Parnautt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Explaining Objective and Subjective Aspects of Musical Sophistication: Insights from General Fluid Intelligence and Working Memory

David John Baker,[1] Emily M. Elliott[2], Daniel Shanahan[3], Juan Ventura[4], Elizabeth Monzingo[5], Brian Ritter[6] and Christopher Young[7]

[1,3,5,6] *School of Music, Louisiana State University, United State of America*
[2,4,7] *Psychology Department, Louisiana State University, United States of America*

[1]dbake29@lsu.edu, [2] eelliott@lsu.edu, [3]daniel.shanahan@gmail.com,
[4]jventu4@lsu.edu,[5]emonzi1@lsu.edu,[6]britte1@lsu.edu,[7]cyoun73@lsu.edu

Abstract

Recent work in music psychology has examined the relationship between individual differences and factors that predict various aspects of musical sophistication. Some of the recent research has begun to model how musical sophistication or aptitude relates to various cognitive measures, ranging from executive functions, to measures of general fluid intelligence. Recent research has also investigated how differences in musical training may lead to differences in working, short-term, and long-term memory capacity. While some of the previously mentioned work uses continuous measures of musical sophistication, many only collect data on years of formalized musical training as opposed to a more multi-faceted view of musical sophistication. The aim of this paper is to share findings from a large study investigating how musical sophistication, as measured by the Goldsmiths Musical Sophistication Index (Gold-MSI), relates to measures of working memory and general fluid intelligence. Results using structural equation modeling (SEM) suggest working memory capacity and general fluid intelligence explain more of the variance in perceptual tasks than self-report measures of musical sophistication. In light of these findings, we suggest that further models of music perception should focus on modeling what processes contribute to a task, rather than using large, composite latent variables.

Introduction

Music and Cognitive Ability

Relative to other sub-disciplines within the field of music psychology, the relationship between musical training and cognitive ability is one of the older and more researched topics of the discipline. Dating back to 1904, Charles Spearman in his publication *General Intelligence: Objectively Determined and Measured* used tests of pitch perception as a measure to relate to general intelligence or *g*. While perhaps more affected by training than some of the other measures of interest to Spearman (e.g. Language and Mathematics), a complex relationship between various aspects of musical ability, intelligence, and factors that might mediate a relationship between the two have yet to be fully understood.

The past three decades have shown a marked increase in attention to the relationship between music and cognitive ability. Largely responsible for the initial interest in this relationship were findings published in *Nature* by Rausher et al. (1993) suggesting that listening to Mozart could have temporary beneficial effects on boosting an individual's ability to do spatial reasoning. These findings were picked up by researchers as well as the general public, even leading to government-lead initiatives to provide access to Mozart for newborn children with the hopes of instilling some sort of long-term boost in cognitive ability (Ethridge, 1998). Since then, the "Mozart Effect" has largely been reframed as findings that are better explained by a temporary increase in mood or arousal, rather than any sort of boost in cognitive ability (Schellenberg, 2014).

Despite evidence for anything resembling a "Mozart Effect", researchers have continued to explore the extent to which musical training might have a far-transfer effect to other cognitive abilities. In general, children who engage with musical training score better on tests of cognitive ability than children who do not engage in any sort of music lessons (Bigbson, Folley, & Park, 2009; Hille & Schupp, 2011; Schellenberg, 2011a; Schellenberg & Mankarious, 2012). The effects on cognitive ability tend to increase as a function of exposure (Dege & Schwarzer, 2011, Corrigall & Schellenberg, 2015; Schellenberg, 2006) and these differences remain apparent in undergraduates who are no longer engaged with music lessons (Schellenberg, 2006, 2011b).

