No causal effect of music practice on music ability. *Psychological Science*, *25*(9), 1795–1803. https://doi.org/10.1177/0956797614541990

Mosing, M. A., Peretz, I., & Ullén, F. (2017). Genetic influences on music expertise. In D. Z. Hambrick, G. Campitelli, & B. N. Macnamara (Eds.), *The science of expertise: Behavioral, neural, and genetic approaches to complex skill* (pp. 272-283). Routledge Psychology Press.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2*:142. https://doi.org/10.3389/fpsyg.2011.00142

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999*, 58-75. https://doi.org/10.1196/annals.1284.006

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. https://doi.org/ 10.3389/fnsys.2013.00030

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3–25. https://doi.org/10.1037/0033-2909.126.1.3

Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. Advance online publication. *Memory & Cognition*. https://doi.org/10.3758/s13421-020-01060-2

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype–>environment effects. *Child Development*, *54*(2), 424. https://doi.org/10.2307/1129703

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology, 98*(2), 457-468. https://doi.org/10.1037/0022-0663.98.2.457

Schellenberg, E. G. (2011a). Examining the association between music lessons and intelligence. *British Journal of Psychology, 102*(3), 283-302. https://doi.org/10.1111/j.2044-8295.2010.02000.x

Schellenberg, E. G. (2011b). Music lessons, emotional intelligence, and IQ. *Music Perception, 29*(2), 185-194. https://www.jstor.org/stable/10.1525/mp.2011.29.2.185Schellenberg, E.

G. (2016). Music and nonmusical abilities. In G. E. McPherson (Ed.), *The child as musician: A handbook of musical development* (2nd ed., pp. 149-176). Oxford University Press.

Schellenberg, E. G. (2016). Music and nonmusical abilities. In G. E. McPherson (Ed.), *The child as musician: A handbook of musical development* (2nd ed., pp. 149-176). Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198744443.001.0001

Schellenberg, E. G. (2019a). Correlation = causation? Music training, psychology, and neuroscience. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. https://doi.org/10.1037/aca0000263

Schellenberg, E. G. (2019b). Music training, music aptitude, and speech perception. *Proceedings of the National Academy of Sciences USA, 116*(8), 2783-2784. https://doi.org/10.1073/pnas.1821109116

Schellenberg, E. G., & Habashi, P. (2015). Remembering the tune and timbre, forgetting the key and tempo. *Memory & Cognition, 43*(7), 1021-1031. https://doi.org/10.3758/s13421-015-0519-1

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 499-550). Elsevier. http://dx.doi.org/10.1016/B978-0-12-381460-9.00012-2

Schneider, W., Niklas, F., & Schmiedeler, S. (2014). Intellectual development from early childhood to early adulthood: The impact of early IQ differences on stability and change over time. *Learning and Individual Differences, 32*, 156-162. https://doi.org/10.1016/j.lindif.2014.02.001

Seashore, C. E. (1915). The measurement of musical talent. *Musical Quarterly, 1*(1), 129-148. https://doi.org/10.1093/mq/I.1.129

Swaminathan, S., & Kragness, H. E., & Schellenberg, E. G. (in press). The Musical Ear Test: Norms and correlates from a large sample of Canadian undergraduates. *Behavior Research Methods*.

Swaminathan, S., & Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychonomic Bulletin & Review, 24*(6), 1929-1934. https://doi.org/10.3758/s13423-017-1244-5

Swaminathan, S., & Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. *Scientific Reports, 8*:9223. https://doi.org/10.1038/s41598-018-27571-2

Swaminathan, S., & Schellenberg, E. G. (2019). Musical ability, music training, and language ability in childhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0000798

Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? *Intelligence, 62*, 119-124. https://doi.org/10.1016/j.intell.2017.03.005

Swaminathan, S., Schellenberg, E. G., & Venkatesan, K. (2018). Explaining the association between music training and reading in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(6), 992-999. http://dx.doi.org/10.1037/xlm0000493

Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology, 5*:658. https://doi.org/10.3389/fpsyg.2014.00658

Ullén, F., Mosing, M. A., Holm, L., Eriksson, H., & Madison, G. (2014). Psychometric properties and heritability of a new online test for musicality, the Swedish Musical Discrimination Test. *Personality and Individual Differences*, *63*, 87–93. https://doi.org/10.1016/j.paid.2014.01.057

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test: A new reliable test for measuring musical competence. *Learning and Individual Differences, 20*(3)*,* 188–196. http://dx.doi.org/10.1016/j.lindif.2010.02.004

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence – Second Edition*. Pearson.

