What is "musical ability" and how do we measure it?

Brooke M. Okada[1]
L. Robert Slevc[1†]
[1] Department of Psychology, University of Maryland, College Park, MD, USA
[†] Corresponding author: slevc@umd.edu
Published 16 December 2021; https://doi.org/10.18061/FDMC.2021.0029

Abstract

There is little consensus on what exactly constitutes musical ability and how to best measure it. Past research has used various tasks; most commonly assessing perceptual skills (e.g., same/different judgments in sequentially presented melodies), but also sometimes production tasks (e.g., singing a series of pitches or tapping along with a musical sequence). Outcome measures have ranged from single indices (e.g., "pitch ability") to composite scores from multiple tasks (e.g., pitch, rhythm, loudness, timbre, etc.). To date, it remains unclear how these different measures/scores relate to one another, limiting the ability to generalize across tasks and results. To address these issues, we assessed 165 participants' performance on 15 representative musical ability tasks to model the unity and diversity of musical abilities. Latent variable model comparisons suggest that musical ability is best represented by related but separable pitch, timing, perception, and production factors.

KEYWORDS: *musical ability, individual differences, factor analysis*

Introduction

Researchers have been measuring musical ability for at least the last hundred years (e.g., Seashore, 1915), and these measures have been used to investigate a variety of topics, including the innate vs. acquired nature of musical skill, relationships between musical and non-musical abilities, and the components of musical ability and their dissociations in amusic patients (e.g., Ayotte, Peretz, & Hyde, 2002; Mosing, Madison, Pedersen, Kuja-Halkola, & Ullén, 2014; Okada & Slevc, 2018). Given the breadth and duration of study, one might assume that we have a pretty good idea what musical ability *is*. However, despite the relatively large literature on individual differences in musical ability, there has been little consensus on what exactly *constitutes* musical ability and how to best measure it (see, e.g., Hallam & Prince, 2003).

Certain debates about the nature of musical ability are perennial. For example, musical ability has often been described as an innate predisposition for music: Seashore (1915) defined it as the "inborn psycho-physic and mental capacities distinguished from skills acquired in training" (p. 129) [1], and Law and Zentner (2012) similarly define it as the "potential for learning music before formal training and achievement" (p. 2). In contrast, others argue that musical ability is experience-based. For example, Shinichi Suzuki claimed that "there is no such thing as an innate aptitude for music" (Hermann, 1981, p. 137).

Other debates are, perhaps surprisingly, less prominent. Most relevant to this paper: does a single factor underlie musical ability, as has been argued for general intelligence (e.g., Carroll, 1997; Spearman, 1904)? Or is what we call "musical ability" actually a disparate set of unrelated (or weakly related) skills?

Most research (including some of our own) has employed a circular definition of musical ability by assuming (at least implicitly) that musical ability corresponds to the task(s) used to measure it. For example, some studies rely on a single score of musical ability, effectively assuming a single underlying factor, while other studies calculate separate scores for melody and rhythm perception tasks, effectively assuming two separable latent perceptual factors. This conflation of task design with underlying constructs is common in individual differences research (take Boring's (1923) oft-cited claim that "intelligence is what is measured by intelligence tests") but is nevertheless problematic.

More broadly, assessing individual differences in musical ability requires the assumption that "musical ability" is a valid construct. However, there may very well be no natural kind (or at least no unitary construct) that corresponds to the term "musical ability." It is easy to assume that musical ability is a psychologically real concept because of the existence of the *label* "musical ability" (and measures thereof) but this, of course, is a fallacy (Brick et al., in press; Malt & Majid, 2013).

Even setting these theoretical / philosophical issues aside, there has been little commonality in the tasks researchers have used to measure (and so define) musical ability. (For short taxonomies of commonly used tasks, see Law & Zentner, 2012; Okada, 2018). On one hand, this is a problem because it makes it difficult to reconcile findings across studies (both practically and because it suggests studies are using different underlying conceptions of musical ability). On the other hand, one advantage of this variety of tasks in the literature is that they may offer a way, albeit indirect, to investigate the unity and diversity of the underlying construct(s) of musical ability.

The work described here is a step in this direction: to assess the unity and diversity of musical ability as defined by the measures typically used in the literature. To this end, we briefly describe an individual differences study investigating the relationships between performance on 15 different musical ability tasks; specifically assessing if performance on these tasks is better explained as an underlyingly unitary ability or as separable abilities for pitch and timing and/or perceptual and production tasks. (Note that methods and results are presented only briefly here; for further details, see Okada, 2018).

Method

We conducted a relatively large, pre-registered (https://osf.io/jwyhu) individual differences study using multiple tasks to measure pitch perception, pitch production, timing perception, and timing production. Because tasks are indirect measures of underlying cognitive constructs (the "task impurity problem" – i.e., even if performance on a task does reflect a specific construct of interest, it involves other processes as well), the analyses below rely on latent variable analysis to estimate the latent ability underlying performance on a set of theoretically related tasks (cf. Miyake et al., 2000; Okada & Slevc, 2018). That is, because a single task measuring a construct of interest may not be indicative of someone's true score (e.g., could include measurement error), multiple measures of each construct of interest were administered. By using latent variable analysis, one can estimate what is common between the tasks measuring a given construct and better estimate the underlying component of interest removed from task-specific effects.

