

Validity of the Montreal Battery of Evaluation of Amusia: An Analysis Using Structural Equation Modeling

Aldebarán Toledo-Fernández

School of Psychology, Anahuac University,
Mexico City
National Institute of Psychiatry Ramón de
la Fuente Muñiz, Mexico City

Leonor García-Gómez

School of Psychology, Intercontinental
University, Mexico City

Luis Villalobos-Gallegos

National Institute of Psychiatry Ramón de
la Fuente Muñiz, Mexico City

Judith Salvador-Cruz

Department of Neuropsychology, National
Autonomous University of Mexico,
Mexico City

The Montreal Battery of Evaluation of Amusia (MBEA) is the gold standard for diagnosing amusia. We aimed to evaluate its factorial and convergent validity. Data were collected for the MBEA and a self-report Amusic Dysfunction Inventory on a non-random sample ($n = 249$), and the following Structural Equation Modeling (SEM) procedures were conducted: confirmatory factor analysis of the theoretical model; exploratory SEM for alternative non-restricted factor solutions; and structural models with each of these solutions as predictors of the inventory's items. The theoretical model did not prove acceptable goodness of fit, and two- and three-factor non-restricted models were better-fitted solutions for Scale, Contour and Interval tests, and Meter and Memory tests, respectively, than the theoretical one-factor model. This may reflect distinct perceptual processes related to neurocognitive demand. The non-restricted models of Scale, Meter and Memory showed to be acceptable predictors of self-reported capacity for melodic perception, vocal production, rhythmic coordination, and memory.

Key words: amusia, music cognition, structural equation modeling, validity

Introduction

Amusia is a neurocognitive impairment of music perception, memory or performance, primarily explained by a deficit in pitch processing

that could be congenital- or lesion-based (Hyde & Peretz, 2004; Peretz, Champod, & Hyde, 2003), and that cannot be explained by hypoacusia, general neurocognitive functioning, education or exposure to music (Cuddy, Balkwill, Peretz, & Holden, 2005; Ayotte, Peretz, & Hyde, 2002).

Acknowledgements. This study was partially supported by a postgraduate scholarship granted by Mexico's National Council on Science and Technology (CONACYT) [Grant number: 322491]. Authors are grateful to MSc. Viviana de los Ríos Olascoaga for her support in the provision of staff. We thank Ángel Álvarez, Luisa María Díaz, Mateo López, Jackie Vargas, and Alina Miranda for assistance with data collection, and to Aliza Brzezinski for revision of this manuscript.

Correspondence concerning this article should be addressed to Aldebarán Toledo-Fernández, Universidad Anáhuac México Sur, Escuela de Psicología, Av. De las Torres 131, Colonia Olivar de los Padres, Delegación Álvaro Obregón, ZIP: 01780, Ciudad de México, MX. E-mail: atoledo@imp.edu.mx

Received October 26, 2017

Assessment of amusia has been almost exclusively performed using the Montreal Battery of Evaluation of Amusia (MBEA) (Pfeifer & Hamann, 2015).

Though not explicitly stated as a formal measurement model in the original theoretical scheme (see Peretz et al., 2003), the MBEA evaluates cognitive processing of music through six first-order factors: scale, contour, and interval, as part of a *melodic organization* second-order factor; rhythm and meter, as part of a *temporal organization* second-order factor; and *memory-recognition*, which is conditioned by melodic- and temporal organization due to the nature of the task, which implies the previous execution of the melodic and temporal tests.

The MBEA theoretical model assumes modularity of music processing in three levels: 1) with regard to other cognitive functions (Peretz & Coltheart, 2003); 2) between second-order factors (e.g., impairment of melodic organization could exist in the presence of non-impaired temporal organization [Hyde & Peretz, 2004]; impairment of memory-recognition could exist in the presence of non-impaired melodic or temporal organization [Peretz et al., 2003]), and 3) between first-order factors (e.g., impairment of interval processing could exist in the presence of non-impaired contour processing, but not vice versa; and impairment of rhythm could exist in the absence of impaired meter [Peretz et al., 2003]).

Research using the MBEA has seen a significant rising; because of this, its validity is being rigorously tested, as it could be expected of such a relatively new and almost unique measure. Several observations can or have been made in this regard. First, the supposed modularity of music processing is in contradiction with the MBEA's composite score ultimately proposed by Peretz et al. (2003) as a unitary measure (a third-order factor) of music perception and memory functioning. Second, the only study that has undertaken the task

of analyzing the items' dimensionality of the MBEA, has reported failure to replicate the assumed unidimensionality of five of its tests (Nunes-Silva & Haase, 2012). Third, one study (Henry & McAuley, 2013) has highlighted the crucial involvement of decision-making processes in the simple behavior of listening and responding to the MBEA items; these processes are not considered in the theoretical model in which the battery is founded (justifiably because of parsimonious reasons), but could have a modifying role in the structure of the theory. Fourth, little is still known regarding the relation of performance of subjects on the MBEA and actual specific amusic behaviors (e.g., ability to distinguish melodies without lyrics, minimal capacity to sing in tune or to follow the basic rhythm of a song), with several studies approaching the phenomena through the inclusion of non-representative samples of so-called "self-declared amusics" with no clear or measured diagnostic criteria other than the MBEA itself. Fifth, there is still some discrepancy concerning the use of the MBEA in terms of the number of subtests, items, scoring, cutoff scores, diagnostic accuracy and consequent estimation of prevalence (Henry & McAuley, 2013; Peretz & Vuvan, 2017; Pfeifer & Hamann, 2015).

For our consideration, in order to reach a better understanding of amusia, and all other neurocognitive features that might be addressed with, a more precise comprehension of the MBEA is needed, especially as it has been taken for granted as the gold standard for proving the existence of the very phenomenon it intends to measure. None of the studies targeting this knowledge gap (Henry & McAuley, 2013; Nunes-Silva & Haase, 2012; Pfeifer & Hamann, 2015) undertakes the task of thoroughly assessing the validity of the theoretical model underlying the MBEA. We believe that this assessment may provide information for further refinement of the theory behind the

MBEA, and could have practical implications concerning the pertinence of each of its tests and items, and the potential relation of specific music perceptual domains (e.g., perception of pitch and perception of rhythm) to other cognitive and neural processes (e.g., decision-making and frontal lobes).

