



Domain-specific and domain-general contributions to reading musical notation

Ting-Yun Chang¹ · Isabel Gauthier¹

Accepted: 5 July 2021 / Published online: 2 August 2021
© The Psychonomic Society, Inc. 2021

Abstract

Musical practice may benefit not only domain-specific abilities, such as pitch discrimination and music performance, but also domain-general abilities, like executive functioning and memory. Behavioral and neural changes in visual processing have been associated with music-reading experience. However, it is still unclear whether there is a domain-specific visual ability to process musical notation. This study investigates the specificity of the visual skills relevant to simple decisions about musical notation. Ninety-six participants varying in music-reading experience answered a short survey to quantify experience with musical notation and completed a test battery that assessed musical notation reading fluency and accuracy at the level of individual note or note sequence. To characterize how this ability may relate to domain-general abilities, we also estimated general intelligence (as measured with the Raven's Progressive Matrices) and general object-recognition ability (as measured by a recently proposed construct *o*). We obtained reliable measurements on our various tasks and found evidence for a domain-specific ability of the perception of musical notation. This music-reading ability and domain-general abilities were found to contribute to performance on specific tasks differently, depending on the level of experience reading music.

Keywords Individual differences · Visual expertise · Music · Object recognition

A surprising number of studies report on the association between musical expertise and cognitive skills that are very different from what musicians practice in the development of their musicianship. It is certainly expected that musicians excel in music tasks such as score reading (e.g., Bean, 1938; Sloboda, 1978) and the discrimination of musical elements including pitch, rhythm, melody, and tone (e.g., Barakat et al., 2015; Bratzke et al., 2012; Magne et al., 2006). Moreover, music training can enhance cognitive abilities like executive function (e.g., Bialystok & DePape, 2009; Bugos, 2019; Clayton et al., 2016; Hannon & Trainor, 2007; Moreno

et al., 2011; Shen et al., 2019), visual/auditory attention (e.g., Clayton et al., 2016; Parbery-Clark et al., 2011; Roden et al., 2014; Strait et al., 2010; Strait & Kraus, 2011), visual/auditory memory (e.g., Degé et al., 2011; Jakobson et al., 2008; Rodrigues et al., 2013, 2014; Strait & Kraus, 2011), intelligence (e.g., Moreno et al., 2011; Schellenberg, 2004, 2006, 2011a, 2011b), and linguistic abilities (e.g., Butzlaff, 2000; Magne et al., 2006; Moreno et al., 2009; Strait & Kraus, 2011). Despite the variety of studies heralding the benefits of music training, the evidence is mainly correlational, making causal inferences imprudent. Some or all of the effects could be due to self-selection (Heckman, 1979). One meta-analytical study investigating the effects on reading ability concluded that the effect size for correlational studies was small but very robust, while that for experimental studies was even smaller and more likely due to publication bias (Butzlaff, 2000). More recent work relied on a longitudinal design to support stronger claims that music training influences both music-specific and domain-general skills (Habibi et al., 2018). Nonetheless, whether or not self-selection is of concern in interpreting a given result, all work in the field of musical expertise is conducted against the backdrop of possible influences of music training on a multitude of cognitive abilities.

This study was supported by a Music, Mind, and Society at Vanderbilt University Booster Grant and the Government Scholarship to Study Abroad from Taiwan Ministry of Education, and by NSF award 1640681. We would like to thank Jackie Floyd and Melanie Kacin for their help in collecting the data.

The data are publicly available online: <https://doi.org/10.6084/m9.figshare.13643732.v1>

✉ Ting-Yun Chang
ting-yun.chang@vanderbilt.edu

¹ Department of Psychology, Vanderbilt University, 111 21st Avenue South, Nashville, TN 37240, USA

The possibility of domain-general advantages (those that are not specific to music, but also apply to music) in musicians can make it particularly difficult to study domain-specific individual differences (variability that only applies to the music domain) in this field. Even when studies cannot lead to causal inferences, they should at least characterize whether advantages observed in musicians are domain specific or domain general. Of interest here is a body of work relating expertise in reading musical notation to a number of behavioral and neural effects. Unlike research that examines musical knowledge or musical ability (e.g., processing of pitch, melody, rhythm; e.g., Law & Zentner, 2012), this line of work pertains specifically to the visual processing of musical notation. Music-reading experience is associated with perceptual advantages. It is positively correlated with the speed of encoding music note sequences, relative to nonmusical stimuli such as digits, shapes, and letters (Chang & Gauthier, 2020; Wong et al., 2019; Wong & Wong, 2018). Music-reading experts experience less visual crowding in the periphery for crowded musical notation than control stimuli (Wong & Gauthier, 2012). This expertise also seems to come with attentional benefits. Music-reading experts show a greater asymmetry during visual search based on their familiarity with the correct configuration of the features (dots and stems) that make out notes (Chang & Gauthier, 2020). They also show enlarged visual span for identifying English letters, but not Chinese characters or novel symbols, suggesting transfer effects due to similarities in the visual stimulus processing involved (e.g., left-to-right reading; Li et al., 2019). Perceptual advantages may arise through top-down influences on early visual cortex (Wong et al., 2014). All of these effects, behavioral and neural, have been suggested to be specific effects of experience reading musical notation on its processing.

A possible limitation to this interpretation is the operational definition of “musical-reading expertise” in this body of work. This definition has been based on a single task, sometimes controlling for only one other task, and thus may not provide a sufficiently strong operationalization of a domain-specific ability. Our work aims to improve on this. Wong and Gauthier (2010a) measured the perceptual thresholds (presentation time required for 80% accuracy) in matching short sequences of four notes. While those who self-identified as music-reading experts could match music sequences based on a shorter presentation time, their thresholds for matching letter sequences did not differ from novices. In addition, the perceptual threshold for notes was related to neural activity for notes in a multimodal network of brain regions. Wong et al. (2020) recently pointed out several advantages of this task. First, it does not require naming or identifying the notes so it can be used to compare individuals with and without music training. Second, it uses “non-sense” sequences of notes, such that any advantage is not due to the visual regularity of patterns that are common in sight-reading material. Third, this task produced

measurements with acceptable reliability (Guttman λ -2 and Cronbach’s $\alpha > .7$) in a large sample study (231 individuals ranging from novice to experienced music readers). Perceptual note thresholds were moderately correlated with self-reported music-reading ability and the number of years of musical training. Some studies improved on the operational definition of musical reading expertise by subtracting the perceptual threshold for letters from that of notes (Wong et al., 2014; Wong & Gauthier, 2012), or by regressing the threshold variability for digits and shapes (Chang & Gauthier, 2020). Thus, studies have increasingly raised the bar to ensure that measurements can isolate the domain-specific expertise associated with reading music. However, many limitations remain, including the limitations to measurements of perceptual fluency in a single matching task and the lack of information about the contribution of domain-general cognitive and visual skills in individual differences in music reading.