More specifically, a series of confirmatory factor analyses (CFA) were used to assess whether individual differences in these abilities fit better with a unitary model of musical ability (cf. general intelligence), a model with separable abilities for pitch- and timing-related abilities, a model with separable perceptual and productive abilities, or some combination of separable pitch/timing and perception/production abilities.

Participants

165 participants were recruited from the University of Maryland's undergraduate research pool and received class credit for participation.

Measures and Procedure

Participants completed a battery of 15 musical tasks measuring both receptive and productive pitch ability and receptive and productive timing ability (see Table 1 for a summary; for additional information, see Okada (2018) and https://osf.io/mp3u7/).

Table 1: Musical ability tasks, split by perception- and production-based, and by pitch- and timing-based tasks.

	Pitch	**Timing**
Perception	Pitch Discrimination[a] Melody Discrimination[a] Tuning Discrimination[a] Chord Analysis[b] Pitch Thresholds[c]	Rhythm Discrimination[a] Tempo Discrimination[a] Beat Perception[d] Timing Thresholds[c]
Production	Familiar Song Imitation[e] Melody Imitation (note, interval, melodies)[e] Tonal Memory[f]	Synchronization to a Metronome[d] Synchronization Continuation[d] Synchronization to Musical Passage[g]

a. PROMS-S (Zentner & Strauss, 2017); **b.** Wing (1962); **c.** Soranzo & Grassi (2014); **d.** Beat Alignment Test (BAT, Iverson & Patel, 2008); **e.** Pfordresher et al. (2010); **f.** Mowrer (1994); **g.** H-BAT (Fujii & Schlaug, 2013)

Results

In general, task reliabilities were acceptable and comparable to past work (but see discussion below). Individual tasks were generally correlated significantly, but the magnitude of these correlations was not especially high (most correlations ranged from about .2 to .5). This suggests that these tasks are tapping related abilities but are not completely redundant measures.

A single-factor model of musical ability (i.e., one latent factor underlying performance on all 15 musical tasks) showed acceptable model fit, but fit was significantly improved with a two-factor model including separate, but correlated Pitch and Timing factors. Model fit was additionally improved by including both pitch/timing and perception/production factors, supporting related, but dissociable abilities in perception and production of both pitch and timing aspects of music (see Figure 1 for the best fitting model).

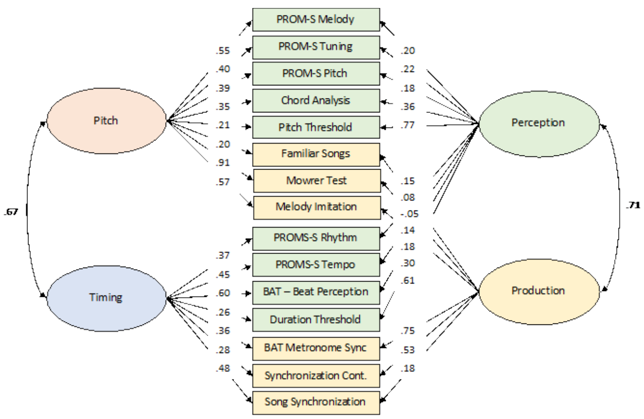

Figure 1: Best fitting CFA model. Observed task scores are represented with rectangles and estimated latent factors with ovals. Single headed arrows from latent factors to measured variables represent standardized factor loadings. Double headed arrows between latent factors represent correlations between the factors.

Discussion

The data presented (briefly) here suggest that individual differences in measures of musical ability are not best explained by a unitary underlying "musical ability" factor. Instead, individuals' performance fit best with a model assuming related but separable underlying latent factors for pitch and timing abilities and for perception and production abilities.

Most research measuring musical abilities has relied on perceptual measures, probably because they are relatively easy to administer and score. However, this may come at a cost: in these data, some of the more commonly used perceptual discrimination tasks had the lowest reliabilities (e.g., the Chronbach's alpha for the PROMS-S measures and the beat perception task from the BAT were below .6). In addition, the "best" individual measure (in terms of being most closely related to the overall latent structure from all tasks) was the Mowrer Test of Tonal Memory, a production task in which participants heard and sung back short melodies (Mowrer, 1996). Of course, production tasks like this are more challenging to administer than perception tasks (e.g., because some participants can be hesitant to sing) and also more challenging to evaluate/score, but we suggest the extra effort may be worthwhile.

Of course, this and all of our conclusions are limited by the tasks we chose to include, which measure only a subset of the types of abilities that are likely relevant to musical ability more broadly (see, e.g., Hallam & Prince, 2003). An additional limitation is that our data come from a group of roughly 20-year-old Western college undergraduates (WEIRD people; Henrich, Heine, & Norenzayan, 2010), and it is not at all obvious that the same pattern would emerge in other demographics.

