Primary School Children's Singing Behaviour in Hunan Province, China

Can Lu

Thesis submitted for the degree of

Doctor of Philosophy,

Institute of Education (Music Education),

UCL

2024

Declaration

I, Can Lu, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicted in the thesis.

Abstract

The current study explores Chinese Primary school children's singing behaviour. *N* = 1,539 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 existed 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. Compared with English data collected from the National Singing programme *Sing Up* reported by Welch et al. (2009a), Chinese participants and English children (who did not receive *Sing Up* training) showed a similar level of development across different age groups. In addition, a smaller cross-section (*n* = 134) of participants' singing performances were analysed using a newly created system of note-by-note pitch analysis, called the Melodic Analysis for Pitch-Matching (*MAPM*) system. It was found that there was a significantly positive correlation between scores measured by the *MAPM* system and related scores measured by the main *VPMD* scale. By using the colour characteristics of the *MAPM* system, it was found that participants tended to make fewer key changes with increasing age during the Primary school period. There was a significant negative correlation between the number of keys used to sing a song and the related scores of vocal register use. Furthermore, the vocal pitch-matching accuracy tended to decrease with increasing size of musical intervals. Older children tended to match larger musical intervals more accurately. Five common pitch characteristics were revealed by the colour characteristics of the *MAPM* system. (i) For a target song starting with a tonic, participants' self-selected starting pitch was generally one to two semitones sharper than the model. (ii) Participants tended to repeat their pitch errors when vocally matching similar melodic phrases within a target song. (iii) When vocally matching a wide upward musical interval in a target song, they tended to match the first pitch of the interval sharply. (iv) Participants were more accurate in singing a high target pitch if approached within the melody by a small interval. If the high pitch required a large upward interval (such as an octave), they tended to be much less accurate in replicating the high pitch target. (v) When vocally matching continuous upward pitches toward a lift point, many participants tended to match them flat, and tended to be sharp when matching descending patterns.

Impact Statement

Based on *N* = 1,193 participants' singing performances collected from six participating schools in varied geographic locations and social economic backgrounds in Hunan Province, China, the current thesis could help music teachers in Hunan Province to understand the variance in Primary school children's singing ability and development in varied circumstances, including different ages, sexes, geographic locations and family incomes. It is assumed that music teachers from other areas in China with similar economic development as Hunan Province could use the results of participants' singing ability and development to predict their students' singing ability and development and how best this might be fostered.

Understanding the influence of physical and economic elements on Primary school children's singing ability and behaviour may help these Chinese music teachers design more appropriate teaching strategies. For instance, the current study found that young Primary school students with good singing teaching resources could reach a relatively high degree of singing ability. Consequently, music teachers should believe that young Primary school students could mostly reach a relatively high degree of singing ability if efficient teaching strategies are used. As girls generally were found to have better singing ability than boys, music teachers should pay more attention to improving Primary school boys' singing ability to reduce the sex/gender gap in singing ability.

The current study also found that participants from less advanced socioeconomic status (e.g., rural areas and lower-income families) tended to show poorer singing ability. Therefore, school managers under this circumstance should try their best to guarantee official time used to teach music within schools. If good quality teaching resources are hard to promise, other informal music resources can be applied in and after classes, such as playing music on the school's radio to students to increase their students' singing experience.

Furthermore, by comparing the singing ability between the current Chinese participants and English students without *Sing Up* training (as a comparable group) in the National Singing Programme in the UK, the current study found that children from both cultures showed similar singing ability and behaviour. This may help to reduce cultural differences in Primary school children's singing ability and development.

Finally, by creating a new system, named the Melodic Analysis of Pitch Matching (*MAPM*) system, the current study analysed characteristics of consistency of key centres and vocal pitch-matching accuracy for different types of musical intervals by age. Five other common features of vocal pitch-matching accuracy for songs for Primary school participants have also been summarised based on the system. These results could help Chinese music teachers understand Primary school children's vocal pitch-matching accuracy for songs in more detail. The teachers can focus on these common issues and then try to solve the common vocal pitch-matching errors to improve students' vocal pitch-matching accuracy for songs.

The system should be able to be applied anywhere in the world in which music utilises equal temperament where researchers (and teachers) are interested in the details of children’s current singing behaviour and how to design strategies to foster development towards singing competency.

Acknowledgements

My thesis would not have reached the current quality if I had not received so many people's support and help!

First of all, I would like to thank Prof. Graham F. Welch for his professional support during my doctoral journey. You helped me to build good academic habits. Being strict and providing support pushed me to be more serious for my research and to ask more questions for myself. I have been flying in the academic world bravely and exploring related questions I raised, as I know you are there and can support me all the time! Thank you very much for your patience to give me space to explore my questions. I know I am more independent than before, and I know you are always there during my doctoral journey, if I need. Thank you very much for encouraging me to attend many conferences and helping me to publish a journal article. These gave me more opportunities to understand our academic world. Thank you for supporting me to ask questions of other scholars. Thank you very much for organizing our doctoral group meeting after I came back to China to finish my thesis. This has not only improved my understanding of academic studies, but also made me feel I am not alone! I would also like to thank Dr. Jo Saunders for her help and support during my doctoral study.

I thank Mr. Howard Pau and my dear father's financial support for my doctoral study when I studied in London, which is a great but expensive city to study and live in. Without the financial support, I wouldn't have had the opportunity to study in the world's greatest university for education. Furthermore, thank you both for your understanding of my difficulties as a foreign student studying in the UK.

I thank Ms. Elizabeth Short. You provided kindness and free support to correct my written English during the last five years! I don't not know how I would have finished my thesis if I had no English support from you! You are very kind and reliable! It would never be too much if I said "Thank you" all the time. Thank you very much for correcting my spoken English and helping me understand the differences in society between China and the UK during our daily conversations, which was important for my understanding for my thesis! It is your encouragement, support and push that made me finally reach the end of my thesis!

I thank my participant children and related teachers for their help during the data collection process of my pilot and main studies. Without all of your support, the thesis wouldn't exist.

During my doctoral journey, I have been in contact with so many academics in the field of research, including Prof. Joanne Rutkowski, Dr. Esther Mang, Dr. Geraldine Leighton, Prof. Bo Wah

Leung, Prof. Bin Wu, Prof. Shengjian Guo and Prof. Ingo R. Titze. Thank you all for your kindness and for providing me with professional answers when I raised a question. I would also like to thank my dear colleagues from our doctoral group for their academic company in our regular group meetings. Thanks to Eunice Tang for organising it for us.

Furthermore, I would like to thank so many people for their help and love when I stayed in the UK. I thank my landlady, Ms. Anna Sherwin, for your care when I stayed with you in your house, especially during the two-year pandemic period! Your knowledge and critical points helped me to understand English society better. Thank you very much for helping me to correct my spoken English. I thank you and your family members' love. Furthermore, I thank my friends Isabella, Erla's family, Yi Wang, Dan Xiao for your company, which made my life happier and more colourful!

Finally, I thank my dear mum's encouragement. Your encouragement helped me get through countless difficulties during my doctoral study! You and my dear sister's countless visual phone calls have been with me wherever I am, either in the UK or Shanghai, China. It is the magic that reduces my aloneness as both of you know I really need it! I also thank my other family members, my two lovely nephews – Qiqi and Changchang, my brother-in-law, my dear husband, and my parents-in-law for all your love and understanding during my doctoral journey and making me feel I am a happy and lucky person! I thank my dear child who is hiding in my belly. Thank you for your support by not making me feel ill during the first six months of pregnancy, so I could finish my thesis smoothly. Lastly, I also thank myself for working hard to finish the thesis.

Table of Contents

List of Figures

List of Tables

Abbreviations

MoEPRC: The Ministry of Education of the People's Republic of China;

MCCEMEC: The Monitoring Centre for the Compulsory Education, Ministry of Education of the People's Republic of China;

SVDM: The Singing Voice Development Measure;

VPMD: The Vocal Pitch-Matching Development;

MAPM: The Melodic Analysis for Pitch-Matching;

NSS: normalised singing score;

CT: cricothyroid muscles;

TA: thyroarytenoid muscles;

PCA: posterior cricoarytenoid muscles;

LCA: lateral cricoarytenoid muscles;

IA: interarytenoid muscles;

F0: fundamental frequency.

Chapter One

Introduction

1.1 Overview

I trained as a music teacher in my undergraduate degree in China, but recognized that I did not have sufficient knowledge to be an effective teacher. Subsequently, I had an opportunity to be a postgraduate student at the University of York, UK on the MA Music Education: Instrumental and Vocal Teaching. This led to an interest in children's singing development and a decision to research this by joining the research team at UCL who had a special focus on this aspect of musical development.

This doctoral study explores the development of children's singing behaviour in the Chinese mainland. Singing is measured using three different approaches. Two of these are established in the research literature and relate to children's use of sung vocal registers (Rutkowski, 1996; 2015) and vocal pitch-matching ability for songs (Welch, 1998). Participants were drawn from one mainland

Chinese area, Hunan Province. The research investigates singing ability and development according to the selected key variables of song task, children's age, sex, socioeconomic group (as measured by geographic location and inferred family income), and school location, using a range of established research tools. In addition, a new assessment tool – the Melodic Analysis of Pitch-Matching (*MAPM*) – was developed to analyse children's vocal pitch in singing behaviour in more detail. The participants were Primary school children in Hunan Province, China numbered over 1,000 and were aged from 6y+ to 11y+.

The structure of the current chapter is as follows: (1.2) a brief description of the population and economy in China, including Hunan Province; (1.3) offers a description of the broad socioeconomic context of Primary school children's music education in China; (1.4) is a description, based on previous studies, of children's singing behaviour and the benefits of singing, and an outline of national policies related to singing education in Chinese Primary schools; (1.5) is a review of international and Chinese studies related to the evaluation of singing behaviours of Primary school students; (1.6) is a description of the research aims and (1.7) the research questions. Finally, (1.8) is a summary of the structure of the other chapters of the thesis.

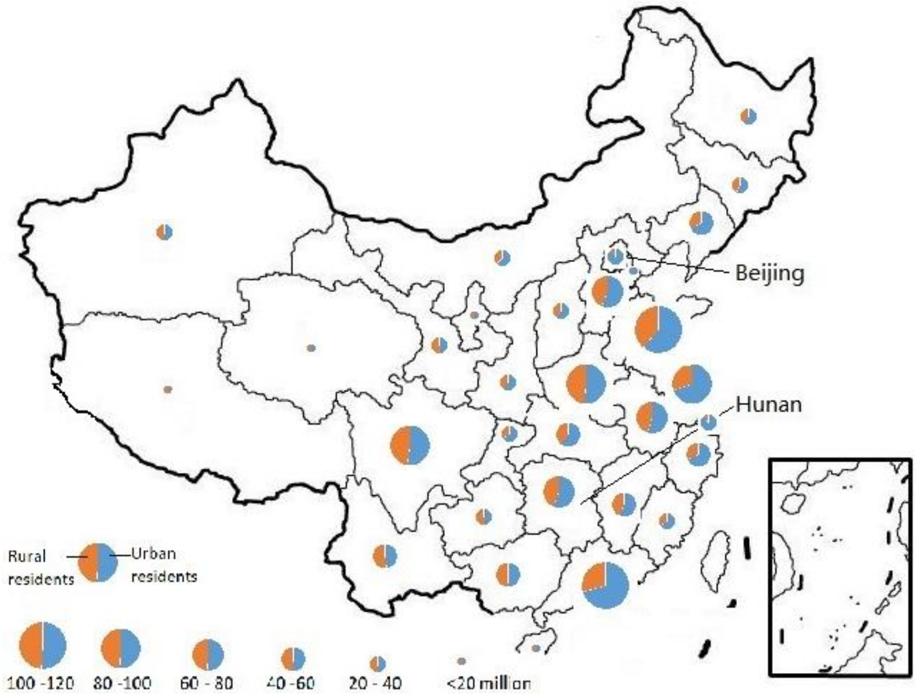
1.2 A Brief Description of Population and Economy in China, including Hunan Province

In the current section, the population and economy based on the difference between rural and urban areas in China, including Hunan Province, are described to provide the contextual understanding of the fieldwork and participants in this research project.

1.2.1 China

China covers approximately 9,600,000 square kilometres in East Asia and had a population of around 1.4 billion in 2019 (National Bureau of Statistics of China, 2019), which makes it the world's most populous country. Eastern China is more densely populated than western China, which relates to both geographical differences and economic development since 1978. The number of urban residents in each province (including municipalities) in China is similar or greater than that of rural residents in the same province (see Figure 1.1).

