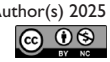
Original Article

ISME INTERNATIONAL SOCIETY FOR MUSIC EDUCATION

International Journal of Music Education
1–19


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/02557614251393885
journals.sagepub.com/home/ijm

Sage

Chinese Primary School Students' Singing Behaviour by Age, Sex and Socio-economic Status

Can Lu[1] and Graham F Welch[1]

Abstract

Limited studies have paid attention to describing Chinese Primary school students' singing behaviour. Consequently, to understand Chinese Primary school students' singing behaviour, singing performances were collected from $N = 1{,}193$ children aged from 6+ to 11+, drawn from six schools in Hunan Province, China. All participants sang three songs: *Twinkle, Twinkle, Little Donkey* (a Chinese nursery song), and *Happy Birthday*, with vocal products analysed against two existing rating scales: the Singing Voice Development Measure (*SVDM*) scale and the Vocal Pitch-Matching Development *(VPMD*) scale. Older participants and girls tended to have more developed singing behaviour than younger participants and boys. Urban children and children from higher-income families tended to show better singing behaviour. However, the differences by sex, geographic location and income were reduced for the oldest participants. The study might help future and current music teachers who are teaching the music curriculum in Primary schools in China to understand the likely development of Primary school students' singing behaviour across different age groups, sex and socio-economic statuses, in order to support development in a more nuanced way which is differentiated by likely need.

Keywords

children's singing, China, age, sex, socio-economic status

Background and Literature Review

Singing has been observed to be an essential activity in music lessons in many Primary schools across the world, such as in the USA (Nichols, 2016), China (Guo, 1999), England (Welch et al., 2009a) and Portugal (Pereira & Rodrigues, 2019). Overall, in some countries, there has been a growing body of research on how children's singing develops with age and experience, such as in

[1]University College London (UCL) Institute of Education, London, UK

Corresponding author:
Graham F Welch, University College London (UCL) Institute of Education, London, WC1E 6BT, UK.
Email: graham.welch@ucl.ac.uk

the UK (e.g., Welch et al., 2009b, 2010), the USA (e.g., Moore, 1991; J. Rutkowski, 1996, 2018), Australia (Barrett, et al., 2020), mainland Europe (e.g., Mecke & Sundberg, 2010), Japan (Welch & Murao, 1994; Azechi, 2016) and Latin America (IIari, 2006). However, published research data on children's singing in China is limited. Exceptions are a new national research project that includes music (The Monitoring Centre for the Compulsory Education, Ministry of Education of the People's Republic of China, 2020), and a large Government-funded study led by Guo (1999) about music education in Primary and Secondary schools in China.

The Chinese national music test in 2016 and 2018 randomly selected Grade 4 and Grade 8 students from varied economic backgrounds in each Province in China. This was to provide a general picture of Primary and Secondary school students' musical ability at both national and regional levels, and thereby by implication infer the quality of the teaching received by these students (The Monitoring Centre for the Compulsory Education, Ministry of Education of the People's Republic of China, 2018, 2020). The national studies defined the general singing performance by the mean scores of five perspectives (vocal pitch-matching accuracy, rhythm, fluency, expression and clarity of lyrics), as measured by a four-rating scale (1–4) for each perspective. Both national tests reported that around 80% of Grade 4 students' general singing behaviour was the second rating ($\geqslant$2). However, the mean scores for the vocal pitch-matching ability for a target song in both testing years (M = 77.4% in 2016 and M = 73.0% in 2018), were the lowest ratings among the five musical perspectives. These findings imply that vocal pitch-matching accuracy for songs likely required more attention than the other four perspectives for Chinese Primary school students. Furthermore, while the national tests provided an overview of Primary school students' singing behaviour in China, these did not involve students from other compulsory school Grades (ages) for reasons of cost and the large population size in China.

Previous studies conducted in Western countries have also concluded that many children may have a problem with singing in-tune (e.g., Welch, 1979a, 1979b; Çenberci & Kalkanoğlu, 2023; Pereira & Rodrigues, 2019), although an alternate interpretation of the data is that singing is a developmental process and that in-tune singing emerges with appropriate experience (Welch, 1985, 2015). The literature also suggests that development could be slower or even decline if appropriate support or engagement is not maintained (Demorest & Pfordresher, 2015; Welch et al., 2010). Many previous studies have reported children's vocal pitch-matching accuracy for songs in varied settings. Variables include children's age, sex, singing individually or within a group and singing with or without text (e.g., Estis et al., 2011; Leighton & Lamont, 2006; Welch et al., 2009a, 2009b). Similarly, many previous studies have reported children's vocal register use for songs with variables such as age (e.g., Mang, 2006; Pereira & Rodrigues, 2019; Welch et al., 2009a) and sex (Dansereau, 2005; Mang, 2006; Welch et al., 2009a).

Mang (2006) and Welch et al. (2009b) defined singing behaviour by both vocal register use and vocal pitch-matching accuracy. These authors combined measures of these two singing perspectives to describe the singing behaviour of Primary school children. Both Mang and Welch et al. studies applied J. Rutkowski's (1996, 2015, 2018) Singing Voice Development Measure (*SVDM*) scale to assess vocal register use, alongside Welch's Vocal Pitch-Matching Development (*VPMD*) scale (Welch, 1998) to measure vocal pitch-matching accuracy for song(s). Mang (2006) and Welch et al. (2009b) reported a positively significant correlation between vocal register use and vocal pitch-matching accuracy for songs. The following section compares the results of these two studies.

Firstly, in terms of the age difference, although the mean score for older children in Mang's (2006) study was slightly higher than that for younger children, the age difference in the overall singing behaviour was not statistically significant. However, the mean scores of older children were statistically significantly higher than those of younger children in Welch et al.'s (2009b)

study. It is uncertain whether the different results by age between these two studies were because of the different complexity levels of target song(s),[1] the wide difference in numbers of participants, or the diverse ranges of age groups.

Secondly, in terms of the sex variable, both Mang (2006) and Welch et al. (2009b) agreed that girls' singing behaviour generally was better than that of boys. When the National Singing Programme *Sing Up* provided professional singing training for boys and girls in the study of Welch et al. (2009b), they still observed a sex difference in their singing behaviour, with a continued developmental bias towards girls (Welch et al, 2010).

Thirdly, regarding the age and sex interaction effect on children's singing behaviour, Mang (2006) reported no statistically significant impact. However, Welch et al. (2009b) reported a statistically significant difference. In their study, the sex difference for younger English Primary school children was smaller. It increased for the middle age group of Primary school children before it decreased for the oldest. It was unsure whether the different results reported in the two above studies were because of diverse complexities of target songs, numbers of participants or age groups. Mang's study included only ages 7 and 9, while Welch et al.'s (2009b) study included ages 6 to 11.

Furthermore, Welch et al. (2009b) reported that Primary school children in England who received *Sing Up* experience generally showed better singing ability than peers who did not receive such an enrichment programme. While the baseline of the initial *Sing Up* data found a significant age difference between Year 2 and Year 3 (Welch et al., 2009a), the age difference disappeared in a longitudinal data comparison for the same children 1 year later. We believed that the impact of *Sing Up* training contributed to the non-age difference in the second data collection, which for Year 2 children dramatically improved their vocal pitch-matching accuracy of the song. This suggested a positive correlation between a positive singing experience and an improvement in rated singing ability (*cf.* Pereira & Rodrigues, 2019).

