The Chinese Version of the Gold-MSI: Adaptation and Validation of an Inventory for the

Measurement of Musical Sophistication in a Taiwanese Sample

Although the question of what musicality is or who is regarded as being musical has

fascinated authors since the beginning of the 19th century (Michaelis, 1805), there is still no

consensus as to what constitutes "musicality" or "excellent musical skills" (Gembris, 1997;

Hallam, 2016). As early as the end of the 19th century, the Austrian surgeon Theodor Billroth

(1829–1894), who was working on the book *Who is Musical?* (*Wer ist musikalisch?*), pointed

out the difficulty of defining the term musicality in a letter to the Austrian music critic Eduard

Hanslick (1825–1904):

> How complex this term is! One has predominantly rhythmic talent and sensation
>
> (…); the other has mainly melodic talent (…), yet another presents himself
>
> musically through an outstanding technical and mechanical talent (…); another
>
> performs musically through the transmission of his intense character into dramatic
>
> expression (…), yet another through enormous memory of tones and rhythms, yet
>
> another through the dedication to the sensual effects of hearing, etc. In me
>
> everything is chaos! (Billroth, 1895, p. 3, translated by the first author)

The "chaos," as Theodor Billroth called it, reveals the high complexity of the concept

musicality, which is also reflected in the development of tests for musical ability or musical

potential that measure different components of musical ability using different subtests. Since

the development of the first standardised test of musical talents (Seashore, 1919; Seashore,

Lewis, & Saetveit, 1960), various test batteries have measured musical talent, aptitude, or

giftedness, including the music aptitude tests and measures of music audiation by Gordon

(1965, 1986, 1989), the Standardised Tests of Musical Intelligence by Wing (1939, 1961), and

the Measures of Musical Ability by Bentley (1966, 1969). These tests have been widely used

in the field of music education to identify the degree of musical aptitude of a student so that

1

the teacher can adapt the teaching content to serve the need of every child in accordance with his or her degree of musical aptitude (Gordon, 1986).

In recent years, researchers have continued to develop various standardised test batteries for the measurement of musical abilities and skills in adults, such as the Swedish Musical Discrimination Test (Ullén, Mosing, Holm, Eriksson, & Madison, 2014), the Karma Music Test (Karma, 2007), and the Profile of Music Perception Skills (Law & Zentner, 2012).
In the above-mentioned tests, musical ability is considered the "potential for learning music" that exists before any kind of formal training occurred and is operationalised further as discrimination ability for auditory stimuli with subtle differences in one specific musical parameter, such as melody, rhythm, pitch, timbre or loudness. Other recently-developed tests focus on specific types of musical achievement, such as the Musical Ear Training Assessment (META) by Wolf and Kopiez (2018) and the Notation-Evoked Sound Imagery test (NESI) by Wolf, Kopiez, and Platz (2018). In contrast with the tests following the tradition of musical aptitude tests since Seashore (1919), the latter two test batteries are intended to measure abilities or skills that could be acquired in formal training (e.g., in music theory or ear training courses) at university level. The META measures a person's ear training and analytical hearing skills whereas the NESI focuses on the listener's ability to manipulate sound imagery by asking test takers to detect the match or mismatch between a musical theme and its variation after reading the music score of the variation and listening to the sound of the theme. Furthermore, a broader definition of the concept "musical ability" has been discussed in the literature (Hallam, 2010; Hallam & Prince, 2003; Levitin, 2012), emphasizing the multidimensional nature of musicality. In line with this conceptual change, a new instrument measuring musicality that comprises subjective self-assessment as well as objective musical listening abilities, the Goldsmiths Musical Sophistication Index (Gold-MSI), was developed (Müllensiefen, Gingras, Musil, & Stewart, 2014). Compared to the traditional term musical

ability, the term "musical sophistication"' is much broader in scope and includes formal as well as informal musical learning processes (see the next section for a definition of the construct of musical sophistication). Arguably, psychometric testing of musical ability has flourished in recent years and vital developments continue to be made.

Nevertheless, these recent developments regarding the assessment of musical ability have barely influenced the Chinese-speaking research community despite the growing interest in research on musical development and music psychology (Wen & Tsai, 2017; Yang, Ma, Gong, Hu, & Yao, 2014). To identify available test batteries for the measurement of musical ability, we searched for research literature on the two most comprehensive databases of academic Chinese publications, namely, the National Digital Library of Theses and Dissertations in Taiwan (NDLTD-Taiwan http://ndltd.ncl.edu.tw) and the CEPS in the Airiti Library (http://www.airitilibrary.com), using the keywords *musical ability*, *musical aptitude*, *musical talent*, *musical performance, musical achievement* and *music test*. The NDLTD-Taiwan is the only database that has consistently collected theses and dissertations of colleges and universities in Taiwan from 1956 to the present whereas the CEPS (the Chinese Electronic Journals Database) collects academic journals published in Taiwan and China as well as journals in Chinese or English that have been published in the U.S.A., Hong Kong, Malaysia and other countries since 1991.

As a result of this database search, we found that in the Chinese research literature, the most widely used test batteries are those developed by Edwin Gordon (Gordon, 1986), namely, the *Primary Measures of Music Audiation* (PMMA) or the *Intermediate Measures of Music Audiation* (IMMA), which are designed for the use with different age groups. However, the validity of Gordon's measures of music audiation has been questioned in recent studies (Gembris, 1997; Mota, 1997; Oreck, Owen, & Baum, 2003; Platz et al., 2015), and their

correlations with other criterion variables tend to be weak ($r < .20$, Rutkowski, 1996; r = ,45–.53, Hanson, 2019). In addition, several aptitude tests have been developed for the identification of musically gifted and talented children or for supporting pupils to make decisions regarding their career development (Chen, Lin, Hsu, & Yen, 2007; Chen, 2003; Lu, Chen, & Chen, 1981). Such tests have only been applied in the selection processes of school programs for musically gifted students and have not been used within the Chinese academic community. Additionally, a few standardised music achievement tests have been published for use in primary schools (Chao, 2013; Lin, 2013) or secondary schools (Ho, 2004). We assume that Chinese researchers have only limited access to suitable test batteries for musical ability and thus have had to develop their own measures for use in specific studies, which makes it difficult to compare outcomes across different studies, for example, in meta-analyses.

Besides objective test instruments, various self-report inventories inspired by the theory of multiple intelligences (Gardner, 1985) have been created, including the Chinese version of the Multiple Intelligences Developmental Assessment Scales (Wu, 2008), the Inventory of Preschoolers' Multiple Intelligences (Yeh, Peng, & Yeh, 2009), and the Eight Multiple Intelligences Questionnaire (Chou, 2006). As these inventories were designed according to the theory of multiple intelligences, musical ability was regarded as a unidimensional factor. Thus, the content of such inventories might not sufficiently capture the multidimensional nature of musical skills and abilities. In addition, correlations between the above-mentioned inventories and other objective measures have not been examined. Therefore, it seemed necessary to introduce a standardised instrument for the measurement of musicality to the Taiwanese research community as well as for wider use with Mandarin speaking populations.

The Gold-MSI is an ideal candidate for this purpose as it comprises subjective self-assessment and objective measures and is already available in English (Müllensiefen et al., 2014), German (Schaal, Bauer, & Müllensiefen, 2014), French (Degrave & Dedonder, 2018), Portuguese (Lima, Correia, Müllensiefen, & Castro, 2018), and Danish (see http://www.gold.ac.uk/music-mind-brain/gold-msi/). Although it was originally developed for adults, research has shown that the factor structure of the Gold-MSI self-report inventory could be replicated in a sample of German school students aged 10 to 17 (Fiedler & Müllensiefen, 2015). Since its publication in 2014, the Gold-MSI has gained popularity in the field of music psychology and music education. For example, researchers used the Gold-MSI inventory to classify participants as musically trained or untrained (Bauer, Kreutz, & Herrmann, 2015; Farrugia, Jakubowski, Rhordi, & Stewart, 2015), to identify the possible influence of musical sophistication on research results (Kopiez, Wolf, Platz, & Mons, 2016; Wolf, Kopiez, Platz, Lin & Mütze, 2018), and to keep track of the musical development trajectories of students in a school setting (Fiedler & Müllensiefen, 2017). Due to the applicability of the Gold-MSI inventory in various research settings, we are convinced that the adaptation of the Gold-MSI into Traditional Chinese for the Taiwanese population will make an essential contribution to the research literature and future cross-cultural research.

**Brief introduction to the Goldsmiths Musical Sophistication Index**

The Goldsmiths Musical Sophistication Index (Gold-MSI) is an instrument that includes one self-report inventory and several musical listening tests, the results of which, taken together, measure musical sophistication in a comprehensive way by considering a wide range of facets of musical expertise in the general population (Müllensiefen et al., 2014). Unlike other tests of musicality, musical talent or abilities intended to assess innate aptitude or musical potential that is influenced only minimally by musical training, the Gold-MSI measures musical sophistication, which can refer to musical ability, skills, expertise, achievements, attitudes

toward music, and music-related behaviours. According to this conceptualisation, the distinction between what is innate and what is learned does not play a role. In other words, this definition makes no assumptions as to how musical sophistication is acquired and whether it stems from natural talent, genetic predispositions, formal or informal learning processes or any combination of these factors. Instead, it is assumed that active engagement with music (consisting of instrumental as well as non-instrumental activities such as concert attendance) can facilitate the development of musical sophistication. According to Müllensiefen et al. (2014), high levels of musical sophistication are characterised by

> a) higher frequencies of exerting musical skills or behaviours, b) greater ease, accuracy or effect of musical behaviours when executed, and c) a greater and more varied repertoire of musical behaviour patterns. (Müllensiefen et al., 2014, p. 2)

In this sense, musically sophisticated individuals are able to deal with a greater range of musical situations, respond to such situations with more flexible approaches, and possess effective means of achieving their goals when engaging with music. In other words, playing an instrument and having formal music lessons are not the decisive criteria.

There are several special features that distinguish the Gold-MSI from existing instruments measuring musical abilities or music-related behaviours. First, the definition of musical sophistication by Müllensiefen et al. (2014) stresses the multidimensional nature of musical ability. Based on previous studies by Hallam and Prince (2003) as well as Ollen (2006), the items of the Gold-MSI self-report inventory take a wide range of behaviours into consideration, including aural skills, receptive responses, different abilities to make music, emotional responses to music, and the motivation to engage in music-related activities. This concept acknowledges the broad spectrum of musical ability beyond aural skills (Gembris, 1997; Hallam & Prince, 2003; Hallam, 2010) and implies that musical sophistication could be developed through various possibilities besides formal musical training. Second, the Gold-

MSI can be applied equally to all music-related domains: for example, to performing musicians of all styles, music educators, producers, and music engineers. This enables the Gold-MSI to assess musical sophistication in the general population where a broad variety of forms of engagement in music-related activities is commonly observed. Third, the Gold-MSI incorporates subjective self-assessment and a battery of objective listening tests that include complex and ecologically valid musical materials stemming from diverse musical styles. The bias toward European art music is thus avoided, which makes the Gold-MSI a research tool that complements most musical ability tests currently available.