Figure 1.1 Population of each Province (including Municipalities) in China, and Proportion of Urban and Rural Habitants in each Province (National Bureau of Statistics of China, 2018) (author translated)

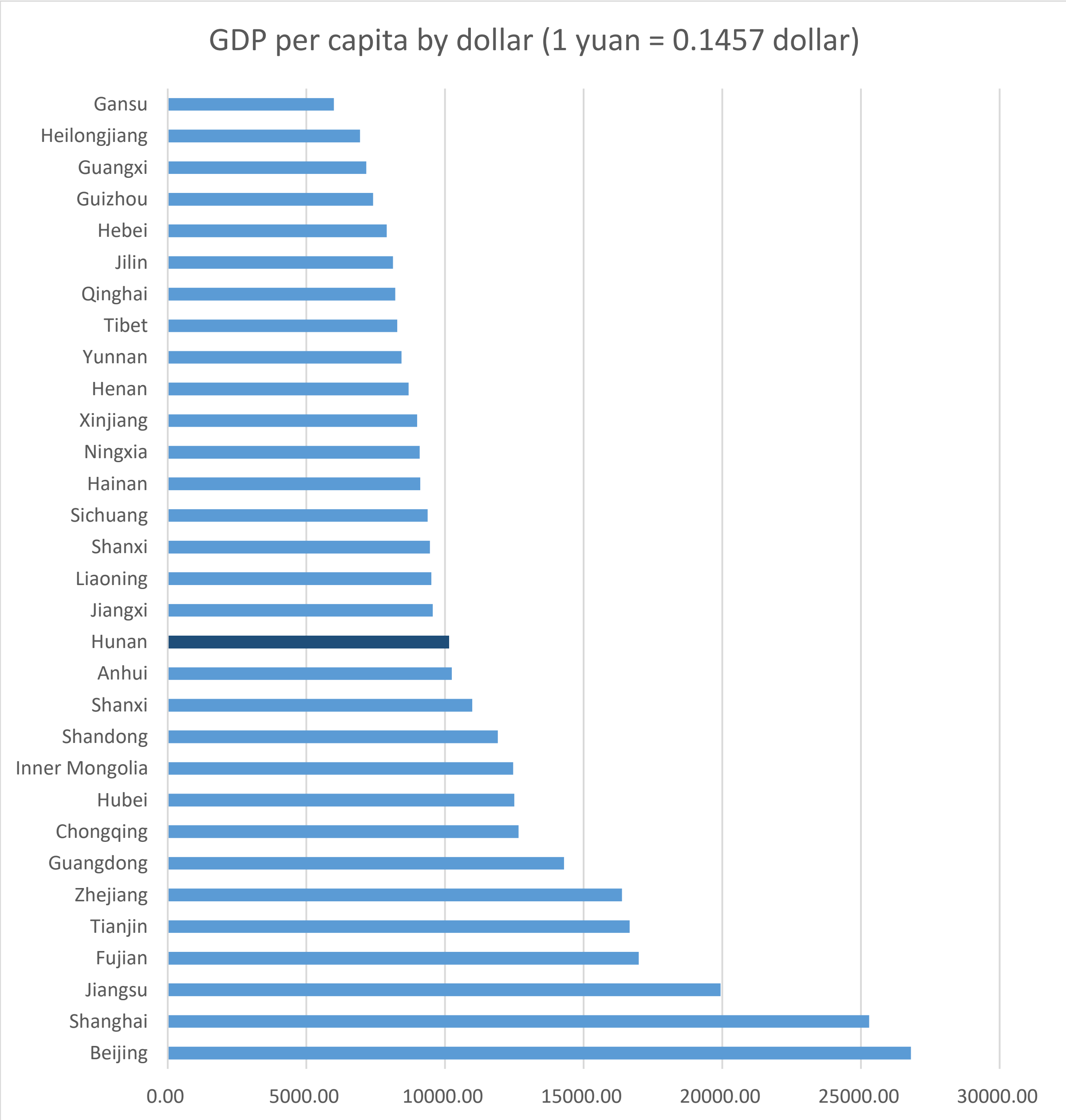


After President Deng Xiaoping introduced a 'Reform and Opening-up reform' policy in 1978, China's economy has been reported as growing significantly. Since 2010, China has been the world's second-largest economy by nominal GDP[1]. Consequently, it is thought that many Chinese people generally have a better quality of life today compared with 20 years ago, although it has been estimated that 370 million Chinese live below the upper-middle-income poverty line (World Bank, 2020)[7]. Figure 1.2 indicates that provinces in most of China showed a similar economic development by disposable income per capita in 2017, with the exception of the provinces of the East coast which tend to be wealthier (Hong, 2022). According to official data, the mean for the urban populations in terms of disposable income per capita per year of the 20 provinces (including municipalities) (being around two-thirds of all provinces of China), and spread across China, apart from the east, was between US $4000 - US $5000. That of the rural population in 20 similar provinces (around two-

[1] https://www.worldbank.org/en/country/china/overview

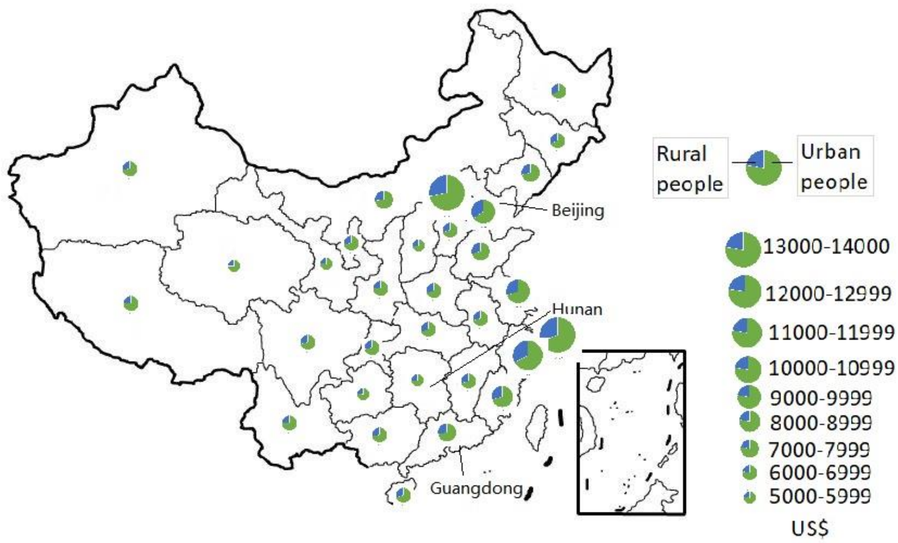
thirds of all of the provinces of all provinces of China) was between US $1000 - US $2000. These data suggest similar economic development across these provinces (see Figure 1.2).

Figure 1.2 GDP Per Capita by Administrative Division in China in 2021 (author translated)



There is a positive relationship reported between urban people's disposable income per capita and that of rural people from the same province. The disposable income per capita of urban people was between 1.86 and 3.40 times bigger than that of rural population in the same 31 provinces (National Bureau of Statistics of China, 2018) (see Figure 1.3).

Figure 1.3 Rural and Urban Peoples' Disposable Income Per Capita of 31 Provinces (Including Municipalities) in China in 2018 (National Bureau of Statistics of China, 2018) (author translated)



One of the reported differences between Western and other societies, such as Chinese society, relates to their different economic development classification. For instance, based on a report[2] from the United Nations that measured country classifications by fuel exporters and fuel importers and per capita gross national income in 2020, the UK is a developed country, while China is classified as a developing country. In the UK, the majority of people are living in urban areas. For instance, in England, 9.7 million (17.14%) and 46.9 million (82.86%) of people were living in rural and urban areas in 2020[3], respectively. Comparatively, the relative gap between the numbers of rural and urban populations is smaller in China, in which there were 551.62 million (39.4%) rural people and 848.43 million (60.60%) urban people in 2019, respectively[4]. As a result, the comparative economic and related educational situation of rural and urban areas between the UK and China is likely to be different. Whether living in a rural or urban setting has an impact on children's singing competency is unknown, although there is some evidence from England that there may be little difference (Welch et al., 2009b). However, this may not be the case in China because of the country being in a different economic development stage.

---

2 https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2020_Annex.pdf

3 Population - Summary - GOV.UK (www.gov.uk)

4 http://data.stats.gov.cn/easyquery.htm?c*n* = C01&zb=A0301&sj=2019

1.2.2 In Hunan Province, China

Hunan Province is the location focus for the thesis fieldwork and is in the southern-middle area of China (see Figure 1.4). The population of Hunan Province was 65,683,722 (about 5% of the whole national population) in 2011, rising to 68,990,000 in 2018 (National Bureau of Statistics of China in 2018)[5]. According to the latest (2023) official data, the number of inhabitants in urban areas (*n* = 39,830,000, 60.31%) was greater than that in rural areas (*n* = 26,210,000, 39.69%)[6] due to various reasons, such as looking for more work opportunities and education resources in urban areas, which was promoted by the national policy of the process of urbanisation[7]. The urban-rural residences' population ratio (1.52:1) in Hunan Province in 2018 was similar to that in many other provinces in China (see Figure 1.3).

GDP per capita in Hunan Province was US $10134.99[4] (14th out of 31 regions) in 2021, which was nearly at the medium level within 31 national provinces (see Figure 1.2). The overall per capita disposable income of residents in Hunan Province in 2022 was $4959.05[8] (using 1 yua*n* = 0.1457 dollars). Per capita disposable income of its urban residents ($6891.76, using 1 yua*n* = 0.1457 dollars) was 2.42 times higher than that of rural residents ($2847.85, using 1 yua*n* = 0.1457 dollars) in 2022. This was similar with that of other provinces in China (see Figure 1.3). Overall, the similar situations in the population ratio between the urban and rural residents and their per capita income ratio among Hunan Province and many other provinces in China are likely to have enhanced the generalisability of the collected data in the current study from Hunan Province.

There were 11,573,757 (17.62% of the population of Hunan Province) people aged 0+ to 14+ (National Bureau of Statistics of the People's Republic of China, 2012)[9]. The doctoral study was undertaken in Hunan Province as it is the researcher's home location, and it was easier to access than other parts of the country. Although data from the study is not intended to represent pupils' singing behaviour towards singing across the whole province, it is intended to be sufficiently large to help music teachers in Chinese Primary schools understand pupils' singing behaviours better.

---

5 国家统计局 (stats.gov.cn)

6 2022 年末湖南常住人口 6604 万人-湖南省人民政府门户网站 (hunan.gov.cn)

7 【"十四五"新型城镇化实施方案】-国家发展和改革委员会 (ndrc.gov.cn)

8 2022 年湖南省民生调查情况发布词-湖南国调信息网,国家统计局湖南调查总队 (stats.gov.cn)

9 国家统计局 (stats.gov.cn)

Figure 1.4 Location of Hunan Province in China

1.3 The Socioeconomic Context of Primary School Children's Education and Music Education in China

To understand Primary school children's singing behaviour in China, socioeconomic contexts related to the compulsory education system and music education within the system are provided below.

1.3.1 The Compulsory Education System in China

In China, education in Primary schools, which is from Grade 1 to Grade 6[10], is compulsory. It is free of charge for all Chinese students since 2008. Around 99.91% of children attend Primary schools, as part of nine years of free compulsory education. The difference in educational opportunities between boys and girls had been reduced due to expanded national educational coverage (Liu, 2008) and the decline in the birth rate of the population (Ye & Wu, 2011). Teachers' salaries have also improved (The Central People's Government of the People's Republic of China, 2010), suggesting that teaching is more professionalised.

Two central problems of the Chinese economy are reported to be the unequal development between eastern and western China, and between urban and rural areas (The Central People's

[10] A small part of school systems use five-years programme of Primary school (OECD, 2016).

Government of the People's Republic of China, 2010). The unequal economic development between rural and urban areas was also reported to be a common problem among developing countries (Liu, 2008). This unequal economic development has resulted in dissimilar levels of educational attainment. For instance, the level of central funding for rural Primary schools tends to be around two-thirds of that provided for all schools[11]. Eastern and urban areas of the country tend to have more local funding[12] and, therefore, are seen to have developed faster in their education provision than western and rural areas[13].

In order to address these educational challenges, the Chinese government has increased financial support to students in western and rural areas[14] (The Central People's Government of the People's Republic of China, 2010). Nevertheless, due to the overall size of the population, funding in western and rural areas is still considered to be insufficient to ensure equal educational standards across the whole country[22].

Furthermore, rapid economic development, especially in the east, has brought more working opportunities, such as in retail, public transport or restaurants. Consequently, a significant number of rural working people (n = 172,660,000, being 59.88% of the total number of rural working people), are reported to having given up their land and moved to cities to work (National Bureau of Statistics of China, 2018), with or without taking their children. Around two-thirds, 60.29%, of rural people have moved to the East of China, of whom around 70% were male, with 68.1% married. Overall, 51.5% of these, whose children tend to be Primary or Secondary school students, were born after 1980 (National Bureau of Statistics of China, 2018).

In 2018, there were 10.5m Primary school students[15] who were originally rurally-based children and who had moved to an urban area with their parents to have a better educational opportunity (National Bureau of Statistics of China, 2018). These represented a new category of students, in addition to the original urban-born children, being approximately 10.14% of total number of Primary school students in 2018 (N = 103,392,500)[23]. Their parents were reported to have found that sending their children to an advanced urban school was difficult, and that living and extra study fees for children were expensive[16].

---

[11] http://data.stats.gov.cn/easyquery.htm?cn = C01&zb=A0M0Y01&sj=2019

[12] See 'Per capita disposable income of residents' from the National Bureau of Statistics

[13] http://www.gov.cn/jrzg/2010-07/29/content_1667143.htm

[14] http://www.gov.cn/jrzg/2010-07/29/content_1667143.htm

15 http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/201907/t20190724_392041.html

16 http://www.gov.cn/shuju/2019-04/30/content_5387773.htm

In terms of comparative numbers, approximately 20% of the Primary school population are located in rural areas (for example, in 2012, *n* = 19,530,000[17], or 20.14% of total Primary school students, *N* = 96,959,000, in China in 2012). One or both of parents had transferred to a city without them, and so they remained in a rural location and were looked after either by their mother or grandparents[25]. Furthermore, the difference in economic development within a district also exists (*cf.* Chen & Fang, 2007).

As a result, in the context of China's current economic development, Primary school students can be divided nationally into four groups: (1) rural children who are left behind when their parents move and live with relatives; (2) rural children who live with both their parents in a non-urban setting; (3) the new urban group of children who have migrated from rural areas, and (4) the original urban-based children.

According to Chen and Fang (2007), different socioeconomic groups tend to have varied education (such as to degree level), income, consumption level, lifestyle, and value orientation. People from the same socioeconomic status were more likely to be gathered in a geographic area due to a similar living consumption level (Chen & Fang, 2007). According to Chen and Fang (2007), an urban area might be divided into middle- and upper-class residential areas and lower-class residential areas. Schools in middle- and upper-class residential areas were more likely to receive more financial support from local regions (Li & Qiu, 2016). This deviation might be more obvious in more developed cities in China (Chen & Fang, 2007). According to Liu (2008), the difference in Primary school students' education by income in urban areas was smaller than in rural areas.