To increase domain coverage, we administered two behavioral tests that measure musical notation reading fluency and accuracy at both the individual note level and the note sequence level. Moreover, for each test, we used better control stimuli to better control for domain-general variance in each test. Several tasks with distinct demands were used to prevent effects due to specifics of the same tasks, improving result generalization to various kinds of tasks. We examined how much domain-specific variance is present once we account for domain-general abilities, such as general intelligence g and general object-recognition ability o_v . Like g , o_v is a statistical construct referring to a broad visual object-recognition ability that influences performance across a variety of visual tasks, with both novel and familiar objects (Richler et al., 2019; Sunday et al., 2021). We also explored the contribution of these domain-general abilities to musical notation reading in people varying in experience in this domain. In the domain of chest radiograph reading, once radiological experience level was controlled, people higher in g and o_v were better at detecting lung nodules (Sunday et al., 2018). Interestingly, o_v contributed to lung-nodule detection performance more for novices than for radiologists. Similarly, here, we examined whether performance on tasks with musical notation may be more influenced by domain-general abilities like g and o_v in music-reading novices than in music-reading experts.

To preview the results: Our behavioral tests produced reliable measurements (see Results, Part 1 section), and performance on tests that used musical notation correlated well with each other and with self-reported music-reading experience level (see Results, Part 2 section). We show that this positive correlation remains after controlling for domain-general effects or variances in the control versions (see Results, Part 3 section). We find evidence of contribution from general intelligence g and general object-recognition ability o_v to performance on music-reading tasks, with different contributions for novices and experts (see Results, Part 4 section).

Method

Participants

We conducted a power analysis based on an effect size of $r = .3$, which we consider a reasonable minimum correlation for tasks that tap into the same ability but differ in the specifics of task demands. At an alpha of .05 (one-tailed), 64 participants are required to detect this effect with 80% power (note, however, that we chose to analyze the results with Bayesian tests, for which a priori power is less important). However, more than *detecting* correlations, we were interested in estimating common variance across tasks and we recruited more participants to increase precision. A total of 96 participants (26 males and 70 females) completed the study, which was approved and conducted in compliance with the Vanderbilt Institutional Review Board. To recruit people with a wide range of music training experience for the purpose of individual difference methods, we recruited participants from the Vanderbilt University Community and the Blair School of Music. Informed consent was obtained at the beginning of the study. Upon completion, participants received monetary compensation or extra credits for course purposes.

Procedure

Participants came in for two sessions, each lasting between 60 and 90 min. The first session included a short survey about the participants' experience with music reading and a series of sequential matching tests (SM). In the second session, participants performed the following: working memory tests (WM), the Raven's Progressive Matrices (RPM), the novel object memory test (NOMT), and an object matching test (OM). In SM and WM tests, we used both musical as well as nonmusical stimuli (e.g., simple shapes and colors) to estimate visual abilities associated with music reading, controlling for non-specific effects. To prevent potential confounding effects due to task order differences across participants, all participants completed the tests in the same order. Trials within each test were presented in a fixed order, except for the SM tests, in which an adaptive method was used.

Measuring music-reading experience

Since one's musical training background does not only reflect visual fluency with musical notation, or the ease with which one processes musical notation, it may not be ideal to use years of music lessons (Brochard et al., 2004; Stewart et al., 2004) or profession (Salis, 1980) as an index of one's music-reading proficiency. Here, we chose to explicitly ask participants to indicate whether they considered themselves a music-reading expert or not (which may differ from their expertise in music playing, for example). They were also asked, "In terms

of reading musical notation, which best describes you?" with four response options on a scale of 1 to 4 (1 = *I do not read music notation*; 2 = *I have some music-reading skills that I do not use regularly*; 3 = *I can read music scores, and I do it regularly*; and 4 = *even among musicians, I may be above average in my ability to read musical notation*). Since the second survey question was included later in the process of data collection, 12 participants completed the two survey questions at the same time when they came in for the first session. Those who had already completed the study took the second survey via phone or email.

Sequential matching tests (SM-notes on staff, SM-notes on endless staff, SM-shapes, SM-bullseyes)

The SM tests measure perceptual fluency with musical notation and control stimuli (for similar designs, see Chang & Gauthier, 2020; Wong et al., 2014; Wong et al., 2020; Wong & Gauthier, 2010a, 2010b, 2012; Wong & Wong, 2018). Each trial begins with a 200-ms fixation and then a 500-ms mask, followed by a reference item presented for a varied duration determined by performance on previous trials. Presentation duration is decreased by 30% following three correct trials or otherwise increased by 35%, starting from an initial duration of 700 ms. After another 500-ms mask, two test items are presented side by side, and participants make an unspeaked choice as to which of the two items matches the reference they just saw. Participants are informed that the distractor test item would be very similar to the reference, and that they should aim to be as accurate as possible. We record every change in the direction of duration adjustment as a reversal (e.g., change from shortening to lengthening the duration), and terminated the testing sequence after the fifth consecutive reversals following at least 10 reversals).

Participants completed this task twice, for two separate estimations for each type of musical/control stimulus, with the number of trials for each estimation varying according to individual performance stability. The two estimated presentation duration thresholds were averaged to index stimulus-specific perceptual fluency.

The *SM-notes on staff* test uses short music sequences composed of four quarter notes on a staff of five horizontal lines (see Fig. 1). Each note could be placed on a staff line or in between two staff lines. In each pair of reference and distractor music sequences, only one of the four notes differs, being randomly chosen to move up/down by one step (e.g., from a line to an adjacent space above/below) or two steps (e.g., from a line to the next higher/lower line). The *SM-notes on endless staff* test uses the same constraints but notes are placed on a staff that has 10 more staff lines, fading in all directions, to reduce the usefulness of verbal coding with note names (see Fig. 1). To avoid ceiling performance for those who are skilled

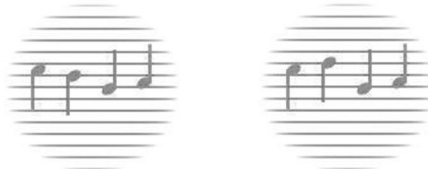
SM-notes on staff



SM-shapes



SM-notes on endless staff



SM-bullseyes

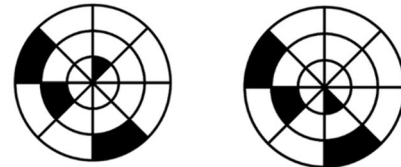


Fig. 1 Sample stimuli (shown in pairs) used in the SM tests

at music reading, contrast of the music sequences was reduced by 30% following previous studies.