Previous studies have documented good psychometric properties of the Gold-MSI inventory. Müllensiefen et al. (2014) examined internal consistency, test-retest reliability, convergent validity, discriminant validity, and the factor structure of the Gold-MSI inventory in a series of studies. With the data from an extensive online sample of English-speaking participants ($N = 147,633$), the Gold-MSI self-report inventory showed a 5+1 bifactor structure with a general factor, General Musical Sophistication, and five group factors, Active Engagement, Perceptual Abilities, Musical Training, Singing Abilities, and Emotions. All subscales of the Gold-MSI inventory possessed good to very good internal consistency (Cronbach's $\alpha = .791–.926$) and test-retest reliability ($r_{tt} = .857–.974$). These findings were later replicated very closely with data from a sample of American college students (Baker, Ventura, Calamia, Shanahan, & Elliott, 2018), German speaking adults (Schaal et al., 2014) as well as adolescent participants (Fiedler & Müllensiefen, 2015). The correlations between the Gold-MSI inventory and the Advanced Measures of Music Audiation (AMMA) (Gordon, 1989) were in the range of $r = .30$ to $.51$, whereby the factors Perceptual Abilities, Musical Training, Singing Abilities, and General Musical Sophistication were correlated more strongly with AMMA than were the factors Active Engagement and Emotions. Because the factors Perceptual Abilities, Singing Abilities, and Musical Training measure self-perceived musical

abilities or self-report on formal learning experiences that should facilitate the development of diverse musical skills, stronger correlations between these subscales and AMMA scores were to be expected and could be seen as evidence for convergent validity. Furthermore, weaker correlations between the two affective dimensions of musical sophistication, namely Active Engagement and Emotions, provided evidence for divergent validity: Compared to cognitive factors or self-perceived abilities, affective and motivational factors usually relate to performance scores to a lesser extent as reported in studies on school achievement (Jansen, Lüdtke, & Schroeders, 2016; Kriegbaum, Becker, & Spinath, 2018; Lau & Roeser, 2002). The correlations between the Gold-MSI inventory and two music perception tests of the Gold-MSI test battery, namely the Melody Memory Task and the Beat Alignment Perception Task, also revealed similar patterns.

The objective listening tests of the Gold-MSI are currently still in development. Among the Gold-MSI test batteries, the psychometric properties of the Melody Memory Task and the Beat Alignment Perception Task have been investigated more thoroughly. They were validated using Rasch modelling and proved to be unidimensional (Müllensiefen et al., 2014). The Melody Memory Task measures melody memory, probably the skill tested most often, in which test takers had to memorise two or more melodies and decide whether the stimuli were same or different. The capacity of memory for memory is associated with a range of musical skills such as sight-reading (Kopiez & Lee, 2008) and aural skills (Wolf et al., 2018). For an earlier 11-item version of the Melody Memory Task, Müllensiefen et al. (2014) reported reliability and validity of the task. Both internal consistency (Cronbach's $\alpha = .61-.68$) and test-retest reliability ($r_{tt} = .57$, calculated using $d'$ scores) were acceptable (Müllensiefen et al., 2014). Furthermore, Harrison, Musil, & Müllensiefen (2016) reported correlations between the $d'$ scores for the Melody Memory Task and the AMMA melody scores ($r = .488$) as well as AMMA rhythms scores ($r = .541$), providing evidence for its convergent validity.

The Beat Alignment Perception Task measures the ability to infer an underlying pulse or beat from an extract of music. For a 15-item version of the test, Müllensiefen et al. (2014) reported good internal consistencies (Cronbach's α = .87–.92) and test-retest correlation ($r_{tt}$ = .70). However, it should be noted that the Gold-MSI does not allow scores from different listening tests to be combined with those for the self-report inventory because it is still unclear whether a common factor of musical ability drives the correlations between different listening tests, as Wing (1971) assumed, or if melodic memory, beat perception and other musical abilities are largely unrelated to each other, as argued by Seashore et al. (1960).

**Aim of the present study**

The purpose of the present study was to adapt the Gold-MSI inventory to Traditional Chinese and to examine the psychometric properties of the Chinese version of the Gold-MSI using a Taiwanese sample. The central question of this study was whether the factor structure of the inventory would remain invariant after translation. In addition, we investigated reliability and validity of the Chinese version of the Gold-MSI in terms of classical test theory, whereby internal consistency and test-retest reliability were evaluated and validity was assessed by correlations between the Gold-MSI inventory scores and two perceptual music tests from the Gold-MSI test battery (the Melody Memory Task and the Beat Alignment Perception Task) and the Musical-Rhythmic Intelligence subscale from the Eight Multiple Intelligences Questionnaire (Chou, 2006). Finally, we explored whether socio-economic status (SES) variables had an influence on musical sophistication in the Taiwanese sample.

=========

Fig. 1 about here

=========

**Method**

**Translation and adaptation of the Gold-MSI inventory**

We translated the Gold-MSI inventory from English into Traditional Chinese following the *ITC Guidelines for Translating and Adapting Tests* (International Test Commission, 2017) as well as the recommendations in the methodological literature on cross-cultural research (Tran et al., 2017; van de Vijver & Leung, 2011). The aim of the translation was to ensure the comparability of the original and translated versions in terms of the underlying theoretical construct and semantic meaning, taking into consideration cultural differences. During the process of translation and adaptation, we engaged experts of both sexes to avoid possible gender bias. The translation process was divided into seven phases (see Figure 1).

*Phase I. Translation.* The English Gold-MSI inventory was translated independently by three translators, two of them women. As the Gold-MSI targets the general population, technical jargon was to be avoided in the questionnaire; therefore, we recruited only musically untrained translators. They all had bachelor's degrees in English linguistics and literature as well as at least two years of professional experience in translation. They were paid approximately 37 euros for each draft translation.

*Phase II. Creating the first version of the inventory.* The first complete draft of the inventory was developed on the basis of the three independent translations in a meeting between the first author and an expert in test development. The three translations were reviewed anonymously. Items were randomised and scrutinised independently. The aim was to assess the extent to which the same meaning were conveyed in the three versions. Where semantic meaning was identical, the best version in terms of linguistic quality was selected. In case of semantic inconsistencies, the wording was compared with the original English version, and

the most accurate version was selected. Finally, the wording of each item was examined, and 21 of 38 items were optimised with regard to linguistic quality.

*Phase III. Back-translation.* Subsequently, the first complete draft of the inventory was translated back into English by a professional translator who had more than five years of experience in the field and was musically untrained. He was paid about 58 euros for his translation.

*Phase IV. Comparing the back-translation with the original inventory.* As a next step, the agreement between the back-translation and the original English version of the inventory was assessed by the fourth author together with another expert in music psychology, fluent in English and familiar with the German and English versions of the Gold-MSI inventory. To evaluate the extent to which the semantic meaning of the items remained unchanged after translation, we developed a worksheet in which the back-translation was juxtaposed with the original version for each item. Discrepancies were found between the back-translation and original for 16 items.

*Phase V. Creating the second version of the inventory.* The 16 items for which discrepancies were found in Phase IV were inspected by a Taiwanese linguist. At this stage, the Chinese translation was compared with the original English version and the linguist indicated whether the meanings of both versions were identical. Four items were then marked as modifiable. Subsequently, the first author and the Chinese linguist revised these items together. The revision was then examined by the expert in test development. This resulted in the second version of the translated inventory.

*Phase VI. Pilot study.* In order to examine the comprehensibility and linguistic complexity of the second version, we conducted an online pilot study using the platform SoSci Survey (Leiner, 2014). The aim of the pilot study was to identify items that were difficult to read. We recruited participants both with and without professional musical backgrounds to ensure the feasibility of items for participants with various musical ability levels. Six Chinese native speakers, two of them women, with academic backgrounds in ethnomusicology ($n = 2$), composition, ($n = 1$), sociology ($n = 1$), social work ($n = 1$) and English ($n = 1$) completed the Chinese Gold-MSI inventory and indicated any items that seemed difficult in terms of their wording or that were phrased ambiguously. This step revealed two items in need of improvement.

*Phase VII. Creating the final version of the inventory.* The two items from Phase VI were revised by the first author according to the suggestions of the linguist and the expert in test development. This revision focused exclusively on word order and resulted in the final version of the inventory, which was validated in the main study.

**Development of the instructions for the music tests**

The instructions for the Melody Memory Task and the Beat Alignment Perception Task were written by the first author on the basis of the instructions in the *PsychoPy* implementations of the tasks developed by the Rhythm and Movement Lab at the University of Vienna (available from http://www.gold.ac.uk/music-mind-brain/gold-msi/download/). To verify that the tests could be implemented through SoSci Survey and that task instructions were comprehensible, we conducted a pilot study with nine participants (age $M = 32.78$ years, $SD = 3.15$). Six participants from the first pilot study were asked to indicate explicitly whether they had any problems understanding the tasks. Furthermore, three participants in Taiwan completed the tests using several different devices (smartphone, tablet, laptop, and PC) and operating

systems (Android, iOS, Windows, and macOS) under various circumstances (using private internet connection, mobile network connection or WiFi). The results of the pilot study indicated that some participants had trouble loading sound files and some problems with performance on Android devices. Therefore, we transferred all sound files to another server (http://www.wikidot.com) and amended the instructions by recommending the use of a laptop or PC.

**Main study**

The main study was created as an online survey and distributed through social media, mainly Facebook, a bulletin board system (BBS, telnet://ptt.cc), which has more than 1.5 million registered users and is the biggest Chinese BBS in the world, and through personal contacts of the first author. We contacted music critics, professors of musicology and schoolteachers to recruit participants. For the study, pupils aged at least 12 years and adults were recruited regardless of their sex, educational background or place of residence. All studies were performed in accordance with the relevant national guidelines and regulations (German Psychological Society, 2016). Informed consent was obtained from all participants. Anonymity of participants and confidentiality of their data were ensured.

**Procedure**

Participants were informed that the study aimed to investigate the reliability and validity of a test of musicality. In addition, the instructions mentioned that technical errors might occur if Android devices were used and that the use of headphones, laptop, or a PC in a quiet environment was highly recommended when taking the survey. Participants gave informed consent and indicated the type of device and operating system they were using and whether they wore headphones while answering the questionnaire. Demographic variables and socioeconomic status (SES) were then collected. Subsequently, participants completed the

Gold-MSI inventory and the Musical-Rhythmic Intelligence subscale of the Eight Multiple Intelligence Questionnaire (Chou, 2006). At this stage, we asked participants whether they had studied or were studying music and if they had pursued or were pursuing any music-related careers. Participants then started either with the Melody Memory Task or with the Beat Alignment Perception Task. For both tests, participants first went through two to three training items to ensure they understood the task. After listening to the sound file of each test item, participants had to advance to the next page to respond and to indicate the level of certainty regarding their answer. The page on which the music excerpts were presented was blocked after 15 to 20 seconds so that the excerpts could not be repeated. At the end of each test, participants indicated whether they had listened to the music excerpts in full.