For English children who came from higher indices of multiple deprivation (IMD) areas with *Sing Up* experience, Welch et al. (2009b) reported that their singing behaviour generally was more advanced than children from lower IMD areas without *Sing Up* experience. This suggests that socioeconomic status need not be a barrier if schools provided appropriate singing support. Instead of the impact of socioeconomic influence, Welch et al. (2009b) suggested that the school management's attitude towards singing might be the key factor in the degree of children's singing behaviour development.

However, while the English study by Welch et al. (2009a, 2009b) provided an overview of Primary school students' singing behaviour, it is unsure whether singing data in another country with big rural and urban income differences, such as China, can apply the findings.

Limited studies have investigated Chinese Primary school students' singing behaviour, and some (e.g., Feng et al., 2013; Zhao, 2020a, 2020b) focused on kindergarten ages. Although the research led by Guo (1999) reportedly measured Chinese Primary school students' singing ability across 20 provinces, their assessment was based on the students' singing using a broad definition of counting the number of songs that students could sing from memory, which is not necessarily an indication of the quality or competence of the participant children's singing. Consequently, there is a need for a more detailed definition of singing behaviour, as well as development, if Chinese data are to be compared to other studies elsewhere in the world. The singing tests in the Chinese national tests (The Monitoring Centre for the Compulsory Education, Ministry of Education of the People's Republic of China, 2020) could provide an overview of Chinese Primary school students' singing abilities by selected Grade 4 students. However, they will not report on the development of singing behaviour from across the age range of Grades 1 to 6 for Chinese Primary school students.

Consequently, the current article reports on a range of findings from a recent Chinese study that aimed to explore Chinese children's singing behaviours and development according to age and sex,

and also how these might be influenced by socio-economic status. The current study uses vocal register use and vocal pitch-matching accuracy for songs as two key measures to define singing behaviour. It uses two variables, geographic location and inferred parents' income to identify socio-economic status. One application of the findings is that this could allow a more suitable match between appropriate pedagogy and curricula for singing and children's current competency levels. The research questions to be addressed in this article are: What is the general singing behaviour of song singing from the perspectives of vocal registers application and vocal pitch-matching accuracy for different songs, ages, sexes and socioeconomic groups for Chinese Primary school students?

Methodology

Participants

In 2017-2018, following university ethical approval, the current authors assessed $n=453$ children in Grade 2 (age 7+) and Grade 4 (age 9+) in the six research schools from a southern-middle area of China, Hunan Province[2] (see Figure 1). In addition, in 2018-2019, we assessed another $n=740$ participants from Grades 1, 2, 4 and 6, to build a picture of children's singing behaviour across all Primary school grades. Across the 2 years, we recorded singing performances from the $N=1,193$ participants, drawn from Grades 1 to 6, aged between 6+ and 11+.

Table 1 shows the number of participants for each age group (by School year) in each visiting year. In total, $n=655$ boys[3] and $n=538$ girls participated. Overall, $n=810$ and $n=383$ participants came from rural and urban areas, respectively. Also, $n=610$ and $n=583$ participants were from lower- and higher-income families, respectively.

We selected the schools based on seeking to contrast those in rural and urban areas, drawing initially on personal contacts, and then checking how they might match against the focus variables related to location, likely parental occupations and their interest in participating in the research. Four schools (Schools A–D) were in four rural areas – one school in each area, while two schools (Schools E and F) were in an urban area (see Table 2). The number of participant rural schools was greater than that of urban schools to allow for the size of an urban Primary school being usually greater. Students at the schools came from local areas. The degrees of remoteness from town centres for the four rural schools, such as a village surrounded by mountains (Schools A and B) or a town centre (Schools C and D), were diverse. We also recognised that, compared with students living in the villages, it might be easier for students who lived in the town centres to access music training or other related musical activities taking place in the town.

Furthermore, the authors explored the impact of unequal socioeconomic development within a geographic area on children's singing behaviour in China by relative analysis of parents' income at a school level. The authors defined income by inferring the parents' occupations from informal interviews with students, teachers and other school staff to create a general sense of the local area's economic development (official data on income was not available). For instance, in rural areas, if parents stayed at home and worked on the land, or went to the city to work in infrastructure construction, we categorised them as a group of rural parents with relatively lower income (i.e., children at Schools A and B). If rural parents owned a shop in a local town centre, we categorised them as a group of rural parents with relatively higher income (*cf.* children at Schools C and D). In urban areas, if parents worked in a factory, we categorised them as a group of urban parents with relatively lower income (*cf.* children at School E). If urban parents were professionals, such as doctors, we recognised them as higher-income parents (*cf.* children at School F). Although the description

Figure 1. Location of Hunan province in China (vectorstock.com/28773929).

Table 1. The Number of Participants by Age in Whole Years Across Two Visiting Years (2017–2019) $N = 1{,}193$ (453 + 740).

Age in whole years	Participants in 2017–2018	Participants in 2018–2019
6+		76
7+	99	158
8+	124	158
9+	123	144
10+	105	116
11+	2	88
Total	453	740

might not be suitable for every participant within a school, informal conversations with teachers and pupils suggested that these inferences were appropriate at a school level.

In terms of the difference in music resources in terms of staffing (see Table 2), although each of the four rural participating schools (Schools A–D) employed music specialists, six of the seven teachers were not able to focus solely on teaching the music curriculum because part of the teachers' assigned responsibility was to cover other subjects as well. In contrast, all the music teachers

Table 2. Information on Each of the Participating Schools in Hunan Province, China.

School	Rural/urban	Lower/higher income	School size from grades 1 to 6	No. of music specialists	No. of music specialists teaching music only
A	Rural	Lower	≈120	1	0
B	Rural	Lower	≈250	2	0
C	Rural	Higher	≈260	2	0
D	Rural	Higher	≈500	2	1
E	Urban	Lower	≈600	3	3
F	Urban	Higher	≈700	4	4

in the two urban participating schools (Schools E and F) were music specialists who were only expected to teach music. Also, compared with the four rural participating schools, music teachers in the two urban schools used the music classrooms regularly. One difference in the musical environment between the two urban schools was that only School F had an orchestra.

Singing Tasks

To test Primary school children's ability in song singing, we asked participants to sing three short, familiar criterion songs, *Twinkle, Twinkle*, *Little Donkey* (a typical Chinese nursery song) (see the tune in Figure 2), and *Happy Birthday*, with lyrics provided. These songs were nursery songs, so it is assumed that Primary school participants already knew these target songs.

