

Validation of the Music Empathizing inventory in China

Psychology of Music

1–17

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03057356211044218

journals.sagepub.com/home/pom



**Qing Zhao^{1,2,3} , Xuejing Lu^{1,2} , Huijuan Zhang^{1,2},
Sandra Djekic⁴, Li Wan³ and Li Hu^{1,2}**

Abstract

The current study aims to cross-culturally validate the Music Empathizing (ME) inventory in China. In total, 515 Chinese participants completed the Chinese version of ME in tandem with two supporting scales, namely, the Interpersonal Reactivity Index (IRI) and the Music USE (MUSE) questionnaire. Confirmatory factor analyses, correlation analyses, *t*-tests, and multiple regressions were conducted. The current results illustrated an innovative structural model for ME (i.e., the Agreement–Disagreement model), revealed significant predictors of ME (i.e., optional music training, interpersonal empathy, and the affective and social reasons for approaching music), and indicated a non-significant sex difference in ME with the Chinese participants. To conclude, the current study showed the validity of ME in evaluating music empathy in a Chinese setting. More importantly, since the previous studies of ME were restricted to Western populations, the current study presented the first observation of ME in Asian cultures.

Keywords

Music Empathizing inventory, the Agreement–Disagreement structural model, Music USE questionnaire, Interpersonal Reactivity Index, confirmatory factor analysis

The Music Empathizing¹ (ME, 9-item) inventory is a self-report scale for testing people's ability to appreciate “the emotional content of a piece of music” (Kreutz et al., 2008, p. 57). It is the first half of the Music Empathizing–Music Systemizing inventory² (ME-MS; Kreutz et al., 2008), and the latter half focuses on the music systemizing. The ME-MS is designed to verify Baron-Cohen's (2009) empathizing–systemizing (E-S) theory in the context of music. According to

¹CAS Key Laboratory of Mental Health, Institute of Psychology, Beijing, China

²Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

³Department of Pain Management, the State Key Clinical Specialty in Pain Medicine, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

⁴School of Applied Arts, University of Arts in Belgrade, Belgrade, Serbia

Corresponding author:

Xuejing Lu, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China.

Email: luxj@psych.ac.cn

the E-S theory (Baron-Cohen, 2009), human brain types ranged on a spectrum from “extreme female brain” (i.e., empathizing; decoding the affective and psychological information of society members) to “extreme male brain” (i.e., systemizing; handling the physical and logical information of non-social objects). Empathy is an essential social skill (Zhao et al., 2021), and several mental disorders have comorbid empathy-related deficiencies (e.g., high functioning autism; Gleichgerrcht et al., 2012). To date, the knowledge of ME (the first and the only self-report scale of music empathy) has been restricted to Western cultures (Table 1). This type of Western-biased investigation has been criticized for decades as it distracts people from cross-cultural understanding and mutual respect (Yang, 2002). The current study aims to validate the ME *ad hoc* in China.

Music empathy consists of two main components: emotional music empathy (i.e., the automatic physical and affective reactions toward music, or the “felt” emotion of music; Sittler et al., 2019) and cognitive music empathy (i.e., the emotional understanding and perspective-taking of music/musicians, or the “perceived” emotions of music; Sittler et al., 2019). ME (9-item) is composed of one item of emotional music empathy (i.e., ME05, “*I often experience physical sensations such as tears, shivers, etc. when listening to certain pieces of music.*”) and eight items of cognitive music empathy (e.g., ME06, “*I never guess the emotions of the performer.*”; negatively scored). Kreutz et al. (2008, p. 59) were concerned that ME might mainly evaluate the “cognitive styles of music process” (i.e., cognitive music empathy).

For ME’s structural validity, Kreutz et al. (2008) recommended treating all ME items as convergently tapping into the overall music empathy (i.e., a one-factor model). To date, this one-factor model of ME has been statistically verified in the United Kingdom (Kreutz et al., 2008), Germany (Linnemann et al., 2018), and Australia (Garrido & Schubert, 2011). Nevertheless, both Kreutz et al. (2008) and Garrido and Schubert (2011) observed an interesting phenomenon; that is, participants’ raw responses on ME’s agreement and disagreement items (before an item scoring³) were inversely loaded on the one factor (i.e., “+” or “-” loaded). On the one hand, the inverse loading phenomenon emphasized that all items should be properly scored before conducting the factor analyses. On the other hand, the above results illuminated the necessity to check the psychometric heterogeneity of ME’s agreement and disagreement items after these items were properly scored. Allison et al. (2011) designed an Agreement–Disagreement model to check the psychometric heterogeneity due to the response direction (see Figure 1). Intriguingly, this Agreement–Disagreement psychometric structure was initially proposed for Empathy Quotient (EQ), a self-report scale embodying the E-S theory in the area of interpersonal empathy (Baron-Cohen & Wheelwright, 2004).

For ME’s conceptual validity, researchers deem that the underpinning concept of ME (i.e., music empathy) should be convergent with interpersonal empathy (Lyvers et al., 2018). Theoretically, music empathy and interpersonal empathy are same-rooted on *Einfühlung* (a German word of aesthetic empathy for arts; Lipps, 1906; Wallmark et al., 2018). Empirically, two groups of researchers investigated the conceptual convergence between ME and interpersonal empathy (Table 1). Dahary et al. (2018) confirmed a moderate positive correlation, $r = .28$, $p = .011$, between ME and EQ with Canadian college students ($N = 81$). In contrast, Garrido and Schubert (2011) identified non-significant correlations (r s range = $[-.05, -.02]$) between ME and two subscales of Interpersonal Reactivity Index (IRI, a self-report scale of interpersonal empathy; M. H. Davis, 1980) with Australian college students ($N = 59$). However, IRI consists of four subscales, namely, fantasy (IRI-FS), perspective-taking (IRI-PT), personal distress (IRI-PD), and empathic concern (IRI-EC; M. H. Davis, 1980). Each subscale has a different obligation: IRI-FS and IRI-PT tap into cognitive empathy for imaginary and real-life characters, respectively; IRI-PD and IRI-EC tap into the self- and other-orientated affective reactions

Table 1. A Review of Studies of the Music Empathizing (ME) Inventory.