A broader limitation of this type of work is that it assumes that performance on a task reflects an underlying ability (i.e., a stable *trait*) but of course task performance is also influenced by dynamic changes in the individuals' current *state*. For example, people regularly show temporary attention lapses during cognitive tasks (e.g., Smallwood et al., 2008) and task engagement can be influenced by a variety of factors such as motivation, mood, and task difficulty (e.g., Chiew & Braver, 2014). Thus, a task score from a single point in time may not accurately reflect an individual's ability even on that specific task.

Despite all these issues, it seems likely that researchers will continue to measure musical ability. Of course, we do not suggest that researchers should conduct hours of cognitive testing to extract latent factors (as we did here). We do, however, hope that the diversity among measures that we highlight will encourage more careful consideration of what measures to include should one decide to measure musical ability. In addition, we hope this type of work motivates future research working to better understand the complex nature of musical ability.

Conclusion

Despite over a hundred years of research on individual differences in musical ability, we still do not have a very good idea what it is. Here, we suggest that performance on most current measures of musical ability do not reflect a single underlying factor, but instead reflect (at minimum) separable, but related, musical pitch, timing, perception, and production abilities. While this is certainly not a comprehensive model, we hope it is a useful step toward a better understanding of the structure and diversity of musical abilities.

End Notes

[1] It is worth noting that Seashore's interest in measuring musical ability dovetailed with his interests in eugenics (see, e.g., Koza, 2007), paralleling the

somewhat ignoble history of intelligence testing in the USA (e.g., Kamin, 1974).

References

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125(2),* 238-251. https://doi.org/10.1093/brain/awf028

Boring, E. G. (1923). Intelligence as the tests test it. *New Republic,* 35–37.

Brick, C., Hood, B., Ekroll, V., & de-Wit, L. (in press). Illusory essences: A bias holding back theorizing in psychological science. *Perspectives on Psychological Science.* https://doi.org/10.31234/osf.io/eqma4

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence, 24(1)*, 25-52. https://doi.org/10.1016/S0160-2896(97)90012-X

Chiew, K. S., & Braver, T. S. (2014). Dissociable influences of reward motivation and positive emotion on cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 14(2),* 509-529. https://doi.org/10.3758/s13415-014-0280-0

Fujii, S., & Schlaug, G. (2013). The Harvard Beat Assessment Test (H-BAT): A battery for assessing beat perception and production and their dissociation. *Frontiers in Human Neuroscience, 7,* 1-16. https://doi.org/10.3389/fnhum.2013.00771

Hallam, S., & Prince, V. (2003). Conceptions of Musical Ability. *Research Studies in Music Education, 20(1),* 2-22. https://doi.org/10.1177/1321103X030200010101

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33(2-3),* 61-83. https://doi.org/10.1017/S0140525X0999152X

Hermann, E. (1981). *Shinichi Suzuki: The Man and His Philosophy.* Miami, FL: Summy-Birchard Inc.

Iversen, J. R, & Patel, A. D. (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. In K. Miyazaki et al. (Eds.), *Proceedings of the 10th International Conference on Music Perception & Cognition (ICMPC10).* Adelaide, Australia: Causal Productions.

Kamin, L. J. (1974). *The science and politics of IQ.* Psychology Press.

Koza, J. E., (2007). In sounds and silences: Acknowledging political engagement. *Philosophy of Music Education Review, 15(2),* 168-176. https://doi.org/10.2979/PME.2007.15.2.168

Law, L. N., & 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

Malt, B. C. and Majid, A. (2013). How thought is mapped into words. *WIREs Cognitive Science, 4,* 583-597. https://doi.org/10.1002/wcs.1251

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41(1),* 49-100. https://doi.org/10.1006/cogp.1999.0734

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

Mowrer, A. (1996). *Tonal Memory as an Audition Factor for Choral Ensembles* (Doctoral dissertation). Temple University.

Okada, B. M. (2018). *Towards a Comprehensive Model of Musical Ability* (Doctoral dissertation). University of Maryland. https://doi.org/10.13016/cajm-89aw

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

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M. Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America, 128(4),* 2182-2190. https://doi.org/10.1121/1.3478782

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

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *Journal of Cognitive Neuroscience, 20(3),* 458-469. https://doi.org/10.1162/jocn.2008.20037

Soranzo, A., & Grassi, M. (2014). Psychoacoustics: A comprehensive Matlab toolbox for auditory testing. *Frontiers in Psychology,* 5:712. https://doi.org/10.3389/fpsyg.2014.00712

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology, 15*, 201−293. https://doi.org/10.2307/1412107

Wing, H. D. (1962). A Revision of the "Wing Musical Aptitude Test". *Journal of Research in Music Education, 10(1),* 39-46. https://doi.org/10.2307/3343909

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences, 1400,* 33-45. https://doi.org/10.1111/nyas.13410