As the expected survey completion time was comparatively long for an online survey (25–30 minutes) for both listening tests and all questionnaires, participants could decide to end the survey after completing the first listening test. It should be noted, however, that 95% of the participants completed both music tests. For all participants who completed the full survey, we provided feedback on the General Musical Sophistication score from the self-report inventory and the results of the listening tests. Furthermore, participants could register for a second survey, which enabled us to investigate test-retest reliability. After successful registration, confirmation was sent via e-mail, and they received the link for the second retest survey two weeks later. Only the questions regarding the device used, the Gold-MSI inventory and the two music tests were administered in the second survey.

**Measures**

*Gold-MSI Inventory*. The full Gold-MSI inventory comprises 39 items; 31 statements are judged on a seven-point Likert scale. Seven further questions were constructed using seven ordered response categories. The 38 metrical items are grouped into five subscales (Active

Engagement, Perceptual Abilities, Musical Training, Singing Abilities, Emotions) and one general factor (General Musical Sophistication). In addition, participants had to indicate their main instrument. The number of items, scale minimum, scale maximum, and item examples for each subscale are reported in Table 1.

==========

Tab. 1 about here

==========

*Musical-Rhythmic Intelligence.* Musical-Rhythmic Intelligence was measured by means of a subscale of the Eight Multiple Intelligences Questionnaire (Chou, 2006), which was based on the theory of multiple intelligences (Gardner, 1985) and was developed according to the self-report inventory by Pérez, Beltramino, and Cupani (2003). Accordingly, each of the eight multiple intelligences is reflected in five to eight items. In a validation study with a sample of Taiwanese students aged 16 to 20 ($n = 1,499$), it could be seen that the Eight Multiple Intelligences Questionnaire had good internal consistency (Cronbach's $\alpha = .90$–$.96$), and the theoretical dimensions of the eight multiple intelligences could be confirmed by exploratory and confirmatory factor analysis (Chou, 2006). The subscale of Musical-Rhythmic Intelligence contains six items, namely "compose," "sing a second voice to a melody sung by others," "tune musical instruments," "play an instrument in an ensemble,", "notate simple melodies," and "arrange melodies.". Using a ten-point rating scale, participants provide information on their self-assessed abilities with regard to the above six items.

*Melody Memory Task.* The Melody Memory Task is part of the Gold-MSI test battery and measures memory for melody. The version of the test used by Harrison et al. (2016) and Allen et al. (2017) contains 13 items. The paradigm aligns closely with the test used by Halpern, Bartlett, and Dowling (1995). In each trial, two versions of the same melody are presented with the second version always transposed into another key. The second melody can either be

identical to the first melody (disregarding the transposition), or one to three notes are changed by maintaining or violating the melodic contour. Participants have to decide whether the two versions presented in a trial are identical despite the transposition in the second melody. Additionally, they have to indicate their level of certainty (1−3) for each answer. The stimuli were derived from well-known folk songs or tunes in the Western cultural region. However, the order of the intervals is altered so that all melodies are novel to participants while the compositional structure and melodic style corresponds to European folk music to ensure the ecological validity of the stimuli.

*Beat Alignment Perception Task*. The Beat Alignment Perception Task is part of the Gold-MSI test battery and measures the ability to perceive beats in music. The version of the test used in this study contains 17 items. The paradigm was developed by Iversen and Patel (2008). Beep tones that either correspond exactly to the beat or deviate from the beat of the music are superimposed in the music excerpts. Participants have to decide whether the beep tones are on the beat or off the beat and indicate their level of certainty (1−3) for each response. The music excerpts used in this test were obtained from a royalty-free sound and music library (http://www.audionetwork.com). There are three different types of item manipulation: The beep tones are consistently placed before the beat of the music, or the tempo of the beep tones are faster or slower than that of the music.

*Socioeconomic status (SES).* In the present study, the participants' level of education and occupational status served as indices for SES. We used the classification guidelines of the ISCED 2011 (International Standard Classification of Education), published by the UNESCO Institute for Statistics (UNESCO Institute for Statistics, 2012, 2015), to determine the level of education attainment. We divided the level of educational attainment into eight categories:

namely, "less than primary education," "primary education," "lower secondary education," "upper secondary education," "short-cycle tertiary education," "bachelor's or equivalent level," "master's or equivalent level," and "doctoral or equivalent level." For occupational status, we applied the five-point scale of the socioeconomic status developed by Hwang (2009). We investigated occupational status by means of 14 categories, which were classified into five levels. According to Hwang (2009), the correlation (Pearson's *r*) between the five-point scale and the finer New Taiwan Occupational Socio-Economic Scale, in which the careers must be coded manually on the basis of several criteria, and each occupation can thus be assigned to a differentiated value, was high ($r = .72$). It should be noted that we offered an alternative option on the scale ("not applicable") to avoid termination of the survey due to difficulties with SES-related variables. For this reason, there are six and 36 missing values for levels of educational attainment and occupational status, respectively.

**Participants**

Totals of 1,108 and 170 participants took part in the first and second survey, respectively. Preceding the analysis, valid cases were identified according to the criteria described in the following sections. Data were cleaned separately for the Gold-MSI inventory, the Melody Memory Task, and the Beat Alignment Perception Task in order to maximise the number of valid observations included in the data analysis.

**Data preparation**

For the Gold-MSI inventory, we first examined whether the participants checked the same category on all items. To this end, we calculated the variance of responses and the MAD (median absolute deviation from the median) of the variance of responses for each participant. The doubled MAD was then subtracted from the median of the variance of responses. Cases with variance of responses smaller than this value should be excluded from the dataset (Leys,

Ley, Klein, Bernard, & Licata, 2013). No cases were removed on these grounds. Second, we controlled for processing time by screening for normal processing speed. The Chinese Gold-MSI inventory has approximately 1,000 characters. Therefore, it seemed implausible to read through the questions and give reliable responses within two minutes while a processing time of more than 15 minutes indicated problems with concentration. Accordingly, all participants who took less than two minutes or more than 15 minutes to complete the 39 questions of the Gold-MSI inventory were excluded from the sample. This was the case for 43 people from the first survey (1,065 participants remaining) and 10 people from the retest survey (160 participants remaining).

For the two music tests, we controlled first for the processing time. The length of the music excerpts was between 10 and 20 seconds. Although we blocked the browser pages in which music excerpts were played after 15 to 20 seconds to ensure that the music excerpts could not be played repeatedly, this may not have been successful in all instances. Therefore, taking possible internet connection problems into account, we excluded those cases of participants who listened to the excerpts for more than 35 seconds. Moreover, participants who needed more than 60 seconds for answering a question were considered inattentive and were accordingly discarded. We then identified the participants who reported not having listened to all the excerpts in full and those whose device did not supported JavaScript, which was essential to the function of blocking the pages for playing music. As a result, we removed 310 cases from the results of Melody Memory Task in the first survey (leaving 769 cases) and 51 cases from the results of the retest survey (leaving 112 cases). We also applied the same criteria on the results of the Beat Alignment task and removed 187 cases from the first survey (leaving 894 cases) as well as 32 cases from the second survey (leaving 137 cases).

The mean age of the 1,065 participants (770 female, 289 male, 6 did not disclose their sex) who completed our version of the Gold-MSI inventory was 26.55 years ($SD$ = 8.22 years, range = 12–68). The majority (97.7%) had grown up in Taiwan and were of Taiwanese nationality when they took part in the survey; 91.8% of the participants were living in Taiwan. In addition, most had not studied or were not currently studying music (85.7%), nor did they have professional working experience in music-related fields (78.0%). Due to the fact that the participant sample had a homogeneous cultural background and were mostly non-musicians, we are convinced that the sample was appropriate for testing the validity of the Chinese Gold-MSI inventory although it was not representative of the Taiwanese population as a whole with regard to age, educational background and occupational status. More specifically, 95% of the participants were under 43 years old and 73% had a bachelor's degree or higher, with more than half of them having careers requiring high qualifications, as indicated by their SES status. In summary, the younger age groups and those with higher socioeconomic status appeared to be over-represented in the sample of the current study compared to the general population of Taiwan (see Figures S1, S2, and S3 in the Supplemental Material Online section for the distribution of the variables age, level of education, and occupational status).

**Data analysis**

All statistical analyses were conducted with the *R Project for Statistical Computing* (R Core Team, 2017) using *RStudio* (RStudio Team, 2016). We carried out the analysis in several steps: First, the descriptive statistics of all items of the Gold-MSI inventory including mean, standard deviation, skewness and kurtosis were calculated using the *moments* package (Komsta & Novomestky, 2015). Subsequently, we examined the assumption of multivariate normality for the Gold-MSI items using the *MVN* package (Korkmaz, Goksuluk, & Zararsiz, 2014). The factor structure of the Gold-MSI inventory was then investigated through

confirmatory factor analysis with the help of the *lavaan* package (Rosseel, 2012). Reliability and validity indices of the Gold-MSI inventory were calculated using the *psych* package (Revelle, 2017). Finally, we explored the relation between the SES variables and the musical sophistication by means of structural equation modelling realised by *lavaan*.

## Results

**Factor structure of the Chinese Gold-MSI inventory**

The skewness of all items was in the range between $-1.03$ and .73, with 30 items being negatively skewed (for statistical details and the translated inventory see Table S1 in the Supplemental Material Online section). This indicates that the participants tended to score highly on most of the items. The kurtosis was between 1.50 and 4.48, suggesting that some response categories predominated. Therefore, we examined the distribution of frequencies of answer categories for every item of the Gold-MSI inventory to check if all categories were selected in all items. A visual inspection showed that no categories remained unselected despite the fact that distributions were mostly left-skewed, and the values might not have been normally distributed. To examine the assumption of multivariate normality, we carried out three tests available in the *MVN* package: the Henze-Zirkler test, the Mardia test, and the Royston test. All three tests indicated that the assumption of multivariate normality was violated to some degree. However, the departure from normality is considered less problematic with a large sample (see Tabachnick & Fidell, 2013, p. 80). In addition, robust maximum likelihood estimators were employed in the confirmatory factor analysis, which can provide unbiased estimates under violations of distributional assumption (Beaujean, 2014).

A confirmatory factor analysis was conducted using the *lavaan* package to examine the factor structure of the Chinese Gold-MSI inventory. We adapted the same analytic procedure as in Müllensiefen et al. (2014). Four models were specified: Model 1 was a hierarchical model

where the general factor is supposed to influence the five groups factors, which in turn have an impact on the individual items associated with them; Model 2 was a bifactor model in which all the individual items are influenced by the general factor, on the one hand, and one of the group factors, on the other hand (the preferred model for the original English Gold-MSI inventory); Model 3 assumed that the general factor does not exist, and the five factors are not correlated with each other; in Model 4, the general factor was not specified either, but the five factors were allowed to be inter-correlated. The following fit-indices are reported in Table 2: $\chi^2$, Bayesian Information Criterion (BIC), Tucker-Lewis Index (TLI), Bentler's Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardised Root Mean Square Residual (SRMR).