In addition to correlational research, experimental designs have also been carried out that report mixed results. For example, Schellenberg (2004) assigned 6-year-old Canadian students to either music, voice, drama, or no lessons, and found that while no specific music group outperformed the others, the combined music groups outperformed the non-musical groups on measures of IQ and more specific abilities like attention, processing speed, verbal ability, and spatial ability. Other studies have attempted to replicate findings that show an increase in cognitive ability, but often fall short when they either do not last long enough to instil any sort of effect (Francois et al., 2013; Moreno et al., 2009) or fail to control for pre-existing differences in the participants (Mehr, Schachner, Katz, & Spelke, 2013). In his summary review on Music Training and General Cognitive Abilities, Schellenberg suggests his theory may explain a lot of these phenomena, stating that "the available evidence indicates that high-functioning individuals are likely to take music lessons, and that music lessons may exaggerate slightly their pre-existing advantages" (Schellenberg, 2016, p. 421).

If it is the case that higher functioning individuals are more likely to engage with musical activity, which then is magnified by training, it would be important to investigate which cognitive abilities are driving these relationships and to what extent. Some recent research has attempted to isolate the effects of certain cognitive abilities in reference to tasks of musical perception.

Investigating executive function Slevc et al., (2016) reported that individuals with five or more years of musical training outperformed their non-musical counterparts in both verbal and tonal working memory performance. Working memory capacity (WMC) has also been shown to contribute significantly in models of musical perception. Using a tapping paradigm and measure of two tasks of WMC (backwards digit span and operation span) Colley, Keller and Halpern (2017) reported that WMC predicted performance above and beyond temporal imagery and an auditory image self report measures of musical training. Additionally, Meinz and Hambrick (2010) found that measures of WMC additionally contributed to an individual's ability to perform sight reading at the piano in a sample of 57 musicians. A recent meta-analysis by Talamini et al., 2017 also reported that musicians tend to outperform people without musical training. Recently Swaminathan et al. (2017) reported being able to predict general fluid intelligence using musical aptitude and failed to find any unique contribution of parent's education-- a variable shown to be related to socioeconomic status (SES) that has been shown by previous research to explain individual differences in musical ability.

Nonetheless, these findings still cannot determine causality. There is a need to establish to what degree cognitive ability plays into any sort of model of musical perception.

Predictions from Process Overlap Theory

In addition to all of this research suggesting a relationship between musical and cognitive ability, additional new theoretical frameworks from cognitive psychology have suggested that in order to better understand cognitive processes, future research should model cognitive activity in terms of processes rather than using composite, latent variables as representations of theoretical constructs. Kovacs and Conway (2016) advocate for this position and suggest that cognitive tests, such as those that generate findings from the positive manifold, tap domain-general executive processes identified in working memory research, as well as other domain-specific processes. They suggest these processes are tapped in an overlapping manner across tests such that the general ones are demanded more than specific ones. If this is true, we would then suspect that measures of WMC/Gf should play a significant role in any sort of task that resembles a high executive load task such as the melodic memory task of the Gold-MSI, which requires an individual to retain information and compare it with new incoming information in a new key context. This task should then be more markedly demanding than being able to detect onset asynchrony in a stimulus, which would be more of a base perceptual ability, such as that required in the beat perception task of the Gold-MSI.

Hypotheses

Given the previous literature and theoretical background, we tested several hypotheses.

- H1: We will replicate results from Müllensiefen et al. (2014) and be able to predict scores on both the beat perception and melodic perception objective tests from the self reported sub-scales of the Gold-MSI.
- H2: We will be able to predict variance above and beyond that of the self report Gold-MSI by adding measures of Working Memory Capacity and General Fluid Intelligence into the model.
- H3: In line with predictions from Process Overlap Theory, we believe that due to the nature of the Melodic Perception Task, that WMC/Gf will better predict scores on the Melodic Perception Task than the Beat Perception Task using structural equation modeling.

In order to test these hypotheses, we used structural equation modeling (SEM) to examine the relative contribution of each of our hypothesised variables using a nested models approach (e.g., Shelton et al., 2010).

Methods

Participants

Two hundred fifty-four students enrolled at Louisiana State University completed the study. We recruited students, mainly in the Department of Psychology and the School of Music. The criteria for inclusion in the analysis were no self-reported hearing loss, not actively taking medication that would alter cognitive performance, and univariate outliers (defined as individuals whose performance on any task was greater than 3 standard deviations from the mean score of that task). Using these criteria, eight participants were not eligible due to self reporting hearing loss, one participant was removed for age, and six participants were eliminated as univariate outliers due to performance on one or more of the tasks of working memory capacity. Thus, 239 participants met the criteria for inclusion. The eligible participants were between the ages of 17 and 43 ($M = 19.72$, $SD = 2.74$; 148 females). Participants volunteered, received course credit, or were paid $20.