Prior work used either letters or digits as a control category for musical stimuli in the SM tests (Chang & Gauthier, 2020; Wong et al., 2019; Wong & Wong, 2018), as they are universally familiar to participants. However, based on recent reports that individual differences in processing digits are much less variable compared to musical notation (e.g., Experiment 1 in Chang & Gauthier, 2020), we chose to use shape sequences instead. Shape sequences lead to performance more comparable to musical sequences in their variability (e.g., Experiment 2 in Chang & Gauthier, 2020), even though the variability cannot be attributed to expertise differences. It therefore provides a more effective control for non-music-specific variance on the SM-note test. The *SM-shapes* test uses sequences of three shapes selected from 10 common shapes (see Fig. 1). In each pair of reference and distractor shape sequences, one of the three shapes is replaced with its mirrored form, rotated form, or with a visually similar shape. Finally, the *SM-bullseyes* test is designed to measure individual differences in domain-general spatial coding for visual patterns, which may contribute to performance on the *SM-notes* test. Bullseye patterns were created by randomly darkening four locations of a pattern composed of polar grids of three concentric circles divided in 45-degree sections (see Fig. 1). On each trial, the reference and distractor patterns only differ in one of the four locations.

Working memory tests (WM-notes, WM-colors)

Studies on visual working memory suggest higher capacity with more familiar objects. For instances, a sequence of words is remembered better than a sequence of nonsense syllables (Jones & Macken, 2015), and car experts can encode more

cars compared to car novices (Curby et al., 2009). Therefore, we designed the WM tests to measure working memory capacity for musical notation and, as a control, colors. It was modeled after prior work (Xu et al., 2018), with some modifications. Participants first performed *WM-colors* followed by the *WM-notes*. Each test included five blocks of 40 trials (with set size varying from two to six), totaling 200 trials. Arrays were generated by randomly selecting two to six items from a set of eight possible musical notation elements (see Fig. 2) or colored squares, without repetition within a trial. Trials start with a fixation presented at the center of the screen for 1,000 ms, followed by an array of two to six items in a single horizontal line presented for 150 ms, a 500-ms blank, and finally a probe item such that participants could judge whether the probe was part of the initial array or not. Accuracy was emphasized over speed but a maximum of 10 s was allowed for a response. We showed musical notation (and therefore colored squares) in a line for ecological validity (we are interested in individual differences relevant to music reading and not in position effects). Accuracy averaged across the five blocks was used to index memory capacity for musical notation and colors. Based on preliminary analyses, for the WM-note test we excluded Set Size 6 trials because performance on these was poorly correlated with performance on the other set sizes (r_s from .09 to .28, depending on level, whereas all other correlations between levels were $>.4$).

General intelligence, g (RPM)

A version of the Raven's Progressive Matrices (Raven, 2000) was used to estimate participants' general intelligence. On each trial, a geometric pattern is presented in the form of a 3×3 matrix, with the bottom-right piece missing. Given the partial pattern, the task is to identify the one diagram that

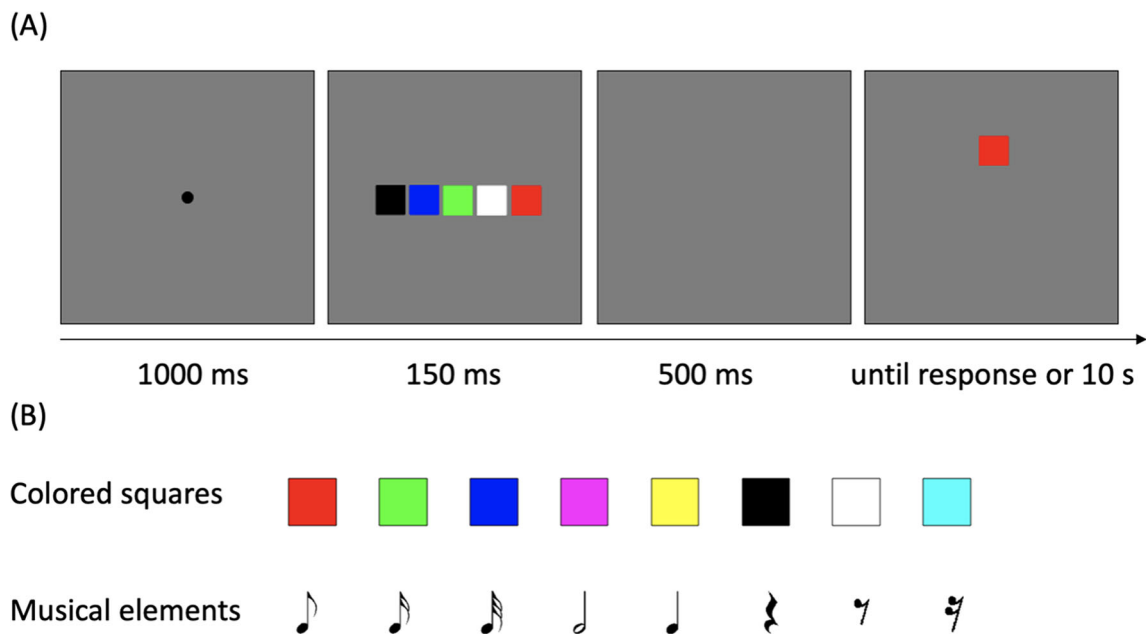


Fig. 2 a Trial procedure and (b) sets of eight possible objects for the WM tests

completes the pattern among eight alternatives. Participants are provided with three practice trials with feedback during instructions. Afterwards, they are timed for 10 minutes and asked to complete as many trials as possible, out of 18 total trials. Trials are ordered from least to most difficult. The total count of correct responses is used as an index of general intelligence.

General visual object-recognition ability (o_v)

A latent variable approach uncovered evidence for a domain-general object-recognition ability, or o_v (Richler et al., 2019; Sunday et al., 2021). In those studies, o_v is a higher-order factor in a structural equation model that accounted for a majority of the variance in lower-order factors, based on performance for three different tasks and five novel object categories. Following others (Sunday et al., 2018), we estimate o_v using the aggregate of performance on two different tasks (NOMT and OM; see below), one with a single novel object category and the other with five novel object categories.

Novel object memory test (NOMT)

The NOMT (Richler et al., 2017) is a learning test with an artificial object category (vertical Ziggerins; Wong et al., 2009; see Fig. 3). The test is divided into two blocks that consist of a study phase and a 24-trial test phase. In the first block, participants study six target Ziggerins presented simultaneously for unlimited time, then on each test trial select which of three Ziggerins is any of the six target Ziggerins. After the first six test

trials, the target Ziggerins are shown again to facilitate learning and ensure instruction. In the second block, participants review the target Ziggerins for as long as they like, then complete 24 test trials where target Ziggerins are presented in novel unstudied views. Across the entire test, visual noise is added to object images in eighteen trials. Participants are informed about visual noise trials and that a target would be present on all trials. Accuracy is emphasized over reaction time. Mean accuracy for all 48 trials is used as an index of performance on this test.