Although we thought *Twinkle, Twinkle*, and *Happy Birthday* to be two typical Western children's songs but also well-known in China, we selected them to be able to compare their singing data between Chinese participants in the current study and English participants evaluated in the *Sing Up* programme (Lu & Welch, 2025). People studying abroad introduced *Twinkle, Twinkle*, and *Happy Birthday* to China from overseas during the 20th century. They kept the melodies of the songs, but they translated lyrics into Chinese. Previous studies agreed that children learned the words of a song before melodies (including rhythm, contour, pitch and key stability) (Davidson et al., 1981; Welch et al., 1996). However, it seems that the text-melody relationship has a limited impact on children's vocal pitch-matching accuracy of song singing (Chen-Hafteck, 1999). In addition, we choose the traditional Chinese song *Little Donkey* to explore any possible difference in singing when the place of origin of a song is different.

When combining musical elements for a target song, we thought *Twinkle, Twinkle* to be the easiest song to match, based on its narrow vocal range (a major sixth) and small musical intervals (⩽ perfect 5th). We judged the complexity of *Little Donkey* to be at a medium level across the three target songs (Lu, 2024) due to its wider vocal range (an octave), but small musical intervals (⩽ perfect 5th). On the other hand, we thought *Happy Birthday* as the most difficult song to sing because of its wide vocal range (an octave) and generally wide and more numerous musical intervals.

Measurement

We assessed all singing performances against two existing rating scales: (a) J. Rutkowski's (1990, 1996, 2015) *SVDM* scale measures of vocal register use; and (b) Welch's (1998) *VPMD* scale to assess vocal pitch-matching accuracy for songs (see Figure 3).

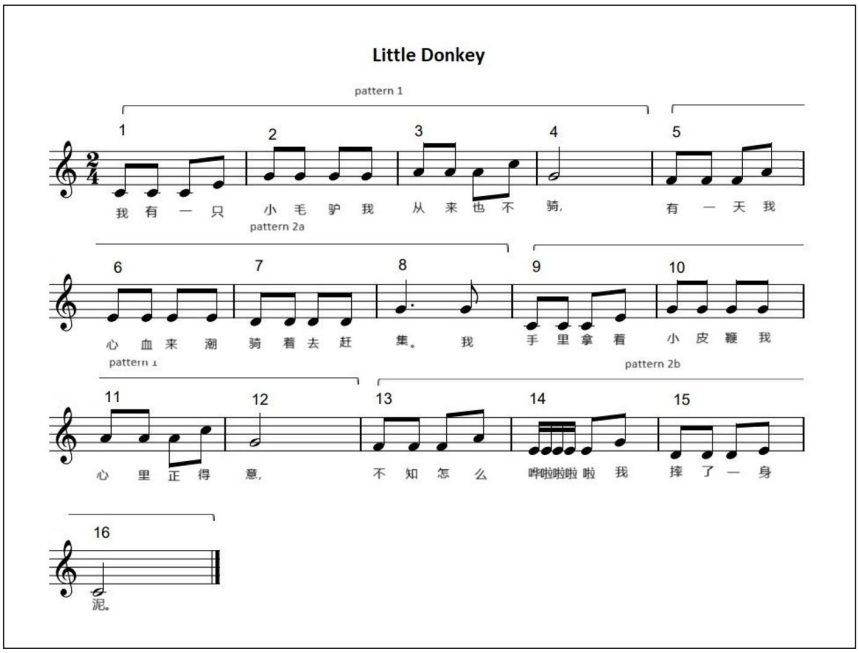


Figure 2. Melodic contour, pitches with patterns, and lyrics of *Little Donkey*.

Figure 3 shows that the latest version of the *SVDM* scale (J. Rutkowski, 1996, 2015, 2018) used specific register shifts and pitch ranges to define a particular type of vocal register used in its nine-rating scoring system. Many previous studies have applied the scale to test children's use of vocal registers in song singing, mainly in Western cultures, such as America (Guerrini, 2006; Levinowitz et al., 1998; J. Rutkowski, 1990, 2004; J. Rutkowski et al., 2002, 2007; J. Rutkowski & Miller, 2003), Asia (Hong Kong – H. Rutkowski & Chen-Hafteck, 2001), Europe (Portugal – Pereira & Rodrigues, 2019; Turkey – Çenberci & Kalkanoğlu, 2023) and England (Welch et al., 2009a).

However, in Mang's (2006) study, half of the Primary school participants were Cantonese monolinguals, suggesting that other cultures, including Asia, could apply the *SVDM* scale as earlier evidenced in the H. Rutkowski and Chen-Haftek (2001) study. High consistency and interrater reliability coefficients were stated for the *SVDM* scale (see J. Rutkowski, 1990, p. 90; 1996, p. 358; 2018; Guerrini, 2006). However, Levinowitz et al. (1998) reported lower reliability for Grade 6 students.

The *VPMD* scale is arranged with a hierarchical order from (i) singing like speaking, (ii) evidence of macro-changes in sung melodic contour, (iii) greater singing accuracy, but still shifts in the key centre and (iv) singing completely in tune (see Figure 3). Many previous studies used the scale to assess the vocal pitch-matching accuracy of song singing for children from diverse cultures, such as Asia (e.g., Mang, 2006) and Western cultures (e.g., Welch et al., 2009a, 2009b).

We applied each rating of the *SVDM* and *VPDM* scales when measuring participants' vocal register use and vocal pitch-matching accuracy for the three target songs. The first author measured all singing performances by both existing rating scales three times to ensure more consistency of judgment in the third and final iteration. To further improve the reliability of raw singing data measured by the two scales, two authors together measured the singing performances of $n = 20$ randomly selected participants. Kendall's tau statistics, which are designed to test interrater

Rutkowski (1996, 2015) *Singing Voice Development Measure (SVDM)*

1 "Presinger" does not sing, but chants the song text.

1.5 "Inconsistent Speaking-Range Singer" sometimes chants, sometimes sustains tones and exhibits some sensitivity to pitch but remains in the speaking voice range (usually A3 to C4).

2 "Speaking-Range Singer" sustains tones and exhibits some sensitivity to pitch but remains in the speaking-voice range (usually A3 to C4).

2.5 "Inconsistent Limited-Range Singer" wavers between speaking and singing voice and uses a limited range when in singing voice (usually up to F4).

3 "Limited-Range Singer" exhibits use of limited singing range (usually D4 to F4).

3.5 "Inconsistent Initial Range Singer" sometimes only exhibits use of limited singing range, but other times exhibits use of initial singing range (usually D4 to A4)

4 "Initial Range Singer" exhibits use of initial singing range (usually D4 to A4).

4.5 "Inconsistent Singer" sometimes only exhibits use of initial singing range, but other times exhibits use of extended singing range (sings beyond the register lift: B4-flat and above).

5 "Singer" exhibits use of extended singing range (sings beyond the register lift: B4-flat and above).

Note. A4 = 440 Hz

Welch (1998) A revised model of *vocal pitch-matching development (VPMD)*

1 The words of the song appear to be the initial centre of interest rather than the melody, singing is often described as 'chant-like', employing a restricted pitch range and melodic phrases. In infant vocal pitch exploration, descending patterns predominate.

2 There is a growing awareness that vocal pitch can be a conscious process and that changes in vocal pitch are controllable. Sung melodic outline begins to follow the general (macro) contours of the target melody or key constituent phrases. Tonality is essentially phrase based. Self-invented and 'schematic' songs 'borrow' elements from the child's musical culture. Vocal pitch range used in 'song' singing expands.