Structural equation modeling (SEM) provides a way to address this issue based on systematic fit assessment procedures and estimation of relationships between latent constructs corrected for measurement error (Asparouhov & Muthén, 2009; Morin, Arens, & Marsh, 2016). Factorial validity of the MBEA can be assessed using: a) Confirmatory Factor Analysis (CFA) to empirically test the original theoretical model, assuming that specific indicators are only related to specific factors (e.g., variances in items of the MBEA's Rhythm test are caused by one and only one pre-specified factor measuring rhythm perception); and b) Exploratory SEM (ESEM) to estimate non-restricted relationships between multiple measures in the absence of a valid pre-specified model (e.g., variances in items of the MBEA's Rhythm test may be caused by two or more interrelated factors measuring different features of rhythm perception). For evaluation of the criterion validity, SEM can also be utilized to test the capacity of the best factorial models (restrained or unrestrained) to predict specific outcomes (e.g., a three factor model obtained through ESEM may have better associations with self-reported rhythm difficulties than a one factor model obtained through CFA).

The overall purpose of this study was to comprehensively assess the validity of the MBEA, aiming for three objectives: 1) to test the factorial structure of the MBEA's original theoretical model; 2) to explore an alternative factorial structure for the MBEA using ESEM; 3) to evaluate the convergent validity of the MBEA in relation to specific indicators of self-perceived amusic dysfunction.

Method

Participants

Data were collected in three Mexican cities (Mexico City, Veracruz, and Merida) during July 2014-April 2015, through convenience sampling. Inclusion criteria were: 1) age between 14 and 70 years-old (the range was established considering previous studies with this minimum age [Nunes-Silva & Haase, 2012], and also aiming to reduce the effect of age-associated cognitive decline [Deary et al., 2009]); 2) literacy; and 3) oral informed consent after assuring the participant's complete understanding of confidentiality, research purposes, presence of low risks (possible feelings of lost time, boredom, fatigue, or minimal hearing discomfort), procedures for minimizing risks (comfortable volume level, and breaks between tests), and voluntary completion of activities. Exclusion criteria were checked through the participant's self-report of: history of neuropathology (one or more of the following formal diagnoses: traumatic brain injury, stroke, epilepsy, neuroinfection, dementia, schizophrenia, or other neurodegenerative disease); compromised hearing acuity (formal diagnosis of damage to tympanic membrane or inner ear structures, or subjective complaint of disability); and formal music training (two or more years of study in a music school).

Measures

MBEA items consist of piano melodies that participants listen to through headphones. For Scale, Contour, Interval and Rhythm tests, participants are presented with 31 pairs of monophonic melodies and asked to judge whether the two melodies are the same or different. A catch trial is included randomly within each of the first four tests to assure that the participant is paying attention. For Meter test, participants

are presented with a single homophonic melody, and asked to judge if the presented melody is a march or a waltz. For Memory test, a single monophonic melody is played and participants must judge whether they heard it or not during previous tests. A 30-point scale score is computed for each of the tests, and an average score can be calculated from the six subtests to obtain a global measure of music cognition (Peretz et al., 2003). Though the diagnostic accuracy of this scoring procedure has been questioned (Henry & McAuley, 2011), we decided to retain it for the descriptive analysis because the theoretical factorial structure of the MBEA was originally developed on the evidence provided by this scoring system, under the practical assumption that response to individual items could be added and averaged into broader indexes (e.g., a global score), and because this system is still widely used in research of music perception and amusia (for examples, see: Fujito et al., 2018; Peretz & Vuvan, 2017; Tang et al., 2018). For the purpose of the main analyses (CFA and ESEM), we used the participant's responses to individual items.

A 9-item self-report Amusic Dysfunction Inventory (ADI) was designed ex profeso for this study as a way to explore amusic complaints in four domains: melodic perception (item 1: *"I can tell when an instrument is out of tune"*, item 2: *"I'm capable of clearly noticing an incorrect note in a familiar melody"*); rhythmic coordination (item 3: *"I dance fluently and to the rhythm of music"*, item 4: *"It is difficult for me to follow the rhythm of a song with my hands or my feet"*); vocal production (item 5: *"People say I sing out of tune"*, item 6: *"I sing out of tune"*); and memory (item 7: *"I have difficulty in recognizing the melody of a song when it has no lyrics"*, item 8: *"I have troubles remembering melodies that I have heard several times"*, item 9: *"I can only remember lyrics of the songs and I often forget the melodies"*) (all items' statements are literally translated from

Spanish). The inventory is based on the MBEA's theoretical dimensions of melodic/temporal perception and memory, and on complaints of daily-life impairments frequently reported by amusic individuals, or specific dysfunctions observed by researchers (Cuddy et al., 2005; Peretz et al., 2003). Participants were asked to rate frequency (1 – *"Never"*, 2 – *"Rarely"*, 3 – *"Frequently"*, 4 – *"Always"*) of behaviors and situations that could be related to amusic dysfunction.

A demographic questionnaire included items about gender, age, years of education (completed levels), handedness, and a brief checklist of exclusion criteria.

Procedure

After an oral informed consent and a brief checklist for exclusion criteria, the MBEA and the ADI were computer-administered. Participants were instructed to register their responses on spreadsheets, arranged as follows: two columns for responses of each of the MBEA tests ("Yes" or "No" for all the tests except Meter for which the responses were "March" or "Waltz"), and four columns for the Likert-type responses of the ADI. All evaluations were administered in quiet settings, during one session, and carried out by psychologists trained in the procedures of the study.

All procedures were in accordance with the declaration of Helsinki, and approved by the Research Committee of the Anahuac University.

Statistical Analyses

Gender, age, years of education, and handedness, as well as sum scores of the MBEA and distribution of responses on the ADI were described as mean (standard deviation) or frequencies (percentages) for continuous and categorical data, respectively, in order to charac-

terize the sample of participants. Statistical differences and correlations with MBEA global score were computed for each of the demographics aiming to identify significant ($p < .01$) confounders. All missing values were reported for specific variables. This analytical procedure was performed with IBM SPSS Statistics Version 22.

To assess factorial validity of the MBEA, the three one-factor models corresponding to theoretical dimensions of the MBEA were individually assessed through CFA, including: all 90 individual items from Scale, Contour, and Interval tests as indicators of melodic organization, all 60 individual items from Rhythm and Meter

as indicators of temporal organization, and the 30 individual items of Memory test as indicators of the memory-recognition dimension. A one-factor model using all pooled items was also run to assess the global dimension of the MBEA. For every model, previously identified confounders were controlled. To determine the goodness of fit of the models, the following indices and correspondent cutoff values were taken into account: $p > .05$ in chi square test (χ^2); comparative fit index (CFI) $> .95$; Tucker-Lewis index (TLI) $> .90$; and root-mean-square error of approximation (RMSEA) $< .05$ (Hu & Bentler, 1999; Schreiber, Stage, King, Nora, & Barlow, 2015).