Object matching test (OM)

On each trial of the OM test (Sunday et al., 2018), participants see two objects from the same novel object category, one after the other, and judge whether they are the same or different regardless of size, brightness, and/or viewpoint differences. The first object is presented for 300 ms. Following a 500-ms random pattern mask, a second object is shown for a maximum of 3 s or until response. The novel object categories used were symmetric Greebles, asymmetric Greebles, horizontal Ziggerins, vertical Ziggerins, and Sheinbugs (see Fig. 3). Five practice trials (one for each object category) are given to familiarize the participants with the task before 180 test trials, where categories are mixed. Participants are informed that the two objects could be very similar if different, and are encouraged to keep doing their best and respond as fast and accurately as possible. Performance on this test is indexed by sensitivity (d').



Fig. 3 Artificial objects used in the NOMT and OM test. From left to right, symmetric Greebles, asymmetric Greebles, horizontal Ziggerins, vertical Ziggerins, and Sheinbugs. Vertical Ziggerins were used in both tests

Results

Part 1: Descriptive statistics and reliability

Different numbers of participants were excluded for each test due to missing data or dropout (see Table 1). Among 96 participants, six did not complete the WM tests; one did not complete the RPM; two did not complete the NOMT; and three did not complete the OM test. Outliers on any test were excluded if their performance was below chance only if their performance on various trials of a specific test did not correlate with the trial difficulty estimated based on the other participants. This is a compromise to eliminate participants who were not making an effort or had misunderstood the instructions on a task, but still include those who performed poorly due to low abilities. When a participant's performance is better on the easiest items relative to the most difficult items, it suggests they understood the instructions and tried to answer correctly. Six outliers were excluded on the NOMT, where chance is 33.3%. Five outliers were excluded for the OM test, for showing a negative d' . For the SM tests, we trimmed the estimated thresholds longer than four standard deviations away from the mean value of that estimation or longer than 5,000 ms (18 out of 768 estimates; 2.34%). Therefore, for these participants, we used only a single estimate to index stimulus-specific perceptual fluency instead of averaging the two estimated presentation duration thresholds.

We evaluated the psychometric properties of all the tests as the first step of our measurement evaluation. Descriptive statistics and reliability for each test are presented in Table 1. For SM tests, the Pearson's correlation r was computed between the two presentation duration thresholds and corrected with the Spearman–Brown prophecy formula (r' ; Brown, 1910; Spearman, 1910) to assess the reliability of the estimation procedure. For all the other tests, Cronbach's α was computed to index the reliability as the internal consistency of the measurement. Overall, the tests demonstrated a good range of performance and acceptable reliability. JASP statistics software (JASP Team, 2020) was used to conduct Bayesian analyses.

Part 2: Zero-order correlations between measures

We assessed convergent validity by determining whether our musical tests correlated with one another. Conversely, we assessed divergent validity by determining whether these musical tests do not correlate as strongly with nonmusic notation tests. All correlations are reported with BF_{+0} , using an alternative hypothesis that is positive (because negative correlations are not expected, neither theoretically nor on the basis of prior work) and a stretched beta prior width of 1 (default in JASP). To estimate experience with musical notation, we combined the binary self-report of expertise and the four-level survey question (after standardizing each), because the

Table 1 Descriptive statistics and reliability (as indexed by Spearman–Brown corrected Pearson's correlation r' or Cronbach's α) for all tests

	<i>N</i>	Mean (<i>SD</i>)	Min	Max	Reliability
SM (threshold in ms)					
Notes on staff	96	1,005.3 (620.5)	191.5	4,113.5	$r' = .82$
Notes on endless staff	96	1,374.8 (1040.7)	161.44	4,111.7	$r' = .71$
Shapes	96	907.9 (519.0)	175.42	3,186.5	$r' = .61$
Bullseyes	96	462.3 (291.9)	111.3	1,648.4	$r' = .68$
WM (percentage accuracy)					
Colors	90	.87 (.07)	.63	.97	$\alpha = .89$
Notes	90	.77 (.07)	.49	.93	$\alpha = .82$
RPM (total correct)	95	11.26 (2.71)	3	16	$\alpha = .65$
NOMT (percentage accuracy)	88	.60 (.14)	.31	.88	$\alpha = .78$
OM (d')	88	1.05 (.56)	0	2.32	$\alpha = .90$

two measures were strongly positively correlated ($r = .65$, $BF_{+0} = 1.56e + 10$). To estimate individual general visual recognition ability as o_v (Richler et al., 2019), we first confirmed the strong positive correlation between the NOMT and OM scores ($r = .47$, $BF_{+0} = 1567$) as predicted from prior reports (e.g., Richler et al., 2019; Sunday et al., 2018), then we standardized the NOMT and OM scores across participants, and averaged the two standardized scores for each participant to obtain the aggregate index o_v . Reliability for o_v is .90 (corrected Cronbach’s α as suggested by Wang & Stanley, 1970). We confirmed that, as in prior work (; Richler et al., 2019), the correlation between these two visual tasks survived after controlling for g ($r_{\text{partial}} = .39$, $BF_{+0} = 159$).

Table 2 summarizes the zero-order correlations and corresponding BF_{+0} between tests for musical abilities, o_v , and g . Scatterplots were examined for bivariate outliers and five of the correlations (based on three participants) led to one value with an externally studentized residual (jackknife) value above 3. While the correlations were all larger without the outliers, the change was generally small (the most significant change was the correlation between SM-bullseyes and WM-colors, which went from inconclusive to strong support for a correlation (Jeffreys, 1961). Values in the table with a superscript are those without bivariate outliers, with the original values in the note.

All measures pertaining to music were positively related, with substantial to extreme evidence relative to the null hypothesis. We note that we find no support for a correlation of music experience with any of the nonmusical measures (and

in some cases, for SM-shapes, WM-colors, and o_v , support for the null). Performance on the same tasks (SM and WM) for musical and control stimuli were positively correlated, with extreme evidence relative to the null hypothesis. This confirms the importance of controlling for domain-general and task-related variance in the interpretation of the correlations, which we address in the next section. Finally, both o_v and g were positively correlated with all performance measures (with substantial to extreme support relative to the null hypothesis, depending on the measure).