Cognitive Measures

All variables used for modeling approximated normal distributions. Processing errors for each task were positively skewed for the complex span tasks similar to Unsworth, Redick, Heitz, Broadway, and Engle (2009). Positive and significant correlations were found between recall scores on the three tasks measuring working memory capacity (WMC) and the two measuring general fluid intelligence (Gf). The WMC recall scores negatively correlated with the reported number of errors in each task, suggesting that rehearsal processes were effectively limited by the processing tasks (Unsworth et al., 2009).

Procedure

Participants in this experiment completed eight different tasks, lasting about 90 minutes in duration. The tasks consisted of the Gold-MSI self-report inventory, coupled with the Short Test of Musical Preferences, and a supplementary demographic questionnaire that included questions about socioeconomic status, aural skills history, hearing loss, and any medication that might affect their ability to perform on cognitive tests. Following the survey they completed three WMC tasks: a novel Tonal Span, Symmetry span, and Operation span task; a battery of perceptual tests from the Gold-MSI (Melodic Memory, Beat Perception, Sound Similarity) and two tests of general fluid intelligence (Gf): Number Series and Raven's Advanced Progressive Matrices.

Each task was administered in the order listed above on a desktop computer. Sounds were presented at a comfortable listening level for the tasks that required headphones. All participants provided informed consent and were debriefed. Only measures used in modeling are reported below.

Measures

Goldsmiths Musical Sophistication Index Self Report (Gold-MSI)

Participants completed a 38-item self-report inventory and questions consisted of free response answers or choosing a selection on a likert scale that ranged from 1-7. (Müllensiefen, et al., 2014). The complete survey with all questions used can be found at goo.gl/dqtSaB.

Tone Span (TSPAN)

Participants completed a two-step math operation and then tried to remember three different tones in an alternating sequence (based upon Unsworth et al., 2005). We modelled the three tones after Li, Cowan, Saults (2005) paper using frequencies outside of the equal tempered system (200Hz, 375Hz, 702Hz). The same math operation procedure as OSPAN was used. The tones was presented aurally for 1000ms after each math operation. During tone recall, participants were presented three different options H M and L (High, Medium, and Low), each with its own check box. Tones were recalled in serial order by clicking on each tone's box in the appropriate order. Tone recall was untimed. Participants were provided practice trials and similar to OSPAN, the test procedure included three trials of each list length (3-7 tones), totalling 75 letters and 75 math operations.

Operation Span (OSPAN)

Participants completed a two-step math operation and then tried to remember a letter (F, H, J, K, L, N, P, Q, R, S, T, or Y) in an alternating sequence (Unsworth et al., 2005). The same math operation procedure as TSPAN was used. The letter was presented visually for 1000ms after each math operation. During letter recall, participants saw a 4 x 3 matrix of all possible letters, each with its own check box. Letters were recalled in serial order by clicking on each letter's box in the appropriate order. Letter recall was untimed. Participants were provided practice trials and similar to TSPAN, the test procedure included three trials of each list length (3-7 letters), totalling 75 letters and 75 math operations.

Symmetry Span (SSPAN)

Participants completed a two-step symmetry judgment and were prompted to recall a visually-presented red square on a 4 X 4 matrix (Unsworth et al., 2005). In the symmetry judgment, participants were shown an 8 x 8 matrix with random squares filled in black. Participants had to decide if the black squares were symmetrical about the matrix's vertical axis and then click the screen. Next, they were shown a "yes" and "no" box and clicked on the appropriate box. Participants then saw a 4 X 4 matrix for 650 ms with one red square after each symmetry judgment. During square recall, participants recalled the location of each red square by clicking on the appropriate cell in serial order. Participants were provided practice trials to become familiar with the procedure. The test procedure included three trials of each list length (2-5 red squares), totalling 42 squares and 42 symmetry judgments.