Table 1

Pairwise Correlations Between Musical-Ability Variables at Time 1 and Time 2 (Age Held Constant)

	Time 1—MBEMA			
Time 2	Melody	Rhythm	Memory	Principal Component
MBEMA-Melody	.262	.496*	.290	.440*
MBEMA-Rhythm	.306*	.490*	.319*	.466*
MBEMA-Memory	.367*	.549*	.473*	.580*
MET-Melody	.551*	.580*	.229	.558*
MET-Rhythm	.418*	.433*	.373*	.505*
Principal Component	.497*	.670*	.441*	.668*

Note. $^*p < .05$ (two-tailed)

Table 2

Pairwise Correlations Between Non-Musical Predictors and Musical Outcome Variables (Age Held Constant)

Predictor Variable (Time 1)	Musical Ability Time 1	Musical Ability Time 2	Music Training Time 2
Auditory Short-Term Memory	.398*	.568*	.246
General Cognitive Ability	.282	.174	.449*
Openness	.344*	.405*	.455*

**p* < .05 (two-tailed)

Table 3

Results from Linear Multiple Regression Predicting Musical Ability at Time 2

	Musical Ability Time 2			
Predictor Variable	β	t	p	*Tolerance*
Musical Ability Time 1	.435	3.32	.002	.584
Music Training (Gold-MSI) Time 2	.249	1.82	.077	.538
Auditory Short-Term Memory Time 1	.338	3.05	.004	.815
General Cognitive Ability Time 1	-.167	-1.21	.232	.526
Openness Time 1	.094	0.81	.424	.745
Age Time 1	.092	0.71	.481	.602
Increase in Age	-.026	-0.24	.809	.849
Model				
$R^2 = .640$, Adjusted $R^2 = .570$				
$F(7, 36) = 9.13$, $p < .001$				

Note. Tolerance measured the proportion of the variance for each predictor variable that was independent of all other predictors.

Figure 1

Scatterplot Depicting the Correlation Between Musical Ability at Time 1 and Musical Ability at Time 2—the Arrow Points to an Outlier

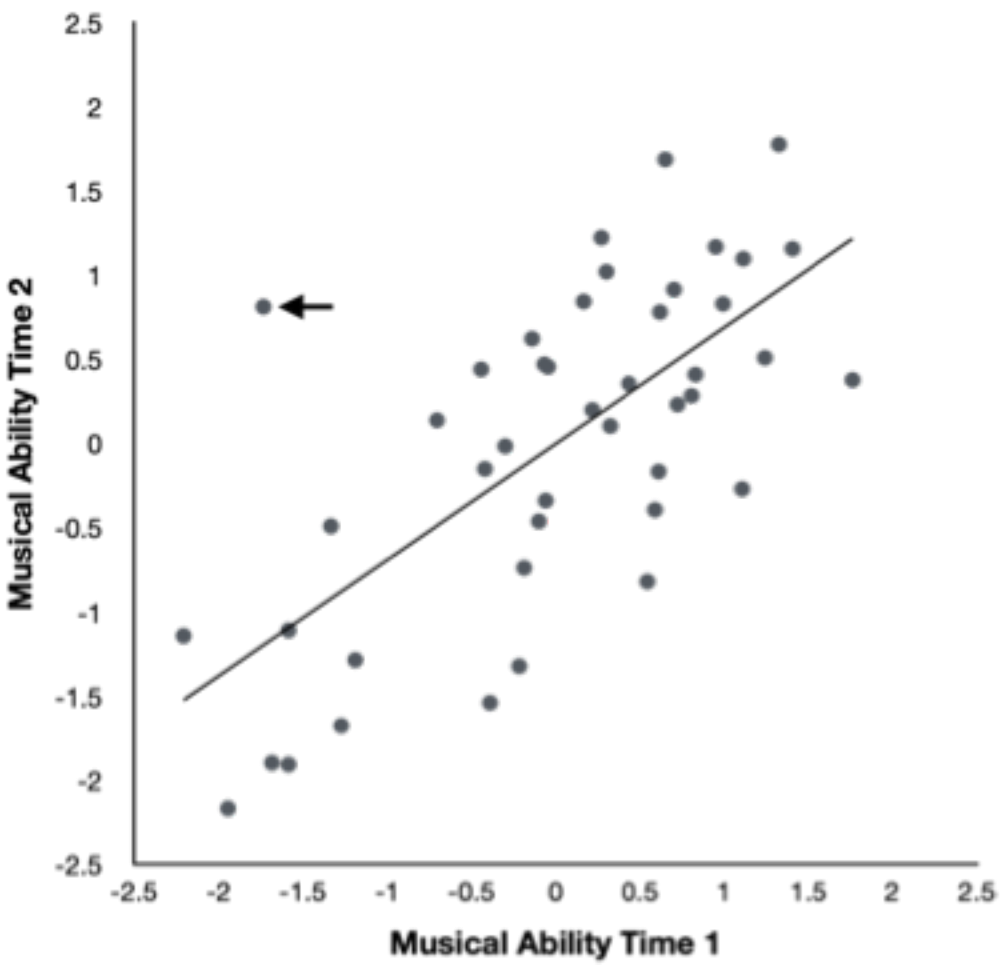

Figure 2

Standardized Slopes and 95% Confidence Intervals from Linear Multiple Regression Predicting Musical Ability at Time 2: Only Musical Ability at Time 1 and Auditory Short-Term Memory Made Independent Contributions to the Model