Values before bivariate outlier exclusion: ¹ $r = .35$, $BF_{+0} = 71$, ² $r = .61$, $BF_{+0} = 1.263e + 8$, ³ $r = .29$, $BF_{+0} = 11$, ⁴ $r = .20$, $BF_{+0} = 1.5$, ⁵ $r = .26$, $BF_{+0} = 4.74$.

Part 3: Estimating domain-specific variance in musical tests

We examined whether participants who had more experience with music reading would show superior performance on the musical SM and WM tests, even after controlling for performance in the nonmusic control conditions. As expected, the two musical SM tasks (*SM-notes on staff* and *SM-notes on endless staff*) were strongly positively correlated across all participants ($r = .65$, $BF_{+0} = 1.57e + 10$). We therefore averaged them to index the performance on SM for notes (abbreviated hereafter as *SM-notes*).

After controlling for domain-general abilities and task-specific variance common in the SM tasks as assessed in *SM-shapes*, *SM-bullseyes*, general object-recognition abilities

Table 2 Zero-order correlations among tests with BF_{+0}

	1	2	3	4	5	6	7	8
1. Music experience	-							
2. SM-notes on staff	.60	–						
	3.047e+8							
3. SM-notes on endless staff	.50	.65	–					
	162254	1.57e+10						
4. WM-notes	.28	.45	.40	–				
	8.92	4665	558					
5. SM-shapes	.10	.35	.48	.33	–			
	0.32	123	39254	43				
6. SM-bullseyes	.14	.41	.48	.46 ¹	.53	–		
	0.54	1253	33896	6404	1.24e+6			
7. WM-colors	.07	.27	.22	.64 ²	.29	.31 ⁴	–	
	0.24	6.99	2.09	1.43e +9	13	20		
8. o_v	.03	.25	.31	.46	.34	.39	.40	–
	0.17	5.12	23	6125	60	518	524	
9. RPM	.17	.28	.30	.34 ⁵	.29	.30	.32 ⁵	.31
	0.88	10.35	17	39	13	21	25	21

Note. For all tests, the alternative hypothesis specifies that the correlation is positive.

o_v , and general intelligence g scores, we found strong evidence for a positive correlation between *SM-notes* scores and music experience (see Fig. 4a; $r_{\text{partial}} = .65$, $\text{BF}_{+0} = 3.54e + 9$). Similarly, there was moderately strong evidence for a positive correlation between *WM-notes* and music experience, after controlling for performance on *WM-colors*, o_v , and g (see Fig. 4b; $r_{\text{partial}} = .29$, $\text{BF}_{+0} = 9.40$).

Substantial support was obtained for a positive correlation between performance residuals for *SM-notes* and *WM-notes*, controlling for performance on non-music tasks, o_v and g (see Fig. 4c; $r_{\text{partial}} = .25$, $\text{BF}_{+0} = 3.89$). This suggests that there is variance shared across these tasks that is specific to music-reading expertise.

Part 4: How music-reading experience, o_v and g relate to performance in music-reading novices and experts

We hypothesized that performance on tasks with musical notation is more likely to be supported by domain-general abilities (either cognitive or visual) in novices than in music-reading experts. Both the *SM-notes* and *WM-notes* tests were designed so as not to require explicit music-reading skills. Indeed, the goal was to create tests that can be used to measure the effects of music-reading expertise across the entire population, at any level of experience. Such a continuous measure allows for more powerful analyses than dichotomizing into groups of novices and experts (MacCallum et al., 2002). However, in the extreme, given no or very little experience reading music, musical notation are no more than non-sense shapes. With little to no experience, we might expect that performance on these tasks are more dependent on o_v and g . If experience reading music results in a domain-specific ability—say, as new features or new attentional strategies are acquired by experts—then these domain-general effects should be smaller.

To test this hypothesis, we split our sample into two groups according to their answers to the music survey (level of

experience 1 or 2, vs. 3 or 4). This resulted in a group of 56 music-reading novices and 40 music-reading experts. Experts showed more variability in music-reading experience than novices (variance = .34 vs .17), $F(39, 55) = 2.0$, $p = .02$, but they did not differ significantly in their variability in o_v (variance = .78 vs .82), $F(39, 55) = 1.05$, $p = .88$, or in g (variance = 6.03 vs 8.06), $F(39, 55) = 1.34$, $p = .34$.

Support for the predicted pattern was found for the *SM-notes* task (see Table 3). Music experience was correlated with performance in experts and not in novices (a difference that was not significant), while o_v was correlated with performance in novices and not in experts (a difference that was significant; see Fig. 5). In contrast, for *WM-notes*, music experience was not a significant predictor in either group alone, while o_v was related to performance in both groups, and g only in experts. Again, this must be interpreted in the context of analyses including all participants, which found support for all three measures (music experience, o_v and g) for both tasks (see Table 2). In several cases, a correlation was supported for one subgroup and not the other (note that even though there are fewer experts than novices here, we find at least as much evidence for correlations in experts). In the case of o_v , which we found to be a stronger predictor of performance in the *SM-notes* tasks for novices than experts, we suspect this reflects a qualitative change in the ability that supports performance.

It is worth pointing out that the comparison of groups was not perfect, as participants with some modest level of music-reading expertise (Level 2) were included among novices, to produce samples large enough to explore and compare correlations. These analyses should be repeated in much larger groups, ideally to compare a group of individuals with no music-reading experience at all to a group with varying expertise, but we find more support for a qualitative change in the ability that support performance for the *SM-notes* test than for the *WM-notes* test. Given that o_v is a fairly recent construct (Richler et al., 2019), it is interesting to see that it is a strong predictor of performance in both music tests in novices,

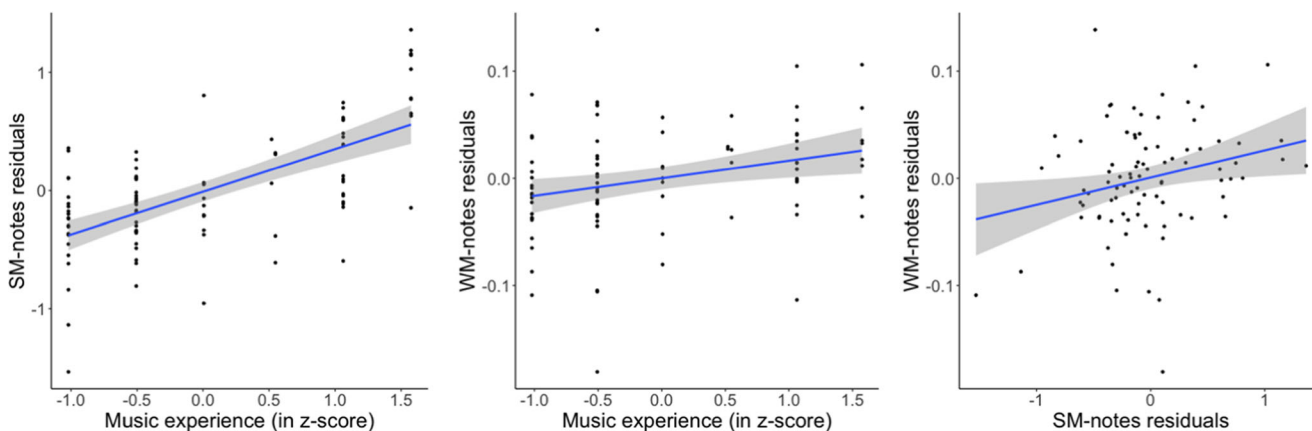


Fig. 4 Scatterplots of music experience and performance residuals of the musical tasks, after controlling for the corresponding controls, o_v and g . Left to right panel: Music experience and *SM-notes*, music experience and *WM-notes*, and *SM-notes* and *WM-notes*