Gold-MSI Beat Perception

Participants were presented 18 excerpts of instrumental music from rock, jazz, and classical genres (Müllensiefen et al., 2014). Each excerpt was presented for 10 to 16s through headphones and had a tempo ranging from 86 to 165 beats per minute. A metronomic beep was played over each excerpt either on or off the beat. Half of the excerpts had a beep on the beat, and the other half had a beep off the beat. After each excerpt was played, participants answered if the metronomic beep was on or off the beat and provided their confidence: "I am sure", "I am somewhat sure", or "I am guessing". The final score was the proportion of correct responses on the beat judgment.

Gold-MSI Melodic Memory Test

Participants were presented melodies between 10 to 17 notes long through headphones (Müllensiefen et al., 2014). There were 12 trials, half with the same melody and half with different melodies. During each trial, two versions of a melody were presented. The second version was transposed to a different key. In half of the second version melodies, a note was changed a step up or down from its original position in the structure of the melody. After each trial, participants answered if the two melodies had identical pitch interval structures.

Number Series

Participants were presented with a series of numbers with an underlying pattern. After being given two example problems to solve, participants had 4.5 minutes in order to solve 15 different problems. Each trial had 5 different options as possible answers (Thurstone, 1938).

Raven's Advanced Progressive Matrices

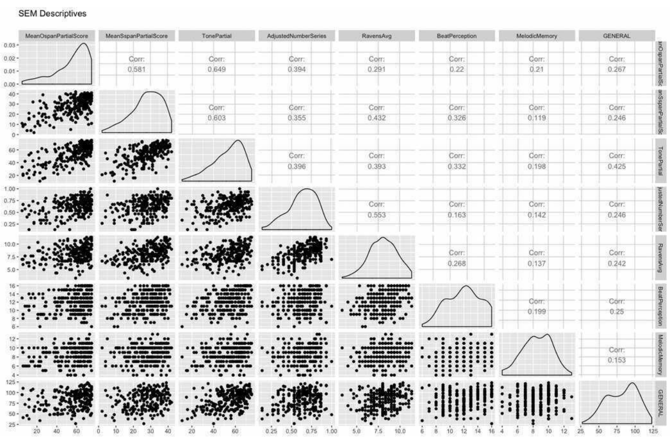
Participants were presented a 3 x 3 matrix of geometric patterns with one pattern missing (Raven et al., 1998). Up to eight pattern choices were given at the bottom of the screen. Participants had to click the choice that correctly fit the pattern above. There were three blocks of 12 problems, totalling 36 problems. The items increased in difficulty across each block. A maximum of 5 min was allotted for each block, totalling 15 min. The final score was the total number of correct responses across the three blocks.

Results

Descriptive Statistics and Data Screening

The goal of the analyses was to examine the relationships among the measures and constructs of WMC, general fluid intelligence, musical sophistication (operationalized as the General score from the Gold-MSI), in relation to the two objective listening tests on the Gold-MSI. Before running any sort of modeling, we inspected our data to ensure in addition to outlier issues as mentioned above, the data exhibited normal distributions. We report both our correlation values, as well as visually displaying our distributions in Figure 1.

Figure 1: Descriptives in Sample



Before running any modeling, we checked our data for assumptions of normality since violations of normality can strongly affect the covariances between items. While some items in Figure 1 displayed a negative skew, many of the individual level items from the self report scale exhibited high levels of Skew and Kurtosis beyond the generally accepted ± 2 (Field, Miles, & Field, 2012), but none of the items with the unsatisfactory measures are used in the general factor.

Measurement Model

We then fit a measurement model to examine the underlying structure of the variables of interest used to assess the latent constructs (general musical sophistication, WMC, general fluid intelligence) by performing a confirmatory factor analysis (CFA) using the *lavaan* package (Rosseel, 2013) using *R* (R Core Team, 2017). Model fits in can be found in Table 3. For each model, latent factors were constrained to have a mean of 0 and variance of 1 in order to allow the latent covariances to be interpreted as correlations. Since the objective measures were on different scales, all variables were converted to z scores before running any modeling.