3 Melodic shape and ıntervals are mostly accurate, but some changes in tonality may occur, perhaps linked to inappropriate singing register usage. Overall, however, the number of different reference pitches is much reduced.

4 No significant melodic or pitch errors in relation to relatively simple songs from the singers' musical culture.

Figure 3. The singing voice development measure (SVDM) and the vocal pitch-matching accuracy (VPMD), based on J. Rutkowski (1996, 2015) and Welch (1998).

reliability (Kinnear & Gray, 2009), showed that the agreement on raw singing scores measured by each of the two rating scales for each target song was relatively high (for Rutkowski's *SVDM – Twinkle, Twinkle*: Kendall's tau = 0.801; *Little Donkey*: Kendall's tau = 0.921; *Happy Birthday*: Kendall's tau = 0.736; for Welch's *VPDM: Twinkle, Twinkle*: Kendall's tau = 0.743; *Little Donkey*: Kendall's tau = 0.606; *Happy Birthday*: Kendall's tau = 0.855). Overall, the test indicated that the first author's scores, derived from ratings of singing performances based on two existing scales, demonstrated reliability.

Process

The University College London's ethical research procedures approved the research in March 2017. Before the date of data collection, three tasks were undertaken: Firstly, we prepared audio recording models for the three target songs without accompaniment, sung by a Chinese adult female, to ensure that each of the six participating schools had the same basic learning materials. We then sent these recording models to participating schools around 2 weeks before the authors' initial visits. The authors asked each teacher to rehearse the three target songs with the participant students a few times (around 3–4 times for each target song) to help the children become more familiar with the target songs and to ensure similar familiarity between participants.

Then, we prepared an information letter and ethics form as part of the university's standard ethical requirements. We wrote initially in English for ethical approval, then translated for fieldwork use in China, and distributed to and signed by the participant music teachers. We informed participating children about the process by their teachers who gave proxy permission from parents and carers under the schools' local ethical practices.

To relax the participant children, the first author met them in their classroom during recess. We invited all children in the focus classes to participate, with no teacher selection involved, thus enabling a wide range of developmental profiles to be included. Furthermore, this could avoid any negative feeling for a student of being ignored (Welch et al., 2009b). We told participants that they could withdraw from the test at any time for any or no reason in line with British Educational Research Association (BERA) ethical guidelines (2011[4]).

On the day of the data collection, we recorded all singing performances in a quiet but familiar room within their school to ensure a good quality audio recording and to reduce the likelihood of stress on participants when singing in front of a group of classmates.

A group of around five to seven participants went to the testing room. When an individual participant sang, the other participants in the group stayed in the testing room but were asked to keep quiet. This helped them to be familiar with the process. Each child sang the three songs in the same order, based on a consideration of the musical material becoming progressively more complex from *Twinkle, Twinkle* to *Little Donkey* to *Happy Birthday*. To avoid any possible negative influence of forgotten lyrics on the scores for song singing, we told participant children that they could see a copy of the lyrics of the songs written on a sheet of paper if they wished. We encouraged them to sing the three target songs at a normal speed in a natural manner with no starting pitch to provide a better opportunity for them to sing within a comfortable singing range. Previous research had reported that participants matched the pitches of songs more accurately when they sang in a key of their choice (Welch et al., 2009b). This was also the protocol followed by Mang (2006). We audio recorded all singing performances by two recording machines, an iPhone 6s (model: iOS 14.8) and a Philips voice tracer (model: VTR5000). As revealed by the earlier pilot study (Lu et al., 2019), participants seemed to be more relaxed when we took an audio recording compared with a video recording.

Data Analysis

We analysed the singing performance of each participant. When assessing the use of vocal registers by the *SVDM* scale, it may be difficult to determine vocal register shifts when no starting pitch is given. However, in the current study we explored the vocally matched pitches–including the chosen starting pitch – for each participant by reference to a virtual keyboard. This helped the authors to identify the starting pitch, register lift points and singing range for each song. We perceived inferred sung register shifts by noting changes in the sung timbral quality. These changes helped to determine the rater's perceived measure on the *SVDM* scale.

All raw data measured by both rating scales was seen to be normally distributed (see Lu, 2024, Appendix). Consequently, we treated scores measured by the two ratings as scale data (*cf.* Mang, 2006; Welch et al., 2009b).

The authors conducted a Pearson correlation coefficient to assess the relationship for scores measured between the *SVDM* scale and the *VPMD* scale. Results showed a strong positive correlation between the two dependent variables across the three songs ($r = .834$, $n = 3579$, $p < .001$). Consequently, we analysed the subsequent data based on the combined data from the two rating scales, taking the means of two existing scales for the three target songs.

The calculation of this normalised singing score (NSS) took a mean score of vocal register use (*cf.* Rutkowski) and a mean score of vocal pitch-matching (*cf.* Welch) for the three target songs, and converted into a normalised scale out of 100, initially calculated out of 1.00. For instance, if a female participant's scores of vocal registers' application measured by the *SVDM* scale (a full score is 5) of the three criterion songs were 4, 4.5 and 4, her normalised scores of vocal registers' application of each song were 0.80(4/5), 0.90(4.5/5) and 0.80(4/5). In this case, her normalised score of vocal registers' application of the three criterion songs was 0.83 [(0.80 + 0.90 + 0.80)/3]. Similarly, if her scores of vocal pitch-matching accuracy of the three songs analysed by the *VPMD* scale (a full score is 4) were 4, 4 and 3, her normalised scores of vocal pitch-matching accuracy of each song were 1.00(4/4), 1.00(4/4) and 0.75(3/4), and her normalised score of vocal pitch-matching accuracy of song singing was calculated as 0.92 [(1.00 + 1.00 + 0.75)/3]. Her final normalised score was 0.88 [(0.83 + 0.92)/2], noted as 88 for computer calculation.

We entered all data into Excel files initially and uploaded them into SPSS files later. All original audio data were saved with permission from the music teachers. All this information was protected in an online drive by password requirement.

Results

One-way ANOVA revealed that participant children's normalised singing scores (NSS) increased statistically significantly with age $F(5, 3573) = 63.20$, $p < .001$ (see Figure 4(a)). A two-way ANOVA test revealed that there was also a statistically significant impact of an age and sex interaction effect on participant children's NSS $F(5, 3567) = 2.85$, $p = .014$. In general, girls' singing ability was significantly more developed than that of boys across varied ages. However, as Figure 4(b) shows, the sex difference in the NSS was smaller for younger participants (ages 6 and 7); this difference increased during the middle of the Primary school period (ages 8 and 10), but decreased at the end of the Primary school period (age 11).

