

# Chinese Version of the Short Geneva Emotional Music Scale: Cross-cultural Adaptation and Psychological Measurement

Yanying Chen<sup>1,a</sup>, Tong Luo<sup>1,b</sup>, and Yuhui Wang<sup>1,c</sup>

<sup>1</sup> School of Architecture and Art, Central South University, China

<sup>a</sup> chenyaning0429@csu.edu.cn, <sup>b</sup> 2876242597@qq.com, <sup>c</sup> 346235912@qq.com

**Abstract.** In this study we developed the Chinese Geneva Emotion Music Scale (GEMS-C) and a simplified version (S-GEMS-C) verified by Chinese native speakers. As the GEMS is a one-word-item scale without context, we performed an expert revision process after forward-backward translation. The results were optimized through structured interviews to obtain the final Chinese GEMS (GEMS-C). We reduced the number of GEMS items to form a short version (S-GEMS-C). The GEMS-C and the S-GEMS-C were shown to have high reliability, construct validity, sensitivity, concurrent validity, and discriminant validity. Additionally, this study found a high correlation between the GEMS-C, S-GEMS-C, and OMES-M, and the difference between familiarity with the music and the degree of emotional arousal.

**Keywords:** Music emotion; Music scale; Cross-cultural adaptation; GEMS-C; S-GEMS-C

## 1. Introduction

The translation of scales has become an important research topic as cross-cultural research has increased. A scale is more than just a set of words. It contains nuanced social meanings, which make translation difficult [1]. When scales are translated into other languages, cultural differences can lead to biases in understanding. Cultural differences also affect the understanding and interpretation of the same terms. Therefore translators must consider whether the same meaning is expressed in the translated scale as in the original. As scales' application increases, simplifying scales is a necessary task [2].

There is a close relationship between cognition and emotion in music [3]. Some studies have demonstrated listeners' emotional responses to music using behavioral, physiological, and neurological measures [4-6]. Therefore many experts have studied emotions generated by music using measures such as the Absorption in Music Scale (AIMS) to assess differences between individual emotional responses to music and predict their intensity [7]; the affective value scale of music (AVSM) to gauge the quality and intensity of emotions [8]; and the Music Mood-Regulation Scale (MMRS) to assess the extent to which participants use music to alter emotional states [9].

The Geneva Emotional Music Scale (GEMS) was developed by Zentner et al. to test their domain-specific model [10]. The scale has nine sub-scales [11]. There are several versions of GEMS, with the 40-word version being more applicable to a broader range of musical genres; there are nine subscales containing nine different affective dimensions, using a 5-point Likert scale ranging from 1 (not at all) to 5 (very much) [10]. The original French GEMS was adapted into English by Zentner et al. for use in an international context [10]. Hedder simplified the GEMS-33 and developed a non-linguistic scale for measuring various musical styles [12]. All GEMS items are translated into Dutch [13]. In a German version, Lepa has adapted GEMS-25 as GEMS-28-G [20]. Polish scholars translated the GEMS into Polish (GEMS-pl) [14, 15] A 40-word version was used in this application.

The GEMS has become a widely used scale for measuring emotions brought by music. It has been modified and translated into several languages in different cultural backgrounds. However, the GEMS has yet to be a cross-cultural adaptation in Chinese. As a result, this research has a dual purpose. First, we used a forward-backward method to translate the GEMS and examined the psychometric characteristics of the GEMS-C. Second, we streamlined the items of the GEMS-C and examined the S-GEMS-C.

## **2. Study 1**

### **2.1 Translation Process**

Forward-backward translation is a widely accepted method. It has been used to translate a variety of Chinese scales.

### **2.2 Methodology Design**

DQF (Dynamic Quality Assessment Framework) [16] was used as the translation results in the evaluation method.

### **2.3 Forward Translation**

We selected five native Chinese translation experts, who had passed English Professional Grade 8 (the highest level of English proficiency test in China), obtained Level 2 translation qualifications, and had more than one year of translation experience. They claimed to have no music testing experience.

### **2.4 Backward Translation**

Two other translation experts translate Chinese to English and compare the differences with the original scale. Twenty percent of the words were significantly different from the original scale. The results were provided to UX and music experts for screening and adjustment.

### **2.5 Expert Processing Adjustments**

Two UX experts and two music professionals were selected to modify the translation results. Experts review and evaluate the preliminary translation results. If consensus can be reached on the item, the result is reserved. If the experts cannot agree, two alternative versions of the item will still be reserved for screening in subsequent interviews.

### **2.6 Structured Interview**

#### **2.6.1 Participants**

Twenty participants who were native Chinese speakers were selected, six of whom had a master's degree or higher and twelve of whom had a bachelor's degree. Their ages range from 22 to 50. They reported that they had not completed this type of test.

#### **2.6.2 Results**

During the interviews, we found four items that were not understood more frequently. According to the participants' feedback, the final version of GEMS is formed after modification. The final version of GEMS is shown in Table 1.