Regarding the family's influence on students' study performance, in line with English studies (e.g., Caro et al., 2009; Cheadle, 2008; Sirin, 2005), many Chinese scholars (e.g., Fang & Feng, 2008; Li & Qiu, 2016; Sun, 2009) reported a positive relationship between Chinese family economic resources and children's academic performance, perceived as being due to high economic investment providing more educational opportunities. In rural areas where parents generally had fewer economic resources, the quality of students' academic performance was likely to be dominated by their level of study commitment (Li & Qiu, 2016). Families with higher socioeconomic status might be more likely to strive for their children to get into an advanced school which had gathered excellent teachers during this compulsory education period (Li, 2006; Li & Qiu, 2016; Liu, 2008; Zhao & Hong, 2002). Li and Qiu (2016) summarised that there was a positive relationship between parents' socioeconomic status and the school's social status (*cf.* Chen & Fang, 2007; Li, 2008; Wen, 2006; Wu, 2013). It has also been reported that parents' high investment in education

[17] http://www.hubei.gov.cn/mzgjc/gjcgk/201305/t20130517_449285.shtml

for their children could increase the likelihood of the children inheriting high socioeconomic status from their parents (Wu, 2013).

Schools built in these different areas had varied socioeconomic groupings with unequal economic development and official support (Chen & Fang, 2007). Financial support of a Primary and Secondary school in China is provided by national and local funding (Fang & Feng, 2008). The national funding is seen as important in order to increase more equal educational opportunity to Chinese children, while the local funding has more power to influence the degree of extra financial support (Fang & Feng, 2008).

Under the context of varied economic development, school variation is also evidenced by the difference in teaching facilities and the degree of teachers' specialization (Chen & Fang, 2007). For instance, the IT hardware and software facilities of a school serving new urban children who originally came from rural areas with their working parents were more likely to be poorer than in a school that is located in a city's financial centre. A 'good' Primary school – based on students' academic performance – was likelier to be in an area of higher socioeconomic status (Chen & Fang, 2007). This was seen as the key point when people selected a school for children when buying a local house. Students in areas of higher socioeconomic status were more likely to go to a 'good' Primary school than students from lower socioeconomic backgrounds (Chen & Fang, 2007). Also, students studying at a 'good' Primary school with advanced teaching resources were more likely to receive a better examination result (Wu, 2013). Urban students were more likely to go to a better high school or university than rural students (Wu, 2013).

Overall, the rapidly developing Chinese economy since 1978 has allowed more support for the education system in China. However, the imbalanced economic development between eastern and western China, between rural and urban areas, by family income and school has also created an unequal development of educational attainment across the country (Chen & Fang, 2007). This educational attainment includes music education in Primary schools (Cao, 2017; Guo, 1999).

1.3.2 A Brief Context of Music Education in Primary Schools in China

Under the latest guidance of the Ministry of Education of the People's Republic of China in 2022 (MoEPRC)[11], Primary school students should receive moral, academic, physical, aesthetic, and work skills education. Among these five targets areas, academic education has received the most attention from Primary school leaders and parents because subjects designed for academic education have a major influence on the national college entrance examination (NCEE) (see curricula of the NCEE[18]). Limited space in universities has increased the competition in the national college entrance examination, which occur at around 18 years of age (Li & Qiu, 2016). Chinese parents tend to have a high ambition for their children (Ma, 2010). Although students take the entrance examination at the age of 18+, they will have been influenced since their Primary school period by the competitive ethos of academic attainment within the classroom and between schools (Li & Qiu, 2016). Although the MoEPRC policy places equal emphasis on the five target areas of education[9], academic education still appears to play the most crucial role for students and their parents, which may be partly because of the intense competition of the global economy (Wang et al., 2020).

In the context of the national education policy and the college entrance examination, music lessons in Primary schools in China designed to develop aesthetic education seem to have a less prominent position than subjects perceived as more 'academic', especially in rural Primary schools (Li, 2010), as evidenced by the comparatively lower number of lessons[13] being offered. This opinion appears to have found political support at a time when many countries are overly focused on science, technology, engineering and mathematics (STEM) subjects (such as in England[19], China[20], and Australia[21]), even though the cultural sector is a multi-billion pound (equivalent) industry in the same countries (the UK[22], China[23], and Australia[24]) and whilst, paradoxically, often being officially encouraged by Government policy. This anti-arts bias has led to music becoming more marginalised in school curricula, such as in schools in England through the introduction of the so-called 'English Baccalaureate'[25], where the arts are replaced by more 'useful' subjects, such as mathematics and literacy. However, such an argument is countered by research into the intrinsic and extrinsic benefits of music in general (e.g., Hallam, 2010; Hallam & Himonides, 2022) and singing in particular (e.g., Tamburri et al., 2021; Welch, 2012; Welch et al., 2010).

However, personal experience as part of the doctoral study fieldwork suggests that this may not

---

18 http://gaokao.neea.edu.cn/html1/category/1509/6212-1.htm

19https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/425601/PRIMARY_national_curriculum.pdf

20 http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/moe_711/201006/xxgk_88602.html

21 https://www.education.gov.au/australian-curriculum-0

22 https://www.ukmusic.org/assets/general/Music_By_Numbers_2019_Report.pdf

23 https://www.mct.gov.cn/ggfw/whzyzg/ys/yl/201111/t20111121_511124.htm

24 https://www.nme.com/en_au/news/music/ilostmygig-australian-music-industry-nearly-50-million-lost-income-bushfire-coronavirus-cancelled-events-2625767

25 https://www.gov.uk/government/publications/english-baccalaureate-ebacc/english-baccalaureate-ebacc

be true for Primary schools which achieve high academic results in China. Greater financial support is made available for schools achieving excellent academic attainment (Guo, 1999), so these schools may attract higher-skilled music teachers (*cf.* Yao & Zhang, 2008).

As a result, the map of the provision of music teachers nationally may correlate with the map of the Chinese economy, where richer areas have better music education[26], especially for musical instrument instruction which requires more financial support than that of singing. Arguably, compared to instrumental teaching, singing teaching has more potential for spreading music education more equally in Primary schools because no fee is required to purchase a musical instrument and every child already has a vocal instrument. This is in line with the national education policy that providing music education to every child is one of five objectives of the latest Standards for music education for Primary and Secondary schools in China (see[27]).

The music lesson schedule in Chinese Primary schools usually consists of two 35-40-minute music lessons every school week[28]. However, the time may be squeezed in some rural schools due to a lack of music teachers, or the preparation for examinations of core subjects, such as literacy (The Central People's Government of the People's Republic of China, 2010). According to the latest official Standards for the music curriculum (Ministry of Education of the People's Republic of China, 2022), Grades 1 (age 6+) to 2 (age 7+) are considered to be the first years of schooling, and Grades 3 (age 8+) to 5 (age 10+) are a more senior age group. Grade 6 is the most senior age group in the Primary school phase. The teaching aims of music education for one of the stages are meant to become the foundation for those of the next stage (Ministry of Education of the People's Republic of China, 2022).

The MoEPRC has suggested that the educational priority should be students, rather than teachers, and school education should follow the development of students[29]. Nevertheless, this implies a competent and appropriately qualified professional workforce.

Under official guidance, a range of strategies have been undertaken to improve the music ability of music teachers in the last 20 years across China. For instance, the China Music Education Journal[30] has been publishing a range of articles to guide music teachers in Primary and Secondary schools since 1989 (e.g., Chen, 2019; Yan, 2020; Yin, 2020). According to Guo (1999), music teachers across

---

26 http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_795/201401/163173.html

27 http://mat1.gtimg.com/edu/pdf/edu/xkb2011/20120130160216344.pdf

28 新课程标准小学各科周课时安排新版 - 百度文库 (baidu.com)

29 http://www.gov.cn/jrzg/2010-07/29/content_1667143.htm

30 http://mall.cnki.net/magazine/magadetail/ZYJA202001.htm

the country have been taking national training to enhance their musical ability and update their teaching skills, although—somewhat paradoxically—music teachers with existing higher skills have more priority for this training. As part of the training system, some international methods, such as Orff, Kodaly and Dalcroze, have been introduced. In addition, more specific departments for music education have been established; more Primary schools have received increased support for the purchase of hardware, such as computers and percussion instruments; and more textbooks for the music curriculum are available (Guo, 1999). Furthermore, a national online competition of music activities for music teachers in Primary and Secondary schools, including singing, piano playing, choir conducting, and micro-course online video-making for a music curriculum, has been taking place since 1996[31]. Also, a range of online talking groups using the Wechat app have been set up for free communication between music teachers across the country, which is believed to help teachers share their resources and experiences more efficiently and widely (Wei, 2018). A range of free online lectures, such as a series of lectures provided by Deqing public welfare[32], are available for music teachers. These are designed to support an improvement in the different musical skills of music teachers. Overall, there is evidence that music education in compulsory education has been seen as more of a priority and improved in the last 20 years in China (Guo, 1999; Zhang, 2016).

As reported above, in China, educational investment across different areas is diverse between rural and urban areas, regions and schools (Li & Qiu, 2016; Yang, 2006). This diversity is evidenced by differences in the ratio of students and teachers, and also hardware and software facilities (Yao & Zhang, 2008). This unequal investment is demonstrated in music education (e.g., Cao, 2017; Li, 2010; Liu, 2018). Problems with the provision of the music curriculum in Primary schools are compounded by uneven economic development across different regions and a weak foundation for music education (Guo, 1999). These problems include the replacement of music lessons by lessons in the core curriculum; lack of appropriate equipment; uneven quality of music teachers; and limited financial support, especially in rural areas (e.g., Du & Yang, 2019; Guo, 1999; Li, 2010; Zhang, 2019).

These common problems have been widely reported across China, and they can be evidenced by reports from the China National Knowledge Infrastructure (CNKI)[33], such as in Helongjiang

---

31 音乐与舞蹈学院音乐学系教师教育专业 2021 年"五项全能"基本功比赛拉开序幕-音乐与舞蹈学院 (hbuas.edu.cn)

32 Deqing public welfare was established in 2004. It has paid attention to improving the quality of children's music education, especially for rural children, by providing high-quality visual lectures for music teaching, including singing. See more from the link 基金会简介_德清基金会 (dqjjh.com).

33 www.cnki.net is the key platform to share academic work, led by Tsinghua University and supported by PRC Ministry of Education.

Province (Li, 2010), in Xinjiang Uyghur Autonomous Region (Cao, 2017), in Henan province (Wang, 2017), in Hubei province (Liu, 2018). These studies agree that a greater proportion of music teachers are music specialists in urban compared with rural areas, and many generalist (non-music specialist) teachers are having to teach music lessons in rural Primary schools due to the limited number of specialist music teachers available locally. The situation for non-specialist Primary teachers is likely to be similar to that found in the UK and elsewhere (Hennessy, 2012; Welch & Henley, 2014). As the generalist teachers are not music specialists and are teaching other subjects, they are likely to give less emphasis to music. Also, a lack of confidence in teaching music is likely to be another reason (*cf.* Welch & Henley, 2014). Consequently, there may be a difference in the degree of teachers' attention to music in urban compared to rural Primary schools in China today (Guo, 1999; Zhang, 2019).

As mentioned above, the music lesson in Primary schools in China formally serves the strand of aesthetic education. This need for aesthetic awareness in China is defined as reflecting the individual's moral personality and tastes, and a positive attitude (MoEPRC, 2011, 2022). Music education in Primary schools in special administrative regions of China, such as Hong Kong and Macao, also keeps in line with the Chinese mainland, which sets aesthetic sensitivity as the prime focus; however, the definition of the aesthetic sensitivity may be different because of context. For instance, in Hong Kong, the definition of aesthetic sensitivity seems to be more related to the development of creativity, the ability to appreciate music personally and for wider community purposes[34].

In China, the current policy of asserting the aesthetic purpose for music may derive from an ancient Chinese concept. For instance, one of the aesthetic aims of Li Yue (礼乐), raised in the Zhou dynasty more than 3000 years ago (Xiang, 2010), was for music to be a formal activity. It was to reach harmony between society, the soul and physical body. More recently, this ancient Chinese philosophy was endorsed in the writing of the American music educator and philosopher, Bennett Reimer, who has had a significant influence on music education policy for compulsory education in China in the last several decades (Zi & Zhu, 2006). Some researchers have reported that 'students like music, but do not like music lessons' (e.g., Pu, 2016, p. 11), in which the 'music lessons' are thought to include knowledge about music. This may be because aesthetic awareness has been set as the main aim of music education in China today (Ministry of Education of the People's Republic of China, 2022; Zi & Zhu, 2006), although Guo (1999) argued that it was not because the musical content was difficult for Primary students to understand, but because the strategy used by some music teachers to teach it

[34] https://www.edb.gov.hk/attachment/en/curriculum-development/kla/artsedu/references/music%20complete%20guide_eng.pdf

was inappropriate.

The functions of music education have been widely agreed by Western researchers (e.g., Hallam, 2010; Welch, 2012; Welch et al., 2014), including physical, psychological, social, musical and educational benefits (Welch, 2017), but countries have different policies due to their different cultural and historical contexts, and political priorities. Perhaps the most significant difference is where education is focusing on the music itself (education in music, such as in the UK), rather than focusing on education through music (as in China), although there will be overlap in these two experiences (e.g., Ockelford, 2000; Pitt & Welch, 2020).

1.4 Children's Singing Behaviour, including Policy, Development and Benefits

In the current section, national policies regarding singing education in the Chinese Primary school system are reviewed initially to provide the Chinese context for participants' singing behaviour. To understand Primary school children's singing behaviour better, a developing process of singing behaviour is reviewed from the infant stage to the later Primary school period below. The benefits of singing reported by previous studies are also reviewed.

1.4.1 National Policies Related to Singing Education in Chinese Primary Schools

In China, singing has always been a popular musical activity, and was especially so in the early part of the 20th century when musical instruments were not widely available (Ma, 2002). Singing is one of four elements of performance of the National Standards for music (National Bureau of Statistics of the People's Republic of China, 2011, 2022)[35]. The other three elements are playing musical instruments, comprehensive art performance, and reading (sheet) music. Singing is still a main music activity in many rural Primary schools, due to a lack of music specialists (Wang, 2004).