Similarly, any possible age and geographic location and the age and income interaction effects on the NSS were explored. Both interaction effects were statistically significant. For the age and geographic location interaction effect $F(4, 3568) = 2.88$, $p = .021$ (see Figure 4(c)), and for the age and income interaction effect $F(5, 3567) = 11.16$, $p < .001$ (see Figure 4(d)). Urban participants (those attending urban schools) and participants from higher-income families tended to have better

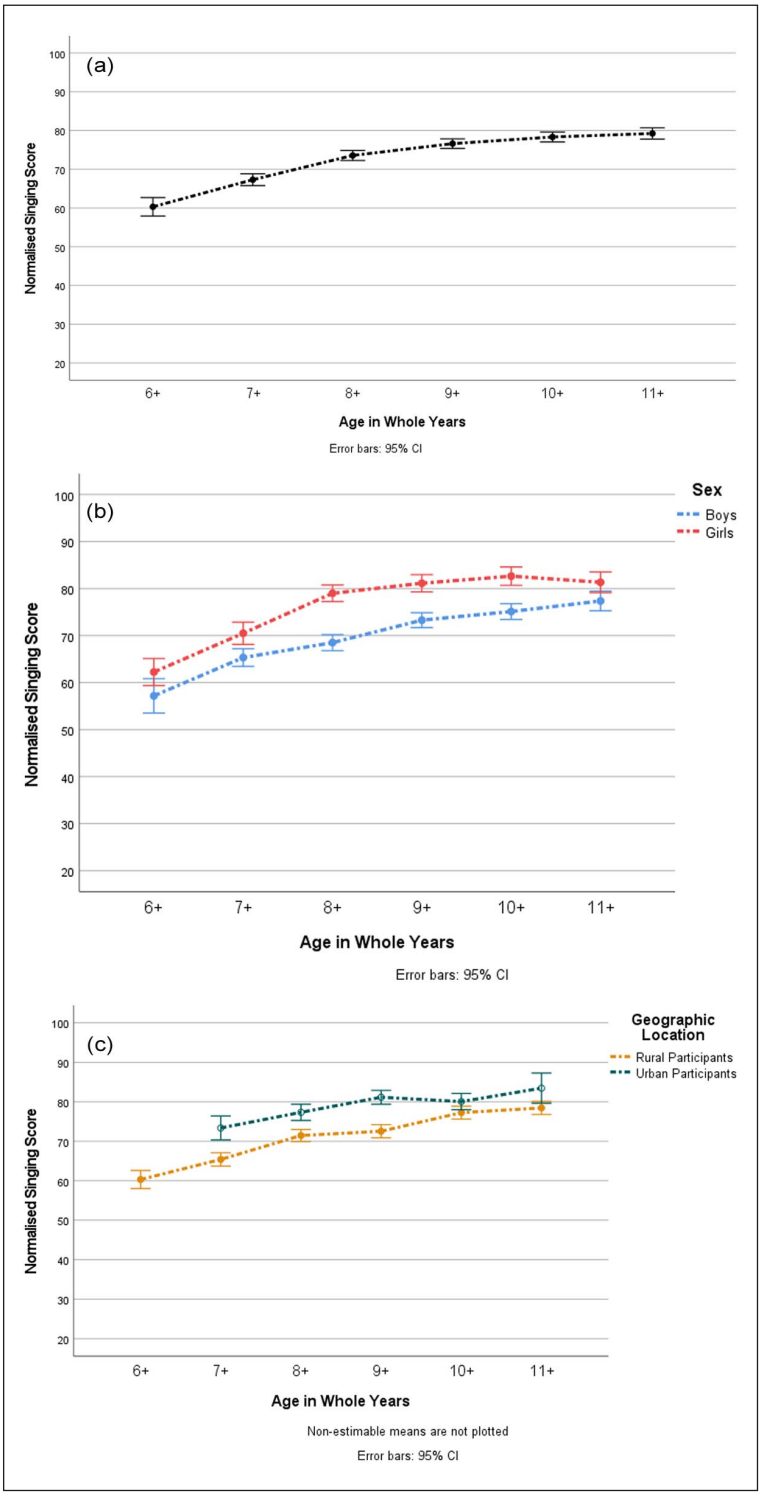
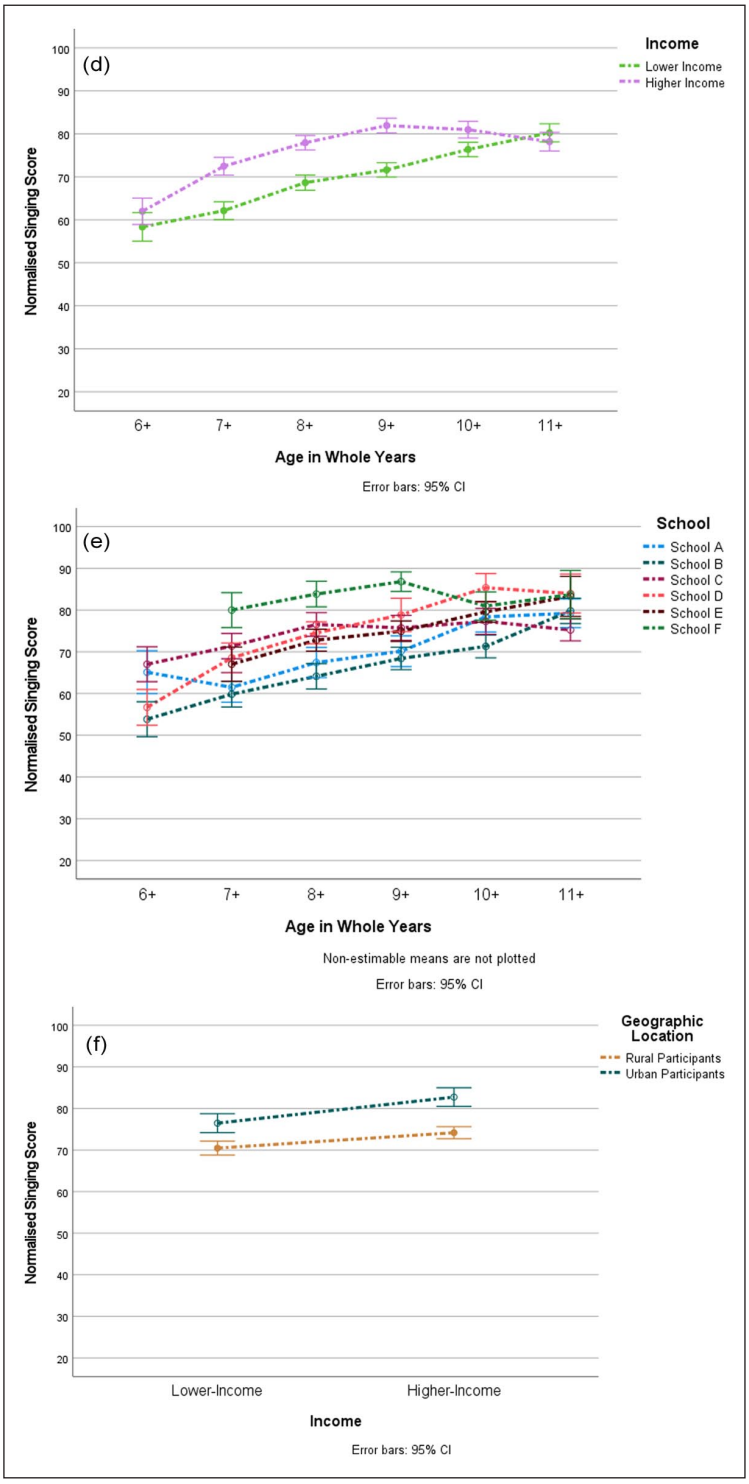


Figure 4. *(Continued)*



Figures 4. (a–f) Normalised singing scores for the three target songs by independent variables and interactions (by age in whole school years)

singing behaviour than rural peers and peers from lower-income families. However, the difference in NSS scores related to geographic location and income appeared to disappear for the oldest participants (ages 10 and 11).