Table 1. The final version of Chinese GEMS

Dimensional words	Items				
奇妙 (Wonder)	快乐的 (happy)	充满好奇的 (filled with wonder)	迷人的 (allured)	绚烂的 (dazzled)	感动的 (moved)
超凡脱俗 (Transcendence)	鼓舞人心的 (inspired)	超凡的 (feeling of transcendence)	灵性的 (feeling of spirituality)	兴奋的 (thrills)	
柔情 (Tenderness)	恋爱感的 (in love)	感性的 (sensual)	深情的 (affectionate)	温柔的 (tender)	柔和的 (mellowed)
怀旧 (Nostalgia)	感伤的 (sentimental)	梦幻的 (dreamy)	思乡的 (nostalgic)	忧郁的 (melancholic)	
平和 (Peacefulness)	平静的 (calm)	放松的 (relaxed)	宁静的 (serene)	舒缓的 (soothed)	沉思的 (meditative)
力量 (Power)	精力充沛的 (energetic)	欢欣鼓舞的 (triumphant)	火热的 (fiery)	强大的 (strong)	不畏艰难的 (heroic)
快活 (Joyful activation)	亢奋的 (stimulated)	喜悦的 (joyful)	有活力的 (animated)	兴高采烈的 (feel like dancing)	被逗乐的 (amused)
紧张 (Tension)	不安的 (agitated)	焦虑的 (nervous)	紧张的 (tense)	不耐烦的 (impatient)	恼怒的 (irritated)
伤感 (Sadness)	难过的 (sad)	悲痛的 (sorrowful)			

## 2.7 Back Interview

We conducted a follow-up survey of 20 participants who participated in structured interviews. They were asked to reread the final version of the Chinese GEMS and were asked if they did not understand it after revision. All participants reported understanding all the items.

## 3. Study 2

### 3.1 Participants

Three hundred participants were selected from different professional backgrounds. They were recruited from the university website. Including undergraduates, postgraduates, and teachers, there were 146 males and 154 females (Mean=20.71; SD=3.151; Max=50; Min=18). A total of 772 questionnaires were collected.

### 3.2 Experimental Design

The experiment was within the subjects' experimental design. The independent variable was music, including three types. The dependent variable was the GEMS-C assessment results.

### 3.3 Materials

To avoid the emotional impact of lyrics [17, 18], the four experts selected seven pieces of music from the pure music library of 'NetEase,' based on the ranking of the world's classical piano pieces and composed context (Table 2).

Table 2. Music and the reason have been chosen by experts

Music	Background to the composition/expert interpretation	Composer	Instruments played	Performer	Duration	Release Time
Turkish March	According to the experts: the music is relaxed, lively, and rhythmic.	Beethoven	Piano	Mitsuko Uchida	3'34'	1778
Moonlight Sonata	Composed amid the grief of lost love and ear disease.	Achille-Claude Debussy	Piano	Emil Gilels	3'04'	1801
Dream Wedding	In the context of this music, the dream wedding is sad and romantic.	Paul DeSenneville & Olivier Toussaint	Piano	Richard Clayderman	2'40'	1979
Hedwig's Theme	This tune is a great interpretation of the prosperity and flourishing of a medieval magical academy.	John Williams	Piano	Christoph Eschenbach & Wiener Philharmoniker	2'25'	2006
He's a Pirate	The original is an impassioned and adventurous piece, and its performance of it in an original score adds to its sense of wonder.	Klaus Badelt & Hans Zimmer	Thumb piano	Abao Yang	2'20'	2003
Laputa Castle in the Sky	Experts found the overall music layered and long without being annoying.	Kazuhiro Morita	Piano	Dylan	3'00'	2017
Turangalila-Symphony: I-Introduction	Modern music is full of uncertainty, new trends, new sound materials, and new ways of writing.	Oliver Messiaen	Piano	Oliver Messiaen	3'22'	1993

Because there is no Chinese scale that has gone through cross-cultural adaptation, some of the items on the Online Music Engagement Scale [19] were modified because they did not match the testing environment, forming the adapted version of the Online Music Engagement Scale (OMES-M) (Table 3). In the test, the order of the two scales was balanced. At the same time, due to the long GEMS, the Latin square method (9\*9) [20, 21] was also used to balance the order of internal items. Because the labeling of the scale and the number of response items also affect the response bias [22, 23]. Just fill in according to true feelings. ' To study the influence of familiarity with music on emotion, we added the test of familiarity with music at the end of the questionnaire.

Table 3. An adapted version of the Online Music Engagement Scale (OMES-M)

No.	Items
T1	在听该音乐时,我感到自己迸发出能量 (I feel a burst of energy when I listen to this music)
T2	在听该音乐时,我感到自己充满活力 (I feel energized while listening to this music)
T3	我对该音乐富有热情 (I am passionate about this music)
T4	该音乐激发了我的灵感 (This music inspires me)
T5	时间合适,我会去听该音乐 (When the time is right, I will listen to that music)
T6	当听该音乐的时候,我会感到快乐 (I feel happy when I listen to this music)
T7	在听该音乐时,我会沉浸于其中 (When listening to this music, I get lost in it)
T8	我在听该音乐时会忘了自己 (I forget myself when I listen to that music)

### 3.4 Procedures

(1) Every 20 participants were divided into groups and seated in the laboratory according to their position number.