According to the National Standards for the music curriculum (MoEPRC, 2011, 2022), singing, which is seen as the easiest music activity to be received and enjoyed by pupils, should provide the

[35] 国家统计局 (stats.gov.cn)

fundamental content of music education of Primary schools in China. Singing is reported to not only develop students' confidence, but also to help students feel happiness.

Music teachers should teach singing carefully. For instance, they should pay attention to gesture in singing, a way of breathing, rhythm and pitch matching abilities; teachers also should provide opportunities for singing performance and create appropriate settings, based on the content of songs for students. Furthermore, the music teachers should notice voice change in students to guide children how to use their voice in a healthier way. Choral singing is an opportunity for pupils to work with other children; this should be encouraged by music teachers as well (MoEPRC, 2011, 2022).

Since 2001, the National Standards for the Music Curriculum for nine-year compulsory education proposed that singing should merge with singing games, playing musical instruments and dancing, and this was extended by the MoEPRC into the latest Standards (MoEPRC, 2011, 2022). This proposal is in line not only with modern conceptions of musical performance in ancient China, particularly in the prosperous period of the Tang dynasty (Yang, 2019), but also with the idea of the 'comprehensive music education' of the Kodaly method, which is a popular method of teaching singing to children, and used internationally (Houlahan & Tacka, 2015, p. 20), including in China[36].

Overall, these set of principles imply that China has built a system to teach singing in Primary schools, but the system still continues to be reviewed and updated (Xie, 2005). It also seems that the National Standards requires appropriately competent teaching strategies to teach singing. But the teaching quality of music as a subject in Primary schools is unequal within areas in China (MoEPRC, 2011). In the real world, where economic development is unevenly spread across eastern and western China, and between urban and rural areas, the national policy on education has been harder to implement in some Primary schools (e.g., Xu, 2019).

1.4.2 Introduction to Singing Development from Infancy to Primary School

Previous studies have agreed that singing is usually a developmental process (Papageorgi et al., 2022; Pfordresher, 2022) and this perspective is reviewed briefly here to provide a general context for the singing behaviour of song singing for Primary school students.

[36] 因"乐"相伴，快乐同行——河北省唐山市开平区幼儿园骨干教师柯达伊教学法线上培训活动实录（二十） - 北京音协柯达伊音乐教育 (bj-kodaly.com)

Before a child is born, they can hear a melodic shape and rhythm from the mother's voice while in the womb during the last three months of pregnancy (e.g., Lecanuet, 1996; Lecanuet et al., 1992; Woodward, 2019). When infants crawl, they might listen to music from family members, musical toys, or media at home (Barrett, 2019; Reynolds, 1960; Williams et al., 2015; Wu, 2005; Wu & Welch, 2022). Once infants can walk and run, they are also likely to have greater access to the wider local society (Marsh, 2012), and be exposed to more diverse music embodying the dominant local culture, including song lyrics (Saltari & Welch, 2022).

During the Kindergarten and early Primary school period, research into pre-school families in Australia revealed that parents' musical behaviours were primarily song-based (Williams et al., 2015). Previous studies (e.g., Aherne, 2011; Kirkpatrick, 1962; Michaud, 2014; Shelton, 1965) agreed that a richer maternal home musical environment tended to help young children to develop their singing better compared to homes where there is less opportunity and encouragement to sing.

In the middle to older Primary school period, while younger children may continue to enjoy singing with their family members, older Primary school students are reported to prefer to sing alone in a safe environment, such as their bedroom (Welch et al., 2009b). They are also more likely to vocally match pitches of songs more accurately than younger children (Welch et al., 2009b). In this period, the local Primary school community, involving peers and teachers, could positively or negatively impact on children's self-identity as singers, based on their experience of collective singing in their schools (Welch, 2017). Furthermore, negative experience in childhood could impact subsequently on their self-identity as a singer in their later life (Cuddy et al., 2005; Whidden, 2009). However, people's self-judgement of their singing ability might not necessarily match their real singing ability (Demorest et al., 2017; Knight, 2010; Welch, 2017; Welch et al., 2009b), although their musical self-concept could be seen as a prediction of their perceived and actual vocal pitch-matching accuracy (Demorest et al., 2017).

In terms of the relationship between pitch discrimination and vocal pitch matching, for Kindergarten and young Primary school children (< 8-year-old), Geringer (1983), Cooper (1992) and Phillips and Aitchison (1997) agreed that there was a non-significant or low correlation between pitch discrimination ability and vocal pitch-matching. For older Primary school children who matched pitches precisely, they tended to discriminate pitches correctly (Cooper, 1992; Geringer, 1983; Pedersen & Pedersen, 1970), but the correlation was not significant for older pupils with lower pitch-matching accuracy (Geringer, 1983; Porter, 1977).

1.4.3 Benefits of Singing

Many examples of singing have been widely reported by the media during the lockdown period in various countries. In China, for example, songs have been composed by music teachers and sung by students to show their respect and love for health professionals, such as doctors and nurses[37]. In Italy, a professional singer sang *O sole mio* on his balcony, which was reported as giving heart to the many people who heard it and also encouraged others to imitate[38]. In another example, medical staff in Wales were recorded singing collectively a version of *Bridge Over Troubled Water* to lift their spirits and those of others, becoming an online hit[39], with many other examples being reported by the BBC[40]. Overall, these personal experiences confirm the author's previous biographical experience that singing can be a powerful, valuable and unique tool. The observations confirmed for the author that we should teach children to sing, and – as far as possible – to sing in-tune, as this seems to be a way to encourage children to feel more confident when they are singing.

For the author, an answer concerning the significant of singing was revealed through her personal experiences during the current coronavirus pandemic. During this critical period, she lived in London and saw that her neighbours' young children sometimes sang spontaneously when they were playing in their gardens, and that singing expressed an inner emotional state of happiness. In another example, despite the lockdown and 'staying at home', the author sometimes sang using a singing app when she felt the need to sing, which also (incidentally) made her landlady happy. It became apparent to the author that singing may be beneficial for mental health – a finding reported elsewhere (e.g., Saarikallio et al., 2020; Vaudreuil et al., 2019; Welch et al., 2020).

1.5 Studies Related to Primary School Students' Singing Behaviour

Although developing aesthetic value as the key target has enhanced the status of music education in Primary schools in China (MoEPRC, 2011, 2022), it is also achieved by other music perspectives, such

---

37 http://bj.people.com.cn/n2/2020/0409/c82846-33938122.html

38 https://www.bbc.co.uk/news/video_and_audio/headlines/51886547/coronavirus-italians-sing-from-their-windows-to-boost-morale

39 https://www.bbc.co.uk/news/uk-wales-politics-52644558

as developing music ability by understanding varied musical elements (including pitch-matching accuracy) and meaning of lyrics (MoEPRC, 2011, 2022). The current section reviews international and Chinese studies related to Primary school children's singing behaviour and development to identify research gaps.

1.5.1 International Studies

The study of children and young people's singing behaviour has been a long established research topic in Western academia, which has included studies from varied perspectives, such as vocal range (e.g., Buckton, 1977; Jersild & Bienstock, 1934; Moore, 1991; Pabon et al., 2014; Paulsen, 1895; Wilson, 1970; Young, 1971); comfortable singing range (Cobes, 1969; Geringer et al., 1980; Joyner, 1971; Welch et al., 2009b); and attitudes towards song singing (e.g., Mizener, 1993; Welch et al., 2009b). These studies have been based on diverse viewpoints which can assist us with understanding singing from a broader perspective.

Although singing has been an essential activity of music lessons in many Primary schools across the world (e.g., Canada[41], China[42], England[43]), nevertheless, previous studies have concluded that many children may have a problem with singing in-tune (e.g., Welch, 1979a, 1979b), although an alternate interpretation of the data is that singing is a developmental process and that in-tune singing emerges with experience (Welch, 1985, 2006, 2015), but 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 from varied perspectives. 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). While these studies have created a general view of children's vocal pitch-matching accuracy for songs, they tend to be summative in their judgements, but with little reporting of the detail of any errors.

Overall, in Western-focused countries, there has been research available on how children's

---

41 http://www.edu.gov.on.ca/eng/curriculum/elementary/arts18b09curr.pdf

42 http://www.moe.gov.cn/srcsite/A26/jcj_kcjcgh/200106/t20010608_167343.html

43 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/425601/PRIMARY_national_curriculum.pdf

singing develops, such as in the UK (e.g., Welch et al., 2009b, 2011), the USA (e.g., Moore, 1991; Rutkowski, 1996, 2018), Australia (Barrett, et al., 2020), mainland Europe (e.g., Mecke & Sundberg, 2010), Japan (Welch & Murao, 1994), and Latin America (IIari, 2006). However, until now, there are limited published research data on children's singing in China, especially in the context of the economic migration from rural to urban settings, although there is a new national research project (The Monitoring Centre for the Compulsory Education, Ministry of Education of the People's Republic of China, 2020), and there was a large Government-funded study led by Guo (1999) about music education in Primary and Secondary schools in China.

1.5.2 Chinese Studies

Based on Chinese journal reports from music teachers teaching at Primary schools in China, many teachers (e.g., Cui, 2018) also found their students could not sing in-tune. It has been suggested that this limited singing ability may impact negatively on their aesthetic development, both during the Primary school period (Zi & Zhu, 2006) and in their later musical education.

Many music specialists, who are usually school music teachers, have published Chinese papers to explain why students sang out-of-tune – based on their personal teaching experience (e.g., Mei, 2017; Yang, 1982a, 1982b; Ye & Xing, 2010a, 2010b). Many Chinese music teachers appear to think that not hearing pitches accurately is the key to children's inaccurate pitch-matching (e.g., Hei, 2015; Lei, 2006; Liu, 2001; Mei, 2017; Xu, 2018; Xu, 2021; Yang, 2015; Ye & Xing, 2010a, 2010b), although fewer teachers have argued that most students could discriminate higher or lower pitches and direction of pitches (Ye & Xing, 2010b), especially for matching models provided by their music teachers or peers. However, compared with an explanation of elements influencing vocal pitch-matching accuracy of song singing, limited Chinese studies have reported Primary school students' vocal pitch-matching accuracy of song singing among varied school Grades and sex, although a few limited studies have focused on Chinese kindergarten's singing behaviour (e.g., Feng et al., 2013; Zhao, 2020a, 2020b).

The research led by Guo (1999) focused on general music education for Primary, Secondary and High schools (6y+ – 18y+), including singing ability, with participants from across 20 provinces, cities and municipalities in China. The study concluded that most students liked their music lessons; music in a social environment had a significant and relatively positive influence on school music education, especially for older Primary school students. Overall, the study provided a valuable context for understanding music education in Primary schools in China, based on fieldwork data at that time. The study measured singing ability by the number of songs that students could sing without reading any lyrics and melodies. It reported that around 50% of Primary school students could sing 5-10 songs without any extra help. Older and urban Primary students could sing more songs without extra help than younger and rural-based Primary students. This finding is in line with previous studies about assessed singing behaviour of different aged Primary school students (e.g., Welch et al., 2009b, 2011).

However, more detailed information on singing behaviour is needed in order to understand Primary school students' singing behaviour in China, particularly as the Guo study was completed over two decades previously.

One project recently carried out in China related to describing Primary school students' singing abilities. It is the national test of all subjects taught in Primary and Secondary schools in China. Every subject, including music, is tested every two years. The aim of the project is to evaluate the quality of teaching for every subject in the Chinese compulsory education system. It is a continuing project, in which music has been tested twice, once in 2016 and again in 2018 (The Monitoring Centre for the Compulsory Education, Ministry of Education of the People's Republic of China, 2020). For the music test, solo singing, including vocal pitch-matching accuracy, is one of the perspectives to be measured. Four other perspectives include rhythm, fluency, expression, and clarity of lyrics, so the application of sung vocal registers in song singing is not explicitly involved. The national music test randomly selects Grade 4 and Grade 8 students from varied economic backgrounds in each province in China to provide a general picture of Primary and Secondary school students' musical ability at both national and regional levels and thereby to evaluate the quality of the teaching received by these students. However, the project's data collection did not involve students from other compulsory school Grades for reasons of cost and the large population size in China.

Given the widespread inclusion of singing as a core activity in Primary school, it seems vital to have some understanding of children's singing behaviours and development according to age and sex, as well as how these might be influenced by locality. This should allow a more suitable match between appropriate pedagogy and curricula for singing and children's current competency levels.

Music education policy in China operates at different levels. The central Government produces National Standards for the music curriculum, but these are interpreted at regional and local levels, depending on the available resources and internal school priorities. Consequently, music teachers have some discretion about what to teach, but this will also be limited by the choice of music textbooks in the school. Where there are additional (ex-curricular) musical activities, such as an orchestra, then this will provide a greater choice for repertoire.

1.6 Research Aims

In the context described above, the current PhD study is focused on the singing behaviour of Primary school students from all Primary school Grades in Hunan Province, China. It could also play a role as a case study for a national Chinese study. It is hoped that the research could go some way to filling the gap of in our knowledge of singing behaviour of other Primary School grades in China, which is a gap in limited previous studies reporting Chinese children's singing behaviour.

The **first research aim** is to understand the nature of Chinese children's pitch-matching accuracy within song singing in a micro (detailed) manner using detailed note-by-note analyses, filling a gap in the academic research literature that could lead to a greater understanding of children's pitch-matching accuracy in song singing.

The **second research aim** is to understand the general development of song singing for Chinese Primary school children from perspectives of vocal pitch-matching accuracy and vocal registers' application of song singing. Six independent variables – children's ages, sex, song, school, geographic location and inferred family income – were selected to describe Chinese Primary school children's singing behaviour in terms of vocal pitch-matching accuracy and the application of sung vocal registers in song singing.

The **third research aim** is to compare the pitch-matching accuracy of song singing between the current study and the previous English studies based on the National Singing Programme *Sing Up.* The English programme was aimed at providing a professionally-based singing experience for Primary school pupils in England each week during 2007-2012, funded by the UK Government (Saunders et al., 2011; Welch et al., 2009a, 2010). The aim of the thesis seeks to understand the possible impact of culture on children's singing behaviours and context using similar research tools in both countries.