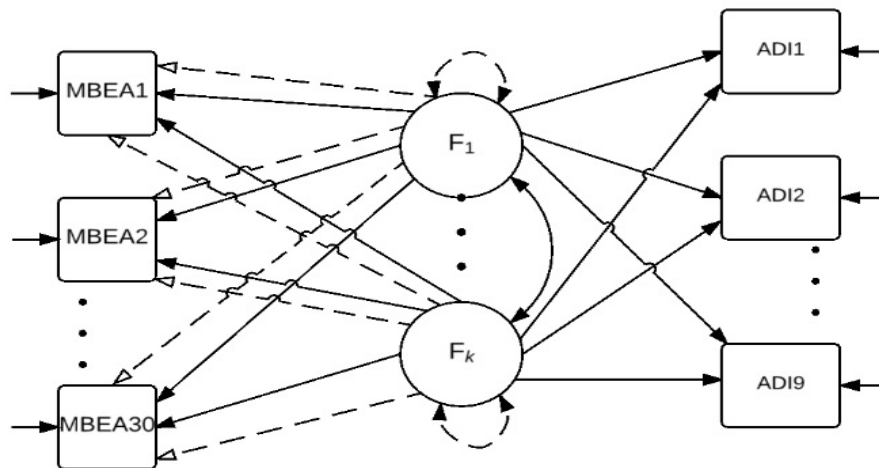


Figure 1 Prototypical representation of ESEM of the MBEA and association with ADI indicators.

Abbreviations: MBEA = Montreal Battery of Evaluation of Amusia; F = Factor; ADI = Amusic Dysfunction Inventory.

Left side of the figure represents prototypical non-restricted models with different number of factors. Right side of the figure represents path analysis, with ESEM factors as predictors of specific ADI's indicators. Confounders are not depicted in the figure.

Squares represent observed variables (indicators) and circles represent unobserved variables (factors). Full unidirectional arrows linked to indicators or factors represent the item uniqueness or factor disturbances. Full unidirectional arrows pointing to indicators represent measurement error. Bidirectional full arrows linking ovals represent factor covariances and correlations. Bidirectional dashed arrows connecting single ovals represent factor variances (Morin et al., 2016). Three-pointed vertical lines indicate ellipsis (e.g., 1, 2, 3...10).

ESEM was conducted to consecutively assess goodness of fit of different non-restricted multifactor models of melodic and temporal dimensions, as well as for individual MBEA tests (Scale-Memory) using Geomin rotation. To choose a specific model, chi square test difference with scaling factor was utilized, considering $p < .01$ to avoid type I error. Significant ($p < .05$) standardized factor loadings and between-factors covariations were identified for each of the retained Scale-Memory models. To evaluate criterion validity of the MBEA, six independent structural models were performed using these factorial solutions as predictors of ADI's indicators, as prototypically depicted in Figure 1. Identified confounders were also controlled for both analytical procedures.

For all SEM analyses, we decided to use individual items as ordered categorical indicators, instead of using scores based on the aggrega-

tion of the correct answers (MBEA scoring system); this procedure allows us to obtain latent variables that represent non-observable traits that underlie the responses of the test (Bollen & Lennox, 1991). The use of linear composites (sum of correct items) was discarded because coefficients estimated with this method may be upwardly or downwardly biased. All SEM procedures were conducted in Mplus Version 6.12 (Muthén & Muthén, 1998-2011) using the weighted least squares (WLSMV) means and variance adjusted estimator.

Results

Demographic Characteristics, MBEA Scores, and Responses to ADI

From 261 participants that concluded the assessment, six individuals were excluded from

Table 1 *Demographic characteristics and MBEA scores*

	Freq. (%) Mean (SD)	Differences or correlations ^a
Gender		$t(247) = 1.86, p = .06$
Male	109 (43.8)	
Female	140 (56.2)	
Handedness		$F(2,248) = .14, p = .86$
Right	229 (92.0)	
Left	17 (6.8)	
Ambidexterity	3 (1.2)	
Age	40.25 (16.37)	$r = -.26, p = .00$
Years of education	15.90 (4.02)	$r = .13, p = .02$
MBEA		
Scale ^b	24.00 (4.34)	
Contour ^c	23.71 (4.11)	
Interval	22.40 (4.09)	
Rhythm ^b	24.99 (3.57)	
Meter ^c	22.93 (5.24)	
Memory ^b	24.77 (4.23)	
Global	23.80 (3.04)	

Notes. $n = 249$; ^aDifferences and correlations are conducted with the MBEA global score as the dependent variable; ^b1 missing value; ^c2 missing values.

Abbreviations: MBEA – Montreal Battery of Evaluation of Amusia.

analysis due to a posteriori self-report of neuropathology history ($n = 5$) and compromised hearing acuity ($n = 1$), and six were excluded due to several discrepancies between item number and correspondent response on one or more MBEA tests. A total of 249 cases were included in the analyses. Table 1 displays the demographic characteristics and MBEA scores for this sample. Mean frequencies and standard deviations of responses on ADI were as follows: item 1 = 2.38(.87); item 2 = 2.35(.92); item 3 = 2.12(.95); item 4 = 1.70(.76); item 5 = 2.58(.99); item 6 = 2.71(1.19) [1 missing value]; item 7 = 2.02(.71); item 8 = 1.73(.64); item 9 = 1.78(.62). Age and years of education showed weak correlations with the MBEA global score (see Table 1), but only the former proved p value $< .01$, and thus was considered as confounder in the main analyses.

Evaluation of Factorial Validity: CFA and ESEM

As shown in Table 2, CFAs of the theoretical one-factor dimensions (melodic organization, temporal organization, memory-recognition, and global) did not prove acceptable goodness-of-fit values or convergence, and neither did the ESEM one-factor alternative solutions for each of the MBEA tests. Alternative multifactor models proved acceptable goodness-of-fit for individual subtests: two-factor solutions on Scale, Contour, Interval, and Rhythm, and three-factor solutions on Meter and Memory. Covariations between F1 and F2 were not observed for any of the two-factor models (Scale: $\beta = .12$, $p = .42$; Contour: $\beta = .15$, $p = .33$; Interval: $\beta = -.11$, $p = .43$;

Table 2 *Goodness-of-fit and difference testing for the MBEA theoretical and alternative model*

	Fit indices					Difference test		
	χ^2	df	RMSEA _s [90% CI]	CFI _s	TLI _s	χ^2_{S-B}	df	p
CFA for theoretical one-factor dimensions ^a								
Global	No convergence.							
Melodic ^b	4739.25	4004**	.02 [.02, .03]	.84	.83			
Temporal ^b	No convergence.							
Memory-recognition ^d	657.58	434**	.04 [.03, .05]	.76	.75			
ESEM alternative solutions for theoretical dimensions ^a								
Melodic								
1	4739.25	4004**	.02 [.02, .03]	.84	.83			
2	4102.03	3914*	.01 [.00, .01]	.95	.95	493.71	90	.00
3	3981.88	3825*	.01 [.00, .01]	.96	.96	137.04	89	.00
4	3869.96	3737	.01 [.00, .01]	.97	.96	123.27	88	.00
5	3765.30	3650	.01 [.00, .01]	.97	.97	115.65	87	.02
Temporal								
1	2416.56	1769**	.03 [.03, .04]	.67	.66			
2	1926.49	1709**	.02 [.01, .02]	.89	.88	321.86	60	.00
3	1777.53	1650*	.01 [.00, .02]	.93	.92	142.97	59	.00
4	1675.37	1592	.01 [.00, .02]	.95	.95	109.97	58	.00
5	1595.92	1535	.01 [.00, .02]	.96	.96	85.29	57	.00
6	1526.24	1479	.01 [.00, .02]	.97	.97	76.20	56	.03