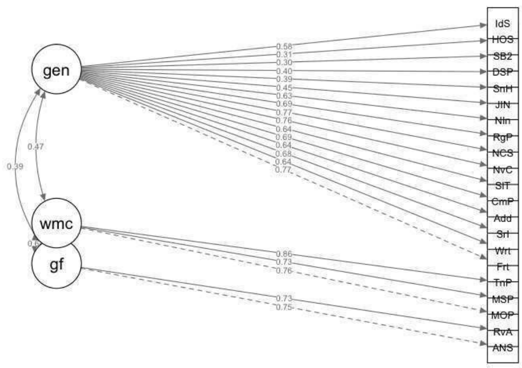


Figure 2: Measurement Model derived from CFA

Variables are defined as follows: gen: general factor latent variable; wmc: working memory capacity; gf: general fluid intelligence; zIS: "Identify What is Special"; zHO: "Hear Once Sing Back"; zSB: "Sing Back After 2-3"; zDS: "Don't Sing In Public"; zSH: "Sing In Harmony"; zJI:"Join In"; zNI: "Number of Instrumetns"; zRP:"Regular Practice"; zNCS: "Not Consider Self Musician"; zNcV: "Never Complimented"; zST: "Self Tonal"; zCP: "Compare Performances"; zAd: "Addiction"; zSI: "Search Internet"; zWr: "Writing About Music"; zFr: "Free Time"; zTP: "Tone Span"; zMS: "Symmetry Span"; zMO: "Operation Span"; zRA: "Ravens"; zAN: "Number Series".

Structural Equation Models

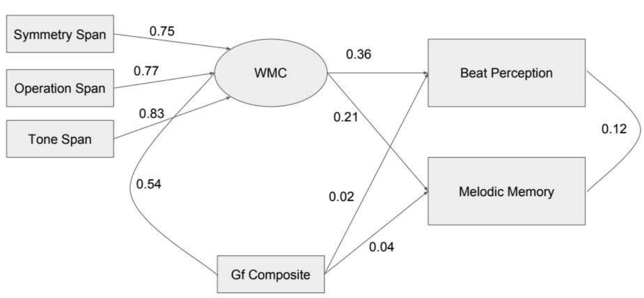
Following the initial measurement model, we then fit a series of SEMs in order to investigate both the degree to which factor loadings changed when variables were removed from the model as well as the model fits. We began with a model incorporating our three latent variables (general musical sophistication, WMC, general fluid intelligence) predicting our two objective measures (beat perception and melodic memory scores) and then detailed steps we took in order to improve model fit. For each model, we calculated four model fits: $\chi 2$, comparative fit index (CFI), root mean square error (RMSEA), and Tucker Lewis Index (TLI). In general, a non-significant $\chi 2$ indicates good model fit, but is overly sensitive to sample size. Comparative Fit Index (CFI) values of .95 or higher are considered to be indicative of good model fits as well as Root Mean Square Error (RMSEA) values of .06 or lower, Tucker Lewis Index (TLI) values closer to 1 indicate a better fit. (Beajean, 2014).

After running the first model (Model 1), we then examined the residuals between the correlation matrix the model expects and our actual correlation matrix looking for residuals above .1. While some variables scored near .1, two items dealing with being able to sing ("I can hear a melody once and sing it back after hearing it 2 – 3 times" and "I can hear a melody once and sing it back") exhibited a high level of correlation amongst the residuals (.41) and were removed for Model 2 and model fit improved significantly (χ^2 (41)=123.39, $p < .001$). After removing the poorly fitting items, we then proceeded to examine if removing the general musical sophistication self-report measures would significantly improve model fit for Model 3. Fit measures for Model 3 can be seen in Table 3 and removing the self-report items resulted in a significantly better model fit (χ^2 (171)=438.8, $p < .001$). Following the rule of thumb that at least 3 variables should be used to define any latent-variable (Beajuean, 2014) we modelled WMC as latent variable and Gf as a composite average of the two tasks administered in order to improve model fit. This model resulted in significant improvement to the model (χ^2 (4)=14.37, $p < .001$). Finally we examined the change in test statistics between Model 2 and a model that removed the cognitive measures-- a model akin to one of the original models reported in Müllensiefen et al., (2014)-- for Model 5. Testing between the two models resulted in a significant improvement in model fit (χ^2 (78)=104.75, $p < .001$). Figure 3 displays Model 4, our nested model with the best fit indices.