1.7 Research Questions

Based on these research aims, three dominant research questions and their sub-questions have been asked:

Research question 1:

- How can participants' singing behaviour of song singing be illustrated from a micro perspective?

Research question 2:

- What is the general singing behaviour in selected song singing for Primary school students in China?
- 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 these participants?
- What are the similarities and differences for data of the general singing behaviour of song singing between the current study from China and studies collected from the National Singing Programme *Sing Up* in England, UK?

Research question 3:

- What is relationship between the micro and macro perspectives in explaining participants' singing behaviour?

The 'micro' perspective is required to provide more details of children's singing behaviour, such as characteristics of consistency of sung key centres and vocal pitch-matching accuracy for different types of musical intervals for songs. This is assessed by the development of a new tool. In contrast, the 'macro' perspective is focused more on a general trend of singing behaviour and development and draws on two established assessment protocols.

Overall, these questions are asked to improve our knowledge of vocal pitch-matching ability of songs from both micro and more macro perspectives for Primary school children in China. Comparing data from the current study and data from the *Sing Up* programme can also help us understand children's singing behaviour across these two countries.

1.8 The Structure of the Thesis

Chapter 2 reviews children's singing behaviour from a micro (detailed elements of the melody) perspective and Chapter 3 reviews it from a more macro (general, whole song) perspective. Chapter 4 describes the research methodology. Chapter 5 and Chapter 6 report the results of the current study from a micro and macro perspective, respectively. The results reported in these two chapters are discussed in Chapter 7. Finally, a summary of findings, potential applications, limitations, and future studies are discussed in Chapter 8.

Chapter Two

Reviewing Singing Behaviour from Musical Elements - a Micro Perspective

In the current chapter, three singing concepts were defined initially: vocal pitch-matching accuracy, vocal register use and the combination of these as singing behaviour. Then, the process of producing a pitch (fundamental frequency) and elements controlling vocal pitch-matching accuracy during singing were reviewed to understand how pitches are accurately matched. Furthermore, vocal pitch-matching accuracy for musical elements, including related singing materials and measurement, were reviewed to understand vocal pitch-matching accuracy for songs.

2.1 Definition of Three Singing Concepts

2.1.1 The Definition of Vocal Pitch-Matching Accuracy

This section focuses mainly on three tuning systems of musical intervals and the labels chosen to describe their level of reproduction accuracy.

(a) The Tuning Systems of Musical Intervals

Nowadays, the three most popular tuning systems of musical intervals are pure intonation, Pythagorean tuning, and equal temperament, which are based on different interval ratios.

Pure intonation is based on whole-number ratios (such as 3:2 or 4:3), and is widely used in Indian music, where an octave is divided into 22 Shrutis (Danielous, 1995, 1999).

The Pythagorean scale is based on the ratio of perfect fifths (3:2) and octaves (2:1) only (Sethares, 2005), and is named after Pythagoras, an ancient Greek philosopher (600 BCE), to whom the scale has been attributed. It can be used to create a simple pentatonic scale of five notes, or a larger scale of eight or more notes, depending on how often you repeat the pattern of perfect fifths. Its Chinese name is *Sanfensunyifa* (三分损益法), attributed to Guan Zhong sometime in the Spring and Autumn period (approximately 771 to 476 BCE) (see 'Guangzi · Diyuanpian' (管子·地员篇) and *Lvshichunqiu · Yinlvpian* (吕氏春秋·音律篇). The Chinese version of the scale contains five notes, named gong (宫), shang (商), jue (角), zhi (徵), and yu (羽), which equate to *do, re, mi, sol*, and *la* in Western music. Pythagorean tuning was widely used in ancient Chinese music, such as guqin music *Guanglingsan* (广陵散) and *Jiukuang* (酒狂), and Chinese folk music, including Tibetan music (Guo, 2013) and Hua'er (a high pitched folk singing music which is popular in the northwest of China) (Yang & Welch, 2016). The music usually also has special gliding tones.

Equal temperament divides an octave into 12 equal semitones, a scheme that emerged around the same time in China and Europe. It is attributed to Zhu Zaiyu in the Ming dynasty of China in 1584 (see *Lvlvjingyi* (律吕精义) and *Yuelvquanshu* (乐律全书)), and in Europe, is attributed to Simon Stevin (ca. 1605) (Lindley, 2020). It has been widely used in Western music since the 18th century. Since the early 19th century, some Western melodies were introduced to China by a group of people who studied in Japan (Guan, 2014). The original lyrics of the songs were replaced with Chinese ones, and were popular with teenagers. Many Chinese songs have subsequently been composed using equal temperament.

In order to be able to compare the findings of the current study with those of previous studies conducted mainly in a Western musical context, this thesis uses equal temperament as its tuning system of musical intervals.

(b) The Labels Describing Pitches Sung Inaccurately

Based on equal temperament, earlier studies commonly used inappropriate labels to describe sung pitches, which did not accurately describe the phenomenon, but rather were dependent on the listener's reaction to the singing. Labels to describe people who could not sing pitches accurately, for example, include 'growlers', 'grunters', 'crows', and 'backward singers' (Bentley, 1968). Such labels can also be found in informal contexts in Chinese society. However, they have since disappeared and more polite descriptive labels have been substituted in research since the 1970s.

The most prevalent current descriptive phrase may be 'poor pitch singer' (Welch, 1979a) which was first used by Roberts in his doctoral thesis of 1972. It is much less offensive than previous words and describes the singing quality of pitch accuracy only. It has been used by scholars from the field of both music education (e.g., Turoy, 2017; Welch, 1983) and cognitive science (e.g., Dalla Bella & Berkowska, 2009; Pfordresher & Brown, 2007). As this phrase cannot pinpoint the level of inaccuracy, scale intervals or cents (1 semitone = 100 cents) have usually been used to supplement it (e.g., Pfordresher & Brown, 2007; Welch, 1998).

'Non-singer' (Murry, 1990) is a similar phrase. As it does not contain the word 'pitch', it is not as specific as 'poor pitch singer', although it also tends to describe inaccurate sung pitches. Another phrase, 'inaccurate singers' (Bradshaw & McHenry, 2005; Watts et al., 2005), similarly means that participants have problems matching pitches, but it additionally contains people who have a problem discriminating pitches.

Like 'tone deafness', 'tune deafness' or 'note-deafness' are other common phrases used to describe people having difficulty with pitch accuracy (e.g., Kalmus & Fry, 1980; Loui et al., 2008; Simcox & Allen, 1878). They seem to be used specifically for discriminating pitches, which may be more appropriate for so-called congenital amusia (e.g., Ayotte et al., 2002; Peretz et al., 2002), which is estimated to affect around 4% of a population (Peretz et al., 2003), but it is not necessarily true if participants do not have a problem with discriminating pitches. Pfordresher and Brown (2007) suggest that other deficits, including those in the sensorimotor system, may better explain sung pitch inaccuracy for people in general.

'Monotone' was used by Bentley (1968) to describe participants who have difficulty in discriminating small pitch intervals, while Joyner (1969) used it to describe people who could not recognisably sing melodies in their original key. The listening and singing tasks from these two studies did not reach an agreement on the definition of 'monotone'. Furthermore, the word 'monotone', which suggests singing melodies using one pitch only, seems unrealistic and misleading.

Furthermore, Pfordresher and Larrouy-Maestri (2015) use the phrase 'vocal pitch imitation deficit' specifically with tasks requiring imitating pitches only.

Overall, these different names reveal different perspectives in identifying the cause of vocal pitch inaccuracy. Research studies have described levels of accuracy of sung pitches; attempted to locate the causes of inaccuracy and set different tasks for testing accuracy (discriminating, matching, and singing). Despite the variety of names and tasks, there is still no standard way to describe pitch sung inaccurately, nor on what counts as 'inaccurate'. Welch (1985), Welch et al. (1991, 1997) moved away from labels that focused on the perceived deficit, but rather saw singing behaviour on a continuum in which improvement was not only possible but commonplace, and so created the term 'developing singers'. This approach has been supported empirically in the evaluation of children's comparative singing behaviour by age for the National Singing Programme *Sing Up* in England (e.g., Welch et al., 2010), covering over 11,000 pupils.

(c) Definition of Pitch Accuracy for the Current Thesis

This thesis focuses on describing the level of pitch accuracy when familiar songs are being sung, not on discriminating or imitating pitches or pitch phrases, so the description 'poor pitch singer' is used to describe matching a criterion song relatively out-of-tune in different ways. In order to more precisely describe the poor levels of pitches' accuracy, a four-point rating scale Vocal Pitch Matching Development (*VPMD*) (Welch, 1998) and an invented system, called the Melodic Analysis of Pitch-Matching (*MAPM)* system, based on an acoustic analysis have been used in this thesis.

Previous studies agreed that children's vocal pitch-matching accuracy of musical elements and of song singing was different. Generally, the development of vocal pitch-matching accuracy of musical elements seems faster than that of song singing. For instance, when comparing vocal pitch-matching accuracy between musical elements and songs, Welch (1998) reported that a significant improvement for musical elements was displayed from five to six years old, but a similar improvement for song singing was not achieved until seven years old, hypothesised as due to song singing requiring lyrics as an additional variable. Consequently, these two singing behaviours, musical elements and song singing, are reviewed separately in the current thesis.

2.1.2 Defining the Use of Vocal Registers

Previous studies did not agree on names to describe the same or different vocal registers. Welch (2022) defined the term vocal registers as follows:

> Vocal registers are associated with different patterns of vocal fold vibration and perceptually distinct vocal qualities. Register terminology is quite varied, the same terms often being used for different registers. However, it is commonly agreed that there are at least three basic vocal registers—modal, falsetto and vocal fry (Sundberg, 2018). Modal register is commonly used in normal speech and is characterized by relatively thicker vocal folds and complete glottal closure during each vibratory cycle. This register has been subdivided in singing into lower and upper sub-registers, often associated with the terms 'chest' and 'head' by singers because of the reported physical sensations of resonance involved (Welch & Sundberg, 2002). (Welch, 2022, p. 386)

The most common name for a low register for speaking and singing was 'chest register' (e.g., Bernardoni et al., 2014; Kochis-Jennings et al., 2012; Large, 1969; Titze, 1979), and other names were also used, such as 'heavy register' (Wurgler, 1990) and 'lower register (Thurman et al., 2004). These various names were based on sensations of vocal tract resonance and timbre in singing, respectively.

When naming a higher vocal register, 'falsetto' (e.g., Kerdar et al., 1987; Titze, 1979; Vennard & Hirano, 1970), 'head' (Kochis-Jennings et al., 2012; McGraw, 1970), 'upper register' (Thurman et al., 2004), and 'light' (Henrich, 2006) have all been applied. Both 'falsetto' and 'head' use part of the vocal folds to vibrate (Henrich, 2006; Thurman et al., 2004).

These different names are partly the result of a disagreement between the fields of music education and physiology on the appropriate nomenclature for various vocal registers. The literature defines a vocal register from an acoustic or physiological perspective and this is mainly based on findings from adults, where pitches of one vocal register share a similar vocal quality or production by the vocal mechanism (Henrich, 2006; Hollien, 1974). In the current study, data exploring the application of vocal registers could be inferred from the participant children's singing of the three target songs, but has not been analysed as part of this thesis. In line with the names applied in many previous studies, the current author uses the conventional terms 'chest' and 'head' registers to represent different configurations of the vocal mechanism for the varied timbres during children's singing in the current study.

Similarly, previous studies did not reach an agreement about the number of vocal registers. Most studies have suggested two to three vocal registers for song singing (Bernardoni et al., 2014; Kochis-Jennings et al., 2012; McAllister et al., 1993; Wurgler, 1990). The transition between two registers might not be an independent register (Van den Berg et al., 1960), but an overlap (Roubeau et al., 2009). Other vocal registers, such as 'pulse' (Titze, 1988; Thurman et al., 2004) and 'whistle' (Thurman et al., 2004; Titze, 1994), are not considered here as they are not commonly used in children's song singing.

2.1.3 Defining Singing Behaviour

The definition of singing behaviour in the current study is defined based on vocal pitch-matching ability and also the use of vocal register for songs. Using these two singing perspectives facilitates the research aim of the thesis, which is to describe the fundamental singing ability of participant children in China.

2.2 Reviewing Elements Influencing Singing Behaviour

To better understand children's singing behaviour, elements influencing each of the two related perspectives (vocal pitch-matching ability and the use of vocal register) are reviewed as follows.

2.2.1 The Physical Perspective of Producing Fundamental Frequency

Overall, previous studies (e.g., Kayes, 2015; Moore et al., 2005; Watts et al., 2005) agreed that the production of sound requires air support from the respiratory system that stimulates vocal folds to vibrate. The original voice is then shaped by the vocal tract to produce the final voice we hear. The current section describes how the fundamental frequency is produced by the vocal mechanism.

The main organs of the respiratory system include the lungs, the rib cage, the diaphragm, the abdominal wall, abdominal contents, and the pulmonary airways (including nose, mouth, larynx, trachea, and bronchial tubing system). They deal with the gaseous exchange by inspiration and expiration (Kayes, 2015). Subglottal pressure created by expiration stimulates the vibration of the vocal folds (Kayes, 2015).

The overall sub-glottal pressure appears to be greater for children than for adults. Sergeant (2015) found that younger children usually used higher percentages of their lung capacity (around 50% to 60%) to sing or speak than adults, although the size of their lungs was smaller than those of adults (Sergeant, 2015). While this high lung capacity can increase vocal loudness, it also increases the challenge for young children to control their pitch level during singing. This is because it is easier to sing a higher pitch with strong air pressure in the sub-glottal area (Baken, 1998; Thurman et al., 2004), especially when the structure of children's vocal mechanisms is less rigid (Sergeant, 2015).