(2) The experimental assistant distributes the introduction and notes.

(3) The experimental assistant played the music. Every time a new piece of music played, a QR code on a projector in the experimental classroom. Each participant completed one to seven tests with different music for each test.

### 3.5 Results

The subjects' completion times were used to clean the data [24]. Five experts with experience were selected in user experience metrics. They completed a testing process, and the completion time ( $M=112$ ,  $SD=9.1$ ). The completion time of 94 seconds (Mean-2SD) was chosen as the cleaning threshold. Therefore, 111 questionnaires were deleted. We also conducted a K-S test on all question items and found that the p-values for all items were less than 0.05, so the data were considered non-normally distributed.

#### 3.5.1 Reliability

The reliability of the Chinese GEMS was 0.878. The reliability of the nine subscales was 0.798, 0.819, 0.895, 0.726, 0.879, 0.911, 0.940, 0.907 and 0.915. The reliability of OMES-M was 0.895. Some dimensions have lower confidence than other language versions [14], but does exceed the lower threshold of 0.7 [25].

#### 3.5.2 Concurrent validity

Concurrent validity is expressed by the correlation between dual questionnaires with a minimum acceptance criterion of 0.30 [25]. The correlation between Chinese GEMS and OMES-M was examined to obtain the concurrent validity of the scale. The correlation between the dimensions of the two scales was calculated respectively.

We performed exploratory factor analysis (EFA) on OMES-M divided into two factors. T1, T2, T3, and T6 were within the first factor. T4, T5, T7, and T8 were within the second factor. We calculated correlations between the three higher-order factors (sublimity, vitality, and unease) in GEMS-C and the two OMES-M factors. We found that the factors of OMES-M and GEMS-C were significant, with factor 1 of OMES-M showing the highest correlation with sublimity in GEMS-C. Factor 2 of OMES-M showed the highest correlation with vitality in GEMS-C.

#### 3.5.3 Distinct validity

Discriminant validity refers to the fact that the observed values should be distinguishable from each other when different methods are applied to measure different constructions. The mean of each dimension was used to calculate the discriminant validity between the scales. There was a significant difference in all items except the 'sadness,' indicating that the scale items were designed to be highly differentiated.

#### 3.5.4 Construct validity

'Construct validity' refers here to the degree of correspondence between a measured value and a structure reflected in measurements, it is usually assessed via factor analysis [26, 27]. The KMO of the GEMS-C was 0.955, and OMES-M was 0.852, both exceeding the minimum acceptable value of 0.6 [28]. Bartlett's test of sphericity for GEMS-C was  $X^2=25693.732$ ,  $df=780$ ,  $p=0.000$ . Thus, the construct validity of the two scales is adequate and suitable for factor analysis.

#### 3.5.5 Sensitivity

'Sensitivity' refers to the appropriate response of the questionnaire to operational changes, which reflects differences in the frequency or usage of the system [26, 27]. The scores of Chinese GEMS have significant differences among music from different compositional contexts. In the Chinese version of the GEMS-9, we found that all nine questions were significant, with Cohen's  $f$  of 0.032, 0.014, 0.034, 0.024, 0.034, 0.033, 0.053, 0.011, 0.052. They are both less than 0.1, which is a small effect size.

### 4. Study 3

#### 4.1 Inter Subscale Correlation

Before conducting the validation factor analysis, the correlations between the nine GEMS-C dimensions were examined. In the correlations of the nine dimensions means, all dimensions except Wonder and Nostalgia, Transcendence and Peacefulness, and Power and Tension dimensions correlated on a scale of 0.067 to 0.865 with -0.077 to -0.559. Among the correlations for each question item in the GEMS-9, all dimensions except Wonder and Tenderness, Nostalgia, and Peacefulness dimensions; Transcendence and Sadness; Nostalgia and Tension; and Joyful Activation and Tension were correlated between 0.617 to 0.066 and -0.071 to -0.56. The correlation between the mean values of the dimensions in GEMS-9 and GEMS-C were all correlated at 0.612 to 0.906.

#### 4.2 Exploratory Factor Analysis

Exploratory factor analysis aims to examine whether the items under each dimension can effectively explain its dimension. It can exclude irrelevant and redundant questions. In the extraction method, principal component analysis was used for extraction. Furthermore, we used the most commonly used maximum variance method on the online version of SPSS for the calculation. After exploratory factor analysis, there is no item of Commonality below 0.4 [29-31]. Scale simplification will be based on confirmatory factor analysis.

#### 4.3 Confirmatory Factor Analysis

A confirmatory factor analysis (CFA) was conducted on the 40 items, and the results are shown in Fig 1. Since the maximum likelihood (ML) method has no restrictions on the independent variables [32], we chose this method for calculation in the online version of SPSS. The multivariate kurtosis and critical ratios were 216.152 and 47.936, and all but three items, amused, homesick, and feeling in love, were non-normal variables. The whole material is Non-normally distributed data.