Table 3: CFA and SEM Model Fits

Model	*df*	χ^2	*p*	CFI	RMSEA	TLI
CFA	186	533.60	*>.001*	.83	.09	.81
1. Full Model	222	586.30	*<.001*	.83	.08	.80
2. Remove High Residuals	181	462.90	*<.001*	.86	.082	.83
3. No Self Report	10	24.11	*<.05*	.97	.08	.94
4. Gf as Observed	6	9.74	*.14*	.99	.051	.97
5. General Only	103	358.16	*<.001*	.83	.102	.801

Figure 3: Model 4



Discussion

Measurement Model

After running a confirmatory factor analysis on the variables of interest, the model fit was below the threshold of what is considered a "good model fit" as shown in Table 1 with references to above model fits. This finding is to be expected since no clear theoretical model has been put forward that would suggest that the general musical sophistication score, when modelled with two cognitive measures should have a good model fit. This model was run to create a baseline measurement.

Structural Equation Model Fitting

Following a series of nested model fits, we were able to improve model fits on a series of SEMs that incorporated both measures of WMC and measures of general fluid intelligence. Before commenting on new models, it is worth noting that the Model 5 does not seem to align with the findings from the original 2014 paper by Müllensiefen et al. While the correlation between the objective tasks is the same (r = .16), the factor loadings from this paper suggest lower values for both Beat Perception (.37 original, .27 this paper) as well as Melodic Memory (.28 original, .18 this paper). Note that two items were removed dealing with melody for memory for this model; when those items were re-run with the data, the factor loadings did not deviate from these numbers.

The first two models we ran resulted in minor improvements to model fit. While difference in models was significant (χ^2 (41)=123.39, $p < .001$), probably due to the number of parameters that were now not constrained, the relative fit indices of the models did not change drastically. It was not until the self-report measures were removed from the model, and then manipulated according to latent variable modeling recommendations, was there a marked increase in the relative fit indices. Fitting the model with only the cognitive measures, we were able to enter the bounds of acceptable relative fit indices that were noted above. In order to find evidence that the cognitive models (Models 3 and 4) were indeed a better fit than using the General factor, we additionally ran a comparison between our adjusted measurement model and a model with only the self report. While both of our nested models were significantly different, the cognitive models exhibited superior relative fit indices. Lastly, turning to Figure 3, we note that our latent variable of WMC exhibited much larger factor loadings predicting the two objective, perceptual tests than our measure of general fluid intelligence. We also note that the factor loading predicting the Beat Perception task (.36) was higher than that of the Melodic Memory task (.21). These rankings mirror that of the original Müllensiefen et al., (2014) paper and merit further examination in order to disentangle what processes are contributing to both tasks. These results align with predictions made with Process Overlap Theory (Kovacs & Conway, 2016), which predict that higher executive loads are needed for tasks of perception. While we failed to predict which task would load higher --we assumed that the ability to maintain and manipulate information in the Melodic Memory task would be better predicted by WMC than the Beat Perception task-- this might be due to the fact that performing well in a melodic memory task demands a certain amount of musical training that is not captured by either cognitive measure.

In the future, we are interested in exploring more theoretically-driven models that use specific, task oriented predictors in order to explain the relationships between the perceptual tasks and the cognitive measures. Given the results here that suggest that measures of cognitive ability play a significant role in tasks of musical perception, we suggest that future research should consider taking measures of cognitive ability into account, so that other variables of interest are able to be shown to contribute above and beyond baseline cognitive measures.

Conclusions

In this paper we fit a series of structural equation models in order to investigate the degree to which baseline cognitive ability was able to predict performance on a musical perception task. Our findings suggest that measures of WMC are able to account for a large amount of variance beyond that of self report in tasks of musical perception.