Furthermore, we undertook a one-way ANOVA to explore any age-related differences in singing development across the six participating schools. The age and school interaction also had a statistically significant influence on the NSS $F(23, 3545) = 4.99$, $p < .001$. As Figure 4(e) shows, the means of the NSS for participant children from age 7 to age 9 in School F were higher than those for peers from the other five schools. Nevertheless, by the age of 11+, the NSS differences evidenced for the youngest participants by school were much reduced and there is greater homogeneity evidenced across the six participating schools.

Any mean difference in singing behaviour for participants from lower- to higher-income families between rural and urban areas was also explored by a two-way ANOVA. No statistically significant interaction was found $F(1, 1189) = 1.68$, $p = .196$ (see Figure 4(f)). Rural participants were less developed in their measured singing competency than their urban peers, irrespective of family income.

Discussion

Generally, the current research found that participant children's singing behaviour (including vocal register use and vocal pitch-matching accuracy) is measured as being more developed (i.e., using more vocal registers and being rated as more accurate in their vocally matched pitches of the target songs) with increasing age during the Primary school period. Girls generally showed better singing ability (in terms of vocal register use and vocal pitch-matching accuracy for songs) than boys. However, the sex difference tended to be smaller for the youngest (ages 6 and 7) and the oldest (age 11) participants. The impact of social-economic status (as measured by geographic location and inferred parents' income) seems to be greater for younger (aged <10) participants' singing behaviour. It tended to disappear in the data for the older participants. Similarly, while the younger (aged <10) participants showed varied singing behaviour, the singing behaviour of older participants was more similar (see standard deviations for older participants in Figure 4(e)). Urban students tended to show more developed singing behaviour (using more vocal registers and more accurate vocal pitch-matching for the three target songs) than rural peers.

The current study found that neither sex, geographic location, nor parental income influenced the age-related improvement. The findings suggest that these Chinese participants' singing behaviour related to developing age was similar to that reported of English Primary school children, both in general terms of a trend, and in particular for those children who were without specialised *Sing Up* programme training at the time of their assessment and who acted as controls for the main *Sing Up* programme intervention evaluation (see Welch et al., 2009b, Figure 13, p. 36). The implication is that age seems to have a greater influence than other cultural variables in the development of Primary school children's singing behaviour (i.e., not counting the possible impact of a specific programme of singing education as in *Sing Up* in England).

However, Mang (2006) did not evidence the development of these singing behaviour by age reported in the current study. She was the pioneer of using the *SVDM* and *VPMD* rating scales in her earlier Hong Kong research to define Primary school children's singing behaviour. We should note that there was a tendency for older children's singing behaviour to be rated as better on these measures, although the mean differences in her participants' singing behaviour between age 7 ($n = 64$) and age 9 ($n = 56$) did not reach a statistically significant level. Compared with Mang's (2006) study, the current research included two simpler songs, *Twinkle, Twinkle* and *Little Donkey* as well as *Happy Birthday* and with a wider age group (ages 6–11), and with a greater number of

participants. These distinctions may explain the different results by age between the current study and Mang's (2006).

The current study found that there was a significant interaction effect between age and sex in participants' assessed singing behaviour. The sex difference was smaller for the youngest and oldest participants, while it was greater for the middle age group of Primary school participants. Although boys' singing behaviour tended to be poorer than that of girls in the current study, boys generally continued to develop their singing behaviour across the Primary school period, while girls tended to build their singing behaviour at a slower speed after reaching a peak at age 9. Both of these current findings align with those of English children without *Sing Up* experience in Welch et al.'s (2009b, see Figure 16) study.

Furthermore, the current study found that the variable 'school' statistically significantly influenced participants' singing behaviour. Young participants of School F generally reached more advanced singing behaviour than their peers from the other five participating schools. This might be because the school provided more extensive singing activities and advanced teaching resources than the other five participating schools. However, the two oldest age groups (ages 10+ and 11+) in School F appear not to have continued to develop their singing behaviour, but sustained a high level of singing behaviour across the whole Primary school period. This might be a ceiling effect.

In terms of the influence of socioeconomic status on children's singing behaviour, the current study reports that urban children and children from higher-income families generally showed more competent singing behaviour than their rural peers and children from lower-income families. This suggests that Chinese children with a higher socioeconomic status (e.g., urban areas and/or higher-income families) tend to have better singing behaviour (e.g., singing more accurately in terms of Western tonalities). Furthermore, the interaction effect between geographic location and income had no statistically significant effect on participants' singing behaviour. This suggests that the singing behaviour of participants from higher-income families, whether in either rural or urban areas, tended to be more advanced than that of their peers from lower-income families in the same areas.

Both geographic location and inferred family income could influence participants' singing behaviour through the availability of specialist music teachers within a school's budget, given that urban schools tend to be larger and have a greater number of music specialists, as evidenced by the fieldwork data (Lu, 2024, Table 4.3). Also, based on the first author's informal interviews with music teachers, urban music teachers tended to have more professional development and training opportunities than rural music teachers. Also, given that six of the seven rural music teachers in this study were expected to teach other subjects, it is likely to be more difficult for the absence from school for training to be covered.

Furthermore, while all the music teachers in Schools E and F, located in urban areas, were music specialists, we noted that the parents in School F, who were generally working in a nearby hospital, were likely to be richer than the parents of School E, who had been rural-based originally and who had moved to the city to work. That the parents in School F were wealthier might also explain why School F had an orchestra, as the instruments were provided by parents. Also, it was discovered that rural participant children from lower-income families in Schools A and B rarely took their music classes because the timetable slots for these classes were often used for main subjects. However, the opportunity to take music classes for rural participants from higher-income families in Schools C and D seemed slightly greater. We made this assumption based on the information provided by music teachers, other teachers within the schools, and participants.

In terms of the age influence on the different development in singing behaviour by socioeconomic status, the current study found that the age and geographic location interaction, and the age and family income interaction had a statistically significant effect on participants' singing behaviour. Whether in both rural or urban areas, and from higher or lower-income families, older

participants tended to have more advanced singing behaviour than their younger peers. However, the difference in singing behaviour by geographic location and income tended to be smaller for older participants (age ⩾10) compared with younger participants. These two above interaction effects suggest that Primary school children from diverse socioeconomic backgrounds are likely to reach a similar level of singing behaviour when they are older (age ⩾10), unless there is a specific singing pedagogy intervention (Welch et al., 2009b). This might be because the underlying vocal mechanism of children from diverse socioeconomic groups reaches a similar level of development before puberty. Simultaneously, participants' singing experience is also enhanced with increasing age. Although some came from lower socioeconomic homes, they may have had more access to music in and out of school as they got older because of modern technology. It suggests that most of the oldest Primary school children can reach a certain level of singing behaviour even if the singing teaching in their school is limited.