==========

Tab. 2 and Fig. 2 about here

==========

The $\chi^2$ values of all models were statistically significant, indicating a deviation from an exact model fit. However, the statistical significance of $\chi^2$ values is affected by sample size. With large samples ($N > 1,000$), the $\chi^2$ values usually become statistically significant and should thus not be used to evaluate the absolute fit of the model to the data (see Keith, 2015, p. 311). We then compared the four models by means of BIC, which showed that the bifactor model had the best fit of all the models. RMSEA and SRMR were below the recommended cut-off values of .08 and .06 (Keith, 2015), providing further evidence for the fit of the model. Although TLI and CFI did not exceed the value of .90 for an adequate fit, we argue that CFA models with more than five factors and five to ten items for each corresponding factor could barely have TLI and CFI values over .90 (Marsh, Hau, & Grayson, 2005; Marsh et al., 2010). Accordingly, in the present study the factor structure of the original Gold-MSI (Müllensiefen et al., 2014) as well as the German version (Schaal et al., 2014) could be replicated successfully. In other words, the original factor structure still held with the Chinese Gold-MSI

inventory, providing evidence that the multidimensional construct of musical sophistication also exists in the Taiwanese culture and can be measured using the Gold-MSI inventory. The factor structure and the standardised factor loadings are reported in Figure 2.

However, it should be noted that there could still be a cultural difference regarding the interpretation of the construct. That is, participants in Taiwan might interpret the construct musical sophistication differently from those who grew up in Western cultures. Müllensiefen et al. (2014) selected 18 items with the highest standardised factor loadings as the subset of the G-factor General Music Sophistication. This procedure would have yielded a different subset in the present study. If the 18 items with the highest factor loadings were to be selected for a short version of the G-factor, Items 05, 11, 18, 20, 22, 26, and 30 should have been selected instead of Items 03, 17, 23, 24, 25, 33, 37. However, we decided to adopt the original subset of items for better comparability between different language versions of the Gold-MSI inventory despite the possible difference in factor loadings. Finally, it should be emphasized that the scores of the Chinese version cannot be compared directly with the scores of the English or German version. For valid cross-cultural comparisons, further investigations regarding the measurement invariance between the different language versions of the Gold-MSI inventory are required (Putnick & Bornstein, 2016).

==========

Tab. 3, Tab. 4 about here

==========

**Reliability of the Chinese Gold-MSI inventory**

We calculated the sum scores for all Gold-MSI subscales. Subsequently, Cronbach's α and test-retest correlations of the Gold-MSI subscales were computed as indicators of internal consistency and test-retest reliability, respectively. Summary statistics and the indicators of

reliability for the Gold-MSI inventory are reported in Table 3. Generally, internal consistency and test-retest reliability of the Gold-MSI subscales were very good ($\geq$ .80); only the subscale Emotions was adequate ($\geq$ .70). The subscale General Musical Sophistication using the same 18 items as in the English version had an excellent internal consistency and test-retest reliability ($\geq$ .90). Indicators of reliability for the Melody Memory Task, the Beat Alignment Perception Task, and the Musical-Rhythmic Intelligence are also reported in Table 3. For the Melody Memory Task and the Beat Alignment Perception Task, $d'$ scores were computed, whereby values below zero were set to chance level ($d' = 0$). Musical-Rhythmic Intelligence was calculated by summing the six corresponding items. Although test-retest correlations of both music tests did not exceed $r = .70$, we still considered their reliability to be acceptable due to the modest test length. The scores for Musical-Rhythmic Intelligence subscale showed a good internal consistency.

**Discrimination analysis**

We analysed the discrimination power of the Gold-MSI subscales and the sensitivity scores ($d'$) of the Melody Memory Task as well as the Beat Perception Alignment Task. First, participants were assigned to either the high- or as low-score group according to their scores on each subscale or listening task. The high-score groups contained participants with a percentile rank greater than 73 whereas participants who scored lower than 27th percentile were classified as low-score groups (DeMars, 2018, p. 59). Subsequently, we investigated the differences in scores of the Gold-MSI subscales, the Melody Memory Task and the Beat Alignment Perception Task between high- and low-score groups by a series of $t$-test. The results are reported in Table S2 (see the Supplemental Material Online section). All mean differences between groups were statistically significant with large effect sizes (Cohen's $d$ = 1.70–7.36), providing evidence for the discrimination power of the test scores.

**Validity evidence based on correlations with other variables**

We examined the correlations of the subscales of the Gold-MSI with the Melody Memory Task, the Beat Alignment Perception Task and the Musical-Rhythmic Intelligence subscale of the Eight Multiple Intelligences Questionnaire. The correlation coefficients are reported in Table 4.

Table 4 shows that the correlations between the subscales of the Gold-MSI inventory and both music tests had weak to medium effects ($r = .14–.42$). We observed stronger effects for the group factors Perceptual Abilities (F2), Musical Training (F3), Singing Abilities (F4) and the general factor General Musical Sophistication (FG) while the correlations were weaker for the group factors Active Engagement (F1) and Emotions (F5). Given that the group factors Active Engagement (F1) and Emotions (F5) should address the affective aspects of musical sophistication (i.e., motivation to participate in music-related events or emotional response to music), it was not surprising that the correlations of these two factors with the music tests were lower than the other three group factors and the general factor, which should measure self-assessed musical skills. Thus, these correlations could be regarded as evidence for convergent and divergent validity of the Chinese Gold-MSI inventory (Abell, Springer, & Kamata, 2009). These findings are in line with the study by Müllensiefen et al. (2014), in which the correlations between the subscales of the Gold-MSI with the music tests also demonstrated weak to medium effects in a large online sample ($N > 100,000$). The correlations between the factors of the Gold-MSI inventory and the Musical-Rhythmic Intelligence subscale also demonstrated a similar pattern but with significantly larger effect sizes ($r = .36–.79$), providing further evidence for validity of the Chinese Gold-MSI inventory.

==========

Fig. 3 and Fig. 4 about here

==========

**Relationship between socioeconomic status and musical sophistication**

Based on two structural equation models, we investigated the possible impact of

socioeconomic status (SES) on musical sophistication. In the structural equation models, the

latent variable SES was estimated in a formative way, with level of education and

occupational status serving as its indicators. The measurement error was set to zero to identify

the model (Brown, 2015, pp. 322–328). We also specified the paths from age to level of

education and occupational status to control for a possible confounding effect. Subsequently,

we specified all paths from age and SES to the five group factors (see Figure 3) and then to

the general factor of the Gold-MSI inventory (see Figure 4) in order to explore the influence

of SES and age on musical sophistication. The correlations between the Gold-MSI subscales

were also estimated. Finally, we removed those paths whose coefficients were not statistically

significant ($p > .05$) and estimated the parameters again. The models were estimated by means

of the R-package *lavaan*, using the Full Information Maximum Likelihood (FIML) method.

We reported standardised coefficients in both models. The method of the graphical

representation applied in these diagrams is in line with the tradition in structural equation

modelling (Ho, Stark, & Chernyshenko, 2012) whereby squares signify observed variables or

variables and circles represent latent variables. In the graphs, the Gold-MSI subscales are

placed in squares because we used sum scores instead of factor scores in the models. The

small "e's" in the circles signify residual variances. The curved bidirectional arrows depict the

correlations between variables, and the related parameters are correlation coefficients. The

directional arrows imply the influence of a predictor on its outcome; the parameters could be

interpreted in a similar manner as the standardised coefficients (β) in multiple regression models. For example, the arrow from SES to Musical Training has a β-value of .24, which implies that if one had achieved an increase of one standard deviation in SES, one would have had .24-standard deviation increase in Musical Training after controlling for the effect of age on SES. It should be noted that such an interpretation is only valid when the causal relationship between the two variables has already been well established. This is also a challenge for all non-experimental studies in which confounding effects could barely be completely excluded. Given the exploratory nature of our analysis, we suggest that the effects in the models should be interpreted with caution.

Figure 3 illustrates the structural equation model depicting the impact of SES on the group factors of the Gold-MSI inventory. The impact of SES on the general factor General Musical Sophistication is demonstrated in Figure 4. Both models showed a good model fit. The $\chi^2$ values did not exceed the significance level ($p =$.124 and .944). In addition, the CFI as well as the TLI were above .95, and the RMSEA as well as the SRMR were below .03. Figure 3 shows that that SES exerted a weak effect on the group factors Perceptual Abilities (F2), Musical Training (F3), and Singing Abilities (F4) as well as on the general factor while age did not have any significant effects. These findings confirmed the impact of SES on the factor Musical Training (F3) found in previous studies (Fiedler & Müllensiefen, 2015; Schaal et al., 2014).

## Discussion

**Psychometric properties of the Chinese Gold-MSI**

In the present study, the Gold-MSI inventory, together with the Melody Memory Task and the Beat Alignment Perception Task, was translated from English into Traditional Chinese for use with a sample of participants from Taiwan. The psychometric properties of the Chinese Gold-

MSI were examined. Results of a confirmatory factor analysis indicate that the original bifactor model could be replicated in the Taiwanese sample, which supports the existence of the construct of musical sophistication in Taiwanese culture. Furthermore, all subscales of the Chinese Gold-MSI inventory had sufficient to very good reliability. The general factor General Musical Sophistication had very good internal consistency ($\alpha = .90$) and test-retest reliability ($r_{tt} = .92$) while the reliability of the group factors was also good or very good ($\alpha = .83–.85$, $r_{tt} = .83–.93$) except for the factor Emotions where reliability was acceptable ($\alpha = .74$, $r_{tt} = .76$). These findings confirm the results of previous studies (Fiedler & Müllensiefen, 2015; Müllensiefen et al., 2014; Schaal et al., 2014) and suggest that the Chinese Gold-MSI inventory can be used to reliably assess musical sophistication in the Taiwanese culture.

In addition, the correlations of the Gold-MSI subscales with the Musical-Rhythmic Intelligence subscale of the Eight Multiple Intelligences Questionnaire provide evidence for the convergent and divergent validity of the Gold-MSI inventory. The correlation between the general factor General Musical Sophistication and the Musical-Rhythmic Intelligence was $r = .75$. The group factors Perceptual Abilities, Musical Training, and Singing Abilities also correlated strongly with the Musical-Rhythmic Intelligence ($r = .58–.79$). As the three group factors and the general factor were designed to measure self-evaluated musical skills and abilities, the strong correlations with the Musical-Rhythmic Intelligence can be interpreted as showing convergent validity (Abell et al., 2009). In contrast, the correlations of the group factors Active Engagement and Emotions with the Musical-Rhythmic Intelligence were weaker ($r = .42$ and $.36$). This can be seen as evidence for divergent validity due to the fact that the two group factors were intended to capture the affective aspect of musical sophistication. Moreover, the correlations of the Gold-MSI factors with the Melody Memory Task and the Beat Alignment Perception Task also demonstrate a similar pattern, providing further evidence for convergent and divergent validity of the Chinese Gold-MSI inventory

although the correlations were essentially weaker. The weaker correlations indicate that the results of the self-evaluation may have differed from those of objective testing. However, the compilation of listening tests of the Gold-MSI is still incomplete. Further tests assessing different listening abilities need to be included in the test battery so that the full factor structure of musical sophistication can be investigated objectively in the future.