Reference	ME inventory information			Participant information		Key findings [¶] and Cronbach's α for ME			
	Version	Statistically identified model [‡] (factor- n -item)	Scoring [*]	Country	N (male%)	Music training	Sex effect [‡]	Correlation with empathy [†]	Other effects [@]
Validation studies of ME									
Kreutz et al. (2008) Survey 1	English	One-factor (ME-13-item)	SUW-17	the UK	442 (32.4%)	(1) professional (2) amateur (3) non-musician	$\times (f = n)$	$1(+s-EQ)$	<i>n.a.</i>
Kreutz et al. (2008) Survey 2	English	One-factor (ME-9-item)	SUW-13	the UK	155 (32.3%)	(1) professional (2) amateur (3) non-musician	$\sqrt{(f > n)}$	<i>n.a.</i>	<i>n.a.</i>
Linnemann et al. (2018)	German	One-factor (ME-9-item)	<i>n.a.</i>	Germany	1014 (52.5%)	<i>n.a.</i>	$1(f > m)$	<i>n.a.</i>	(1, 2)
The current study	Chinese	Two-factor (AG-5-item, DG-4-item)	OSM	China	515 (40.2%)	(1) with optional (2) without optional	$\times (f = m)$	$\sqrt{(+PT); \sqrt{(+ES); 1(+EC); \sqrt{(PD)}}$	(1)
Empirical studies of ME									
Garrido and Schubert (2011)	English	One-factor (ME-9-item)	SUW-13	Australia	59 (55.9%)	(1) formal or non-formal	$\sqrt{(f > m)}$	$\times (s-EC); \times (s-FS)$	(1)
Sandstrom and Russo (2013) Study 1	English	<i>n.a.</i>	<i>n.a.</i>	Canada	42 (4.8% [#])	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>	(1)
Gasselseder (2014)	German	<i>n.a.</i>	<i>n.a.</i>	Denmark	60 (61.7%)	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>	(1)
Poli et al. (2014) Study 2	Italian	<i>n.a.</i>	SUW-13	Italy	23 (73.9%)	(1) professional musicologists	<i>n.a.</i>	<i>n.a.</i>	(1, 2)
Friedman and Taylor (2014) Study 1	English	<i>n.a.</i>	<i>n.a.</i>	America	134 (48.5%)	(1) with 0–17 years of music training	<i>n.a.</i>	<i>n.a.</i>	(1)
Friedman and Taylor (2014) Study 2	English	<i>n.a.</i>	<i>n.a.</i>	America	58 (34.5% [#])	(1) with 0–14 years of group/private music training	<i>n.a.</i>	<i>n.a.</i>	(1)
Kantor-Martynuska and Horabik (2015) Study 2	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>	Poland	107 (45.8%)	(1) musician (2) non-musician	<i>n.a.</i>	<i>n.a.</i>	(1)

(Continued)

Table 1. (Continued)

Reference	ME inventory information			Participant information		Key findings ^a and Cronbach's α for ME					
	Version	Statistically identified model ^b (factor- <i>n</i> -item)	Scoring ^c	Country	N (male%)	Music training	Music training effect ^d	Sex effect ^e	Correlation with empathy ^f	Other effects ^g	α
Kantor-Martynuska and Horabik (2015) Study 3	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>	Poland	81 (44.4%)	(1) musician (2) non-musician	$\sqrt{(.42 > n)}$	$\times (f = m)$	<i>n.a.</i>	(1, 2)	.79
Dahary et al. (2018)	English	<i>n.a.</i>	SUW-13	Canada	81 (35.8%)	(1) five levels of musicianship (from professional to hardly ever play/sing)	$\sqrt{(+)}; \sqrt{(\cap)}$	<i>n.a.</i>	$\sqrt{(+ EQ)}$	(1)	<i>n.a.</i>

Note. *Key findings: $\sqrt{}$ = statistically significant; $\times$ = statistically non-significant; + = positive impact/correlation; † = special results (see below for details).

‡Statistically identified models: structural models of ME identified through factorial analyses.

*Scoring: SUW = simplified unit weighting score [the SUW-17 = $\sum (\text{weight} \times \text{adjusted item score}) - 17$; the SUW-13 = $\sum (\text{weight} \times \text{adjusted item score}) - 13$; OSM = ordinary scoring method [OSM = $\sum (\text{weight} \times \text{item score}) + 20$].

†Music training effect: μ = musicians; η = amateur/participants with optional music training; n = non-musicians/participants without optional music training; For Dahary et al. (2018), + = positive impacts from overall musical training, the number of musical instruments played, and the years of formal instrumental training on the ME score; $\cap$ = participants with a moderate length (0.5 to 3 years) of music theory training had the highest ME score than participants either without (0years) or with an extended length of music theory training (4 or more years).

‡Sex effect: f = female participants; m = male participants; † = Linnemann et al. (2018) did not conduct t-tests for the sex difference, but they found more female than male participants in the high-ME group.

†Correlation with empathy: EQ = the Empathy Quotient; PT = the perspective-taking subscale of the Interpersonal Reactivity Index (IRI); FS = the fantasy subscale of IRI; EC = the empathic concern subscale of IRI; PD = the personal distress subscale of IRI; s- = selective items of the respective scale; † = Using factorial analyses, Survey 1 of Kreuz et al. (2008) reported that items of ME and that of s-EQ loaded on one factor (i.e., indicating a positive correlation); † = The current study showed that for participants with optional music training, ME was not significantly correlated with IRI-EC and was negatively correlated with IRI-PD; for participants without optional music training, however, ME was positively correlated with both IRI-EC and IRI-PD.

@Other effects: (1) the relationship between ME and music-related traits and responses (e.g., absorption and music genre preference); (2) dividing participants into subgroups according to ME and Music Systemizing (MS) scores.

†There were missing values on the variable of sex for these studies, and the presented sex ratio was an estimation according to the available information. In the above table, n.a. = not available.

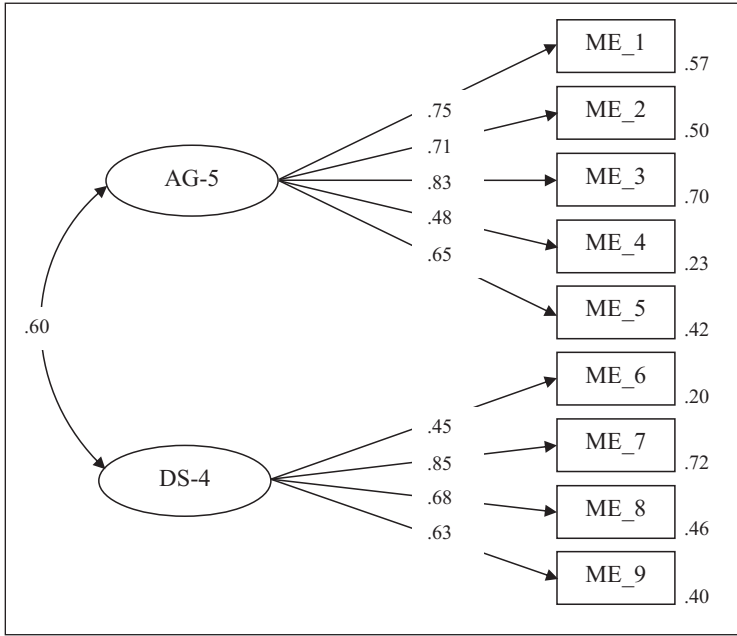


Figure 1. The Standardized Estimates of the Factor Loadings of Music Empathizing (ME, 9-item) Inventory on the Agreement–Disagreement model.