Summary and Conclusions

The current study aimed to explore Chinese Primary school students' singing behaviour and development by age, sex and social-economic status. We found that older participants and girls had more advanced singing behaviour (using more vocal registers and vocally matched pitches of the target songs more accurately) than younger participants and boys. Urban participants and participants from higher-income families tended to show better singing behaviour than rural peers and peers from lower-income families. However, the socio-economic status (measured by geographic location and inferred parents' income) had less impact on the oldest participants' singing behaviour.

The current study suggests that singing ability is a developing process during the Primary school period. Given that middle-Primary school-aged girls (ages 8–10 years) demonstrated statistically significant more superior in singing behaviour compared to boys at these ages, music teachers should provide more targeted encouragement to male students. Music teachers working with students from rural and lower-income families should not be discouraged by their students' singing development. When a rich singing environment is provided, these students can achieve singing behaviour comparable to students from urban and higher-income families.

The implications are that the current findings could help music teachers in Hunan Province to understand the likely variation in Primary school children's singing ability and development in varied circumstances, including different ages, sexes, geographic locations and family incomes. We assumed that music teachers from other areas in China with similar economic development as Hunan Province could apply the findings to understand and support the singing development of their pupils. However, as we collected the current data from Hunan Province only, where the GDP is close to the median level across the country (https://www/stats/gov.cn), data from the Province may not be appropriate to show the singing behaviour of Primary school children from the top or bottom GDP areas in the country.

It is suggested that a future study of variables in children's sung vocal registers should compare ratings evaluated by the *SVDM* scale, both with and without a given starting pitch and with/without a sung model. It is also suggested that future research should be undertaken to discover (a) more detail of the nature of children's singing behaviour and development in different circumstances, and (b) how these behaviours might be impacted by effective singing pedagogy, as well as (c) what might count as effective singing pedagogy in a Chinese Primary school context. Assuming that singing has value, both in itself and for other aspects of development (such as social and emotional development), then there seems to be a clear need to ensure that music and singing are supported in schools and not marginalised, as might be the case in some rural locations.

ORCID iDs

Can Lu https://orcid.org/0009-0003-4626-8838
Graham F Welch https://orcid.org/0000-0003-2258-8039

Author Contributions

Can Lu: Conceptualisation, Investigation, Methodology, Writing - original draft
Graham F Welch: Conceptualisation, Methodology, Supervision, Writing - review & editing

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Mang (2006) used the song Happy Birthday only, while Welch et al. (2009b) applied two target songs, Twinkle, Twinkle, Little Star (called Twinkle, Twinkle hereafter) and Happy Birthday. Happy Birthday includes an octave leap, whereas Twinkle, Twinkle has a much more limited pitch range.
2. The population of Hunan Province was $N=65{,}683{,}722$ (about 5% of the whole national population) in 2011, rising to $N=68{,}990{,}000$ in 2018 (National Bureau of Statistics of China in 2018). The urban-rural residents' population ratio (1.52:1) in Hunan Province in 2018 was similar to that in many other provinces in China. GDP per capita in Hunan Province was US $10134.994 (14th out of 31 regions) in 2021 (stats.gov.cn).
3. The singing performances of participant pubertal boys whose voices were changing at the time of assessment are not included in the current analyses. Evidence of pubertal voice change was in their vocal products, with the overall pitch range being much lower, and the voice quality being perceptually different and less child-like compared with unchanged voices (*cf*. Cooksey & Welch, 1998).
4. BERA-Ethical-Guidelines-2011.pdf, updated 2018 and 2024

References

Azechi, N. (2016). *The influence of mother tongue on young children's rhythmic behaviour in singing* [Doctoral thesis]. University College London.

Barrett, M. S., Zhukov, K., Brown, J. E., & Welch, G. F. (2020). Evaluating the impact of a generalist teacher-led music program on early childhood school children's singing skills and attitudes to music. *Psychology of Music*, *48*(1), 120–136. https://doi.org/10.1177/0305735618790355

BERA. (2018). *Ethical guidelines for educational research* (4th ed.).

Çenberci, S., & Kalkanoğlu, B. (2023). Turkish Adaptation of Singing Voice Development Measure (SVDM) and examination of children's use of singing voice on different variables. *International Journal of Education Technology and Scientific Researches*, *8*(25), 2949–2982. https://doi.org/10.35826/ijetsar.715

Chen-Hafteck, L. (1999). Singing Cantonese children's songs: Significant of the pitch relationship between text and melody. *Music Education Research*, *1*(1), 93–108. https://doi.org/10.1080/1461380990010108

Cooksey, J. M. & Welch, G. F. (1998). Adolescence, singing development and national curricula design. *British Journal of Music Education*, *15*(1), 99–119.

Dansereau, D. R. (2005). *The musicality of 3-year-old children within the context of research-based musical engagement* [Doctoral thesis]. The College of Education Georgia State University. ProQuest Dissertations Publishing.

Davidson, L., Mckernon, P., & Gardner, H. (1981, March). *The acquisition of song: A developmental approach*. In Documentary report of the Ann Arbor Symposium: Application of psychology to the teaching and learning of music (pp. 301–315), Reston, VA. Music Educators National Conference.

Demorest, S. M., & Pfordresher, P. Q. (2015). Singing accuracy development from K-adult: A comparative study. *Music Perception: An Interdisciplinary Journal*, *32*(3), 293–302. https://www.jstor.org/stable/10.1525/mp.2015.32.3.293

Estis, J. M., Dean-Claytor, A., Moore, R. E., & Rowell, T. L. (2011). Pitch-matching accuracy in trained singers and untrained individuals: The impact of musical interference and noise. *Journal of Voice*, *25*(2), 173–180. https://doi.org/10.1016/j.jvoice.2009.10.010

Guerrini, S. C. (2006). The developing singer: Comparing the singing accuracy of Elementary students on three selected vocal tasks. *Bulletin of the Council for Research in Music Education*, *167*, 21–31. https://www.jstor.org/stale/40319287

IIari, B. (2006, August). *Mapping musical development in Brazil: Children's musical practices in Maranhão and Pará*. In M. Baroni, A.R. Addessi, R. Caterina, & M. Costa (Eds.), Proceedings of the 9th International Conference on Music Perception and Cognition (pp. 1794-1801), Bologna, Italy.

Kinnear, P. R., & Gray, C. D. (2009). *SPSS 16 made simple*. Psychology Press.

Leighton, G., & Lamont, A. (2006). Exploring children's singing development: Do experiences in early schooling help or hinder? *Music Education Research*, *8*(3), 311–330. https://doi.org/10.1080/14613800600957461

Levinowitz, L. M., Barnes, P., Guerrini, S., Clement, M., D'April, P., & Morey, M. J. (1998). Measuring singing voice development in the Elementary general music classroom. *Journal of Research in Music Education*, *46*(1), 35–47. https://doi.org/10.2307/3345758

Lu, C. (2024). Primary school children's singing behaviour in Hunan Province, China Unplublished doctoral thesis, University College London, Institute of Education.