The vocal folds are layered (see Figures 1 and 2 of Hirano, 1988, p.52). According to findings on the structure of an adult's vocal folds reported by Hirano (1988) and Kayes (2015), vocal folds are covered by a thin capsule called an epithelium to hold their shape. During the process of vibration of vocal folds, the epithelium moves most noticeably. The lamina propria, including three layers, is below the epithelium. The superficial layer of the lamina propria lacks elastic and collagenous fibres, while the intermediate layer of vocal folds mainly consists of elastic fibres. The deep layer mainly consists of collagen fibres. The intermediate and deep layers are transition layers, called the vocal ligament. The body of vocal folds consists of vocalis muscle under the deep layer. The layered structure of vocal folds allows them to be shaped like a thin, a thick, a long, a short, a stiff, or a soft string (Hirano, 1988).

However, the structure of the vocal folds of children is different from that of adults (Hirano, 1988). When a child is born, the lamina propria is loose (Sato et al., 2001) and is a uniform structure without the ligament (Hirano et al., 1983; Sato et al., 2001). For an infant, the elastic fibres can be seen in the lamina propria at low density, but they develop faster once collagenous and reticular fibres appear (Sato et al., 2001). The second layer of the lamina propria appears in the second month, while the third layer does not appear until around the age of seven (Boseley et al., 2006). At the age of seven, the superficial layer accounts for 22%, which is the same as that of adults. From birth to age 10, the superficial layer tends to be thinner, while the intermediate and deep layers of the lamina propria tend to be thicker (Boseley et al., 2006), which makes the vocal folds to be more adjustable. The lamina propria mature during adolescence (Hirano et al., 1983).

The developing of children's vocal folds means that their geometry changes. For instance, the length of vocal folds increases over time before age 20 (Titze, 1994). Furthermore, the sex difference in the vocal folds' length is small for infants. It then increases with age until age 20, when boys generally have longer vocal folds. Because the geometry (length, depth, and thickness) and the stiffness and stress conditions of vocal folds have a great influence on the fundamental frequencies arising from the vibration of the vocal folds, longer vocal folds suggest lower fundamental frequencies (Titze, 1994) and a wider comfortable pitch range when using a modal register to sing (Kayes, 2015) (see more details about vocal range and comfortable singing range in Section 2.5.1). The stiffness of the vocal folds means the elastic restoring force after deformation, while tension or stress is the vocal folds' mechanical state (Zhang, 2016). When stiffness or tension increases, the fundamental frequencies of vocal folds increases and perceived pitch is higher (Zhang, 2016).

Figure 2.1 The Structure of the Vocal Mechanism

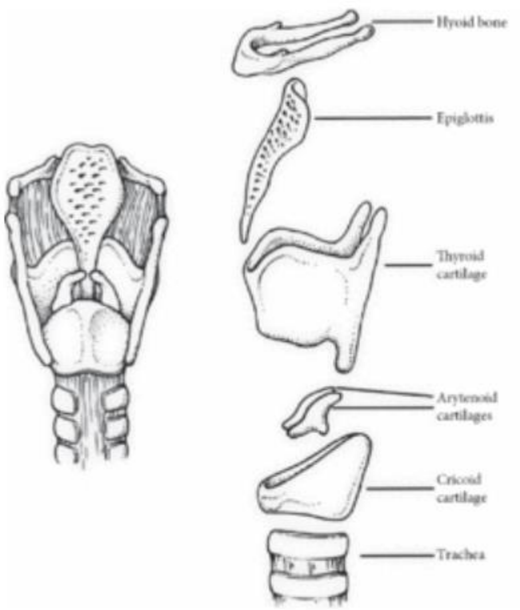


Figure 1.3 Larynx, rear and lateral view, exploded.

Reproduced from Kayes (2004), p. 20. Copyright © 2004, Bloomsbury Publishing PLC.

Fundamental frequency in voicing is determined by the number of vocal fold cycles per second that relates to the geometry of vocal folds (length, tension, and mass) (Zhang, 2016). The tension of vocal folds is controlled by their relative length (elongating or shortening) (Zhang, 2016). Mass is the mass of vocal folds that is set to vibration (Zhang, 2016). Increasing tension, decreasing mass and lengthening of the vocal folds would increase the fundamental frequency (Zhang, 2016).

The increasing or decreasing fundamental frequency for vocal folds needs assistance from cartilages and muscles of the larynx, especially for the three attached cartilages: the cricoid cartilage, the paired arytenoid cartilages, and the thyroid cartilage (see the location and shape of the cartilages in Figure 2.1) and their related muscles, including cricothyroid muscles (CT), thyroarytenoid muscles (TA), and posterior cricoarytenoid muscles (PCA), lateral cricoarytenoid muscles (LCA) and interarytenoid muscles (IA) (Kayes, 2015) (see a figure for the muscles in Figure 1, Zhang, 2016, p.2016).

- The CT is thought of as the biggest inner muscle of a larynx (Kayes, 2015). Activation of the CT can elongate the vocal folds, so they become thinner, and the stiffness and tension of the body and cover layers of vocal folds increases (Kayes, 2015; Zhang, 2016). Consequently, this action leads fundamental frequencies to be increased due to the short cycle of the vocal folds' vibrations for pitch increase (Baken, 1998).

- When activating PCA, LCA, or IA, the length of vocal folds increases significantly, which is good for singing at a high fundamental frequency (Kayes, 2015). However, the greater stretching of vocal folds makes it snap back more quickly (Baken, 1998), which may lead to singing out-of-tune due to more control difficulties.
- On the other hand, activation of the TA muscles can shorten the vocal folds, so they thicken, and their stiffness and tension in the body layer increases (Zhang, 2016). Then, the fundamental frequency decreases because of a long cycle of vocal fold vibrations.

Compared with the respiratory and phonation systems, the resonator system (vocal tract) has a smaller impact on fundamental frequency. The main function of the resonator system is to modify the timbre of the voice. Consequently, its influence on fundamental frequency will not be discussed further.

2.2.2 The System Related to Control of Pitch in Singing

Control of pitches in singing appears to be more complicated than producing a fundamental frequency (a pitch). It needs cooperation between the auditory system and the motor system, and cooperation within the neurological system which connects the two. It is also influenced by the singing environments. Such cooperation and influence are explained below.

(a) The Auditory System

Firstly, to match the pitches of a familiar song, Watts et al. (2005) suggested that people need to hear and discriminate the pitches of a song initially. While previous studies (e.g., Cooper, 1992; Geringer, 1983; Phillips & Aitchison, 1997) did not find a strong correlation between pitch discrimination and vocal pitch-matching ability for younger children (age < 8), they agreed that the correlation was significant for older Primary school children (*cf*. Pedersen & Pedersen, 1970). It is assumed that the reason for the low correlation for younger children might be because their ability to discriminate pitch developed faster than their vocal pitch-matching ability.

Previous studies suggested that Primary school children's vocal pitch-matching ability varied significantly in response to the timbre of a singing model. Overall, previous studies agreed that children matched human voices, especially a child's and women's voices, better than the timbre of musical instruments (Gratton, 1989; Hutchins & Peretz, 2012; Moore et al., 2008).

(b) The Neurological System and the Motor System

After hearing the pitches of a song, working memory temporarily saves them. Matching the pitches of a song with neutral syllables seems easier than matching those with lyrics (Goetze, 1985; Pereira & Rodrigues, 2019). When visual gestures, giving guidance on pitches' height and their melodic contour are provided (Frederickson, 1992; Iacarino, 2018; Liao & Davidson, 2007, 2016; Martin, 1991), children are more likely to match pitches more accurately.

During the process of matching pitches, breathing patterns could be flexibly adjusted based on the length and rhythmic structure of pitches (Kayes, 2015). When matching a high pitch, high subglottal pressure is required, which could be achieved by preparing more breath support (Kayes, 2015).

Furthermore, the length, tension, and mass of vocal folds need to be adjusted under the control of the central nervous system when matching different pitches (Hirano, 1988). As reviewed in Section 2.2.1, the CT muscle is the dominant muscle to be activated when vocally matching a higher pitch. The PCA muscle might be also activated if the pitch was significantly high. On the other hand, the TA would be activated when matching lower pitches (Hirano, 1988; Kayes, 2015; Zhang, 2016).

Furthermore, when singing using different vocal registers, the cooperation among muscles, cartilages, and vocal folds seems to be more complicated. The vocalis muscle (a part of the TA muscle) showed a remarkable change in activity when transferring among vocal registers (Hirano, 1988). In more detail, when using the chest register to sing, its activation was the strongest, but it was the weakest when using the falsetto (Hirano, 1988). In other words, when singing from chest to head registers, the activity of vocalis is found to be reduced, and vice versa (Hirano, 1988). The activity of LCA is slightly stronger when using the chest and head register than the falsetto (Hirano, 1988). However, in one study the activity of the CT muscle was reported to not vary significantly when using different vocal registers (Hirano, 1988) for the same pitch, i.e., singing the same pitch with different tone qualities.

Transferring between vocal registers is reported to be mainly dominated by the ratio of vocalis muscle tension to others (Hirano, 1988). In this study, when the ratio of vocalis changes gradually, the change between vocal registers could not be identified. However, if the ratio changed suddenly, the shift of vocal registers would be identified easily due to the change of timbres (*cf.* related to 'lift points' suggested in Rutkowski's 1996 *SVDM* scale of children's vocal register use).

Another essential concept for vocal register is the lift point, especially for untrained singers. The lift point means a transition pitch point or range between two adjacent vocal registers. It can be influenced by age, an individual's physical vocal folds, personal singing ambitions, experience, and the characteristics of the target sung pitches (e.g., musical interval) (Wurgler, 1990). A focus on characteristic lift points can help in the measurement of the quantity and type of vocal registers (Vennard & Hirano, 1970), such as used in Rutkowski's (1996) *SVDM* scale. If some participants cannot find their upper register, the lift point may be the highest pitch that they can sing comfortably (Paulsen, 1895; Young, 1971). Participants' ability to pass the lift point and transfer from one register to another defines the extent of their vocal range of song singing.

Previous studies suggest that many participants did not know how to transfer between vocal registers. For instance, Wurgler (1990) reported that although 88% of Primary school children could find their chest and head registers, only 40% of them could transfer between the vocal registers. Consequently, transition is a skill that requires systematic experience and is supported by training. For instance, Jones (1979) suggested that singing wider skips could help children gain in the use of the upper vocal register.

As a child's larynx is still developing, the larynx structure is not stable, so the lift point seems not to be a fixed pitch point, but rather a small pitch range. Generally, previous studies agreed that the lift point of children runs from G4 to A4 (A4 = 440 Hz) (Joyner, 1969; Krause, 1983; McAllister et al., 1993; Wassum, 1979; Wilson, 1970; Wurgler, 1990). At the start of adolescence, boys' lift point tends to be lower than that of girls. The lift point of female adults is around D4 – F4, which is reported to be lower but more stable than that of children (Keidar et al., 1987; Titze, 1988).

After vocal pitches are produced, they should be heard by the ears to allow for further adjustment (Watts et al., 2005). Previous studies (e.g., Estis et al., 2011; Watts et al., 2005) agreed that auditory feedback was important for vocal pitch-matching accuracy. They suggested that real-time visual feedback (*cf.* Welch et al., 1989) could help to develop vocal pitch-matching accuracy, possibly because this feedback helped the adjustment of vocal folds.

While previous studies found that vocal pitch-matching accuracy was influenced by the three above processes: hearing pitches, regulating pitches by adjusting the vocal folds controlled by the sensory system and the action of the general motor system (e.g., young children's vocal range might be limited, if they did not use different vocal registers), and auditory feedback, previous studies also agreed that each of the processes could be trained to improve vocal pitch-matching ability (e.g., Estis et al., 2011; Sergeant, 2015; Watts et al., 2005).

(c) The Influence of the Singing Environment

The literature suggests that singing environment at home and in schools can have a positive or negative impact on children's singing behaviour and development. Previous studies have suggested that children's vocal pitch-matching ability has a positive correlation with the number of musical activities experienced within the family, such as singing with members of their family, which can help children to feel prepared for and confident to sing in school (Aherne, 2011; Kirkpatrick, 1962; Michaud, 2014; Shelton, 1965; Welch et al., 2009b). Furthermore, if family income allowed, parents might provide opportunities to develop children's musical (including singing) ability if they thought that their child had potential musical talent (Wu, 2005). Australian research suggests that mothers will take their children to community-based music programmes if these are available because they value the possibility of extending the home musical provision (Abad & Barrett, 2023). Also, mothers often sing songs as part of the socialization process, including to comfort their babies (Trehub & Gudmundsdottir, 2019).

The arrangements made when organising singing activities in a Primary school can also influence children's vocal pitch-matching ability by taking into account the impact of current vocal range, musical task and the auditory environment. For instance, children are more likely to match a song with an appropriate pitch range (Petzold, 1963).

The auditory environment could vary in a singing class. When singing in a group, children were more likely to match the pitches of a song more accurately if there was no accompaniment (Atterbury & Silcox, 1993; Estis et al., 2011; Houlahan & Tacka, 2015; Smith, 1973). This is because singing in a group could both reduce students hearing their own voices during group singing, and could also provide more accurate pitch models from others in the immediate group if they are matching pitch accurately. However, when singing individually, a music teacher's accompaniment might help to improve a child's vocal pitch-matching accuracy (*cf.* Wise & Sloboda, 2008). This is because the child could not only listen to their voice when singing alone, but could also correct their inaccurate pitches by following the main melody played on a piano. Hearing an accurate model of singing in a group might depend on the size of the group and classmates nearby.

For children of various age groups, the influence of singing individually or in a group seems to have a different impact on vocal pitch-matching accuracy. Previous studies (Goetze, 1985; Leighton & Lamont, 2006) tended to agree that young children benefited more from solo singing than group singing. For older children, Nichols (2013) summarised that previous studies (Cooper, 1995; Green, 1994; Nichols, 2013; Smith, 1973) utilizing pitch patterns as singing tasks tend to report that singing individually was more accurate, but the studies involving whole song singing were more likely to report the individual as more accurate when singing within a group. This might be because children who could not sing in-tune could adjust their pitches based on singing from nearby classmates. Consequently, music teachers could mix children who can and who cannot sing in-tune to create a network of support for the latter's vocal pitch-matching ability.