Agreement (AG-5) and disagreement (DS-4) are latent factors clustered according to the response direction of items (i.e., items require participants' agreement and disagreement to indicate empathy, respectively).

From left to right, the value on the error covariance pathway represents the correlation coefficient (r); the values on regression lines represent the standardized factor loading regression weight (b); the values beside rectangles are the squared multiple correlation (R^2).

toward others' emotions, correspondingly (M. H. Davis, 1980). The reason that Garrido and Schubert (2011) only administered IRI-FS and IRI-EC was not given, and the relationship between ME and each IRI subscale should be acquired in the future.

For ME's conceptual validity, researchers also proposed that participants' ME level may be relevant to their purposes for approaching music (Linnemann et al., 2018). For example, Linnemann et al. (2018) reported that German participants with a relatively higher ME score were more likely to listen to music for emotional reasons (e.g., evoking a certain emotion, reducing loneliness, and inhibiting aggression). The above phenomenon is worth evaluating in China—a culture stressing emotional concealment and regulation (Bond, 1993). Chin and Rickard (2012) proposed a self-report questionnaire—Music USE (MUSE) questionnaire—to record one's purpose for approaching music. As per Australian data, Chin and Rickard (2012) presented MUSE with five factors, namely, cognitive and emotion regulation (MUSE-CE, seven items), social connection (MUSE-SC, three items), engaged production/playing (MUSE-EP, nine items), physical exercise (MUSE-PE, three items), and dance (MUSE-DA, two items). Noticeably, the MUSE-DA comprises only two items, which ought to be challenged since an independent factor should consist of three or more reliable items (Ahmed & Stewart, 1981; Cappelleri et al., 2000; McCroskey & McCain, 1974). In all, the best structural model of MUSE waits to be verified, especially in Asian cultures.

For ME's discriminative ability, researchers consider that individuals' ME scores can be relevant to their music training experience (Kreutz et al., 2008). Kreutz et al. (2008) observed that the ME score was the highest with British professional musicians, and was the lowest with non-musicians, with amateurs locating between the two extremes. Similarly, Kantor-Martynuska and Horabik (2015) found that the ME score was higher for young Polish music students recruited from music-specialized high schools than non-music students recruited from general high schools. However, Dahary et al. (2018) identified a more complicated result with Canadian college students. They found that students with a longer formal "music instrumental" training had a higher ME score (Dahary et al., 2018). Nonetheless, they observed a "∩-shape" relationship between the length of formal "music theory" training and the ME score (Dahary et al., 2018). That is, a moderate length of formal "music theory" training (0.5 to 3 years) had the maximum positive impact on students' ME relative to either an absence of (0 years) or an extended length of "music theory" training (≥ 4 years; Dahary et al., 2018).

Remarkably, the aforementioned studies mainly focused on the impact of professional music training on ME. In contrast, the possible benefits of optional music training⁴ on ME have hardly been investigated (Table 1). Noticeably, Dahary et al. (2018) and Sittler et al. (2019) theoretically proposed that music training may help patients with interpersonal empathy-deficiency (e.g., autism spectrum disorders and schizophrenia) to restore their music and interpersonal empathy ability. Relative to professional music training, optional music training is more practical in the clinical setting and more affordable for these clinical patients. Meanwhile, it is indispensable to evaluate the impact of optional music training on ME with the general population before future clinical adoptions.

For ME's discriminative ability, researchers proposed that ME may be sensitive to participants' sex (Kreutz et al., 2008). According to Baron-Cohen's E-S theory (2009) and Extreme Male Brain (EMB) theory (2002), females are expected to be better at empathizing than males. Although the above notion was confirmed with Western participants in terms of interpersonal empathy (female > male; see a review of EQ by Groen et al., 2015), conclusions based on ME were inconsistent (Table 1). With British participants, Kreutz et al. (2008) observed a significant sex difference on a preliminary version of the ME (i.e., ME 13-item, female > male) but not its final version (i.e., ME 9-item, female = male). Hitherto, a significant sex difference on ME (9-item) has been confirmed with Australian participants (female > male; Garrido & Schubert, 2011) but has been denied with Polish participants (female = male; Kantor-Martynuska & Horabik, 2015). These inconsistent findings based on Western populations are intriguing, indicating that the Baron-Cohen's E-S and EMB effect could be less of a trend for music empathy (i.e., ME; Table 1) relative to interpersonal empathy (e.g., EQ; Groen et al., 2015).

Vis-à-vis Asians, the sex difference in ME could be further diminished according to the theory of "Western-Asian culture-sex interaction effect in empathy" (Zhao et al., 2019). Zhao et al. (2019, 2021) observed significant culture-sex interaction effects on interpersonal empathy (EQ and IRI) with Chinese and Australian participants. Specifically, they found that the sex difference in the interpersonal empathy was non-significant for the Chinese participants (female = male) but significant for Australian participants (female > male; Zhao et al., 2019, 2021). They postulated that the interaction effect could be due to the sex-role differentiation in each culture (Zhao et al., 2019, 2021). For example, referencing Confucius' Golden Mean philosophy, Chinese people prefer a balanced manifestation of Yin-Yang (feminine-masculine) in social activities (Zhao et al., 2019, 2021). In contrast, Westerners primarily advocate "emotional female and rational male" (Merten, 2005). Whether the sex difference in ME may also be insignificant in a Chinese setting is innovative to be verified.

The current study aims to address the above research questions in China. Specifically, this study aims to evaluate:

1. The structural validity of a simplified Chinese version of ME using confirmatory factor analyses;
2. The conceptual validity of ME in light of its relationships with IRI and MUSE through Pearson's correlation coefficients;
3. The discriminative ability of ME in terms of both sex (female vs. male) and optional music training (participants with vs. without optional music training) by independent samples *t*-tests;
4. Predictors of ME in a Chinese setting according to regression analyses.

Considering the results of Kreutz et al. (2008) and Garrido and Schubert (2011), the current authors deemed the Agreement–Disagreement model (Allison et al., 2011) as a good-fit structural model of ME. Referencing the theory of “Western–Asian culture–sex interaction effect in empathy” (Zhao et al., 2019), the sex difference in ME was expected to be insignificant with Chinese participants. In line with the design of ME (Kreutz et al., 2008), positive correlations were anticipated between ME and cognitive interpersonal empathy (IRI-PT and IRI-FS). As per Linnemann et al. (2018), ME was expected to be higher for participants who seek affective and social benefits through music (MUSE-CE and MUSE-SC). Finally, since other effects (e.g., the optional music training effect) had not been studied (Table 1), the current authors planned to explore them with no specific postulations.