Table 2 continues

Table 2 continued

Table 2 *Goodness-of-fit and difference testing for the MBEA theoretical and alternative model*

	Fit indices					Difference test		
	χ^2	df	RMSEA _S [90% CI]	CFI _S	TLI _S	χ^2_{S-B}	df	p
ESEM alternative solutions for tests ^a								
Scale ^c								
1	644.11	434**	.04 [.03, .05]	.85	.84			
2	438.55	404	.01 [.00, .03]	.97	.97	166.00	30	.00
3	400.54	375	.01 [.00, .02]	.98	.97	40.83	29	.07
Contour ^d								
1	562.19	434**	.03 [.02, .04]	.88	.87			
2	433.12	404	.01 [.00, .02]	.97	.96	105.53	30	.00
3	390.72	375	.01 [.00, .02]	.98	.98	45.15	29	.02
Interval								
1	651.15	434**	.04 [.03, .05]	.84	.83			
2	451.04	404	.02 [.00, .03]	.96	.96	145.40	30	.00
3	412.46	375	.02 [.00, .03]	.97	.96	41.30	29	.06
Rhythm ^c								
1	602.16	434**	.03 [.03, .04]	.86	.85			
2	422.09	404	.01 [.00, .02]	.98	.98	141.59	30	.00
3	384.80	375	.01 [.00, .02]	.99	.99	37.77	29	.12
Meter ^d								
1	591.69	434**	.03 [.03, .04]	.90	.90			
2	493.05	404*	.03 [.01, .03]	.94	.94	90.09	30	.00
3	432.47	375	.02 [.01, .03]	.96	.95	60.68	29	.00
4	389.03	347	.02 [.00, .03]	.97	.96	44.01	28	.02
Memory ^c								
1	657.58	434**	.04 [.03, .05]	.76	.75			
2	447.32	404	.02 [.00, .03]	.95	.94	166.91	30	.00
3	395.02	375	.01 [.00, .02]	.97	.97	55.05	29	.00
4	360.24	347	.01 [.00, .02]	.98	.98	36.18	28	.13

Notes. $n = 249$; ^aControlling for age; ^b3 missing values; ^c1 missing value; ^d2 missing values; * $p \leq .05$; ** $p \leq .001$.

Bolds indicate the best model based on scaled testing (model solution is retained when the next solution reaches $p > .01$).

Rhythm: $\beta = -.21, p = .08$), nor between the factors of the Memory test (F1 with F2: $\beta = .18, p = .22$; F1 with F3: $\beta = .26, p = .10$; F2 with F3: $\beta = .07, p = .70$). Significant associations were only found for Meter test, between F1 and F2 ($\beta = .41, p = .01$) but not between the rest of its factors (F1 with F3: $\beta = -.17, p =$

$.21$; F2 with F3: $\beta = .10, p = .33$). Not controlling for the age confounder, these results showed a similar pattern, with the exception of Meter test that rendered a correlation between F1 and F3 ($\beta = .38, p = .01$). Factor loadings for each of the MBEA retained ESEM solutions are displayed in Table 3.

Table 3 Standardized factor loadings for the ESEM solutions of each of the tests of the MBEA