Table 3 Zero-order correlations (with 95% credible interval and BF_{+0}) between task performance on SM-notes and WM-notes, and music experience, σ_v and g in music-reading novices and in music-reading experts

		Novices	Experts	Fisher Z test p value
<i>SM-notes</i>	Music experience	.25 (.03, .48) $BF_{+0} = 1.66$.55 (.27, .72) $BF_{+0} = 268^{**}$.09
	σ_v	.55 (.31, .71) $BF_{+0} = 1575^{**}$.07 (.01, .38) $BF_{+0} = 0.28$.01
	g	.21 (.02, .45) $BF_{+0} = 0.93$.29 (.04, .54) $BF_{+0} = 1.87$.69
	Music experience	.26 (.04, .48) $BF_{+0} = 1.88$.12 (.01, .43) $BF_{+0} = 0.40$.50
<i>WM-notes</i>	σ_v	.38 (.12, .58) $BF_{+0} = 15.67^{**}$.59 (.31, .76) $BF_{+0} = 449^{**}$.20
	g	.16 (.01, .41) $BF_{+0} = 0.56$.47 (.16, .68) $BF_{+0} = 25^{**}$.10

Note. *flags BF_{+0} that are above 3, providing substantial evidence for a correlation. ** flags BF_{+0} that are above 10, providing at least strong evidence for a correlation. The p value for a Fisher Z test comparing the two correlations is included.

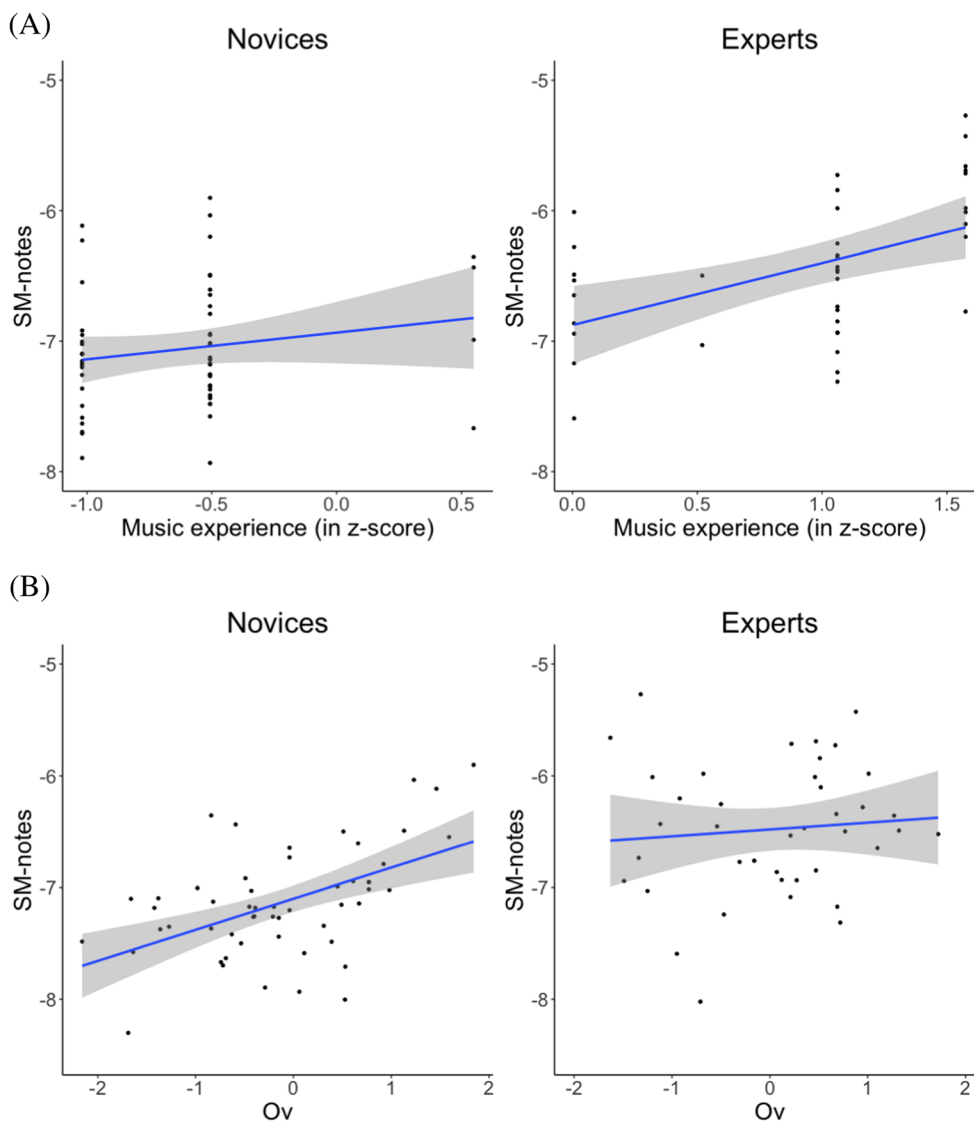


Fig. 5 Correlations between SM-notes performance and (a) music experience and (b) σ_v in music-reading novices (left) and experts (right)

despite the fact that musical notation is quite different from the shaded novel objects used in the o_v tasks for all participants as a whole. This illustrates the importance of accounting for domain-general abilities when assessing individual differences in a specific domain of expertise, and that some aspects of music-reading performance might be sourced from domain-general ability.

General discussion

In this study, we explored the validity of a domain-specific ability for reading musical notation. We created a battery of tests to quantify this ability and began to characterize its nomological network (Cronbach & Meehl, 1955). We measured this ability in a sample of 96 participants varying in their experience with music training and examined how it relates to domain-general abilities, in both skilled music readers and novices.