Methods

Participants

The current participants were recruited online and from a comprehensive university in Dalian, China. Participants were informed about two inclusion criteria of this study: 1. Chinese and 2. who are willing to finish the whole task carefully. Participants were instructed to first register with the experimenters, and then wait for their turns to come to a laboratory to complete the study individually. Referencing the original survey of ME ($N = 155$ to 442 ; Kreutz et al., 2008) and the previous cross-cultural validation of the EQ in China ($N = 588$; Zhao et al., 2018), a reasonable sample size for the current study was around 500.

In total, 517 participants were recruited. However, two had to be excluded: one was unwilling to complete the questionnaire set (for an unknown reason), and the other was unable to finish the ME and MUSE (this participant said that being a farmer, he hardly listened to music). The final sample ($N = 515$, Table 2) consisted of 83.9% Han Chinese (the main ethnicity of Chinese), 40.2% males, 95.3% college students, and 16.7% individuals with optional music training (the mean years of training was 2.95, $SD = 2.82$). The mean age of the final sample was 21.44, $SD = 4.04$. All participants provided their informed consent for this study and got RMB ¥15 as reimbursement. This study was conducted according to a protocol approved by the ethics committee at the Institute of Psychology, Beijing, China.

Measures

All reading materials used in this study were in simplified Chinese characters.

Table 2. Demographic Information of the Current Chinese Participants.

Demographic characteristic	Initial sample with univariate and multivariate outliers* (N = 515)	Sample without univariate and multivariate outliers* (n = 505)
Age (year, <i>SD</i>)	21.44 (4.04)	21.47 (4.07)
Education (year, <i>SD</i>)	14.65 (1.98)	14.66 (2.00)
Sex (<i>n</i> and % for male)	207 (40.2%)	204 (40.4%)
Social role: student/worker (<i>n</i> and % for student)	491 (95.3%)	481 (95.2%)
Ethnicity: Chinese Han/minorities (<i>n</i> and % for Chinese Han)	432 (83.9%)	424 (84.0%)
Music training (<i>n</i> and % for participants with optional music training)	86 (16.7%)	84 (16.6%)
Training length (year, <i>SD</i> for participants with optional music training)	2.95 (2.82)	2.87 (2.66)

SD: standard deviation.

Note. *In this study, confirmatory factor analyses were conducted based on the initial sample ($N = 515$); the other analyses (i.e., *t*-tests, Pearson's correlations, and regression analyses) were conducted based on the sample without outliers ($n = 505$).

Education: participants' overall formal education (from elementary school to the highest education level); Ethnicity: the current sample composes Chinese Han (i.e., the main ethnicity) and 15 Chinese minorities (e.g., Bouyei and Manchu); Music training: dummy coded (0 = *without optional music training*, 1 = *with optional music training*) according to whether participants received any optional music training (i.e., after-school music class, private music tutoring, and other extra-curricular music training beyond the compulsory music courses).

Missing values were identified on two variables: Social role ($n = 3$ missing, which were replaced as one worker and two students based on their other demographic information, including age, study major, and working type) and Music training ($n = 3$ missing, which were replaced by the mode: 0 = *without optional music training*).

Demographic sheet. The current researchers designed a demographic sheet for collecting three-fold information from participants: (1) the basic demographic information (sex, age, ethnicity, and education level), (2) social role (full-time students or worker; in addition, study major for students and working area for worker), and (3) optional music training experience (with or without; if relevant, the years of optional music training).

Music Empathizing (ME). The English version of ME (Kreutz et al., 2008) was translated into simplified Chinese by the current authors following the cross-cultural translation standard (Beaton et al., 2000). First, an English-Chinese bilingual researcher translated the English version of the ME into simplified Chinese. Second, the translation was proofread and agreed on within the current research team. Third, an independent English-Chinese researcher back-translated the simplified Chinese version into English. Fourth, the simplified Chinese translation and its English back-translation were sent to the original authors of ME for checking (Kreutz et al., 2008). Finally, Kreutz et al. (2008) agreed on our translation (see Supplementary Document S1).

Each item of ME is rated on a 4-point Likert scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). According to the current data, Cronbach's α for ME (9-item) was .78. This Cronbach's α value was slightly higher than that of the original English version of ME ($\alpha = .69$; Kreutz et al., 2008) and its translated German version ($\alpha = .75$; Linnemann et al., 2018).

Music USE (MUSE) questionnaire. The current authors translated the English version of the MUSE (Chin & Rickard, 2012) into simplified Chinese, following the same procedure as introduced above.

According to Chin and Rickard (2012), MUSE consists of 24 items divided into five factors (see Supplementary Document S1); namely, MUSE-CE (seven items), MUSE-EP (nine items), MUSE-SC (three items), MUSE-PE (three items), and MUSE-DA (two items). Each MUSE item is rated on a 6-point Likert scale ranging from 0 (*not at all/not applicable to me*) to 5 (*strongly agree*). As per the current sample, Cronbach's α s for the five MUSE factors were .80, .91, .89, .85, and .92, in sequence.

Interpersonal Reactivity Index (IRI). A validated Chinese version of IRI (22-item; Chan, 1986) was administered in this study. The IRI (see Supplementary Document S1) consists of four factors, namely, IRI-FS (six items), IRI-PT (five items), IRI-EC (six items), and IRI-PD (five items) (Chan, 1986). Each IRI item is rated on a 5-point Likert scale ranging from 0 (*does not describe me well*) to 4 (*describes me very well*). Based on the current participants, Cronbach's α s for the four IRI subscales were .66, .69, .62, and .75, respectively.

Data analysis

Confirmatory Factor Analyses (CFAs) were conducted using *Mplus* 8.2 (Muthén & Muthén, 1998–2012). The weighted least squares with mean and variance adjustment estimation was used for these CFAs as per the robustness of this procedure in handling ordered categorical data (Sass, 2011). The best-fit structural model was judged according to the recommendation by Wang and Wang (2012): comparative fit index (CFI) $\geq .95$, Tucker–Lewis Index (TLI) $\geq .90$, standardized root mean square residual (SRMR) $\leq .08$, and root mean square error of approximation (RMSEA) $\leq .08$.

After the CFAs, distributions of self-report scores on ME, MUSE, and IRI were evaluated. In total, six univariate outliers (z -scores > 3.29 ; one on IRI-FS, two on MUSE-CE, and three on IRI-EC) and four multivariate outliers (as per Mahalanobis distance; Allen & Bennett, 2010) were identified. Based on the sample without these outliers ($n = 505$, Table 2), independent samples t -tests were conducted to evaluate both sex effect (*female* vs. *male*) and optional music training effect (*participant with* vs. *without optional music training*) for each self-report score on ME, MUSE, and IRI. Pearson's correlation coefficients among these self-report scores were evaluated separately for participants with and without optional music training. In addition, for participants with optional music training, Pearson's correlation coefficients were also calculated between each self-report score and the length of optional music training.