**The impact of SES on musical sophistication**

Finally, we explored the impact of SES on musical sophistication by means of two structural equation models. Our findings show that SES could facilitate the group factors Perceptual Abilities (F2), Musical Training (F3), and Singing Abilities (F4) as well as the general factor General Musical Sophistication (FG). The fact that the factor Musical Training (F3) was related to SES accords with the findings of previous studies (Fiedler & Müllensiefen, 2015; Müllensiefen et al., 2014; Schaal et al., 2014). Schaal et al. (2014) also report weak effects of SES variables on F2, F3, and F5 in a German adult sample while Fiedler and Müllensiefen (2015) found medium to strong effects of SES variables on F2 and F3 in the sample of German pupils. However, our finding that age does not influence musical sophistication is not in line with previous studies. Schaal et al. (2014) reported small to moderate negative effects of age on all group factors of the Gold-MSI inventory in the German adult sample whereas Fiedler and Müllensiefen (2015) found that age had weak to moderate effects on F1, F2, and F5 in the sample of German pupils. Furthermore, Schaal et al. (2014) did not observe an effect of SES on the general factor. These discrepancies suggest that there might be cultural differences regarding the impact of SES and age on musical sophistication. Nevertheless, it should be noted that the difference observed in the present study cannot be interpreted clearly due to the fact that the SES variables were measured by means of different indicators and also that the samples consisted of participants of different ages. Furthermore, the samples used in the present study as well as in previous studies are not representative regarding the

distribution of the SES variables. Thus, the present study does not resolve to what extent the difference can be attributed to cultural diversity and not to the artefacts caused by dissimilar research designs or by sampling bias.

The role of the educational system in the development of musical sophistication and the role of the family's socioeconomic status during the developmental process should also be of interest for future investigations in sociological and educational studies. Further studies could adapt a longitudinal design to investigate the relationships between SES, musical sophistication and other variables of interest such as musical achievement, motivational factor, personality factors as well as intelligence.

**The use of the Gold-MSI inventory and listening tests**

Since its publication, the Gold-MSI has been widely used in various research projects (e.g., Bouwer, Werner, Knetemann, & Honing, 2016; Fiedler & Müllensiefen, 2017; Jakubowski & Müllensiefen, 2013; Kopiez et al., 2016; Martins, Gingras, Puig-Waldmueller, & Fitch, 2017; Matthews, Thibodeau, Gunther, & Penhune, 2016; Schaal et al., 2015). The subscales of the Gold-MSI inventory can be applied separately according to the subject of interest in each project. For example, the subscale Musical Training served as a grouping criterion to dichotomise participants into musicians and non-musicians for group comparison experiments (Schaal et al., 2015). Kopiez et al. (2016) applied the subscale General Musical Sophistication and identified a positive correlation between musical sophistication and the discrimination of real orchestra sounds from orchestra sample libraries. In a longitudinal study, Fiedler and Müllensiefen (2017) investigated the musical development of German students in secondary schools using the whole Gold-MSI inventory along with other research instruments. Moreover, the Melody Memory Task and the Beat Alignment Perception Task have also been

shown useful in research on music perception and cognition (Bouwer et al., 2016; Jakubowski & Müllensiefen, 2013; Martins et al., 2017; Matthews et al., 2016).

To enhance the norm-referenced interpretations of the test scores, we have provided percentile norms for the Chinese Gold-MSI inventory, the Melody Memory Task, and the Beat Alignment Perception Task on the basis of the Taiwanese sample (see Tables S3, S4, and S5 in the Supplemental Material Online section). For the subscales of the Gold-MSI inventory, test scores were calculated by summing corresponding items after reversing the negatively phrased items (items 09, 11, 13, 14, 17, 21, 23, 25, and 27). Müllensiefen, Gingras, Stewart, and Musil (2013) offer a function in the R programming language for the scoring of the music tests. The scoring function generates *Accuracy*, *d'* (Sensitivity), and Area Under Curve *(AUC)* according to Signal Detection Theory (Sorkin, 1999). We recommend test users to apply *d'* in statistical analysis as *d'* takes both correct identifications as well as false alarms into consideration and is thus a more reliable value than the simple sum scores of accuracies (e.g., percent correct). While conducting statistical analysis for the self-report inventory, researchers may either use raw scores or the percentile rankings reported in the current paper. If research data are collected using different language versions of the Gold-MSI in different cultural regions, however, the use of percentile ranking for the interpretation of the test values is highly recommended. In this context, it should be emphasised again that the direct comparison between the test scores of the Gold-MSI of different language versions is not permissible as the results suggest that there might be cultural differences in the factor loadings of the measurement model. For valid comparisons, the strict measurement invariance between the different language versions of the Gold-MSI inventory would have to be established.

Furthermore, it should be noted that we translated the Gold-MSI inventory into Traditional Chinese and adapted its linguistic properties for the Taiwanese culture. Psychometric

properties were also examined using a Taiwanese sample. Although the inventory could also be applied in other Chinese speaking regions using Simplified Chinese, such as in China or Singapore, after converting the Traditional Chinese characters into simplified ones, the linguistic properties are at times quite different. Some wording may be unnatural for the participants outside Taiwan. Further investigations regarding the applicability of the Chinese Gold-MSI in other Chinese speaking regions may be of interest from the perspective of cross-cultural research in music psychology. As stated by Stevens and Byron (2016), most findings on music perception and cognition stem from a participant pool with European cultural backgrounds and cross-cultural research. It is therefore crucial to collect less culturally-biased knowledge in the discipline of music psychology. Accordingly, the present study has made a valuable contribution to research in the cross-cultural psychology of music by adapting a measurement instrument already widely established in Western research communities for use in Taiwan and potentially other Chinese-speaking countries.

## Supplement

Tables with the index 'S' are available as Supplemental Online Material, which can be found attached to the online version of this article at http://xxxxxxx.

Table 1. Basic information on the Gold-MSI test batteries and item examples of the Gold-MSI inventory

|  | N | Min. | Max. | Example items |
|---|---|---|---|---|
| **F1 Active Engagement** | 9 | 9 | 63 | I spend a lot of my free time doing music-related activities. (FG) |
|  |  |  |  | I don't spend much of my disposable income on music. |
| **F2 Perceptual Abilities** | 9 | 9 | 63 | I usually know when I'm hearing a song for the first time. |
|  |  |  |  | I can compare and discuss differences between two performances or versions of the same piece of music. (FG) |
| **F3 Musical Training** | 7 | 7 | 49 | I can play 0 / 1 / 2 / 3 / 4 / 5 / 6 or more musical instruments. (FG) |
|  |  |  |  | I have had 0 / 0.5 / 1 / 2 / 3-5 / 6-9 / 10 or more years of formal training on a musical instrument (including voice) during my lifetime. |
| **F4 Singing Abilities** | 7 | 7 | 49 | I can sing or play music from memory. (FG) |
|  |  |  |  | I don't like singing in public because I'm afraid that I would sing wrong notes. (FG) |
| **F5 Emotions** | 6 | 6 | 42 | I sometimes choose music that can trigger shivers down my spine. |
|  |  |  |  | I often pick certain music to motivate or excite me |
| **FG General Musical Sophistication** | 18 | 18 | 126 | Items with "FG" in aforementioned examples also belong to the subscale General Musical Sophistication. |
| *d'* **Melody Memory Task** | 13 | 0 | 3.46 | |
| *d'* **Beat Alignment Perception Task** | 17 | 0 | 3.46 | |

*Note*. N = Number of items, Min. = minimum value of the scale, Max. = maximum value of the scale. FG = General Musical Sophistication. The subscales of the Gold-MSI inventory are

scored using scale sums. The Melody Memory Task and the Beat Alignment Perception Task are scored with d' values by means of Signal Detection Theory (Sorkin, 1999).

Table 2. Comparison of the four models from confirmatory factor analysis

| | **Model 1** (Hierarchical) | **Model 2** (Bifactor) | **Model 3** (Simple factors) | **Model 4** (Correlated factors) |
|---|---|---|---|---|
| $\chi^2$ | 3008.43 | 2286.26 | 5115.58 | 2835.18 |
| *df* | 660 | 627 | 665 | 655 |
| BIC | 129682.75 | 129056.00 | 132062.20 | 129516.82 |
| TLI | .835 | .879 | .689 | .846 |
| CFI | .845 | .892 | .706 | .856 |
| RMSEA | .062 | .053 | .085 | .060 |
| SRMR | .069 | .052 | .247 | .064 |

*Note*. BIC = Bayesian Information Criterion, TLI = Tucker-Lewis Index, CFI = Bentler's Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, and SRMR = Standardised Root Mean Square Residual.

Table 3. Summary statistics and indicators of reliability for the Gold-MSI inventory, Melody Memory Task and Beat Alignment Perception Task

| | *M (SD)* | Min. | Max. | Sk. | Kur. | α | $r_{tt}$ |
|---|---|---|---|---|---|---|---|
| **F1 Active Engagement** | 40.24 (8.77) | 9 | 63 | −.07 | −.30 | .83 | .91 |
| **F2 Perceptual Abilities** | 46.56 (7.89) | 9 | 63 | −.25 | .04 | .85 | .83 |
| **F3 Musical Training** | 27.20 (9.57) | 7 | 49 | −.06 | −.70 | .85 | .93 |
| **F4 Singing Abilities** | 32.42 (7.13) | 7 | 49 | −.32 | .18 | .83 | .87 |
| **F5 Emotions** | 32.78 (4.82) | 6 | 42 | −.40 | .39 | .74 | .76 |
| **FG General Musical Sophistication** | 81.22 (16.64) | 18 | 126 | −.17 | −.23 | .90 | .92 |
| ***d' Melody Memory Task (N = 769)*** | 1.73 (0.97) | 0 | 3.46 | −.26 | −.31 | - | .59 |
| ***d' Beat Alignment Perception Task (N = 894)*** | 2.07 (0.97) | 0 | 3.46 | −.63 | .08 | - | .64 |
| **Musical-Rhythmic Intelligence** | 26.32 (14.00) | 6 | 60 | .26 | −.99 | .89 | - |

*Note*. Min. = minimum value of the scale, Max. = maximum value of the scale, Sk. = skewness, Kur. = kurtosis, α = Cronbach's α, $r_{tt}$ = test-retest correlation. $N = 157$ for $r_{tt}$ of the Gold-MSI subscales, $N = 89$ for $r_{tt}$ of the *Melody Memory Task*, $N = 113$ for $r_{tt}$ of the *Beat Alignment Perception Task*. All test-retest correlations were statistically significant ($p < .0001$).