Each of our tests produced reliable scores, allowing us to characterize the ability to encode musical versus nonmusical objects (SM tests), as well as working memory capacity for musical versus nonmusical stimuli (WM tests). We validated the tests by showing that the musical versions of these tests (but not the control versions) were related to self-reported music-reading expertise. Importantly, performance on tests that use musical stimuli correlated well with each other, despite differences in task demands, providing convergent validity. This supports the existence of a latent construct that is domain-specific, one that should be relevant across different tasks and situations in which reading musical notation is required.

In trying to establish reading of musical notation as a domain-specific ability, it was important to characterize its relation to domain-general abilities, thereby providing information on discriminant validity of the construct. We measured general object-recognition ability o_v and estimated general intelligence g alongside our tests for music reading abilities. We found evidence that domain-general ability and domain-specific music reading experience contribute differently to performance on the SM test used in many studies to index music reading expertise (Chang & Gauthier, 2020; Wong et al., 2014; Wong et al., 2020; Wong & Gauthier, 2010a, 2010b, 2012; Wong & Wong, 2018). That is, in music reading experts, performance in matching short sequences of notes was primarily predicted by their music reading training whereas, in novices, it was mostly driven by general visual object-recognition ability, o_v . In other work, people higher in o_v and g performed better in a lung nodule detection task when their radiological training experience was controlled (Sunday et al., 2018). Our findings extend this work by suggesting that with growing experience in a visual domain, the same task that is

originally supported by domain-general mechanisms can come to rely on more specialized processes.

Interestingly, in the present study, g did not predict performance on any of the musical tests (SM or WM tests), in neither music-reading expert group or novice group. It is possible that general intelligence was somewhat restricted in range in this population of Vanderbilt students. However, it is also possible that simple music reading tasks as operationalized here, which do not require translating the notes into their names, sounds, or associated motor patterns, rely more on visual rather than cognitive skills. This may be different from more complex music-learning situations. For instance, when novices learned to perform a simple piece of music from memory, general intelligence was the only significant predictor of their performance (Burgoyne et al., 2019). Of course, no other study has yet measured the contribution of both visual and cognitive domain-general abilities. While cognitive abilities have routinely been considered as an important contribution to musical skill acquisition, we would argue that the visual requirement in both simple and complex musical tasks should not be ignored. Here, o_v was a significant predictor of both matching and working memory tasks with notes, and remained so in experts in the case of working memory. We believe o_v will be a useful addition to future work looking to understand individual differences in the music domain.

This is only the first attempt to relate the recognition of musical notation to general intelligence and general visual abilities. We readily acknowledge the limitations of this work. Our sample was relatively small, and with a larger sample we could have investigated experts with a history of reading musical notation for different purposes (e.g., playing different instruments or singing). We only had three tests with notes, two of them relatively similar in format. Developing a broader range of tests would be useful to better estimate a latent construct related to music reading ability that is independent from task specific requirements, and therefore better able to generalize across tasks and contexts. There are other aspects of music reading expertise besides the recognition of musical notation and short music note sequence processing. However, one ongoing challenge is to create more complex tests that still allow a comparison between experts and novices without any formal training. We started this paper by noting the well-known challenges in drawing causal inferences on the effects of musical training based on correlational evidence. This work is no different, but we have strengthened the evidence for domain-specific perceptual effects with musical notation. In addition, we believe that because our tasks are simple and can be performed by novices and experts alike, they may be sensitive to changes in relatively short training protocols. Other domain-specific effects have been successfully obtained

in short training studies (Chua & Gauthier, 2020) and changes in the perception of musical notation could be among the very first effects of musical training.

References

- Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology*, 25(2), R60–R61. <https://doi.org/10.1016/j.cub.2014.12.011>
- Bean, K. L. (1938). An experimental approach to the reading of music. *Psychological Monographs*, 50(6), 1–80. <https://doi.org/10.1037/h0093540>
- Bialystok, E., & DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 565–574. <https://doi.org/10.1037/a0012735>
- Bratzke, D., Seifried, T., & Ulrich, R. (2012). Perceptual learning in temporal discrimination: Asymmetric cross-modal transfer from audition to vision. *Experimental Brain Research*, 221(2), 205–210. <https://doi.org/10.1007/s00221-012-3162-0>
- Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition*, 54(2), 103–109. [https://doi.org/10.1016/S0278-2626\(03\)00264-1](https://doi.org/10.1016/S0278-2626(03)00264-1)
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296–322.
- Bugos, J. A. (2019). The effects of bimanual coordination in music interventions on executive functions in aging adults. *Frontiers in Integrative Neuroscience*, 13, 68–68. <https://doi.org/10.3389/fnint.2019.00068>
- Burgoyne, A. P., Harris, L. J., & Hambrick, D. Z. (2019). Predicting piano skill acquisition in beginners: The role of general intelligence, music aptitude, and mindset. *Intelligence*, 76, Article 101383. <https://doi.org/10.1016/j.intell.2019.101383>
- Butzlaff, R. (2000). Can music be used to teach reading? *The Journal of Aesthetic Education*, 34(3/4), 167–178. <https://doi.org/10.2307/3333642>
- Chang, T.-Y., & Gauthier, I. (2020). Distractor familiarity reveals the importance of configural information in musical notation. *Attention, Perception, & Psychophysics*, 82(3), 1304–1317. <https://doi.org/10.3758/s13414-019-01826-0>
- Chua, K.-W., & Gauthier, I. (2020). Domain-specific experience determines individual differences in holistic processing. *Journal of Experimental Psychology General*, 149(1), 31–41. <https://doi.org/10.1037/xge0000628>
- Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd, G. (2016). Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PLOS ONE*, 11(7), Article e015763. <https://doi.org/10.1371/journal.pone.0157638>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 94–107. <https://doi.org/10.1037/0096-1523.35.1.94>
- Degé, F., Wehrum, S., Stark, R., & Schwarzer, G. (2011). The influence of two years of school music training in secondary school on visual and auditory memory. *European Journal of Developmental Psychology*, 8(5), 608–623. <https://doi.org/10.1080/17405629.2011.590668>
- Habibi, A., Damasio, A., Ilari, B., Sachs, M. E., & Damasio, H. (2018). Music training and child development: A review of recent findings from a longitudinal study. *Annals of the New York Academy of Sciences*, 1423(1), 73–81. <https://doi.org/10.1111/nyas.13606>
- Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, 11(11), 466–472. <https://doi.org/10.1016/j.tics.2007.08.008>
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161. <https://doi.org/10.2307/1912352>
- Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception*, 26(1), 41–55. <https://doi.org/10.1525/MP.2008.26.1.41>
- JASP Team. (2020). *JASP* (Version 0.14.1)[Computer software]. <https://jasp-stats.org/>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Clarendon.
- Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, 144, 1–13. <https://doi.org/10.1016/j.cognition.2015.07.009>
- 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), Article e52508. <https://doi.org/10.1371/journal.pone.0052508>
- Li, S. T. K., Chung, S. T. L., & Hsiao, J. H. (2019). Music-reading expertise modulates the visual span for English letters but not Chinese characters. *Journal of Vision*, 19(4). <https://doi.org/10.1167/19.4.10>
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40. <https://doi.org/10.1037/1082-989X.7.1.19>
- Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, 18(2), 199–211. <https://doi.org/10.1162/jocn.2006.18.2.199>
- Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, 22(11), 1425–1433. <https://doi.org/10.1177/0956797611416999>
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, 19(3), 712–723. <https://doi.org/10.1093/cercor/bhn120>
- Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PLOS ONE*, 6(5), Article e18082. <https://doi.org/10.1371/journal.pone.0018082>
- Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48. <https://doi.org/10.1006/cogp.1999.0735>
- Richler, J. J., Tomarken, A. J., Sunday, M. A., Vickery, T. J., Ryan, K. F., Floyd, R. J., Sheinberg, D., Wong, A. C.-N., & Gauthier, I. (2019). Individual differences in object recognition. *Psychological Review*, 126(2), 226–251. <https://doi.org/10.1037/rev0000129>
- Richler, J. J., Wilmer, J. B., & Gauthier, I. (2017). General object recognition is specific: Evidence from novel and familiar objects. *Cognition*, 166, 42–55. <https://doi.org/10.1016/j.cognition.2017.05.019>
- Roden, I., Könen, T., Bongard, S., Frankenberg, E., Friedrich, E. K., & Kreutz, G. (2014). Effects of music training on attention, processing