Finally, three multiple regressions of ME were processed. Regression 1 was a standard multiple regression of ME (method = ENTER) based on four demographic predictors; that is, Sex (0 = *male*, 1 = *female*), Ethnicity (0 = *Chinese Han*, 1 = *Chinese minorities*), Social role (0 = *student*, 1 = *worker*), and Music training (0 = *without optional music training*, 1 = *with optional music training*). Regression 2 was a standard regression of ME (method = ENTER) based on eight self-report variables (i.e., scores on MUSE and IRI). Regression 3 was a two-stage hierarchical multiple regression (method = STEPWISE) based on both the demographic predictors (the first stage) and the self-report variables (the second stage). Apart from CFAs, the other analyses were conducted using SPSS Version 27 (IBM Corp).

Results

CFA

CFA results for both ME and MUSE are presented in Supplementary Document S2.

ME. The best-fit model of ME was the Agreement–Disagreement model (Figure 1), CFI = .97, TLI = .96, RMSEA [90% CI] = .070 [.054, .085], SRMR = .038. In contrast, the original one-factor model of ME was unfit for the current data, CFI = .86, TLI = .82, RMSEA [90% CI] = .144 [.130, .158], SRMR = .076. As per Allison et al. (2011), the total score of the Agreement–Disagreement model of ME (9-item) was used in this study.⁵

MUSE. Amongst the structure candidates of MUSE, a 22-item four-factor model was identified as the best model, CFI = .96, TLI = .96, RMSEA [90% CI] = .071 [.066, .077], SRMR = .047 (see Supplementary Document S2).

Independent samples t-test

The mean scores on ME, MUSE, and IRI for overall participants, two music training groups, and two sex groups are illustrated in Supplementary Document S3.

Music training effect. The music training effect (all were *participant with optional music training* > *without optional music training*) was significant on ME, $t(503) = 3.42, p < .001$, 95% CI = [0.79, 2.91], Cohen's $d = 0.41$ and three MUSE subscales, namely, MUSE-CE, $t(503) = 2.11, p = .035$, 95% CI = [0.11, 3.10], Cohen's $d = 0.25$, MUSE-EP, $t(125.99) = 7.08, p < .001$, 95% CI = [6.42, 11.39], Cohen's $d = 0.80$, and MUSE-SC, $t(138.13) = 3.11, p = .002$, 95% CI = [0.42, 1.87], Cohen's $d = 0.32$.

Sex effect. The sex effect (both were *female* > *male*) was significant on IRI-FS, $t(503) = 1.99, p = .047$, 95% CI = [0.01, 1.42], Cohen's $d = 0.18$ as well as IRI-PD, $t(503) = 4.33, p < .001$, 95% CI = [0.81, 2.15], Cohen's $d = 0.39$.

Pearson's correlation

Pearson's correlation coefficients amongst continuous variables were calculated separately for participants with and without optional music training (see Supplementary Document S4).

With optional music training. ME was positively correlated with MUSE-CE, $r(82) = .39, p < .001$, MUSE-EP, $r(82) = .48, p < .001$, MUSE-SC, $r(82) = .40, p < .001$, MUSE-PE, $r(82) = .32, p = .003$, IRI-FS, $r(82) = .31, p = .004$, and IRI-PT, $r(82) = .33, p = .002$, but negatively correlated with IRI-PD, $r(82) = -.25, p = .022$. Meanwhile, ME was not significantly correlated with IRI-EC, $r(82) = .18, p = .109$ or the length of optional music training, $r(82) = .14, p = .205$.

Without optional music training. ME was positively correlated ($p < .001$, unless otherwise specified) with all scores on MUSE and IRI, namely, MUSE-CE, $r(419) = .52$, MUSE-EP, $r(419) = .29$, MUSE-SC, $r(419) = .42$, MUSE-PE, $r(419) = .35$, IRI-FS, $r(419) = .39$, IRI-PT, $r(419) = .29$, IRI-EC, $r(419) = .23$, and IRI-PD, $r(419) = .12, p = .011$.

Multiple regression analysis

Results for the regression analyses are presented in Table 3. The multicollinearity did not interfere with the current result interpretations (all VIFs ≤ 1.70).

Table 3. Regression Analyses of Music Empathizing (ME) Inventory Based on Chinese Participants ($n = 505$).

Regression 1 (Method = ENTER)					Regression 3 (Method = STEPWISE)				
B	95% CI	β	p-value	sr^2	B	95% CI	B	p-value	sr^2
Sex	[-1.24, 0.39]	-.05	.305	.002	/	/	/	/	/
Ethnicity	[-2.12, 0.03]	-.08	.058	.007	/	/	/	/	/
Social role	[-2.96, 0.78]	-.05	.254	.003	/	/	/	/	/
Music training	[0.84, 2.98]	.16	<.001	.024	0.96	[0.11, 1.82]	.08	.028	.006
Total R^2	/	/	/	/	/	/	/	/	/

Regression 2 (Method = ENTER)					Regression 3 (Method = STEPWISE)				
B	95% CI	β	p-value	sr^2	B	95% CI	B	p-value	sr^2
MUSE-CE	[0.12, 0.25]	.25	<.001	.041	0.21	[0.15, 0.27]	.29	<.001	.059
MUSE-EP	[0.01, 0.07]	.10	.018	.007	/	/	/	/	/
MUSE-SC	[0.09, 0.32]	.16	<.001	.015	0.29	[0.19, 0.39]	.22	<.001	.039
MUSE-PE	[-0.02, 0.19]	.07	.108	.003	/	/	/	/	/
IRI-FS	[0.17, 0.34]	.22	<.001	.041	0.26	[0.18, 0.35]	.23	<.001	.046
IRI-PT	[0.06, 0.27]	.12	.002	.012	0.19	[0.09, 0.29]	.14	<.001	.017
IRI-EC	[-0.02, 0.18]	.06	.135	.003	/	/	/	/	/
IRI-PD	[-0.08, 0.09]	<.01	.919	<.001	/	/	/	/	/
Total R^2	/	/	/	/	0.38***	/	/	/	/

Note. Sex (0 = male, 1 = female); Ethnicity (0 = Chinese Han, 1 = Chinese minorities); Social role (0 = student, 1 = worker); Music training (0 = without optional music training, 1 = with optional music training).
MUSE = Music USE Questionnaire; MUSE-CE = total score for MUSE cognitive and emotional regulation subscale; MUSE-EP = total score for MUSE engaged production subscale; MUSE-SC = total score for MUSE social connection subscale; MUSE-PE = total score for MUSE physical exercise subscale; IRI = Interpersonal Reactivity Index (Chinese version, 22-item); IRI-FS = the total score for IRI fantasy subscale; IRI-PT = the total score for IRI perspective-taking subscale; IRI-EC = the total score for IRI empathic concern subscale; IRI-PD = the total score for IRI personal distress subscale.
For R^2 (multiple correlation squared), * $p < .050$. ** $p < .010$. *** $p < .001$.