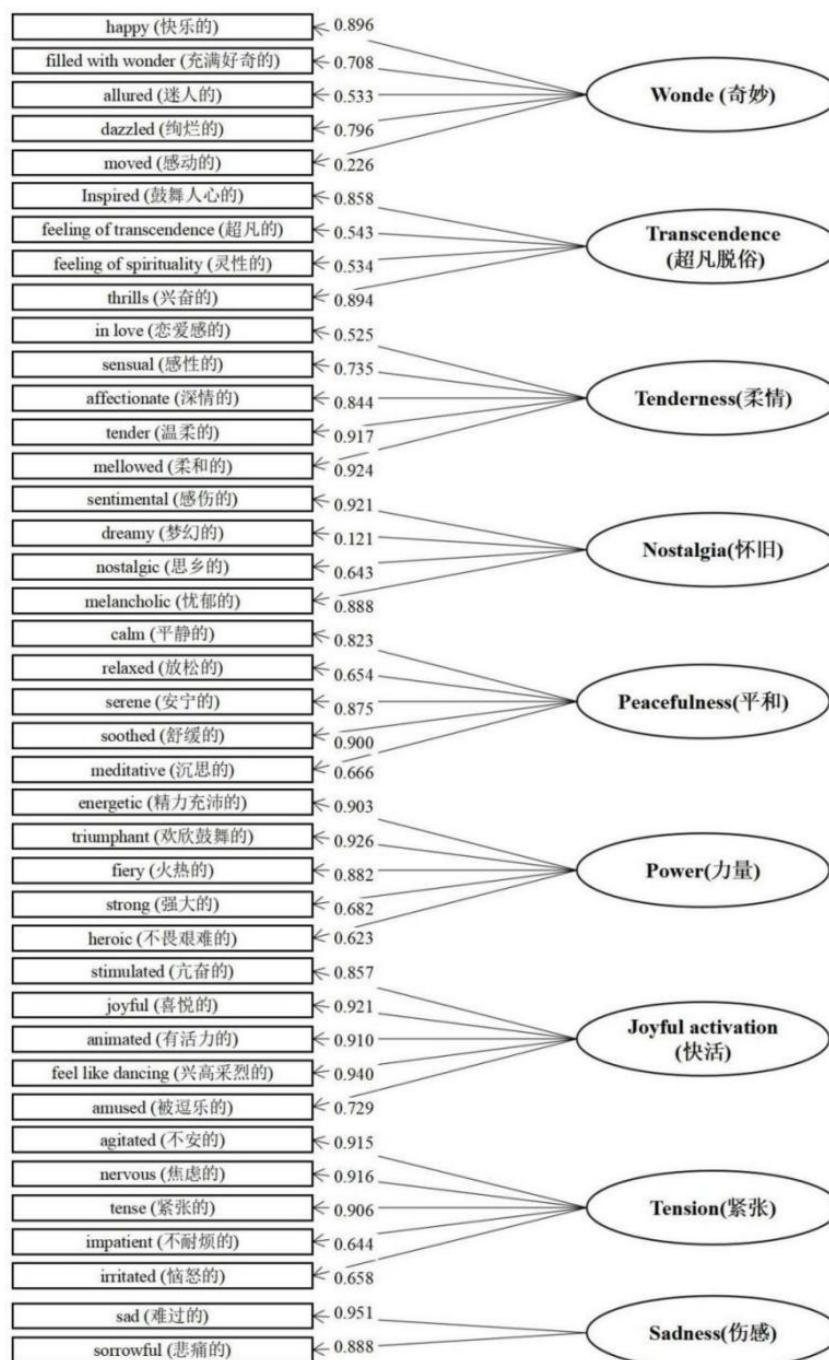


Fig 1. GEMS-C structure coefficients for each dimension (raw)

The initial model was of an inadequate fit:  $\chi^2=5025.805$ ,  $df=704$ ,  $p=0$ ,  $\chi^2/df=7.139$ ,  $GFI=0.809$ ,  $RMSEA=0.096$ ,  $RMR=0.065$ ,  $CFI=0.831$ ,  $NFI=0.809$ ,  $NNFI=0.812$ . As the whole material was non-multivariate normal data, we ran Bootstrapping tests and calculated Bollen-Stine Bootstrap, Number of bootstrap samples=500,  $p=0.002$  ( $p<0.05$ ) and therefore accepted the original hypothesis. The GFI and AGFI should be greater than 0.9, and the RMR should be less than 0.05. Consequently, the model is poorly constructed, and correction is needed.

A standard loadings coefficient value greater than 0.70 would indicate a strong correlation between the item and the factor. However, standard loadings coefficient values are below 0.70 for each dimension, which is unsatisfactory. Removing these items was attempted to improve this initial model while maintaining the scale's reliability (Cronbach's  $\alpha > 0.8$ ). Therefore, the item of '感动的(moved)', '灵性的(Inspired)', '恋爱感的(in love)', '梦幻的(dreamy)', '沉思的(meditative)', '不畏艰难的(heroic)', '被逗乐的(amused)' and '不耐烦的(impatient)' were deleted.