Table 4. Correlations of the subscales of the Gold-MSI with the Melody Memory Task, Beat Alignment Perception, and Musical-Rhythmic Intelligence subscale of the Eight Multiple Intelligences Questionnaire

|  | F1 | F2 | F3 | F4 | F5 | FG |
|---|---|---|---|---|---|---|
| $d'$ Melody Memory Task ($N = 769$) | .14 | .29 | .27 | .25 | .15 | .29 |
| $d'$ Beat Alignment Perception Task ($N = 894$) | .16 | .37 | .42 | .31 | .15 | .41 |
| Musical-Rhythmic Intelligence ($N = 1,065$) | .42 | .65 | .79 | .58 | .36 | .75 |

*Note*. F1: Active Engagement, F2: Perception Abilities, F3: Musical Training, F4: Singing Abilities, F5: Emotions, FG: General Musical Sophistication. All correlation coefficients were statistically significant ($p < .0001$).
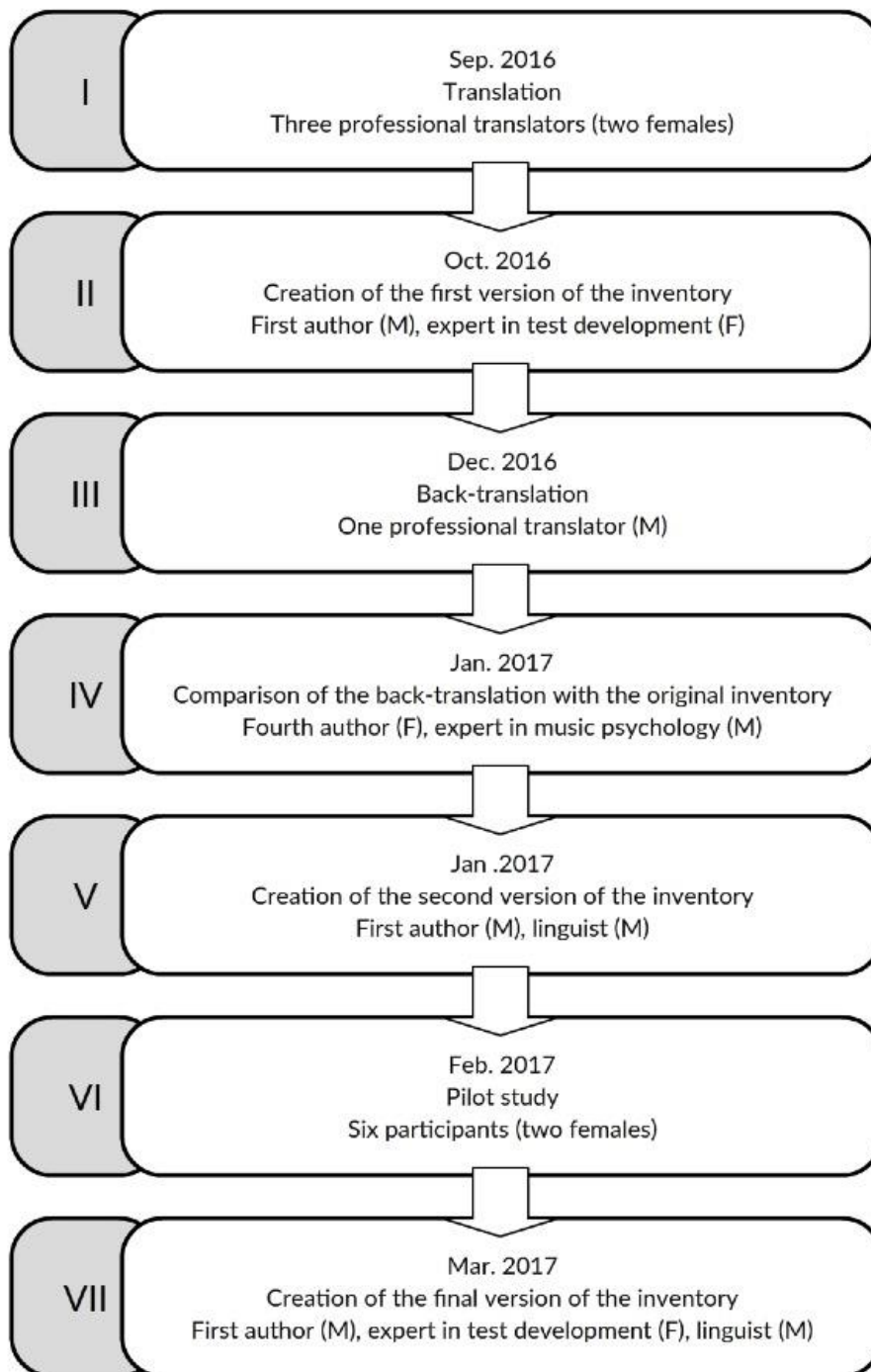
Figure captions



I — Sep. 2016
Translation
Three professional translators (two females)

II — Oct. 2016
Creation of the first version of the inventory
First author (M), expert in test development (F)

III — Dec. 2016
Back-translation
One professional translator (M)

IV — Jan. 2017
Comparison of the back-translation with the original inventory
Fourth author (F), expert in music psychology (M)

V — Jan .2017
Creation of the second version of the inventory
First author (M), linguist (M)

VI — Feb. 2017
Pilot study
Six participants (two females)

VII — Mar. 2017
Creation of the final version of the inventory
First author (M), expert in test development (F), linguist (M)

Figure 1. Procedure of translation and adaptation of the Gold-MSI inventory from English into Traditional Chinese, (M) indicating male and (F) indicating female experts
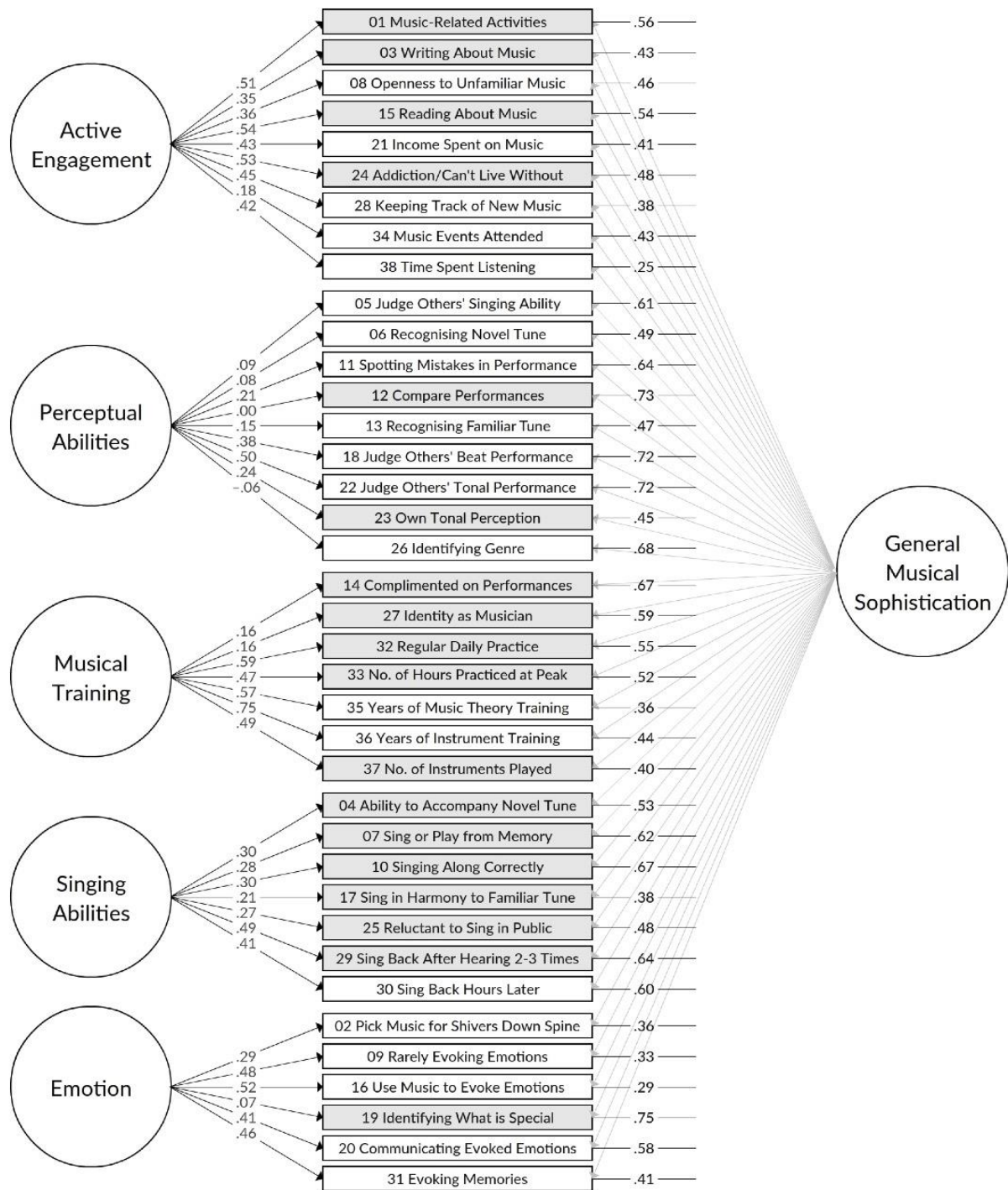
Figure 2. Factor structure of the Chinese Gold-MSI inventory. Standardised factor loadings are reported. The item subset for the G-factor General Musical Sophistication is highlighted in grey. Error terms are not reported so as to reduce the complexity of the graph. The numbers before the abbreviated item names indicate the item numbers of the Gold-MSI inventory. For example, the first item "I spend a lot of my free time doing music-related activities" is