Acknowledgements. The authors would like to thanks RAs that are not authors here for help running participants for this study, as well as Elizabeth Baker for suggestions on improving the model fits for the structural equation modeling.

References

Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.

Colley, I. D., Keller, P. E., & Halpern, A. R. (2017). Working Memory and Auditory Imagery Predict Sensorimotor Synchronization with Expressively Timed Music. *The Quarterly Journal of Experimental Psychology*, , 1-16.

Corrigall, K. A., & Schellenberg, E. G. (2015). Predicting who takes music lessons: Parent and child characteristics. Frontiers in psychology, 6, 282.

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in psychology*, *4*, 222.

Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. Frontiers in psychology, 2, 124.

Ethridge, P (1998, June, 24). Georgia program bringing classics to newborns. *CNN.com*. Retrieved from http://edition.cnn.com/SHOWBIZ/Music/9806/24/beethoven.for.babies/index.html

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage publications.

François, C., Chobert, J., Besson, M., & Schön, D. (2012). Music training for the development of speech segmentation. Cerebral Cortex, 23(9), 2038-2043.

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

Hille, A., & Schupp, J. (2015). How learning a musical instrument affects the development of skills. Economics of Education Review, 44, 56-82.

Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*(3), 151-177.

Li, D., Cowan, N., & Saults, J. S. (2013). Estimating working memory capacity for lists of nonverbal sounds. *Attention, Perception, & Psychophysics*, *75*(1), 145-160.

Mehr, S. A., Schachner, A., Katz, R. C., & Spelke, E. S. (2013). Two randomized trials provide no consistent evidence for nonmusical cognitive benefits of brief preschool music enrichment. PloS one, 8(12), e82007.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological science*, *21*(7), 914-919.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2008). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. Cerebral Cortex, 19(3), 712-723.

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. http://dx.doi.org/10.1371/journal. pone.0089642

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford, England: Oxford Psychologists Press.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). Journal of statistical software, 48(2), 1-36.

Schellenberg, E.G. (2016). Music training and nonmusical abilities. In S. Hallam, I. Cross, & M. Thaut (Eds.), Oxford handbook of music psychology (2nd ed., pp. 415-429). Oxford, UK: Oxford University Press.

Schellenberg, E.G. (2014). Mozart effect. In W.F. Thompson (Ed.), *Music in the social and behavioral sciences: An encyclopedia*(pp. 717-718). Thousand Oaks, CA: Sage.

Schellenberg, E. G., & Mankarious, M. (2012). Music training and emotion comprehension in childhood. Emotion, 12(5), 887.

Schellenberg, E. G. (2011a). Examining the association between music lessons and intelligence. British Journal of Psychology, 102(3), 283-302.

Schellenberg, E. G. (2011b). Music lessons, emotional intelligence, and IQ. Music Perception: An Interdisciplinary Journal, 29(2), 185-194.

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. Journal of Educational Psychology, 98(2), 457.

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

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch, D. Deutsch (Eds.). *The psychology of music* (pp. 499-550). San Diego, CA, US: Elsevier Academic Press. doi:10.1016/B978-0-12-381460-9.00012-2

Shelton, J. T., Elliott, E. M., Matthews, R. A., Hill, B. D., & Gouvier, W. D. (2010). The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 813-820. doi:10.1037/a0019046

Silvia, P. J., Thomas, K. S., Nusbaum, E. C., Beaty, R. E., & Hodges, D. A. (2016). How does music training predict cognitive abilities? A bifactor approach to musical expertise and intelligence. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(2), 184-190. doi:10.1037/aca0000058

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, 152, 199-211. doi:10.1016/j.cognition.2016.03.017

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201-292.

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. doi:10.1016/j.intell.2017.03.005

Talamini F., Altoè G., Carretti B., & Grassi M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLoS ONE*, *12*(10), e0186773. doi:10.1371/journal.pone.0186773

Thurstone, L. L. (1938). Primary mental abilities. Chicago, IL: University of Chicago Press.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505. doi:10.3758/BF03192720

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635-654. doi:10.1080/09658210902998047

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. doi:10.1016/j.lindif.2010.02.004