The first revised model fit was  $\chi^2=2679.603$ ,  $df=428$ ,  $p=0$ ,  $\chi^2/df=6.261$ ,  $GFI=0.879$ ,  $RMSEA=0.089$ ,  $RMR=0.196$ ,  $CFI=0.896$ ,  $NFI=0.879$ ,  $NNFI=0.879$ . The model fit improved somewhat compared to the pre-correction period. As the whole material was non-multivariate normal data, we ran Bootstrapping tests and calculated Bollen-Stine Bootstrap, Number of bootstrap samples=500,  $p<0.05$ , and therefore accepted the original hypothesis.

The indicators of GFI, NFI, and CFI improved, and the RMR and RMSEA decreased. All these indexes are close to the standard of good fit. The model construction can be revised again. After deleting some items, the CFA was conducted once again.

The modified model fit improved (see Fig 2),  $\chi^2=1404.114$ ,  $df=263$ ,  $p=0$ ,  $\chi^2/df=5.339$ ,  $GFI=0.925$ ,  $RMSEA=0.081$ ,  $RMR=0.125$ ,  $CFI=0.938$ ,  $NFI=0.925$ ,  $NNFI=0.924$ . We ran the Bootstrapping and calculated the Bollen-Stine Bootstrap, Number of bootstrap samples=500,  $p<0.05$ . Therefore, the original assumptions are accepted. GFI, NFI, RMR, RMSEA, and NNFI reached common standards.

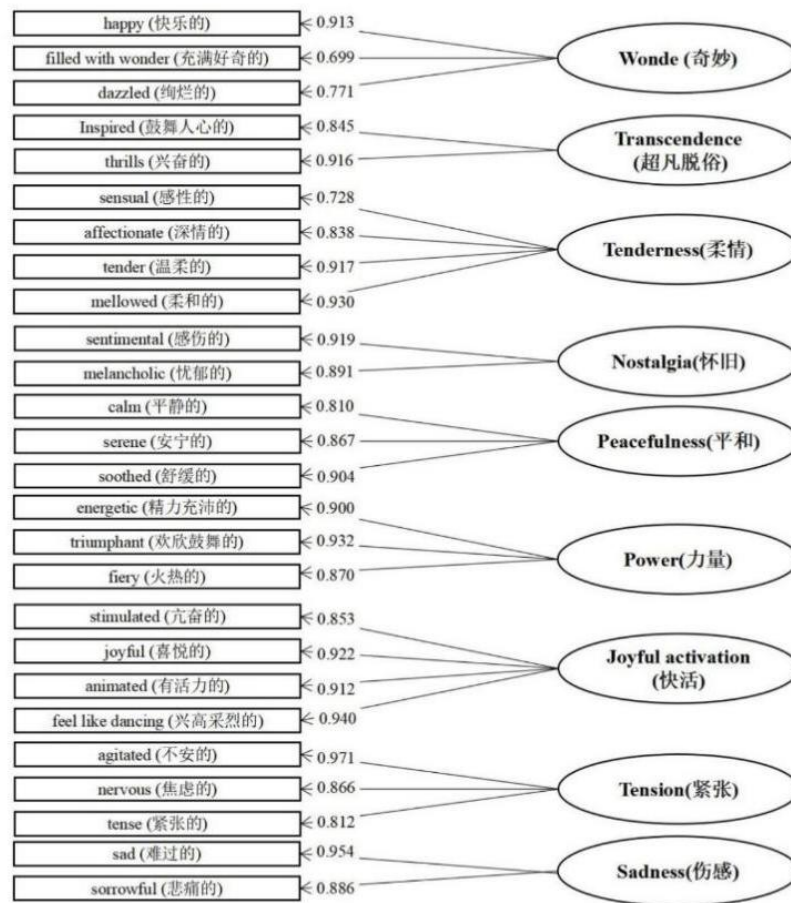


Fig 2. S-GEMS-C structure coefficients for each dimension

## 4.4 Results

### 4.4.1 Reliability

The overall reliability of the S-GEMS-C was 0.787, with the Cronbach alpha coefficients for all nine dimensions above 0.8. The reliability of the nine subscales was 0.840, 0.872, 0.916, 0.900, 0.895, 0.928, 0.948, 0.938 and 0.915. Although the reliability of some dimensions is still lower than that of the other language versions, the reliability of each dimension has improved significantly compared to the GEMS-C and is higher than the generally used standard [25].



#### 4.4.2 Construct validity

The KMO of S-GEMS-C was 0.949, which was lower than that of GEMS-C(0.955) but significantly higher than the minimum acceptable value of 0.6 [28]. In Bartlett's test of sphericity,  $\chi^2 = 18708.287$ ,  $df = 325$ ,  $p = 0.000$ . SRMR=0.059<0.08. It has high structural validity.