An early study by Plumridge (1972) reported that children were more able to sing in-tune when in a small group and this was reported by Welch et al. (2009a; 2009b) as a technique used in gathering their *Sing Up* research data. Children seem to be better able to tune into another voice at the same pitch with a similar timbre rather than an instrument, although there is research to suggest that having a piano melodic model as an accompaniment can help (Jones, 1974; 1979). There is also evidence that lyrics can interfere with vocal pitch accuracy in singing (Levinowitz, 1989; Welch et al., 1997).

While so many variables can influence the development of children's singing behaviour, it seems that musical specialists who teach the music curriculum may have a better chance to set an appropriate singing environment for children because they have more knowledge. In many countries, including China, music is a specialist subject, but it is often taught by general teachers who usually have to teach several subjects in school. Previous studies suggested that general teachers might have some personal musical experience (Bhachu, 2019; Henley, 2017; Hennessy, 2000), but they usually have no formal or only limited music training before their teaching career (Cuadrado & Rusinek, 2016) and, consequently, they might have a relatively limited understanding of the musical elements of effective song singing and singing skills (Bhachu, 2019; Holden & Button, 2006). Therefore, children's vocal pitch-matching ability may vary according to the kind of pedagogical singing environment.

2.3 Singing Materials Used to Assess Aspects of Vocal Pitch-Matching Accuracy for Musical Elements

To test children's vocal pitch-matching ability, previous studies applied varied types of musical elements, mainly focusing on single pitches, musical intervals and patterns. The pitch range of single pitches applied was usually within an octave. The singing material selected to test vocal pitch-matching accuracy of musical intervals in previous studies included two formats: two pitches in a single interval (AB) (e.g., Moore et al., 1995/1996; Wolf, 2005) and four pitches with two repeating pitches (AABB) (e.g., Demorest & Pfordresher, 2015; Demorest et al., 2018; Nichols, 2016). The pitch range of these musical intervals was usually within an octave. The vocal pitch-matching accuracy of these two formats can vary, as perfect unison (AA) may assist in singing the tested musical interval (AB) more accurately because it is hypothesised that the initial perfect unison creates a more stable condition of the larynx.

Previous studies reported that the degree of complexity of vocal pitch-matching for different musical intervals varied[44]. According to synthesis from studies by Jones (1971), Sinor (1984), Moore et al. (1995/1996) and Wolf (2005), musical intervals can be split into three groups based on their complexity level: (i) perfect unisons to major thirds (except minor seconds); (ii) perfect fourths and perfect fifths; and (iii) minor sixths to octaves, and minor seconds. Generally, previous studies (e.g., Jones, 1971; Moore et al., 1995/1996; Sinor, 1984; Wolf, 2005) agree that the musical intervals of the first group, especially unison, are generally easier to match accurately than those of the other two groups, although a minor second is also thought of as a difficult musical interval to match (Moore et al., 1995/1996).

Patterns of pitches were frequently chosen as test material in previous research, especially for studies testing the singing behaviour of kindergarten children and young Primary school students (e.g., Brophy, 1997; Cooper, 1995; Flowers & Dunne-Sousa, 1990; Geringer, 1983; Goetze, 1985; Young, 1971). Most patterns in previous studies were composed by the researchers who usually applied these within a controlled pitch range. For instance, the pitch range of two patterns composed by Goetze (1985) was a perfect fifth, and that of Geringer (1983) and Cooper (1995) was a major third and a perfect fourth, respectively. Almost all the musical intervals of the patterns used to test Primary school children's singing behaviour were from perfect unison to perfect fifths, with a majority ranging from perfect unison to major thirds, such as the patterns employed in Sinor (1984) and Wolf (2005). This corresponds to the proportion of musical intervals in many target songs used in previous studies (such as the three target songs chosen in the current doctoral study, see Table 3.3 in Chapter 3, Section 3.2.1).

---

[44] Augmented and diminished intervals are not included in the current review. The relative ease of matching the direction of musical intervals, being upward or downward intervals, has not been a major focus in the literature, although there is some evidence about the impact of vocal range limiting upward intervals as the child's vocal pitch accuracy is restricted by only using a speech (chest) register (Wassum, 1979; Welch, 1979).

2.4 Measurement Used to Explore Detailed (Micro) Vocal Pitch-Matching Accuracy for Musical Elements

A range of software has been used by researchers in previous studies to test children's vocal pitch-matching accuracy from an external perspective, such as Visi-Pitch (e.g., Alberston, 1979; Cooper, 1995; Goetze, 1985), a Kong AT-12 Auto Chromatic Tuner (Flowers & Dunne-Sousa, 1990), the BBC micro-computer-based SINGAD (Singing Assessment and Development) system (Howard & Welch, 1993), the Real-Time Features of the CSL (4100) software-hardware system (Western, 2002) and MultiSpeech (Hedden & Baker, 2010). Some software was originally designed for correcting the pronunciation of speaking, such as Visi-Pitch, while some was initially considered for improving the vocal pitch-matching ability for singing, such as the BBC micro-computer-based SINGAD. However, all software shares common characteristics, including collecting, visually showing, and analysing the mean fundamental frequency of vocal pitches of a pattern automatically. By using this approach, many characteristics of vocal pitch-matching accuracy could be explored (see review in Section 2.5).

Fundamental frequencies have been analysed by varied approaches. Many previous studies (e.g., Cooper, 1995; Howard & Welch, 1993; Welch et al., 1996; Western, 2002) analysed the mean fundamental frequency of vocally matched pitches of patterns. Goetze (1985) analysed the mean fundamental frequency for the highest and the lowest pitches of short patterns, because she believed that a mean fundamental frequency of all vocally matched pitches could not show accurate vocal pitch-matching accuracy when all pitches were matched with the same pitch height.

Furthermore, Goetze (1985) analysed the deviation between a target pitch and a matched pitch by cents. When all deviations were the same, a pattern was thought to be matched accurately, but in a different tonality. Different deviations meant singing in diverse tonalities. Using the data analyses, a relatively accurate pitch manner could be found. Deviations of cents could also track the change of key centres of vocally matched pitches. This approach could be applied when analysing a pattern with many pitches, such as a song. One drawback was that these absolute cents could not show types of vocal pitch-matching errors (sharpness or flatness).

Instead of analysing the mean fundamental frequency and deviation of vocally matched pitches as in many studies (e.g., Cooper, 1995; Goetze, 1985), Flowers and Dunne-Sousa (1990) analysed the number of pitches or patterns (total n = 20) matched absolutely, relatively accurately and relatively inaccurately, using within +/-50 cents (half a semitone) of a recorded vocal model as a standard to define an accurately matched pitch (*cf.* MultiSpeech used in the study of Hedden and Baker, 2010). They also analysed vocal pitch-matching of a self-selected song and a taught song by a rating scale according to the number of modulations used to sing a song. While using the two types (micro and macro perspectives) of measurements, more details of vocal pitch-matching accuracy for patterns and songs could be revealed. However, details, such as the type of vocal pitch-matching errors, including sharp and flat, could also not be provided in this study.

Similarly, Nichols (2016) also used the standard of +/-50 cents to measure an accurately matched pitch for *N* = 120 fourth-grade students in the USA, coding ‘1’ as an accurate pitch, and ‘0’ as an inaccurate one. The coding system can show the accuracy of each of the vocally matched pitches of musical elements in his study, but the coding ‘0’ cannot show the type and degree of inaccuracy for a vocally matched pitch.

Overall, the diverse approaches to analyses for vocal pitch-matching accuracy by the selected software can help to reveal the vocal pitch-matching accuracy of patterns from different perspectives. Because these previous studies analysed vocal behaviours using so-called objective software (i.e., software which was designed to reproduce assessments in a systematic and reliable manner), the reliability of the results from their diverse views was believed to be relatively high in these previous studies, either in online or a person setting (Honda & Pfordresher, 2023). However, although such software could show the mean of the fundamental frequency of sounds’ pitches, the results might not be in-tune in a musical sense, which also relates to the listener’s perception and relative tolerance for minor pitch errors within the whole performance (Welch, 1994).

The study by Elmer and Elmer (2000) reported a new approach to illustrating vocal pitch-matching accuracy by using diverse symbols. For instance, using a solid dot represented a stable pitch; and using a solid dot with a downward or upward-sloping trend line represented a stable pitch, but ending with an upward or downward glissando. Such an approach seems more musical than the other software approaches reported above, whilst also seeking to remedy the drawback of the currently available software at that time that cannot show the types of vocal pitch-matching errors (sharpness or flatness). However, the approach of Elmer and Elmer (2000) did not have the advantages of other software used in previous studies in that they could not reveal the size of any vocal pitch-matching errors. Consequently, a new approach is needed to show both the type (sharp or flat) and relative size of vocal pitch-matching errors.

It should be noted that several previous studies which tested vocal pitch-matching accuracy for musical elements (e.g., Honda & Pfordresher, 2023; Nichols, 2016; Levinowitz, 1989; Welch et al., 1997) have either explored the difference between with/without lyrics or asked participants to sing using neutral syllables (e.g., the /u/ vowel, or ‘la’) to reduce the potential impact of lyrics.

2.5 Review of Vocal Pitch-matching Accuracy of Musical Elements

Reviewing vocal pitch-matching accuracy for musical elements can help with understanding vocal pitch-matching accuracy for a song. In the current section, the following musical elements are reviewed consecutively: vocal range, vocal pitch-matching accuracy for melodic contour, consistency of key centre, and vocal pitch-matching accuracy for musical intervals and pitches. Finally, the findings of previous studies exploring vocal pitch-matching accuracy for musical elements are reviewed.

2.5.1 Reviewing Vocal Range and Comfortable Singing Range

Previous studies (e.g., Feng et al., 2013; Sinor, 1984) suggested that vocal range influenced the vocal pitch-matching accuracy of musical intervals or patterns. The vocal range has been widely studied over a long period and especially between 1880 and 2000. Based on the research by Titze and Sundberg (1992) and Henrich et al. (2005), the vocal range can be defined as the maximum vocal pitch range that a singer can produce. This maximum vocal pitch range is reported to relate positively to the use of vocal registers, to the number of elastic fibres in the vocal folds, and vocal intensity (Titze & Sundberg, 1992).

Earlier studies (Fieldhouse, 1937; Jersild & Bienstock, 1934; Paulsen, 1895; Wilson, 1970) suggested that Primary school children's vocal range tended to extend with increasing age (see Figure 2.2). For instance, Houlahan and Tacka (2015) suggested an expansion process of the vocal range for Primary school children in general. In more detail, they proposed D4 – B4 (A4 = 440 Hz) (a major 6th) for Kindergarten children and First Grade pupils; C4 – D5 (a major 9th) for Second to Third Graders; and C4 – Eb5 (a minor 10th) for Fourth to Fifth Graders. However, pitches higher than Bb4 would be difficult to match for children if they had not experienced or been educated to use a head register. Furthermore, children's pitch height of the vocal ranges for song singing is usually located in 'the lower part of the vocal range' (Welch et al., 2009b), although they could sing a wider range. Previous studies found that girls and boys tend to have a similar vocal range (see Figure 2.2). Wilson (1970) reported an interaction effect between pitch height and sex on vocal pitch-matching accuracy. Girls tended to sing high pitches more accurately, while boys did the opposite. However, both sexes' vocal range could be expanded by training (e.g., Gould, 1969; Roberts & Davies, 1976; Rutkowski, 2015).

Previous studies suggested that children's vocal range could be different according to the singing tasks. Compared with singing a song, children tended to take more risks to expand their vocal range when echoing patterns (Flowers & Dunne-Sousa, 1990), which might be because of more breathing support than when singing a song. A reason that children did not use a whole range of target singing materials might be that these materials' vocal range was over the limits of 'the child's physical or self-imposed vocal production limits' (Flowers & Dunne-Sousa, 1990, p.112). However, the children may produce a similar vocal range when asked to sing different songs. When Flowers and Dunne-Sousa (1990) asked kindergarten children to sing a self-selected song and a target song with an octave pitch range, they reported that the mean vocal ranges between the two singing tasks were close.

A comfortable singing range has also been paid attention to when exploring the vocal range in previous studies. The definition of a 'comfortable' level is somewhat subjective, usually dependent on the individual participant's personal feelings or the quality of voice as perceived by someone else (Geringer et al., 1980; Welch et al., 2010). Using the term 'a comfortable singing range' rather than a whole vocal range might be more suitable for explaining a normal singing range (Welch, 1979b).

Figure 2.3 (below) presents the comfortable singing ranges established by previous studies. Some original data that could not be traced from original sources were extracted from Welch's study (1979b). According to the figure, older Primary school children generally had a wider comfortable singing range than younger children. However, the sex difference in the comfortable singing range for Primary school children has received less attention in previous studies.

Comparing Figures 2.2 and 2.3, Primary school children's comfortable singing range tended to be narrower than their whole vocal range, although usually closer to the lower part of the whole vocal range.