Items ^b	Scale	Factor loadings ^a											
		Contour			Interval			Rhythm			Meter		
		F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
1	.11(.10)	.23(.11)	.62 ^{**} (.06)	.03(.11)	.83 ^{**} (.05)	.48 ^{**} (.10)	.20(.12)	-.29 ^{**} (.09)	.33 ^{**} (.09)	.26(.11)	.33 ^{**} (.16)	.61 ^{**} (.16)	.04(.15)
2	.73 ^{**} (.06)	.12(.11)	-.19 [*] (.10)	.34 ^{**} (.10)	.67 ^{**} (.07)	.14(.13)	.87 ^{**} (.14)	.66 ^{**} (.07)	.08(.10)	.26(.11)	.78 ^{**} (.09)	.16(.13)	.00(.06)
3	-.23 ^{**} (.09)	.39 ^{**} (.09)	.04(.11)	.21(.11)	.44 ^{**} (.08)	-.31(.12)	.53 ^{**} (.12)	.47 ^{**} (.08)	-.22(.10)	.11(.12)	-.01(.06)	.59 ^{**} (.13)	.26(.13)
4	-.04(.09)	.27(.10)	.09(.09)	.07(.14)	.44 ^{**} (.08)	-.12(.10)	.07(.13)	-.08(.14)	.55 ^{**} (.10)	.38 ^{**} (.10)	.54 ^{**} (.10)	-.02(.11)	-.16(.14)
5	-.13(.10)	.31 ^{**} (.10)	-.08(.10)	.38 ^{**} (.10)	.72 ^{**} (.05)	-.16(.10)	.56 ^{**} (.11)	-.03(.15)	.41 ^{**} (.13)	.39 ^{**} (.13)	.69 ^{**} (.08)	-.15(.12)	.00(.09)
6	.76 ^{**} (.07)	.08(.12)	.57 ^{**} (.07)	.04(.11)	.27(.11)	.48 ^{**} (.11)	.16(.13)	.75 ^{**} (.05)	.00(.06)	.13(.11)	.64 ^{**} (.08)	.02(.06)	-.22(.10)
7 ^c	.66 ^{**} (.06)	.00(.07)	.61 ^{**} (.08)	.32 ^{**} (.09)	.12(.10)	.38 ^{**} (.10)	.19(.12)	.95 ^{**} (.08)	.33 ^{**} (.15)	-.05(.11)	-.06(.15)	.64 ^{**} (.11)	-.08(.18)
8 ^d	.72 ^{**} (.05)	.01(.07)	.07(.09)	.34 ^{**} (.13)	.77 ^{**} (.06)	.00(.06)	.33 ^{**} (.14)	-.00(.03)	.53 ^{**} (.09)	.27(.11)	.04(.14)	.67 ^{**} (.09)	-.12(.12)
9	-.15(.14)	.63 ^{**} (.10)	.60 ^{**} (.06)	.02(.10)	-.04(.12)	.45 ^{**} (.10)	-.01(.12)	.67 ^{**} (.07)	-.01(.13)	.13(.11)	.54 ^{**} (.11)	.30 ^{**} (.09)	-.15(.10)
10	.73 ^{**} (.05)	-.05(.10)	-.00(.02)	.60 ^{**} (.10)	-.01(.12)	.50 ^{**} (.10)	.02(.09)	.72 ^{**} (.06)	-.00(.08)	.28(.12)	.04(.10)	.49 ^{**} (.13)	.50 ^{**} (.14)
11	.67 ^{**} (.07)	.23 ^{**} (.08)	-.43 ^{**} (.09)	.38 ^{**} (.10)	.66 ^{**} (.07)	.19(.11)	.45 ^{**} (.12)	.34 ^{**} (.14)	.50 ^{**} (.11)	.02(.11)	.51 ^{**} (.08)	-.05(.12)	.05(.12)
12	.16(.15)	.76 ^{**} (.08)	.52 ^{**} (.07)	-.01(.08)	.17(.11)	.55 ^{**} (.09)	-.05(.15)	.77 ^{**} (.06)	-.02(.11)	.35 ^{**} (.12)	.56 ^{**} (.06)	.03(.09)	-.08(.11)
13	.73 ^{**} (.05)	.21 ^{**} (.10)	-.09(.11)	.51 ^{**} (.10)	.47 ^{**} (.08)	-.19(.09)	.29(.13)	.71 ^{**} (.08)	-.12(.16)	-.20(.11)	.54 ^{**} (.07)	.00(.09)	.04(.12)
14	-.17(.10)	.33 ^{**} (.10)	.67 ^{**} (.10)	.31 ^{**} (.11)	-.07(.12)	.57 ^{**} (.10)	.00(.09)	.05(.13)	.56 ^{**} (.11)	.80(.09)	.16(.14)	.81 ^{**} (.10)	.00(.04)
15	.10(.13)	.53 ^{**} (.10)	-.25 ^{**} (.12)	.59 ^{**} (.10)	.08(.13)	.49 ^{**} (.14)	.38 ^{**} (.12)	-.21(.11)	.57 ^{**} (.08)	.27(.13)	.57 ^{**} (.07)	.06(.11)	.00(.10)
16	.71 ^{**} (.08)	.27 ^{**} (.09)	-.47 ^{**} (.08)	.18(.09)	-.16(.10)	.54 ^{**} (.08)	.46 ^{**} (.12)	.76 ^{**} (.06)	.02(.10)	-.15(.12)	-.14(.08)	.33 ^{**} (.09)	.06(.12)
17	.11(.15)	.38 ^{**} (.13)	-.02(.11)	.28(.12)	.47 ^{**} (.08)	-.30 ^{**} (.10)	.45 ^{**} (.12)	.67 ^{**} (.07)	-.04(.11)	.18(.14)	.55 ^{**} (.09)	.08(.14)	.25 ^{**} (.12)
18	.63 ^{**} (.07)	.21(.11)	.70 ^{**} (.05)	-.00(.06)	.63 ^{**} (.07)	.01(.10)	.29(.13)	.76 ^{**} (.05)	-.14(.09)	.05(.14)	-.06(.11)	.40 ^{**} (.13)	.09(.15)
19	.00(.05)	.42 ^{**} (.11)	.06(.14)	.82 ^{**} (.09)	.52 ^{**} (.07)	.17(.09)	.22(.12)	-.13(.11)	.49 ^{**} (.10)	-.15(.12)	.54 ^{**} (.06)	.11(.10)	.27(.11)
20	.13(.12)	.51 ^{**} (.14)	.60 ^{**} (.07)	.05(.14)	-.13(.10)	.23(.12)	.49 ^{**} (.13)	-.05(.12)	.51 ^{**} (.09)	.12(.13)	.00(.05)	.64 ^{**} (.15)	.36(.20)
21	-.10(.11)	.42 ^{**} (.10)	.16(.12)	.52 ^{**} (.12)	.06(.10)	.35 ^{**} (.10)	-.07(.13)	.68 ^{**} (.07)	-.19(.11)	.25 ^{**} (.12)	.17(.10)	.71 ^{**} (.12)	.31(.16)
22	.69 ^{**} (.07)	.09(.11)	.80 ^{**} (.05)	-.04(.11)	.72 ^{**} (.06)	.13(.10)	-.13(.15)	.17(.14)	.62 ^{**} (.10)	.46 ^{**} (.12)	-.10(.07)	.47 ^{**} (.11)	.42 ^{**} (.13)
23	.83 ^{**} (.06)	.22(.10)	.69 ^{**} (.05)	-.04(.12)	-.17(.10)	.50 ^{**} (.09)	.23(.13)	.74 ^{**} (.07)	-.01(.08)	-.00(.03)	.69 ^{**} (.11)	-.26 ^{**} (.11)	.21(.13)
24	.12(.16)	.78 ^{**} (.08)	.64 ^{**} (.07)	.22(.12)	.78 ^{**} (.05)	.11(.09)	.19(.18)	.19(.18)	.76 ^{**} (.11)	.12(.12)	.55 ^{**} (.09)	-.00(.01)	.52 ^{**} (.08)
25	.76 ^{**} (.05)	.00(.09)	.63 ^{**} (.07)	-.20(.11)	.07(.12)	.53 ^{**} (.10)	-.01(.10)	-.06(.10)	.47 ^{**} (.09)	.14(.12)	-.02(.11)	.66 ^{**} (.11)	.08(.18)
26	.66 ^{**} (.06)	-.03(.09)	-.22(.12)	.19(.10)	.78 ^{**} (.05)	.00(.06)	.03(.10)	-.03(.12)	.46(.11)	-.02(.10)	-.22(.09)	.35 ^{**} (.12)	.06(.14)
27	.78 ^{**} (.04)	.00(.07)	.66 ^{**} (.05)	-.00(.07)	.00(.01)	.63 ^{**} (.08)	.00(.07)	.75 ^{**} (.06)	.19(.13)	-.12(.11)	.00(.01)	.59 ^{**} (.19)	.86 ^{**} (.13)
28	.77 ^{**} (.04)	-.02(.08)	.12(.14)	.58 ^{**} (.12)	.15(.10)	.46 ^{**} (.09)	.07(.13)	.07(.13)	.28 ^{**} (.12)	-.00(.10)	.76 ^{**} (.07)	-.09(.12)	.00(.19)
29	-.02(.11)	.55 ^{**} (.09)	.64 ^{**} (.06)	-.16(.11)	.77 ^{**} (.05)	-.15(.12)	-.03(.11)	-.04(.09)	.39 ^{**} (.11)	-.00(.10)	.52 ^{**} (.12)	-.18(.12)	.47 ^{**} (.10)
30 ^e	-.02(.12)	.67 ^{**} (.08)	.70 ^{**} (.06)	.14(.08)	.68 ^{**} (.05)	.01(.09)	.07(.12)	.80 ^{**} (.07)	.19(.10)	-.27(.11)	-.07(.09)	.47 ^{**} (.16)	.45 ^{**} (.15)

Notes. $n = 249$; ^aAge confounder is controlled for all the models; ^bItems are different for each test but follow the same numeration. Catch trials aiming to assess alertness of the listener (Scale: trial 7, Contour: trial 28, Interval: trial 28, Rhythm: trial 14) are not taken as effective items and thus are excluded from this numeration. ^c1 missing value in Rhythm; ^dMissing values in: Contour = 1, Meter = 1; ^eMissing values in: Scale = 1, Meter = 1; Memory = 1.