#### 4.4.3 Sensitivity

There are significant differences in S-GEMS-C scores for music compositional contexts. Using Kruskal-Wallis, we found that all were significant at the mean of the nine-dimensional scores, with Cohen's  $f$  of 0.047, 0.051, 0.034, 0.052, 0.034, 0.054, 0.055, 0.02, and 0.051, respectively. All were less than 0.1, with small effect sizes. Also, the correlations between the three higher-order factors in S-GEMS-C and the two OMES-M factors we found were similar to that of Study 2.

The sensitivity of different musical backgrounds was also explored. The results are shown in Table 4.

Table 4. Kruskal-Wallis test analysis results on the relationship between musical familiarity and emotional arousal

Dimension	Category	AVG	p	Cohen's $f$
Wonder	A	2.485	0.000	0.019
	B	3.062		
	C	3.278		
Transcendence	A	2.374	0.000	0.02
	B	2.894		
	C	3.199		
Tenderness	A	2.789	0.895	0.001
	B	2.775		
	C	2.738		
Nostalgia	A	2.681	0.000	0.013
	B	2.290		
	C	2.147		
Peacefulness	A	2.519	0.671	0.002
	B	2.458		
	C	2.428		
Power	A	2.343	0.000	0.018
	B	2.870		
	C	3.093		
Joyful Activation	A	2.263	0.000	0.018
	B	2.934		
	C	3.118		
Tension	A	2.562	0.000	0.016
	B	1.915		
	C	1.935		
Sadness	A	2.500	0.000	0.015
	B	1.988		
	C	1.838		

Note. Categories are divided based on participants' familiarity with the music. A means not heard of the music, B means have heard the music but does not understand the music background, and C means have heard the music and be familiar with it. The bolded ones are the options with the highest average. AVG= Average.

We calculated the discriminant validity between the S-GEMS-C and OMES-M. All questions were significant except for the Nostalgia, Tension, and Sadness dimensions. Except for nostalgia

and sadness, all the dimensions have significant differences. This indicates that the scale design has high discriminant validity.

## **5. Discussion**

### **5.1 Cross-cultural Adaptation Process**

In fact, after the forward and backward translation, it was discovered that many words had different translation results. Compared with the translation results of other scales, our results generated more than 50% divergence. This may be because (a) the translators do not have expertise in using this scale, resulting in many comprehension disagreements. (b) The items in the scale are inspirational words. This complicates the translation process of how emotions are perceived, interpreted, and reacted [33]. (c) in the backward translation process, we found that the translation results differed from the original. Heij et al. point out that conceptual activation of first-language words is more accessible than that of a second language [34]. The context influences translation results in the backward translation process. Therefore, there will be some bias in the backward translation process. Context plays a vital role in comprehending and translating texts [35]. GEMS lacks context, so the translation results are more ambiguous than other scales composed of sentences. Due to the lack of context in word scale, translation experts may need to adjust or add a context reduction process. Therefore, it is recommended that the original translation process of the word scale be modified.

It is vital to conduct cognitive debriefing tests for the translation results. Researchers have used structured interviews to gain insight into participants' cognitive ambiguity regarding the initial translation results [36]. Because of the influence of their musical experience, this could lead to erroneous conclusions. Therefore, cognitive debriefing tests should be conducted in translation to avoid this bias.

### **5.2 Music Selection**

In a study by Rompay et al. [37], It was found that the study of 'feeling of transcendence' should be removed because the meaning of the word was unclear. Likewise, 'Wonder' is also not evident in our study. It is speculated that 'Transcendence' and 'wonder' are not commonly used in music evaluation. This is difficult for the participants who are not experienced in listening to music to complete a clear judgment.

In S-GEMS-C, it was found that the reliability of 'Wonder' and 'Transcendence' reached 0.840 and 0.803, respectively, in the music group. The comprehensive questionnaire's reliability (Cronbach  $\alpha=0.821$ ) and construct validity (KMO=0.949, SRMR= 0.059) met the general use requirements.

### **5.3 Use The Short Version Of The Scale**

Measuring nine-dimensional words as items was attempted, and the Chinese version of the GEMS-9 (9-GEMS-C) was measured with a reliability of 0.524, which does not meet the standard for general use [25]. The construct validity was 0.762. However, the S-GEMS-C reliability and structure validity of 26 item versions were significantly improved. Although the reliability of the short version of the scale decreased compared with the version after cross-cultural adaptation, it was still greater than 0.8. The short scale is more popular than the long one [38]. GEMS is a word-items scale, but it has 40 items. S-GEMS-C has only 26 items, and each dimension's reliability is up to 0.840, which can still meet the use requirements.

### **5.4 Differences In The Emotions Generated By Different Music Experiences**

The S-GEMS-C was used to explore the difference between levels of emotional arousal for different musical experiences. Emotional involvement (i.e., the feeling of emotion, as distinct from

the recognition of emotion) can be moderated by familiarity [39]. A main effect of experience on all dimensions was found except for the Tenderness and Peacefulness dimension (Table 4). The reason for this may be that the emotion of equanimity is a low-arousal emotion [40].