Regression 1. In sum, four demographic predictors (Sex, Ethnicity, Social role, and Music training) accounted for 3.3% variance of the ME, $R^2 = .03$, adjusted $R^2 = .03$, $F(4, 500) = 4.31$, $p = .002$, Cohen's $f^2 = .03$. However, only the prediction of Music training on ME was significant, $p < .001$.

Regression 2. In total, eight self-report scores (the scores on MUSE and IRI) explained 38.9% variance of the ME, $R^2 = .39$, adjusted $R^2 = .38$, $F(8, 496) = 39.54$, $p < .001$, Cohen's $f^2 = .64$. Significant predictors were MUSE-CE, $p < .001$, MUSE-EP, $p = .018$, MUSE-SC, $p < .001$, IRI-FS, $p < .001$, and IRI-PT, $p = .002$.

Regression 3. In all, five significant predictors were included in the regression, taking 38.2% variance of ME explained, $R^2 = .38$, adjusted $R^2 = .38$, $F(5, 499) = 61.77$, $p < .001$, Cohen's $f^2 = .62$. These predictors were Music training, $p = .028$ as well as four self-report scores (all $p < .001$; MUSE-CE, MUSE-SC, IRI-FS, and IRI-PT).

Discussion

In this study, we validated a simplified Chinese version of ME (9-item) in China, and presented a new structural model (the Agreement–Disagreement model) for ME. As anticipated, insignificant sex difference was identified on ME with the current Chinese participants. Meanwhile, based on the current data, optional music training was found to be a benefit to ME. In addition, four psychological traits were revealed as positive predictors of ME, namely, two interpersonal cognitive empathy traits (IRI-PT and IRI-FS) as well as affective and social reasons for approaching music (MUSE-CE and MUSE-SC). Finally, a negative correlation between ME and personal distress (IRI-PD) was observed among the current participants with optional music training. As far as the current authors know, this is the first evaluation of ME beyond Western cultures (Table 1).

The Agreement–Disagreement model was designed by Allison et al. (2011) to account for the psychometric differences between self-report agreement and disagreement items (after all items were properly scored). As per Allison et al. (2011), relative to the agreement items, the disagreement items require participants' extra effort to do an inverse rating on the Likert scale. Similarly, Kreutz et al. (2008, p. 69) commented that although disagreement items may "give rise to some increase of error variance," these disagreement items force "the participants to concentrate more on individual items"; therefore, they designed ME with both agreement and disagreement items. Surpassing the one-factor model, the Agreement–Disagreement model can account for the psychometric heterogeneity due to the response direction, but still suggests the conceptual homogeneity of a scale (i.e., agreement and disagreement items tap into the overall empathy convergently; Allison et al., 2011). To date, the model fitness of the Agreement–Disagreement model with the ME has only been investigated in this study with Chinese participants. An intriguing and important research question is whether the Agreement–Disagreement model of ME is specifically fit for the Chinese population (due to Chinese ethnic, cultural, and linguistic characteristics) or if it is inclusively fit for other cultures as well.

As anticipated, the current Chinese participants' sex difference in ME was insignificant. On the one hand, this result aligns with the diminished sex difference observed in interpersonal empathy with Chinese populations (e.g., EQ and IRI; Zhao et al., 2018, 2019, 2020, 2021). On the other hand, it is consistent with Zhao et al.'s (2019, 2021) theory of "Western–Asian culture–sex interaction effect in empathy." As per this theory, the sex difference in empathy could be smaller if a culture emphasizes less sex-role differentiation (Zhao et al., 2019). Relatively, the

social expectation on males and females could be more convergent for Chinese (see Confucius' Golden Mean and the Yin–Yang philosophies) than Western populations (see the notion of “emotional females and rational males”; Zhao et al., 2018, 2019, 2020, 2021). The current results may indicate a culture–sex interaction effect in music empathy, waiting to be verified by future cross-cultural comparisons (Zhao et al., 2019, 2021).

Nonetheless, it should be stressed that Westerners' sex difference on the ME (Table 1) was also smaller than it was postulated based on interpersonal empathy (e.g., Groen et al., 2015). The current authors considered that relative to externalizing interpersonal empathy (e.g., EQ), males could feel less criticized/self-criticized as losing mannish (i.e., the symbol of authority in male-centered societies) for exhibiting music empathy (i.e., ME). Consequently, it may be sensible to adopt music training to help individuals with empathizing issues (e.g., racists, sexists, and sexual offenders; C. H. Davis et al., 2021; Varker & Devilly, 2007) to lift their empathy without “hurting” their dominant-cohort identities (e.g., ethnocentrism and masculism). The above proposal is in line with some scholars who promote music as the best “weapon” in the battle against of anti-racism (Rastas & Seye, 2019).

Encouragingly, the current results suggested that optional music training fostered ME with the Chinese participants recruited from the general population (Table 3). Previous researchers recommended adopting music training to help clinical cohorts with empathy deficiency (Dahary et al., 2018; Sittler et al., 2019). Apparently, optional music training is practical and financially manageable for these clinical patients over professional music training. Relevant to the clinical adoption, three shreds of innovative findings from this study should be noted. First, the current result indicated that ME was specifically predicted by the cognitive but not emotional components of interpersonal empathy (IRI-FS and IRI-PT vs. IRI-EC and IRI-PD, see Table 3). This result is concurrent with Kreutz et al.'s (2008) design for ME (see Introduction). Noticeably, lacking cognitive interpersonal empathy is the key issue underpinning “empathy disorders” (e.g., high functioning autism; Gleichgerrcht et al., 2012). The current results implied a positive succession among ME, optional music training, and cognitive interpersonal empathy.

Second, the current results suggested that participants' ME was relevant to their purpose for approaching music. The purpose can be multi-fold, including affective (MUSE-CE), social (MUSE-SC), as well as practical reasons (e.g., engaged playing of music and physical exercise along with the music; MUSE-EP and MUSE-PE). The current results revealed that these former affective and social reasons were positive predictors of ME, but these latter practical reasons were not (Table 3). The above results contain important implications on the clinical adaptation of music training; namely, to enhance clients' empathy, physicians and trainers should consciously focus on inducing clients' affective reactions to music and strengthening the social impact of music on clients' life, rather than focusing on clients' playing skills or simply relying on passive listening techniques.