Lu, C., Saunders, J., & Welch, G. (2019). A pilot study of seven-year-old children's singing behaviour, development and engagement in China. *Interdisciplinary Journal for Music and Art Pedagogy*, *9*(2), 23–49.

Lu, C., & Welch, G. F. (2025). A new protocol for revealing the details of children's vocal pitch-matching in song-singing [Manuscript submitted for publication].

Mang, E. (2006). The effects of age, gender and language on children's singing competency. *British Journal of Music Education*, *23*(2), 161–174. https://doi.org/10.1017/S0265051706006905

Mecke, A. C., & Sundberg, J. (2010). Gender differences in children's singing voices: Acoustic analyses and results of a listening test. *The Journal of the Acoustical Society of America*. *127*, 3223–3231. https://doi.org.10.1121/1.3372730

Moore, R. S. (1991). Comparison of children's and adults' vocal ranges and preferred tessituras in singing familiar songs. *Bulletin of the Council for Research in Music Education*, *107*, 13–22. https://www.jstor.org/stable/40318417

Nichols, B. E. (2016). Task-based variability in children's singing accuracy. *Journal of Research in Music Education*, *64*(3), 309–321. https://doi.org/10.1177/0022429416666054

Pereira, A. I., & Rodrigues, H. (2019). The relationship between Portuguese children's use of singing voice and singing accuracy when singing with text and a neutral syllable. *Music Perception*, *36*(5), 468–479. https://doi.org/10.1525/mp.2019.36.5.468

Rutkowski, J. (1990). The measurement and evaluation of children's singing voice development. *The Quarterly*, *1*(1–2), 81–95.

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. https://doi.org/10.2307/3345447

Rutkowski, J. (2004, April). *The validity of the 'Singing Voice Development Measure' for measuring non-American children's use of singing voice*. Paper presented at MENC National Biennial In-Service Conference, Minneapolis, MN, United States.

Rutkowski, J. (2015). The relationship between children's use of singing voice and singing accuracy. *Music Perception: An Interdisciplinary Journal*, *32*(3), 283–292. https://www.jstor.org/stable/10.1525/mp.2015.32.3.283

Rutkowski, J. (2018). Development and pedagogy of children's singing. In S. L. Burton & A. M. Reynolds (Eds.), *Engaging musical practices: A sourcebook for elementary general music* (pp. 33–50). Rowan & Littlefield.

Rutkowski, J., & Chen-Hafteck, L. (2001). The singing voice within every child: A cross-cultural comparison of first graders' use of the singing voice. *Early Childhood Connections*, *7*(1), 37–42.

Rutkowski, J., Chen-Hafteck, L., & Gluschankof, C. (2002, July). *Children's vocal connections: A cross-cultural study of the relationship between first graders' use of singing voice and their speaking ranges*. Paper presented at the Children's Musical Connections: ISME Early Childhood Commission Conference, Copenhagen, Denmark.

Rutkowski, J., Chuang, M. J., & Gluschankof, C. (2007, March). *The singing worlds of children: A cross-cultural study of first graders' use of singing voice when singing songs from their own and other cultural traditions*. Paper presented at the Society for Research in Music Education Symposium on Research in Music Education, Lawrence, KS, United States.

Rutkowski, J., & Miller, M. S. (2003). The effectiveness of frequency of instruction and individual/small-group singing activities on first graders' use of singing voice and developmental music aptitude. *Contributions to Music Education*, *30*(1), 23–38. https://www.jstor.org/stable/24127025

Welch, G. F. (1979a). Poor pitch singing: A review of the literature. *Psychology of Music*, *7*(1). 50–58.

Welch, G. F. (1979b). Vocal range and poor pitch singing. Psychology of Music, 7(2). 13-31.

Welch, G. F. (1985). A developmental view of children's singing. *British Journal of Music Education*, *3*(3). 295–303. https://doi.org/10.1017/S0265051700000802

Welch, G. F. (1998). Early childhood musical development. *Research Studies in Music Education*, *11*(1). 27–41. https://doi.org/10.1177/1321103X9801100104

Welch, G. F. (2015). Singing and vocal development. In G. E. McPherson (Ed.), The child as musician: A handbook of musical development (pp. 441-461). Oxford University Press: New York.

Welch, G. F., Himonides, E., Papageorgi, I., Saunders, J., Rinta, T., Stewart, C., Preti, C., Lani, J., Vraka, M., & Hill, J. (2009a). The National Singing Programme for primary schools in England: An initial baseline study. *Music Education Research*, *11*(1), 1–22. https://discovery.ucl.ac.uk/id/eprint/10003942

Welch, G. F., Himonides, E., Saunders, J., Papageorgi, I., Vraka, M., Preti, C., & Stephens, C. (2009b). Researching the second year of the National Singing Programme in England: An ongoing impact evaluation of children's singing behaviour and identity. London: Institute of Education, University of London.

Welch, G. F., Himonides, E., Saunders, J., & Papageorgi, I. (2010). Researching the impact of the National Singing Programme 'Sing Up' in England. International Music Education Research Centre.

Welch, G. F., & Murao, T. (1994). Onchi and singing development: A cross-cultural perspective, centre for advanced studies in music education. London: Fulton in Association with the Centre for Advanced Studies in Music Education, Roehampton Institute.

Welch, G. F., Sergeant, D. C., & White, P. J. (1996). Listeners' identification of gender differences in children's singing. *Research Studies in Music Education*, *24*(1). 28–39.

Chinese Literature

Feng, W. Y., Yan, Y., & Ge, X. S. [冯婉燕, 晏炎&葛晓穗]. (2013). Characteristics of vocal pitch-matching accuracy of children from three to six years old – analysing musical intervals of song singing from a quantitative perspective [3岁-6岁儿童歌唱音准特点研究-歌曲音程的量化研究视角]. *Early Education*[*早期教育(教科研版)*], *4*. 13–17.

Guo, S. J. [郭声健]. (1999). The current situation of Primary and Secondary school students' music ability [中小学生音乐素质现状管窥]. *China Music Education* [*中国音乐教育*], *1*, 6–10.

The Monitoring Centre for the Compulsory Education, Ministry of Education of the People's Republic of China (MCCEMEC). [教育部基础教育质量监测中心]. (2018). *Chinese compulsory education quality monitoring report* [*中国义务教育质量监测报告*].

The Monitoring Centre for the Compulsory Education, Ministry of Education of the People's Republic of China (MCCEMEC). [教育部基础教育质量监测中心]. (2020). *National Assessment of education quality – arts in 2019* [*2019年国家义务教育质量监测艺术学习质量监测结果报告*].

Zhao, S. [赵苏]. (2020a). Research on pitching in singing of Pre-school children – part 1 [学龄前儿童歌唱音准能力调查研究分析报告（上）]. *China Music Education* [*中国音乐教育*], *6*. 35–40.

Zhao, S. [赵苏]. (2020b). Research on pitching in singing of Pre-school children – part 2 [学龄前儿童歌唱音准能力调查研究分析报告（下）]. *China Music Education* [*中国音乐教育*], *7*. 34–38.