represented as "01 Music-Related Activities." See Table S1 in the Supplementary Material for all items of the Gold-MSI inventory.
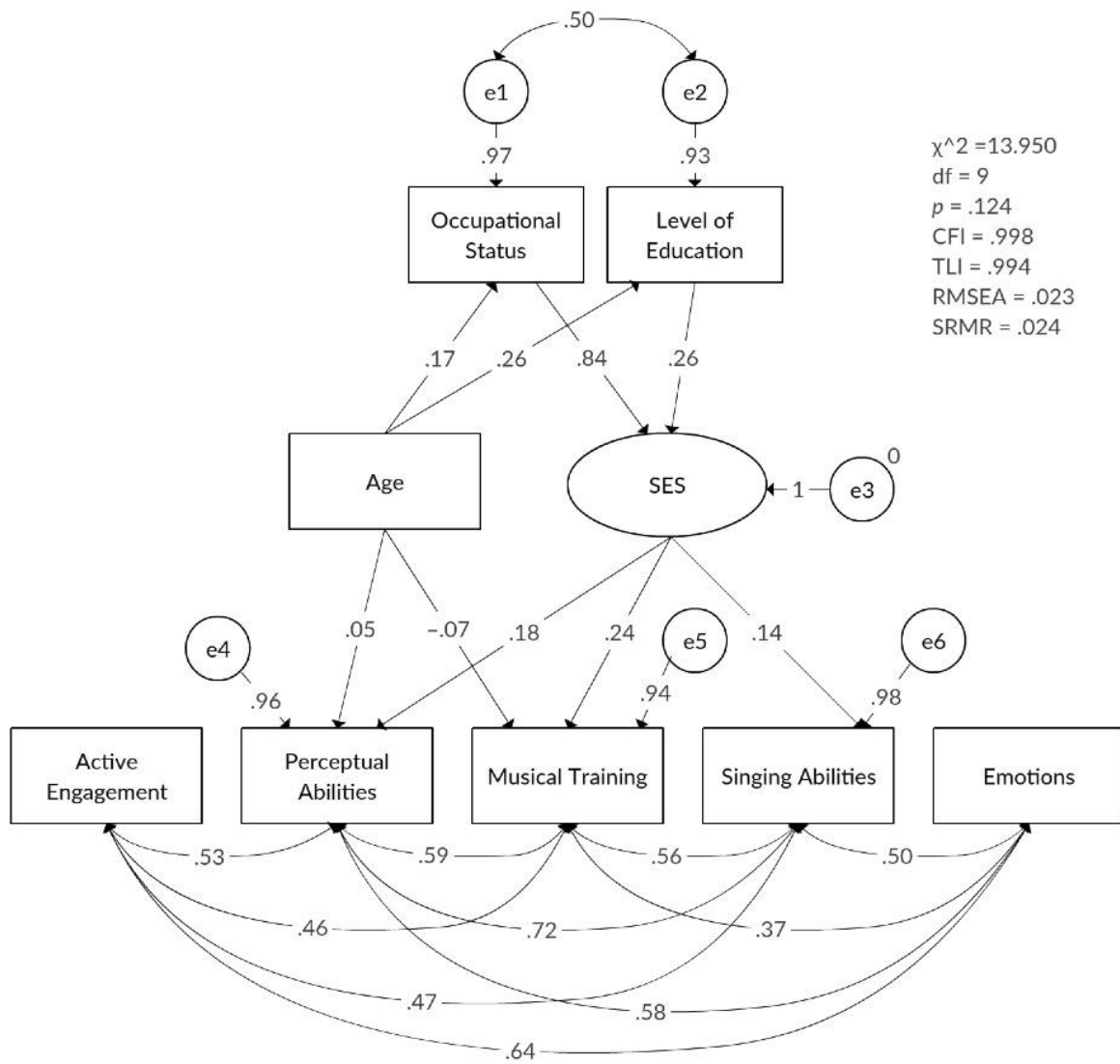
Figure 3. Structural equation model relating SES to the group factors of the Gold-MSI inventory. Standardised factor loadings are reported.
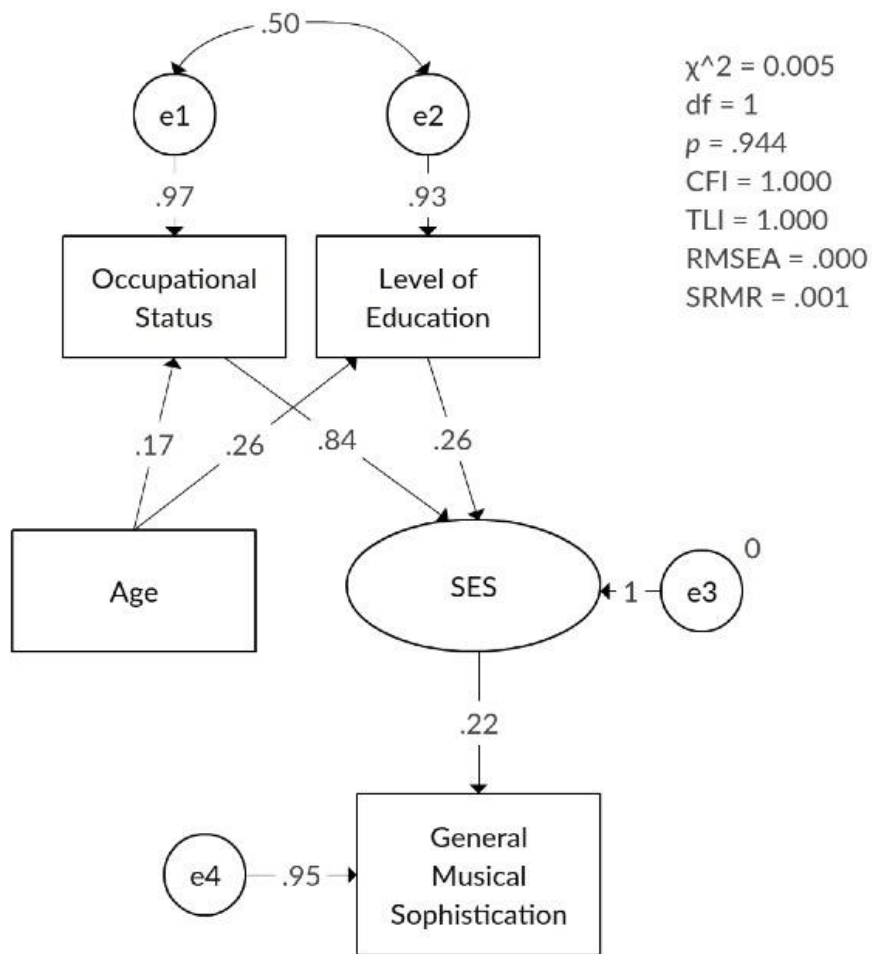
Figure 4. Structural equation model relating SES to the general factor of the Gold-MSI

inventory. Standardised factor loadings are reported.

References

Abell, N., Springer, D. W., & Kamata, A. (2009). *Developing and validating rapid assessment instruments*. New York: Oxford University Press.

Allen, M., Glen, J. C., Müllensiefen, D., Schwarzkopf, D. S., Fardo, F., Frank, D., . . . Rees, G. (2017). Metacognitive ability correlates with hippocampal and prefontal microstructure. *NeuroImage, 149*, 415-423. doi:10.1016/j.neuroimage.2017.02.008

Baker, D. J., Ventura, J., Calamia, M., Shanahan, D., & Elliott, E. M. (2018). Examining musical sophistication: A replication and theoretical commentary on the Goldsmiths Musical Sophistication Index. *Musicæ Scientiæ*. Advance online publication..

Bauer, A.-K. R., Kreutz, G., & Herrmann, C. S. (2015). Individual musical tempo preference correlates with EEG beta rhythm. *Psychophysiology, 52*(4), 600-604. doi:10.1111/psyp.12375

Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. New York: Routledge.

Bentley, A. (1966). *Measures of musical abilities*. London, UK: George A. Harrap.

Bentley, A. (1969). Measurement and development of musical abilities. *Journal of Research in Music Education, 17*, 41-46. doi:10.2307/3344182

Billroth, T. (1895). *Wer ist musikalisch?* Berlin, Germany: Paetel.

Bouwer, F. L., Werner, C. M., Knetemann, M., & Honing, H. (2016). Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm. *Neuropsychologia, 85*, 80-90. doi:10.1016/j.neuropsychologia.2016.02.018

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: The Guilford Press.

Chao, H.-C. (2013). *Arts and humanities of sixth graders: music appreciation computer quiz questions preparation* (master's thesis). National Taichung University of Education, Taichung, Taiwan.

Chen, P.-H., Lin, Y.-L., Hsu, C.-L., & Yen, Y.-C. (2007). *The development of multidimensional computerized adaptive testing for multiple intelligence*. Paper presented at the IMPS 2007, the 15th International and 72nd Annual Meeting for the Psychometric Society, Tokyo, Japan.

Chen, Y.-T. (2003). *Development of music aptitude tests for primary school students* (master's thesis). National Taichung University of Education, Taichung, Taiwan.

Chou, T.-C. R. (2006). Analysis of reliability and validity of Eight Multiple Intelligences Questionnaire. *Bulletin of Educational Psychology, 37*(3), 215-229. doi:10.6251/BEP.20051130

Degrave, P., & Dedonder, J. (2018). A French translation of the Goldsmiths Musical Sophistication Index, an instrument to assess self-reported musical skills, abilities and behaviours. *Journal of New Music Research*. Advance online publication. doi:10.1080/09298215.2018.1499779

DeMars, C. E. (2018). Classical test theory and item response theory. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 49–73). Hoboken, NJ: Wiley Blackwell.

Farrugia, N., Jakubowski, K., Rhodri, C., & Stewart, L. (2015). Tunes stuck in your brain: The frequency and affective evaluation of involuntary musical imagery correlate with cortical structure. *Consciousness and Cognition, 35*, 66-77.

Fiedler, D., & Müllensiefen, D. (2015). Validierung des Gold-MSI-Fragebogens zur Messung Musikalischer Erfahrenheit von Schülerinnen und Schülern der Sekundarstufen an allgemeinbildenden Schulen [Validation of the Gold-MSI questionnaire for measuring

musical sophistication of pupils at secondary level in general-education schools]. *Musikpädagogische Forschung, 36*, 199-220.

Fiedler, D., & Müllensiefen, D. (2017). Musical self-concept and musical sophistication influence the development of interest in music as a school subject. An empirical long-term study of German students at secondary schools. *Research in Music Education, 38*, 217-235.

Gardner, H. (1985). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gembris, H. (1997). Historical phases in the definition of musicality. *Psychomusicology: Music, Mind & Brain, 16*, 17-25. doi:10.1037/h0094070

German Psychological Society. (2016). *Berufsethische Richtlinien* [Guidelines for professional ethics]. Retrieved from https://www.dgps.de/fileadmin/documents/Empfehlungen/berufsethische_richtlinien_dgps.pdf

Gordon, E. E. (1965). *Musical aptitude profile*. Boston, MA: Houghton Mifflin.

Gordon, E. E. (1986). *The nature, description, measurement, and evaluation of music aptitudes*. Chicago, IL: GIA Publications.

Gordon, E. E. (1989). *Manual for the Advanced Measures of Music Audiation*. Chicago, IL: G. I. A. Pubilcations.

Hallam, S. (2010). 21st century conceptions of musical ability. *Psychology of Music, 38*(3), 308-330. doi:10.1177/0305735609351922

Hallam, S. (2016). Musicality. In G. E. McPherson (Ed.), *The child as musician: A handbook of musical development* (2nd ed., pp. 67-80). Oxford, UK: Oxford University Press.