Finally and remarkably, a negative correlation between ME and IRI-PD was identified with the current participants with optional music training (Supplementary Document S4). Personal distress (IRI-PD) is self-orientated inversive feelings during empathy (e.g., nausea and panic; López-Pérez et al., 2014). Experiencing intense personal distress may drive people from altruism to self-protection (Batson et al., 1987). Moreover, empathy-related personal distress is more serious for collectivism than individualism (e.g., Chinese vs. Australians; Zhao et al., 2019). Relative to individualism, collectivism has a blurred boundary between “self” and “others,” triggering people from collectivist cultures to encounter more personal distress but externalize less cognitive empathy (Zhao et al., 2019). In addition, Zhao et al. (2020) found that ordinary school education may lift Chinese college students' cognitive empathy and personal distress tandemly. The current results, in contrast, suggest that optional music training may help

youths enhance cognitive empathy and dampen personal distress (also see Wu & Lu, 2021). This observation is fascinating as people can enjoy sad music without being distressed (Garrido & Schubert, 2011). We reckon that music training may teach people to engage in empathy actively without blurring the psychological distinction between “self” (i.e., the music player) and “others” (e.g., the music and composer) (also see Garrido & Schubert, 2011). This postulation can be good news for emotionally vulnerable cohorts (e.g., Chinese females; Zhao et al., 2019, 2020) as well as clinical populations (e.g., depression; Domes et al., 2016), who are fighting to show empathy for others as well as to protect their own emotional wellness.

Limitations of this study should be noted. First, the current sample was mainly Chinese university students, and the validity of the newly translated version of ME requires replications with the other Chinese populations (e.g., Chinese youths). Second, the sample sizes of the participants with ($n = 86$) and without optional music training ($n = 429$) were biased. Although this biased sample ratio might reflect the popularity of optional music training in China, the uneven sample sizes could diminish the statistical power of the current analyses. Third, in this study, the optional music training was dummy coded. Nevertheless, optional music training can be further sub-grouped (e.g., online, offline, one-on-one, and group training). It is crucial to identify which type of music training has the maximum impact on ME. Fourth, no re-test was conducted in this study, and the reliability of the ME was unclear. Fifth, besides demographic and psychological variables covered by the current study, further researchers may like to explore biomarkers of ME, such as participants' prolactin level (Huron, 2011) and their PRL's genetic polymorphisms (Sittler et al., 2019). Finally, to illuminate an entire image of Baron-Cohen's E-S theory on music, a validation of MS in China is imperative in the future.

Acknowledgements

We would like to thank all participants of the current study. We appreciate the original authors of ME and MUSE permitting us to translate the two scales into the simplified Chinese versions. We would also like to thank Mr. Gang Zhao and Mr. Andrew Mathis for helping us proofread the manuscript.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The National Natural Science Foundation of China to XL (No. 32171077 and 31701000) and LH (No. 31822025 and 32071061), and Guangzhou Medical University Post-doc Research Funding to QZ.

ORCID iDs

Qing Zhao  <https://orcid.org/0000-0001-7163-7779>

Xuejing Lu  <https://orcid.org/0000-0001-7586-5576>

Supplemental material

Supplemental material for this article is available online.

Notes

1. In this study, we adopted “music empathy” and “interpersonal empathy” to concisely distinguish the empathy evaluated by ME and IRI, respectively.
2. A scale introduction is given in Supplementary Document S1.
3. ME consists of five agreement items (i.e., ME01 to ME05) and four disagreement items (i.e., ME06 to ME09). All agreement items should be positively scored, while all disagreement items should be negatively scored (Kreutz et al., 2008).

4. In China, optional music training includes a variety of extracurricular, non-compulsory, and amateur music courses, which are accessible to children, students, workers, and seniors. Its music genres cover traditional Chinese music, Western classical music, popular music, soft rock, and so on; meanwhile, its course format includes online, offline, one-on-one, small group, and ensemble lessons.
5. Allison et al. (2011) stated that agreement and disagreement items are conceptually homogenous in terms of evaluating the overall empathy though these items are psychometrically heterogenous.

References

- Ahmed, S. M., & Stewart, R. A. (1981). Factor analysis of the Machiavellian Scale. *Social Behavior and Personality: An International Journal*, 9, 113–115.
- Allen, P., & Bennett, K. (2010). *PASW statistics by SPSS: A practical guide* (Version 18.0). National Library of Australia.
- Allison, C., Baron-Cohen, S., Wheelwright, S. J., Stone, M. H., & Muncer, S. J. (2011). Psychometric analysis of the Empathy Quotient (EQ). *Personality and Individual Differences*, 51, 829–835. <https://doi.org/10.1016/j.paid.2011.07.005>
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, 6, 248–254. [https://doi.org/10.1016/s1364-6613\(02\)01904-6](https://doi.org/10.1016/s1364-6613(02)01904-6)
- Baron-Cohen, S. (2009). Autism: The empathizing–systemizing (E-S) theory. *Annals of the New York Academy of Sciences*, 1156, 68–80. <https://doi.org/10.1111/j.1749-6632.2009.04467.x>
- Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34, 163–175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>
- Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality*, 55, 19–39. <https://doi.org/10.1111/j.1467-6494.1987.tb00426.x>
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25, 3186–3191. <https://doi.org/10.1097/00007632-200012150-00014>
- Bond, M. H. (1993). Emotions and their expression in Chinese culture. *Journal of Nonverbal Behavior*, 17, 245–262. <https://doi.org/10.1007/BF00987240>
- Cappelleri, J. C., Gerber, R. A., Kourides, I. A., & Gelfand, R. A. (2000). Development and factor analysis of a questionnaire to measure patient satisfaction with injected and inhaled insulin for type 1 diabetes. *Diabetes Care*, 23, 1799–1803. <https://doi.org/10.2337/diacare.23.12.1799>
- Chan, C. (1986). *The relations between age, sex-role, orientation of human and empathy* [Unpublished master's thesis, Department of Education, National Chengchi University, Taipei, Taiwan].
- Chin, C., & Rickard, N. S. (2012). The music USE (MUSE) questionnaire: An instrument to measure engagement in music. *Music Perception: An Interdisciplinary Journal*, 29, 429–446. <https://doi.org/10.1525/mp.2012.29.4.429>
- Dahary, H., Palma Fernandes, T., & Quintin, E. M. (2018). The relationship between musical training and musical empathizing and systemizing traits. *Musicae Scientiae*, 24, 113–129. <https://doi.org/10.1177/1029864918779636>
- Davis, C. H., Krafft, J., Hicks, E. T., & Levin, M. E. (2021). The role of psychological inflexibility and perspective taking in anti-racism and anti-sexism. *Personality and Individual Differences*, 175, Article 110724. <https://doi.org/10.1016/j.paid.2021.110724>
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Journal Supplement Abstract Service Catalog of Selected Documents in Psychology*, 10, 85–85.
- Domes, G., Spenthof, I., Radtke, M., Isaksson, A., Normann, C., & Heinrichs, M. (2016). Autistic traits and empathy in chronic vs. episodic depression. *Journal of Affective Disorders*, 195, 144–147. <https://doi.org/10.1016/j.jad.2016.02.006>
- Friedman, R. S., & Taylor, C. L. (2014). Exploring emotional responses to computationally-created music. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 87–95. <https://doi.org/10.1037/a0034940>