Shaded cells indicate the correct response to the item: "different" for Scale-Rhythm, "march" for Meter, and "No" for Memory. Blank cells indicate the other option of response.

Abbreviations: ESEM – Exploratory structural equation modeling; MBEA – Montreal Battery of Evaluation of Amusia; F – Factor.

* $p \leq .05$; ** $p \leq .001$

Assessment of Convergent Validity: MBEA as Predictor of Self-Perceived Amusic Dysfunction

SEM analyses for testing convergent validity of MBEA alternative models with regard to specific amusic dysfunctions are displayed in Table 4. For each of the structural models, fit indices were as follow: Scale: $\chi^2(656) = 707.49$, $p = .08$; RMSEA = .01 [.00, .02]; CFI = .97; TLI = .97. Contour: $\chi^2(656) = 696.75$, $p = .13$; RMSEA = .01 [.00, .02]; CFI = .97; TLI = .97. Interval: $\chi^2(656) = 704.77$, $p = .091$; RMSEA = .01 [.00, .02]; CFI = .97; TLI = .97. Rhythm: $\chi^2(656) = 689.54$, $p = .17$; RMSEA = .01 [.00, .02]; CFI = .98; TLI = .97. Meter: $\chi^2(618) = 682.43$, $p = .03$; RMSEA = .02 [.00, .029]; CFI = .97; TLI = .96. Memory: $\chi^2(618) = 643.94$, $p = .22$; RMSEA = .01 [.00, .02]; CFI = .98; TLI = .98.

Discussion

The main purpose of this study was to evaluate the validity of the MBEA by thoroughly

analyzing the original and alternative factorial structures via systematic fit assessment procedures, identification of latent constructs, and estimation of relationships between them and with regard to self-report of specific amusic-related impairments.

Factorial Validity

Overall, the findings strongly suggest that the theoretical factorial structure of the MBEA is not a good-fitted measurement model of music neurocognition. More specifically, neither the assumed three second-order factor structure (melodic and temporal organization) of the battery, nor the presumed unifactoriality of each of its tests, stood empirical proof. Instead, the results of ESEM suggest to discard melodic organization and temporal organization as composite measures, and to consider multifactoriality for each of the tests.

For the cases of Scale, Contour, Interval, and Rhythm (Scale-Rhythm), respective two-factor solutions showed better goodness-of-fit when compared to other multifactorial alternatives.

Table 4 *Structural analysis estimates for MBEA tests' ESEM solutions as predictors of ADI*

Item	Scale ^a		Contour ^b		Interval		Rhythm ^a		Meter ^b			Memory ^a		
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F3	F1	F2	F3
Melodic perception ^c														
1	-.19 ^{**} (.07)	-.07(.08)	-.10(.07)	-.16 [*] (.08)	-.13(.07)	-.02(.07)	-.08(.07)	.00(.07)	-.32 [*] (.15)	.03(.13)	.24(.12)	.31(.19)	-.06(.21)	-.42 [*] (.15)
2	-.34 [*] (.06)	-.11(.08)	-.21 ^{**} (.07)	-.24 ^{**} (.08)	-.25 ^{**} (.07)	-.08(.08)	-.18 [*] (.07)	-.15 [*] (.07)	-.42 ^{**} (.14)	.05(.11)	.28 [*] (.12)	.21(.21)	-.13(.17)	-.57 [*] (.16)
Rhythmic coordination ^c														
3	-.10(.07)	-.18 [*] (.07)	-.04(.08)	-.12(.08)	-.12(.07)	-.14(.07)	-.10(.07)	-.14(.08)	-.32 [*] (.13)	.06(.11)	.09(.11)	.12(.12)	.02(.12)	-.26 [*] (.10)
4	-.31 ^{**} (.06)	-.04(.09)	-.23 ^{**} (.07)	-.16(.08)	-.24 ^{**} (.07)	-.08(.09)	-.11(.07)	-.06(.09)	-.56 ^{**} (.14)	.02(.12)	.31 [*] (.13)	.10(.15)	-.17(.11)	-.31 [*] (.13)
Vocal production ^c														
5	-.39 ^{**} (.06)	-.06(.07)	-.29 ^{**} (.07)	-.19 [*] (.07)	-.30 ^{**} (.07)	-.07(.08)	-.15 [*] (.07)	-.00(.08)	.04(.13)	-.25 [*] (.11)	-.08(.12)	.10(.17)	-.06(.11)	-.25(.13)
6	-.42 ^{**} (.06)	-.17 [*] (.08)	-.34 ^{**} (.07)	-.22 ^{**} (.08)	-.39 ^{**} (.07)	-.22 ^{**} (.09)	-.20 ^{**} (.07)	-.05(.08)	-.11(.13)	-.13(.11)	.10(.10)	.09(.21)	.01(.11)	-.35 [*] (.13)
Memory ^c														
7	-.21 ^{**} (.07)	-.05(.09)	-.21 ^{**} (.07)	-.02(.09)	-.25 ^{**} (.07)	-.02(.07)	-.17 [*] (.07)	.06(.08)	-.15(.13)	-.08(.11)	.09(.12)	.18(.11)	-.19(.13)	-.25(.13)
8	-.17 [*] (.07)	.09(.09)	-.09(.07)	-.05(.09)	-.12(.07)	.01(.08)	.01(.08)	.09(.09)	-.24(.15)	-.02(.12)	.09(.12)	.05(.12)	.13(.09)	.10(.12)
9	-.22 ^{**} (.07)	.06(.08)	-.14 [*] (.06)	-.03(.08)	-.11(.07)	-.03(.08)	-.05(.06)	-.04(.08)	-.19(.14)	.08(.11)	.08(.12)	.08(.13)	-.04(.09)	-.30 [*] (.12)

Note. $n = 249$; ^a1 missing value; ^b2 missing values; ^cControlling for age; * $p < .05$; ** $p < .001$.