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

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1995). Aging and experience in the recognition of musical transpositions. *Psychology and Aging, 10*(3), 325-342. doi:10.1037/0882-7974.10.3.325

Hanson, J. (2019). Meta-analytic evidence of the criterion validity of Gordon's music aptitude tests in published music education research. *Journal of Research in Music Education*. Advance online publication.

Harrison, P. M. C., Musil, J. J., & Müllensiefen, D. (2016). Modelling melodic discrimination tests: Desciptive and explanatory approaches. *Journal of New Music Research*, *45*(3), 265-280. doi:10.1080/09298215.2016.1197953

Ho, M.-h. R., Stark, S., & Chernyshenko, O. (2012). Graphical representation of structural equation models using path diagrams. In R. H. Hoyle (Ed.), *Handbook of structural equation modelling*. New York: The Guilford Press.

Ho, Y.-C. (2004). *The development of a music achievement test for junior high school students*. Unpublished master's thesis, National Taiwan University of Science and Technology, Taipei, Taiwan.

Hwang, Y.-J. (2009). The suitability of the New International Occupational Scale in Taiwan's educational research: Consideration of indigenization and internationalization. *Journal of Research in Education Sciences, 54*(3), 1-27.

International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (2nd ed.). Retrieved from https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf

Iversen, J. R., & Patel, A. D. (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population *Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval* (pp. 311-318). Queen Mary University of London, Londen, UK.

Jakubowski, K., & Müllensiefen, D. (2013). The influence of music-elicited emotions and relative pitch on absolute pitch memory for familiar melodies. *Quarterly Journal of Experimental Psychology, 66*(7), 1259-1267. doi:10.1080/17470218.2013.803136

Jansen, M., Lüdtke, O., & Schroeders, U. (2016). Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology, 46*, 116–127.

Karma, K. (2007). Musical aptitude definition and measure validation: Ecological validity can endanger the construct validity of musical aptitude tests. *Psychomusicology: A Journal of Research in Music Cognition, 19*(2), 79-90. doi:10.1037/h0094033

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). New York: Routledge.

Komsta, L., & Novomestky, F. (2015). moments: Moments, cumulants, skewness, kurtosis and related tests (R package version 0.14). Retrieved from https://CRAN.R-project.org/package=moments

Kopiez, R., & Lee, J. I. (2008). Towards a general model of skills involved in sight reading music. *Music Education Research, 10*(1), 41-62.

Kopiez, R., Wolf, A., Platz, F., & Mons, J. (2016). Replacing the orchestra? – The discernibility of sample library and live orchestra sounds. *PLoS ONE, 11*(7), e0158324. doi:10.1371/journal. pone.0158324

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal, 6*(2), 151-162.

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review, 25*, 120–148.

Lau, S., & Roeser, R. W. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment, 8*(2), 139–162.

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PLoS ONE, 7*(12), e52508. doi:10.1371/journal.pone.0052508

Leiner, D. J. (2014). Sosci Survey (Version 2.6.00-i) [Computer software]. Available at http://www.soscisurvey.com

Levitin, D. J. (2012). What does it mean to be musical? *Neuron, 73*(4), 633-637. doi:10.1016/j.neuron.2012.01.017

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764-766. doi:10.1016/j.jesp.2013.03.013

Lima, C. F., Correia, A. I., Müllensiefen, D., & Castro, S. L. (2018). Goldsmiths Musical Sophistication Index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality and music preferences. *Psychology of Music*. Advance online publication. doi:10.1177/0305735618801997

Lin, W.-C. (2013). *Fifth grade arts and humanities computerized music achievement test*. Unpublished master's thesis, National Taichung University of Education, Taichung, Taiwan.

Lu, C.-M., Chen, S.-M., & Chen, L.-C. (1981). A study on the musical aptitude of Chinese student. *Bulletin of Educational Psychology, 14*, 149-160. doi:10.6251/BEP.19810601.10

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics:*

*A festschrift for Roderick P. McDonald* (pp. 275-340). Mahwah, NJ: Lawrence Erlbaum Associates.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*(3), 471-491. doi:10.1037/a0019227

Martins, M. D., Gingras, B., Puig-Waldmueller, E., & Fitch, W. T. (2017). Cognitive representation of "musical fractals": Processing hierarchy and recursion in the auditory domain. *Cognition, 161*, 31-45. doi:10.1016/j.cognition.2017.01.001

Matthews, T. E., Thibodeau, J. N. L., Gunther, B. P., & Penhune, V. B. (2016). The impact of instrument-specific musical training on rhythm perception and production. *Frontiers in Psychology, 7*(69). doi:10.3389/fpsyg.2016.00069

Michaelis, C. F. (1805). Über die Prüfung musikalischer Fähigkeiten [On the testing of musical abilities]. *Berlinische Musikalische Zeitung, 56-68*, 222-230.

Mota, G. (1997). Detecting young children's musical aptitude: A comparison between standardized measures of music aptitude and ecologically valid musical performances. *Bulletin for the Council of Research in Music Education, 133*, 89-94.

Müllensiefen, D., Gingras, B., Musil, J. 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

Müllensiefen, D., Gingras, B., Stewart, L., & Musil, J. J. (2013). *The Goldsmiths Musical Sophistication Index (Gold-MSI): Technical report and documentation v1.0*. Retrieved from http://luna.gold.ac.uk/media/Gold-MSIv10_Documentation.pdf

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings*. Unpublished doctoral dissertation, The Ohio State University. Retrieved from https://etd.ohiolink.edu/

Oreck, B. A., Owen, S. V., & Baum, S. M. (2003). Validity, reliability, and equity issues in an observational talent assessment process in the performing arts. *Journal for the Education of the Gifted, 27*(1), 62-94. doi:10.1177/016235320302700105

Pérez, E., Beltramino, C., & Cupani, M. (2003). Inventario de autoeficacia para inteligencias múltiples: Fundamentos teóricos y estudios psicométricos [Self-report inventory of multiple intelligences: Theoretical foundations and psychometric studies]. *Evaluar, 3*, 35-60.

Platz, F., Kopiez, R., Lehmann, A. C., Hasselhorn, J., Büring, M., Wolf, A., . . . Estrada Rodriguez, L. A. (2015, 17-22 August 2015). *Requiem for Edwin Gordon's "Advanced Measures of Music Audiation" (AMMA)? - A critical evaluation.* Paper presented at the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music, Manchester, UK.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future direction for psychological research. *Developmental Review, 41*, 71-90.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Revelle, W. (2017). psych: Procedures for personality and psychological research (Version 1.7.5). Evanston, IL: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. doi:10.18637/jss.v048.i02

RStudio Team. (2016). RStudio: Integrated development for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Rutkowski, J. (1996). The effectiveness of individual/small-group singing activities on kindergartners' use of singing voice and developmental music aptitude. *Journal of Research in Music Education, 44*(4), 353-368. doi:10.2307/3345447

Schaal, N. K., Bauer, A.-K. R., & Müllensiefen, D. (2014). Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musikalischer Erfahrenheit anhand einer deutschen Stichprobe [The Gold-MSI: Replication and validation of a questionnaire instrument for measuring musical sophistication, based on a German sample]. *Musicæ Scientiæ, 18*(4), 423-447. doi:10.1177/1029864914541851

Schaal, N. K., Krause, V., Lange, K., Banissy, M. J., Williamson, V. J., & Pollok, B. (2015). Pitch memory in nonmusicians and musicians: Revealing functional differences using transcranial direct current stimulation. *Cerebral Cortex, 25*(9), 2774-2782. doi:10.1093/cercor/bhu075

Seashore, C. E. (1919). *The psychology of musical talent*. Boston, MA: Silver, Burdett and Company.

Seashore, C. E., Lewis, D., & Saetveit, J. G. (1960). *A second revision of the manual of instructions and intrepretations for the Seashore Measure of Musical Talents (1939 revision)*. New York: The Psychological Corporation.

Sorkin, R. D. (1999). Spreadsheet signal detection. *Behavior Research Methods, Intruments, & Computers, 31*(1), 46-54. doi:10.3758/BF03207691

Stevens, C. J., & Byron, T. (2016). Universals in music processing. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 19–31). Oxford, UK: Oxford University Press.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Perason.

Tran, T. V., Nguyen, T. H., & Chan, K. T. (2017). *Developing cross-cultural measurement in social work research and evaluation* (2nd ed.). New York: Oxford University Press.

Ullén, F., Mosing, M. A., Holm, L., Eriksson, H., & Madison, G. (2014). Psychometric properties and heritability of a new online test for musicality, the Swedish Musical Discrimination Test. *Personality & Individual Differences, 63*, 87-93. doi:10.1016/j.paid.2014.01.057

UNESCO Institute for Statistics. (2012). *International Standard Classification of Education ISCED 2011*. Retrieved from http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf doi:10.15220/978-92-9189-123-8-en

UNESCO Institute for Statistics. (2015). *ISCED 2011 Operational manual: Guidelines for classifying national education programmes and related qualifications*. Retrieved from http://dx.doi.org/10.1787/9789264228368-en doi:10.15220/978-92-9189-174-0-en

van de Vijver, F. J. R., & Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analytic procedures. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 17-45). New York: Cambridge University Press.

Wen, Y.-C., & Tsai, C.-G. (2017). The effect of harmonization on cortical magnetic responses evoked by music of rapidly changing tonalitites. *Psychology of Music, 45*(1), 22-35. doi:10.1177/0305735616639386

Wing, H. D. (1939). *Standardised tests of musical intelligence*. The Mere, UK: National Foundation for Educational Research.

Wing, H. D. (1961). *Standardised tests of musical intelligence*. The Mere, UK: National Foundation for Educational Research.

Wolf, A., & Kopiez, R. (2018). Development and validation of the Musical Ear Training Assessment (META). *Journal of Research in Music Education, 66*(1), 53-70. doi:10.1177/0022429418754845

Wolf, A., Kopiez, R., & Platz, F. (2018). Thinking in music: An objective measure of notation-evoked sound imagry in musicians. *Psychomusicology: Music, Mind and Brain, 28*(4), 209-221.

Wolf, A., Kopiez, R., Platz, F., Lin, H.-R., & Mütze, H. (2018). Tendency towards the average? The aesthetic evaluation of a quantitatively average music performance: A successful replication of Repp's (1997) study. *Music Perception, 36*(1), 98-108. doi:10.1525/mp.2018.36.1.98

Wu, W.-T. (2008). *Chinese version of multiple intelligences developmental assessment scales form-C, revised from C. Branton Shearer's Multiple Intelligences Developmental Assessment Scales.* Taipei, Taiwan: Psychological.

Yang, H., Ma, W., Gong, D., Hu, J., & Yao, D. (2014). A longitudinal study on children's music training experience and academic development. *Scientific Reports, 4*, 5854. doi:10.1038/srep05854

Yeh, Y.-C., Peng, Y.-Y., & Yeh, Y.-H. (2009). The development of "the Inventory of Preschoolers' Multiple Intelligences". *Early Childhood Education, 296*, 2-21. doi:10.6367/ECE.200912.0002