- speed and cognitive music abilities—findings from a longitudinal study. *Applied Cognitive Psychology*, 28, 545–557. <https://doi.org/10.1002/acp.3034>
- Rodrigues, A. C., Loureiro, M. A., & Caramelli, P. (2013). Long-term musical training may improve different forms of visual attention ability. *Brain and Cognition*, 82(3), 229–235. <https://doi.org/10.1016/j.bandc.2013.04.009>
- Rodrigues, A. C., Loureiro, M., & Caramelli, P. (2014). Visual memory in musicians and non-musicians. *Frontiers in Human Neuroscience*, 8, 1–10. <https://doi.org/10.3389/fnhum.2014.00424>
- Salis, D. L. (1980). Laterality effects with visual perception of. *Perception & Psychophysics*, 28(4), 284–292. <https://doi.org/10.3758/BF03204387>
- Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, 15(8), 511–514. <https://doi.org/10.1111/j.0956-7976.2004.00711.x>
- 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. *The 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://doi.org/10.1525/mp.2011.29.2.185>
- Shen, Y., Lin, Y., Liu, S., Fang, L., & Liu, G. (2019). Sustained effect of music training on the enhancement of executive function in pre-school children. *Frontiers in Psychology*, 10, 1910–1910. <https://doi.org/10.3389/fpsyg.2019.01910>
- Sloboda, J. A. (1978). Perception of contour in music reading. *Perception*, 7(3), 323–331. <https://doi.org/10.1068/p070323>
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 1904–1920, 3(3), 271–295. <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>
- Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception & Psychophysics*, 66(2), 183–195. <https://doi.org/10.3758/BF03194871>
- Strait, D. L., & Kraus, N. (2011). Playing music for a smarter ear: Cognitive, perceptual and neurobiological evidence. *Music Perception*, 29(2), 133–146. <https://doi.org/10.1525/mp.2011.29.2.133>
- Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, 261(1), 22–29. <https://doi.org/10.1016/j.heares.2009.12.021>
- Sunday, M. A., Donnelly, E., & Gauthier, I. (2018). Both fluid intelligence and visual object recognition ability relate to nodule detection in chest radiographs. *Applied Cognitive Psychology*, 32(6), 755–762. <https://doi.org/10.1002/acp.3460>
- Sunday, M. A., Tomarken, A., Cho, S.-J. & Gauthier, I. (2021). Novel and familiar object recognition rely on the same ability. *Journal of Experimental Psychology: General* (in press).
- Wang, M. W., & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, 40(5), 663–705. <https://doi.org/10.3102/00346543040005663>
- Wong, A. C.-N., Ng, T. Y. K., Lui, K. F. H., Yip, K. H. M., & Wong, Y. K. (2019). Visual training with musical notes changes late but not early electrophysiological responses in the visual cortex. *Journal of Vision*, 19(7), 8. <https://doi.org/10.1167/19.7.8>
- Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert – but which type? *Psychological Science*, 20(9), 1108–1117. <https://doi.org/10.1111/j.1467-9280.2009.02430.x>
- Wong, Y. K., & Gauthier, I. (2010a). A multimodal neural network recruited by expertise with musical notation. *Journal of Cognitive Neuroscience*, 22(4), 695–713. <https://doi.org/10.1162/jocn.2009.21229>
- Wong, Y. K., & Gauthier, I. (2010b). Holistic processing of musical notation: Dissociating failures of selective attention in experts and novices. *Cognitive, Affective, & Behavioral Neuroscience*, 10(4), 541–551. <https://doi.org/10.3758/CABN.10.4.541>
- Wong, Y. K., & Gauthier, I. (2012). Music-reading expertise alters visual spatial resolution for musical notation. *Psychonomic Bulletin & Review*, 19(4), 594–600. <https://doi.org/10.3758/s13423-012-0242-x>
- Wong, Y. K., Lui, K. F. H., & Wong, A. C. N. (2020). A reliable and valid tool for measuring visual recognition ability with musical notation. *Behavior Research Methods*. Advance online publication. <https://doi.org/10.3758/s13428-020-01461-w>
- Wong, Y. K., Peng, C., Fratus, K. N., Woodman, G. F., & Gauthier, I. (2014). Perceptual expertise and top-down expectation of musical notation engages the primary visual cortex. *Journal of Cognitive Neuroscience*, 26(8), 1629–1643. https://doi.org/10.1162/jocn_a_00616
- Wong, Y. K., & Wong, A. C.-N. (2018). The role of line junctions in object recognition: The case of reading musical notation. *Psychonomic Bulletin & Review*, 25(4), 1373–1380. <https://doi.org/10.3758/s13423-018-1483-0>
- Xu, Z., Adam, K. C. S., Adam, K. C. S., Fang, X., Fang, X., Vogel, E. K., & Vogel, E. K. (2018). The reliability and stability of visual working memory capacity. *Behavior Research Methods*, 50(2), 576–588. <https://doi.org/10.3758/s13428-017-0886-6>

Open practices statement The study was not preregistered. The data sets are available online (<https://doi.org/10.6084/m9.figshare.13643732.v1>), and the materials are available upon request.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.