Abbreviations: MBEA – Montreal Battery of Evaluation of Amusia; ESEM – Exploratory Structural Equation Modeling; ADI – Amusic Dysfunction Inventory.

Observing the factor loading patterns of these four models, a clear tendency can be detected: F1 was mostly loaded by items containing altered stimuli as the second melody (scale-, contour-, interval-, and rhythm-violated conditions, according to Peretz et al. [2003]), and F2 was mainly comprised by non-violated conditions. We suggest labeling these factors: *assessment of difference* (AoD), and *assessment of sameness* (AoS), respectively. Both factors might be reflecting distinct neurocognitive processes or degrees of ability needed to perceive, retain, and compare the two melodies of each trial, as well as different degrees of decision bias partially founded on the difficulty to process each melody.

As already noted by Henry and McAuley (2013) using the signal detection theory, this response process is the result of the additive contributions of the sensitivity of the listener to correctly discriminate between same versus different melodies and his/her response bias, which in the case of the MBEA seems to originate from a systematic measurement error founded in the dichotomous options of response for each of its tests. Thus, the patterns of response may be an effect of common-method bias attributable to the MBEA (and thus a serious limitation of the battery) and not the expression of a critical feature of music cognition itself. However, taking into account the nature of the ESEM models (latent variables are free of measurement errors [Bollen, 1984]), our results suggest that the obtained models are independent of the response bias, and the source of the responses for same versus different is the consequence of two independent variations within the process of perception-retention-comparison-decision, which determine the final response of the listener. Knowing which of these processes is more critical for diagnosis of amusia falls beyond the scope of this study.

Cross-loadings between AoD and AoS further shed light on this process, as most of the

covariates shared negative directions, possibly meaning that in order to produce an accurate assessment of the difference or similarity between melodies, individuals ought to execute one process of perception-retention-comparison-decision while restricting the use of the other (e.g., involvement of AoS – in theory more suitable for less neurocognitively demanding trials – for responding to trials with violated conditions might produce a higher rate of inaccurate responses).

In the case of the Memory test, a two-factor solution, though it showed acceptable goodness-of-fit, was not statistically better informative than a three-factor model. Many of the items loaded distinctively in F1 and F2 in a similar pattern to Scale-Rhythm tests, suggesting a similar distinction of assessment processes for discriminating between recently learned and newly perceived musical material, possibly founded on our proposed assessment processes related to neurocognitive demand. Nonetheless, this three-factor model displayed several heterogeneous cross-loadings, making it harder to parsimoniously support this interpretation. Dependence of Memory test on perception and performance on the rest of the previous tests might have influenced this pattern of modeling.

Convergent Validity

Exploring convergent validity, the value of these two-factor solutions is further stretched by the fact that mostly AoD for each Scale-Rhythm test was moderately associated (β ranging from -.14 to -.42) with ADI's indicators. From all of these tests, Scale's AoD associated better with expected outcomes related to vocal production and melodic perception, whereas Contour and Interval displayed more heterogeneous covariation of AoD and AoS with all of the outcome indicators, nonetheless biased to AoD. This may be explained by the fact that Scale

test seems to have better accuracy for diagnosis of amusia (Peretz & Vuvan, 2017; Goulet, Moreau, Robitaille, & Peretz, 2012) (for this reason, it has been used as a screening measure of amusia in some studies [McDonald & Stewart, 2008; Peretz et al., 2008]).

Contrary as might be expected, no association between Rhythm test's factors and ADI's indicators of rhythmic coordination was found, but only low associations with some of the rest of the ADI. Furthermore, low associations between ADI's memory items and Scale-Interval's AoD were found, but only one ADI's memory indicator (self-perceived capacity to remember song lyrics but not melodies) proved a meaningful covariation with F3 of MBEA's Memory test. Interestingly, indicators of melodic perception, rhythmic coordination, and vocal production (signing out of tune), also displayed moderate covariations with Memory's F3.

This pattern of relationships may move forward the hypothesis of AoD and AoS as two different assessment processes strongly related to task complexity, meaning that Scale's AoD and Memory's F3 items may be fairly more neurocognitively demanding, and thus could be closer to the complexity of real-life musical behaviors. Other studies have already noted the sensitivity and specificity of the Scale and Memory test to detect amusia (Peretz & Vuvan, 2017; Pfeifer & Hamann, 2015; Henry & McAuley, 2013), although one study has reported scale processing as a rather automated function of the brain (Brattico, Tervaniemi, Näätänen, & Peretz, 2006).

Special Considerations for Meter test

For this test, a three-factor solution showed a better goodness-of-fit, though with several cross-loadings between two or even all its factors, and no clear profile of the items as to characterize latent constructs. We hypothesize that, similar to the other MBEA tests, the distinction

between these factors might rely on different degrees of neurocognitive demand, perhaps with regard to interaction between meter pulse and more detailed musical features of the pieces comprising the test, such as tempo, saliency of rhythmic chords, or complexity of note durations (e.g., more frequency of crotchets or quavers). Insufficient number of items and thus non-representativeness of musical features, however, prevent this analysis.

Concerning its convergent validity, Meter's F1 and F3 (both containing fewer items than F2) were significantly associated with expected ADI's indicators pertaining to rhythmic coordination; particularly, F1 displayed inverse association with self-perceived difficulty to follow rhythm with hands or feet and to dance fluidly, whereas F2 showed positive association with the latter indicator. Interestingly though, very similar patterns were also observed in relation to ADI's indicators of perception of melody. These patterns may support our hypothesis (e.g., F1 reflects a different degree of neurocognitive demand than F2); however, not having any more data to support this assumption, it is beyond the scope of this work to speculate on the matter. Thoughtful questioning about the inclusion of the Meter test in the MBEA total scoring can be stated though, as has been somewhat evidenced by other results that signal its particular behavior in contrast with the rest of the MBEA tests (Henry & McAuley, 2013; Nunes-Silva & Haasse, 2012; Paraskevopoulos, Tsapkini, & Peretz, 2010; Toledo-Fernández & Salvador-Cruz, 2015).

Limitations

First, the use of non-random sampling might limit the external validity of the results due to lack of methodological control of common confounders in neuropsychological testing, such as age and years of education. This issue was addressed in the case of age via statistical con-

trol within the tested models. We avoid statistical control of education based on its very weak correlation with MBEA's global score, and because the concept of amusia itself excludes it as a confounder (Ayotte et al., 2002).

Second, the sample size could also limit the reach of the findings because of the number of MBEA's indicators included in the analyzed models. Particularly, CFA models require considerably large sample size to attain an admissible solution when complex models are assessed (e.g., models with high-order factors such as the one proposed by Peretz et al. [2003]). However, the flexibility of ESEM models decreases due to the lack of restriction in the covariance matrices and the use of Geomin rotation (Asparouhov & Muthén, 2009).