The perceived level of the four dimensions of wonder, transcendence, power, and joyful activation rose according to the participants' familiarity with the music. For those who had heard the music before the experiment or knew about the background of the music's creation had higher arousal levels. This is in line with the findings of [41]. This may be because these dimensions are positive emotions, and familiarity with the music may increase expectancy and predictability. At the same time, different levels of familiarity, expectancy, and predictability may play an essential role in emotional arousal and pleasurable experience [42].

Similarly, there was a main effect of familiarity in three dimensions: nostalgia, tension, and sadness. However, these scores decreased as music familiarity increased (Table 4). The likely reason is that participants with higher experience developed mechanisms to avoid such negative emotions. They avoid these negative emotions by avoiding listening intently. However, participants unfamiliar with the song listened more intently to successfully label the emotion in the test [43]. Thus the degree of negative emotion evoked would be increased. Whereas both positive and negative emotions are the focus of emotional expression in music, many sad themes even permeate many modern pop songs [44], and participants with more musical experience may be familiar with them.

## 6. Limitations and future work

Since our subject group was recruited on campus, the degree of education may also contribute to the sincerity of the experiment. However, the use of the scale is not limited to young groups, even though their musical experience and emotions may be rich and sensitive. Therefore, further validation among more Chinese users is still needed.

Due to the large number of items in the GEMS, the number of our subjects needs to be increased, and we will need to obtain more validation in future studies.

Finding one piece of music that could evoke nine dimensions of emotion simultaneously was challenging, so we selected multiple pieces of music as material. This may have affected the normally distributed state of our data. We did not delete or add dimensions to maintain a structure similar to the original GEMS. In future research, we can optimize the structure of the GEMS.

There is no standardized music rating scale available in the Chinese environment. Therefore, we temporarily adopted OMES-M to report. In the follow-up study, there is hope for more cross-cultural adaptation of this scale and the corresponding relationship of S-GEMS-C.

## Acknowledgements

This work was supported by the Humanities and Social Sciences Research Project of the Ministry of Education (Grant Number: 20YJCZH012). Innovative Province Construction Special Science Project (Grant Number: 2023ZK4008)

## References

- [1] Mills C Wright. Language, logic, and culture. *American Sociological Review*, 1939, 4(5): 670-80.
- [2] Nagy Mark S. Using a single - item approach to measure facet job satisfaction. *Journal of occupational and organizational psychology*, 2002, 75(1): 77-86.
- [3] Krumhansl Carol L. Music: A link between cognition and emotion. *Current directions in psychological science*, 2002, 11(2): 45-50.
- [4] Krumhansl Carol L. An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 1997, 51(4): 336.

- [5] Mitterschiffthaler Martina T, Fu Cynthia Hy, Dalton Jeffrey A, et al. A functional MRI study of happy and sad affective states induced by classical music. *Human brain mapping*, 2007, 28(11): 1150-62.
- [6] Witvliet Charlotte Vo, Vrana Scott R. Play it again Sam: Repeated exposure to emotionally evocative music polarises liking and smiling responses, and influences other affective reports, facial EMG, and heart rate. *Cognition and Emotion*, 2007, 21(1): 3-25.
- [7] Sandstrom G. M., Russo F. A. Absorption in music: Development of a scale to identify individuals with strong emotional responses to music. *Psychology of Music*, 2013, 41(2): 216-28.
- [8] Taniguchi T. Construction of an affective value scale of music and examination of relations between the scale and a multiple mood scale. *Japanese Journal of Psychology*, 1995, 65(6): 463-70.
- [9] Hewston R. M., Lane A. M., Karag C. I. Development and initial validation of the Music Mood-Regulation Scale. *E-Journal of Applied Psychology*, 2008, 4(1).
- [10] Zentner M., Grandjean D., Scherer K. R. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, 2008, 8(4): 494-521.
- [11] Irrgang M. Integrating Embodied Music Cognition into Music Recommendation Services. 2015.
- [12] Hedder Maïke K. Chartistic-A new non-verbal measurement tool towards the emotional experience of music; University of Twente, 2010.
- [13] Vuoskoski J. K., Eerola T. Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, 2011, 15(2): 159-73.
- [14] Chelkowska-Zacharewicz M., Janowski M. Polska adaptacja Geneva Emotional Music Scale (GEMS)–badania wstępne. *Polskie Pismo Muzykoterapeutyczne*, 2016, 3: 44-63.
- [15] Chelkowska-Zacharewicz M., Janowski M. Polish adaptation of the Geneva Emotional Music Scale: Factor structure and reliability. *Psychology of Music*, 2021, 49(5): 1117-31.
- [16] Junsong Wang. New Directions in Translation Quality Assessment: DQF Dynamic Quality Assessment Framework. *Chinese Scientific Translation*, 2019, 32(3): 27-9.
- [17] Barradas G. T, Sakka L. S. When words matter: A cross-cultural perspective on lyrics and their relationship to musical emotions. *Psychology of Music*, 2021, 50(2): 650-69.
- [18] Mori Kazuma, Iwanaga Makoto. Pleasure generated by sadness: Effect of sad lyrics on the emotions induced by happy music. *Psychology of Music*, 2014, 42(5): 643-52.
- [19] Ziqing Ye . Development and examination of the online music perceived value scale for college students. *Journal of Fujian Medical University: Social Science Edition*, 2020, 21(2): 5.
- [20] Kazdin Alan E. Acceptability of child treatment techniques: The influence of treatment efficacy and adverse side effects. *Behavior Therapy*, 1981, 12(4): 493-506.
- [21] Richardson John Te. The use of Latin-square designs in educational and psychological research. *Educational Research Review*, 2018, 24: 84-97.
- [22] Xinin Chen, Hongyan Yu. Number of best response items on rating scales:evidence from eye movement studies. *Psychological and Behavioral Research*, 2014, 12(4): 551-7.
- [23] Greenleaf Eric A. Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, 1992, 29(2): 176-88.
- [24] Yuhui Wang, Tian Lei, Xinxiong Liu. Reliability of perceived usability assessment via crowdsourcing platform: retrospective analysis and novel feedback quality inspection method. *International Journal of Human–Computer Interaction*, 2020, 36(11): 1070-84.
- [25] Nunnally J. C. Psychometric theory. *American Educational Research Journal*, 1978, 5(3): 83.
- [26] Alghannam Bareeq A, Albustan Sana A, Al-Hassan Abeer A, Albustan Lamees A. Towards a standard arabic system usability scale: Psychometric evaluation using communication disorder app. *International Journal of Human–Computer Interaction*, 2018, 34(9): 799-804.
- [27] Blažica Bojan, Lewis James R. A slovene translation of the system usability scale: The sus-si. *International Journal of Human-Computer Interaction*, 2015, 31(2): 112-7.
- [28] Kaiser Henry F. An index of factorial simplicity. *psychometrika*, 1974, 39(1): 31-6.
- [29] Kim Eunhyang, Koh Eunyoung. Avoidant attachment and smartphone addiction in college students: The mediating effects of anxiety and self-esteem. *Computers in Human Behavior*, 2018, 84: 264-71.