- Garrido, S., & Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: A literature review and experiment. *Music Perception*, 28, 279–295. <https://doi.org/10.1525/mp.2011.28.3.279>
- Gasselseder, H.-P. (2014). Re-scoring the game's score: Dynamic music, personality and Immersion in the ludonarrative. *IADIS International Journal*, 12, 17–34.
- Gleichgerricht, E., Torralva, T., Rattazzi, A., Marengo, V., Roca, M., & Manes, F. (2012). Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. *Social Cognitive and Affective Neuroscience*, 8, 780–788. <https://doi.org/10.1093/scan/nss067>
- Groen, Y., Fuermaier, A. B. M., Den Heijer, A. E., Tucha, O., & Althaus, M. (2015). The Empathy and Systemizing Quotient: The psychometric properties of the Dutch version and a review of the cross-cultural stability. *Journal of Autism and Developmental Disorders*, 45, 2848–2864. <https://doi.org/10.1007/s10803-015-2448-z>
- Huron, D. (2011). Why is sad music pleasurable? A possible role for prolactin. *Musicae Scientiae*, 15, 146–158. <https://doi.org/10.1177/1029864911401171>
- Kantor-Martynuska, J., & Horabik, J. (2015). Granularity of emotional responses to music: The effect of musical expertise. *Psychology of Aesthetics, Creativity, and the Arts*, 9, 235–247. <https://doi.org/10.1037/a0039107>
- Kreutz, G., Schubert, E., & Mitchell, L. A. (2008). Cognitive styles of music listening. *Music Perception*, 26, 57–73. <https://doi.org/10.1525/MP.2008.26.1.57>
- Linnemann, A., Kreutz, G., Gollwitzer, M., & Nater, U. M. (2018). Validation of the German version of the Music-Empathizing-Music-Systemizing (MEMS) Inventory (Short Version). *Frontiers in Behavioral Neuroscience*, 12, Article 153. <https://doi.org/10.3389/fnbeh.2018.00153>
- Lipps, T. (1906). *Aesthetics: Psychology of beauty and art. Volume 2: The aesthetic view and the visual arts* [German Edition]. Hamburg und Leipzig: Verlag von Leopold Voss.
- López-Pérez, B., Carrera, P., Ambrona, T., & Ocejá, L. (2014). Testing the qualitative differences between empathy and personal distress: Measuring core affect and self-orientation. *The Social Science Journal*, 51, 676–680. <https://doi.org/10.1016/j.soscij.2014.08.001>
- Lyvers, M., Cotterell, S., & Thorberg, F. A. (2018). “Music is my drug”: Alexithymia, empathy, and emotional responding to music. *Psychology of Music*, 48, 626–641. <https://doi.org/10.1177/0305735618816166>
- McCroskey, J. C., & McCain, T. A. (1974). The measurement of interpersonal attraction. *Speech Monographs*, 41, 261–266. <https://doi.org/10.1080/03637757409375845>
- Merten, J. (2005). Culture, gender and the recognition of the basic emotions. *Psychologia*, 48, 306–316. <https://doi.org/10.2117/psych.2005.306>
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.).
- Poli, G. D., Canazza, S., Roda, A., & Schubert, E. (2014). The role of individual difference in judging expressiveness of computer-assisted music performances by experts. *ACM Transactions on Applied Perception*, 11, 1–20. <https://doi.org/10.1145/2668124>
- Rastas, A., & Seye, E. (2019). Music and anti-racism: Musicians' involvement in anti-racist spaces. *Popular Music and Society*, 42, 592–610. <https://doi.org/10.1080/03007766.2018.1526527>
- Sandstrom, G. M., & Russo, F. A. (2013). Absorption in music: Development of a scale to identify individuals with strong emotional responses to music. *Psychology of Music*, 41, 216–228. <https://doi.org/10.1177/0305735611422508>
- Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, 29, 347–363. <https://doi.org/10.1177/0734282911406661>
- Sittler, M. C., Cooper, A., & Montag, C. (2019). Is empathy involved in our emotional response to music? The role of the PRL gene, empathy, and arousal in response to happy and sad music. *Psychomusicology: Music, Mind, and Brain*, 29, 10–21. <https://doi.org/10.1037/pmu0000230>
- Varker, T., & Devilly, G. J. (2007). Types of empathy and adolescent sexual offenders. *Journal of Sexual Aggression*, 13, 139–149. <https://doi.org/10.1080/13552600701661573>

- Wallmark, Z., Deblieck, C., & Iacoboni, M. (2018). Neurophysiological effects of trait empathy in music listening. *Frontiers in Behavioral Neuroscience*, 12, Article 66. <https://doi.org/10.3389/fnbeh.2018.00066>
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.
- Wu, X., & Lu, X. (2021). Musical training in the development of empathy and prosocial behaviors. *Frontiers in Psychology*, 12, Article 661769. <http://doi: 10.3389/fpsyg.2021.661769>
- Yang, R. (2002). University internationalisation: Its meanings, rationales and implications. *Intercultural Education*, 13, 81–95. <https://doi.org/10.1080/14675980120112968>
- Zhao, Q., Neumann, D. L., Cao, Y., Baron-Cohen, S., Yan, C., Chan, R. C., & Shum, D. H. K. (2019). Culture–sex interaction and the self-report empathy in Australians and Mainland Chinese. *Frontiers in Psychology*, 10, Article 396. <https://doi.org/10.3389/fpsyg.2019.00396>
- Zhao, Q., Neumann, D. L., Cao, X., Baron-Cohen, S., Sun, X., Cao, Y., Yan, C., Wang, Y., Shao, L., & Shum, D. H. (2018). Validation of the empathy quotient in Mainland China. *Journal of Personality Assessment*, 100, 333–342. <https://doi.org/10.1080/00223891.2017.1324458>
- Zhao, Q., Neumann, D. L., Yan, C., Djekic, S., & Shum, D. H. K. (2021). Culture, sex, and group-bias in trait and state empathy. *Frontiers in Psychology*, 12, Article 760. <https://doi.org/10.3389/fpsyg.2021.561930>
- Zhao, Q., Ren, Q., Sun, Y., Wan, L., & Hu, L. (2020). Impact factors of empathy in Mainland Chinese youth. *Frontiers in Psychology*, 11, Article 688. <https://doi.org/10.3389/fpsyg.2020.00688>