Third, items of the ADI were not rigorously developed through a formal methodological process (e.g., items' pooling, piloting, or review by judges) but rather designed *ex profeso* based only on previous questionnaires, reports of amusics' common complaints of daily-life impairments and laboratory-observed dysfunctions (Cuddy et al., 2005; Ayotte et al., 2002). It is interesting to note that recent questionnaires with similar evaluation of melodic perception, vocal production, rhythmic ability and memory were developed contemporarily to our ADI (Müllensiefen, Gingras, Musil, & Stewart, 2014; Pfeiffer & Hamann, 2015), and could be used for further validation. Lastly on this matter, the MBEA being the gold standard for diagnosis of amusia, the criterion validity of the ADI could not be tested *a priori* either. We believe that, in view of the results and the lack of another plausible measure of self-report amusic dysfunction, this study stands as a first examination of the validity of the ADI's items.

Fourth, dependence of observations might have influenced the associations between MBEA's and ADI's outcomes, since participants may have judged their self-perceived musical capacities predisposed by their perceived self-

efficacy on the MBEA tests. This assessment procedure, however, is not rare in studies using the MBEA (Cuddy et al., 2005; Pfeiffer & Hamann, 2015). Ulterior studies could employ random or alternate sequences of the assessment procedure, or use ecologically-valid behavioral measures (e.g., singing, hand coordination, dancing, etc.), aiming to avoid this bias.

Lastly, and thought not exactly a limitation of our study's design *per se*, it is important to highlight the possible influence of acculturation in the observed performance of our sample in the MBEA. Evidence for the effects of musical acculturation on the MBEA has been reported for the Greek population, since their musical system significantly differs from that of the Western world (Paraskevopoulos et al., 2010). Considering that most of the Mexican popular music is founded on the Western tonal system, that for centuries it has received influences from European and North American music, and that Mexicans are currently highly exposed to international music which is also based on this musical system, we believe that our findings could be transferred to other populations. Further cross-cultural studies are needed in order to test this assumption.

Conclusion

Our analyses showed that alternative good-fitted models unveil an even more complex phenomenon of music neurocognition as measured by neuropsychological testing, and the insufficiency of the current, most used measurement model for diagnosing it. Further neuropsychological research on the MBEA should be conducted using latent variable modeling to evaluate these measurement models in accurately pre-diagnosed amusic individuals, as well as in different groups with plausibly-related brain pathologies (e.g., temporal epilepsy, Alzheimer, Parkinson, schizophrenia), while rigorously testing the diagnostic precision of the alternative

factorial models, their assumed dissociation with other neurocognitive functions, and their ecological validity.

References

- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397–438. doi: 10.1080/10705510903008204
- 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. doi: 10.1093/brain/125.2.238
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305–314. doi: 10.1037/0033-2909.110.2.305
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Brattico, E., Tervaniemi, M., Näätänen, R., & Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Research*, 1117(1), 162–174. doi: 10.1016/j.brainres.2006.08.023
- Cuddy, L. L., Balkwill, L.-L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare. A study of “tone deafness” among university students. *Annals of the New York Academy of Sciences*, 1060, 311–324. doi: 10.1196/annals.1360.026
- Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., ... & Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, 92(1), 135–152. doi: 10.1093/bmb/ldp033
- Fujito, R., Mine, M., Hatada, S., Kamimura, N., Morinobu, S., Lang, D. J., ... & Sawada, K. (in press). Musical deficits and cortical thickness in people with schizophrenia. *Schizophrenia Research*. doi: 10.1016/j.schres.2018.01.027
- Goulet, G. M., Moreau, P., Robitaille, N., & Peretz, I. (2012). Congenital amusia persists in the developing brain after daily music listening. *PLoS One*, 7(5), e36860. doi: 10.1371/journal.pone.0036860
- Henry, M. J., & McAuley, J. D. (2013). Failure to apply signal detection theory to the Montreal Battery of Evaluation of Amusia may misdiagnose amusia. *Music Perception*, 30(5), 480–496. doi: 10.1525/MP.2013.30.5.480
- Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, 15(5), 356–360. doi: 10.1111/j.0956-7976.2004.00683.x
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. doi: 10.1080/10705519909540118
- McDonald, C., & Stewart, L. (2008). Uses and functions of music in congenital amusia. *Music Perception: An Interdisciplinary Journal*, 25(4), 345–355. doi: 10.1525/mp.2008.25.4.345
- Morin, A. J. S., Arens, K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*, 23(1), 116–139. doi: 10.1080/10705511.2014.961800
- 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. doi: 10.1371/journal.pone.0089642
- Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus User's Guide (6th Ed)*. Los Angeles, CA: Muthén & Muthén.
- Norman-Haignere, S. V., Albouy, P., Caclin, A., McDermott, J. H., Kanwisher, N. G., & Tillmann, B. (2016). Pitch-responsive cortical regions in congenital amusia. *The Journal of Neuroscience*, 36(10), 2986–2994. doi: 10.1523/JNEUROSCI.2705-15.2016
- Nunes-Silva, M., & Haase, V. G. (2012). Montreal Battery of Evaluation of Amusia. Validity evidence and norms for adolescence in Belo Horizonte, Minas Gerais, Brazil. *Dementia e Neuropsychologia*, 6(4), 244–252. doi: 10.1590/S1980-57642012DN06040008
- Paraskevopoulos, E., Tsapkini, K., & Peretz, I. (2010). Cultural aspects of music perception: Validation of a Greek version of the Montreal Battery of Evaluation of Amusia. *Journal of International Neuropsychological Society*, 16(4), 1–10. doi: 10.1017/S1355617710000494
- Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, 999, 58–75. doi: 10.1196/annals.1284.006
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6(7), 688–691. doi: 10.1038/nn1083
- Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics*, 25, 625–630. doi: 10.1038/ejhg.2017.15
- Pfeifer, J., & Hamann, S. (2015). Revising the diagnosis of congenital amusia with the Montreal Battery of Evaluation of Amusia. *Frontiers in Human Neuroscience*, 9, 1–15. doi: 10.3389/fnhum.2015.00161
- Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2015). Reporting structural equation

- modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, 99(6), 323–337. doi: 10.3200/JOER.99.6.323-338
- Tang, W., Wang, X. J., Li, J. Q., Liu, C., Dong, Q., & Nan, Y. (in press). Vowel and tone recognition in quiet and in noise among Mandarin-speaking amusics. *Hearing Research*. doi: 10.1016/j.heares.2018.03.004
- Toledo-Fernández, A., & Cruz, S. (2015). Exploración de las propiedades psicométricas de la Bateria Montreal de Evaluación de Amusia en una muestra de sujetos con epilepsia de lóbulo temporal. *Salud Mental*, 38(5), 311–319. doi: 10.17711/SM.0185-3325.2015.043