- [30] Lee Wangoo, Jeong Chul. Effects of pro-environmental destination image and leisure sports mania on motivation and pro-environmental behavior of visitors to Korea's national parks. *Journal of destination marketing & management*, 2018, 10: 25-35.
- [31] Pappas Nikolaos. COVID19: Holiday intentions during a pandemic. *Tourism Management*, 2021, 84: 104287.
- [32] Kleinbaum David G, Klein Mitchel. Maximum likelihood techniques: An overview. *Logistic regression*, 2010: 103-27.
- [33] Porter Richard E, Samovar Larry A. Cultural influences on emotional expression: implications for intercultural communication. *Handbook of communication and emotion*. Elsevier. 1996: 451-72.
- [34] Heij Wido, La, Andre, Hooglander, Robert, Kerling, Velden Esthervan, Der. Nonverbal Context Effects in Forward and Backward Word Translation: Evidence for Concept Mediation. *Journal of Memory & Language*, 1996, 35(5): 648-65.
- [35] Xianchun Zhu. No Context, No Text --- The importance of Context in Translation. *Teaching English as a Foreign Language in China and the United States: English Version*, 2006, 3(9): 3.
- [36] Arredondo E. M, Mendelson T., Holub C., et al. Cultural adaptation of physical activity self-report instruments. *Journal of Physical Activity and Health*, 2012, 9(s1): S37-S43.
- [37] Rompay Tjl, Verhoeven Jwm, Hout M. Chartistic - A new non-verbal measurement tool towards the emotional experience of music. 2010.
- [38] Eriksson Mårten, Boman Eva. Short is beautiful: dimensionality and measurement invariance in two length of the Basic Psychological need Satisfaction at work scale. *Frontiers in psychology*, 2018, 9: 965.
- [39] Pereira Carlos Silva, Teixeira João, Figueiredo Patrícia, et al. Music and emotions in the brain: familiarity matters. *PloS one*, 2011, 6(11): e27241.
- [40] Lim Nangyeon. Cultural differences in emotion: differences in emotional arousal level between the East and the West. *Integrative medicine research*, 2016, 5(2): 105-9.
- [41] Loui Psyche, Bachorik Justin, Li Hui C, Schlaug Gottfried. Effects of voice on emotional arousal. *Frontiers in psychology*, 2013, 4: 675.
- [42] Bosch Iris Van Den, Salimpoor Valorie, Zatorre Robert J. Familiarity mediates the relationship between emotional arousal and pleasure during music listening. *Frontiers in human neuroscience*, 2013, 7: 534.
- [43] Thammasan Nattapong, Moriyama Koichi, Fukui Ken-Ichi, Numao Masayuki. Familiarity effects in EEG-based emotion recognition. *Brain informatics*, 2017, 4(1): 39-50.
- [44] Schellenberg E Glenn, Von Scheve Christian. Emotional cues in American popular music: Five decades of the Top 40. *Psychology of Aesthetics, Creativity, and the Arts*, 2012, 6(3): 196.