Figure 2.2 Vocal Ranges Analysed by Age and Sex Reported by Previous Studies

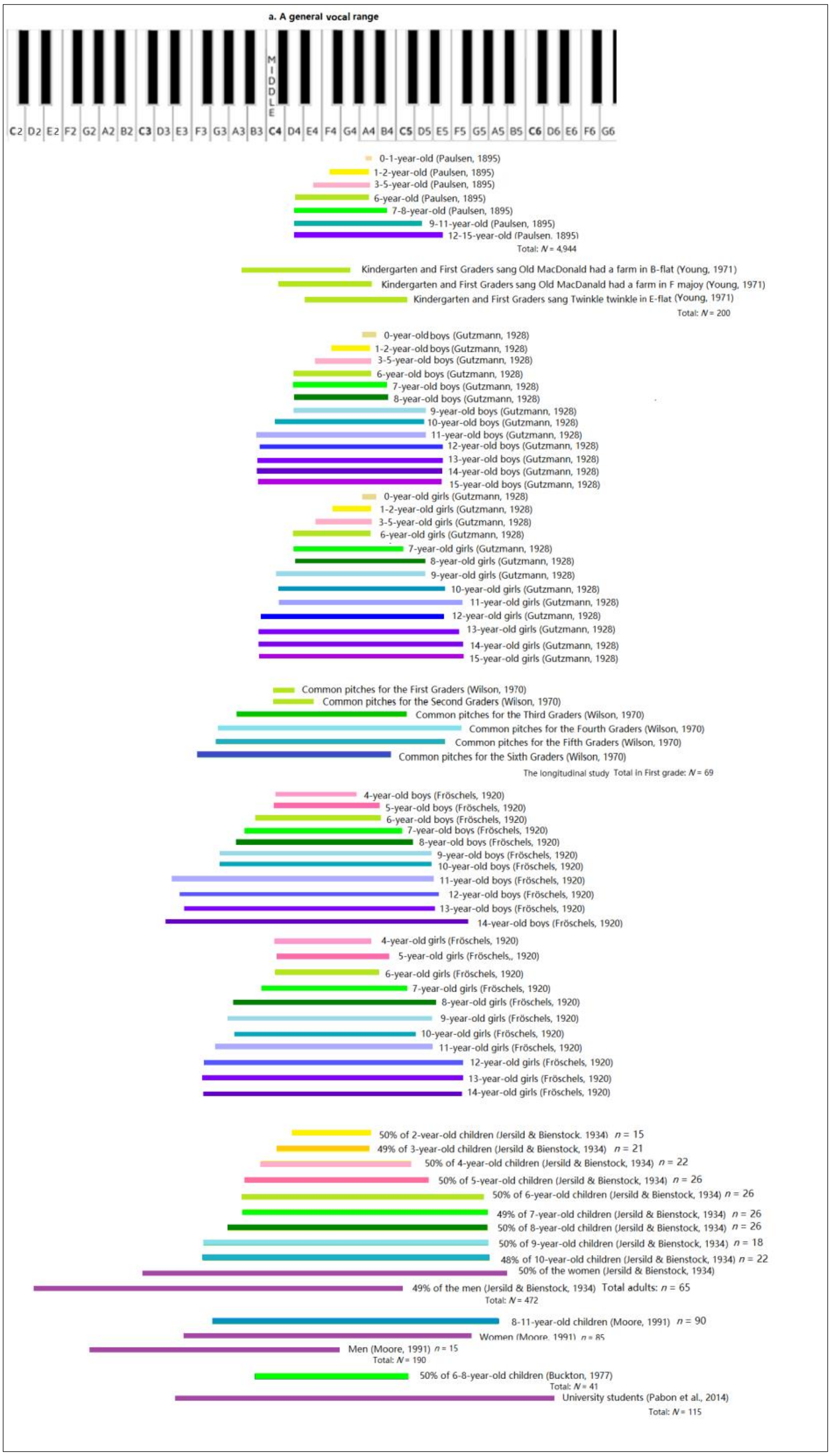
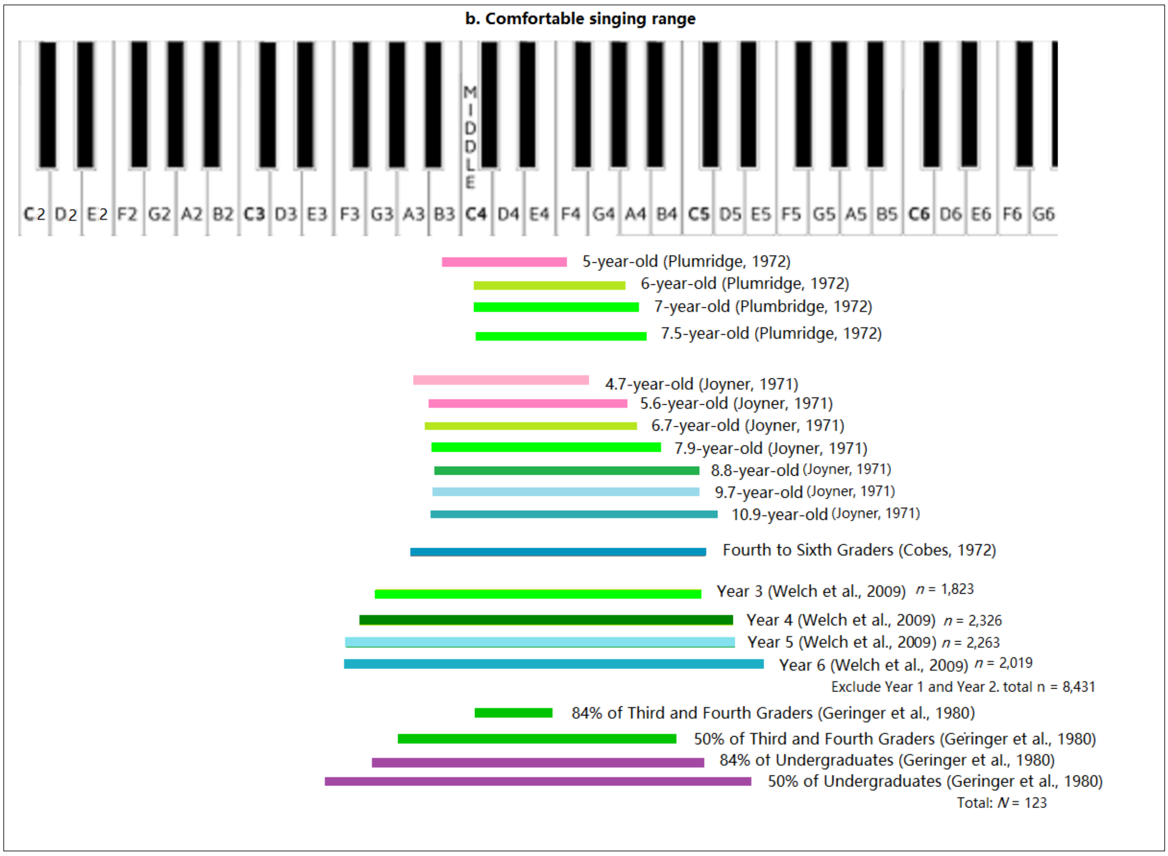


Figure 2.3 Comfortable Singing Range Reported by Previous Studies



Note. Fourth to Sixth Graders in the study of Cobes (1972) would be around ages 9y+ to 11y+. Years 3 to 6 in the study of Welch et al. (2009) were aged from 7y+ to 10y+. Third and Fourth Graders in the study of Geringer et al. (1980) were aged from 8y+ to 9y+.

2.5.2 Accuracy for Melodic Contour

Overall, previous studies suggested that children generally could match melodic contour accurately, especially for short singing materials (e.g., Demorest et al., 2018; Leighton & Lamont, 2006; Nichols, 2016; Welch et al., 1996; Wolf, 2005). For instance, Flowers and Dunne-Sousa (1990) invited kindergarten children to echo 20 pitch patterns with ascending, descending, combined ascending and descending, and single-pitch contours. An accurate pitch was defined as within +/- 50 cents compared to a target pitch. They reported in a general summary that most 4- to 5-year-old children could match a melodic contour accurately. This finding was supported by Moore et al. (1995/1996) who reported that *N* = 480 children aged 6 and 9 from Argentina, Poland, Spain, and the United States generally matched 97% of melodic contours of musical intervals accurately. However, it should be noted that related singing materials used in previous studies were musical intervals or patterns, so it is uncertain whether children can match melodic contour accurately in terms of song singing, although this was reported by Welch (1986) in his own data on children's singing.

Furthermore, children seem to show different vocal accuracy when matching varied melodic contours. Sinor (1984) reported that *N* = 96 American kindergarten children from diverse socioeconomic backgrounds when asked to match 4-note patterns (simple ascending and descending contours and V-shaped contours) were less accurate.

2.5.3 Consistency of Key Centre

Previous studies (e.g., Flowers & Dunne-Sousa, 1990) agreed that vocal pitch-matching accuracy for children in song singing was related to their ability to maintain a key centre. Children who changed tonal centres more frequently generally matched pitches and melodic contour of patterns less accurately than children who changed fewer tonal centres.

The inconsistency of sung key centres seems to relate to vocal range. Jones (1979) reported that when pitches' height was above children's comfortable vocal range, children tended to change their key centres. They might or might not maintain the new key centre. Participants in Jones' (1979) study were *N* = 36 children who could not sing in-tune from Grades 2, 3, and 4 in the United States. Jones (1979) assumed that the change of tonalities due to the children's limited vocal range might be related to a lack of experience in transferring between vocal registers.

Previous studies (e.g., Flowers & Dunne-Sousa, 1990; Welch et al., 1996) suggested that the ability to maintain a key centre during matching musical elements or songs seems to improve with increasing age. Flowers and Dunne-Sousa (1990) reported that kindergarten children aged 4 and 5 tended to use fewer key centres to sing a self-selected song than children aged 3. However, the age difference was not statistically significant when children were asked to sing a target song. This might be because the target song *The Little White Duck* with an octave pitch range was a relatively difficult song for all kindergarten children to match. The study by Mang (2006) supported the view that any observable age difference could be less when children were asked to sing, for them, a difficult target song.

2.5.4 Accuracy for Musical Intervals or Patterns

Previous studies agreed that matching melodic contour accurately was easier than matching relative pitches, while relative pitches were simpler to match than absolute pitches (Flowers & Dunne-Sousa, 1990; Moore et al., 1995/1996). A pattern usually involves ≥ 4 pitches or melodic phrase. It might involve several bars of the melody. See annotated examples in the updated Figure 4.4. Vocal pitch-matching accuracy for musical intervals, which relates to the accuracy of relative pitches, is reviewed below.

(a) Vocal Pitch-Matching Accuracy for Diverse Types of Musical Intervals

Previous studies agreed that vocal pitch-matching accuracy for different types of musical intervals was diverse. They concurred that unison was one of the relatively easiest intervals to match, irrespective of the types of singing tasks, age groups, and culture. Matching unison musical intervals for *N* = 480 aged 6 and 9 participants from Argentina, Poland, Spain, and the United States in the study of Moore et al. (1995/1996) explored this concept, and matching unison in two songs for *N* = 130 aged 3 to 6 Chinese children was also researched in Feng et al.'s (2013) study.

A falling minor third is thought to be a relatively easily matched musical interval as reported in previous studies (e.g., Jones, 1979; Sinor, 1984). It has also been frequently used spontaneously by children and adults in musical conversational singing, such as calling for the dog to come back (Anderson, 1991). The related task in the studies of Jones (1979) and Sinor (1994) was intervals or patterns. Jones (1979, p.181) even suggested that 'accuracy increased when the stimulus pitches were at least a minor third apart'. His suggestion was based on training for improving vocal pitch-matching accuracy by gaining the use of what he called the upper vocal register.

Minor 2nd and an octave were reported as two difficult types of musical intervals to match accurately, irrespective of the direction of the interval (ascending or descending), age group, type of singing tasks, and culture (Feng et al., 2013; Moore et al., 1995/1996; Nichols et al., 2023; Tillotson, 1972). This might be related to limited skills in a micro adjustment of vocal folds for the minor 2nd and poor skill in transferring between vocal registers for the octave.

Previous studies (e.g., Chen-Hafteck, 1999; Moore et al., 1995/1996; Tillotson, 1972) were inclined to agree that vocal pitch-matching accuracy tended to decrease with increasing size of musical intervals. However, they also revealed that vocal pitch-matching accuracy for different types of musical intervals might be not linear, especially when the impact of directions (ascending or descending) on the pitch accuracy of musical intervals were considered in song singing (Feng et al., 2013). Welch et al. (1996) and Moore et al. (1995/1996) agreed that children usually matched flat (under the target pitch) when they could not match a musical interval accurately. While the singing materials of most of the previous studies were musical elements, it is uncertain whether these findings would apply to singing a song.

On the other hand, modality (major or minor) (Wolf, 2005), repeated notes (Sinor, 1984), stepwise motion (Sinor, 1984), and successive skips in the same direction (Sinor, 1984) were not found to have a significant influence of vocal pitch-matching accuracy on patterns. These findings were based on matching patterns, so their reliability needs to be checked when a singing task is more complex, such as in a song.

(b) Reviewing Vocal Pitch-Matching Accuracy for Musical Intervals or Patterns by Age, Sex, Training and Country

As vocal pitch-matching accuracy for musical intervals or patterns has received much attention from previous studies, this allowed the author to analyse them by key independent variables in the thesis: age, sex, training and country.

(i) Age

Previous studies tended to agree that older children matched musical intervals (Feng et al., 2013; Moore et al., 1995/1996; Sinor, 1984) and patterns (Cooper, 1995) more accurately than younger children. These studies collected data from diverse contexts with different singing tasks. For instance, *N* = 130 kindergarten Chinese children in the study of Feng et al. (2013) were asked to match two composed songs; *N* = 480 aged 6 and 9 children from four Western countries (Argentina, Poland, Spain, and the United States) were asked to echo 16 musical intervals in Moore et al.'s (1995/1996) study; and *N* = 96 American kindergarten children from diverse socioeconomic backgrounds were asked to match 12 patterns randomly selected from 48 patterns with four pitches in each pattern in Sinor's (1994) study. That these studies, undertaken using diverse age groups, singing tasks and measurements, had similar findings improves the perceived reliability of their individual results.

While the above findings are based on an overall singing task, Feng et al. (2013) reported that the accuracy of a type of musical interval was similar across age groups (ages 3 to 6). Given that the number of previous studies is limited, further studies are needed to explore the vocal pitch-matching accuracy for different types of musical intervals by age.

Furthermore, Moore et al. (1995/1996) remind us that the age difference in matching musical intervals can be smaller when musical instruction was limited. This was found based on the data from the Polish participants, while it was not found in the data of children in the other three countries, Argentina, Spain and the United States. Instead of the cultural difference, Moore et al. thought that it was because of different frequencies of weekly musical instruction.

Finally, the longitudinal study of Welch et al. (1996) reminds us that vocal pitch-matching accuracy for musical intervals or patterns is generally easier than for a song. Their later study (Welch et al., 1997) reported that children's comparative vocal pitch-matching accuracy for songs did not display a significant improvement during the first two years up to age 7, although the same children showed a steady improvement in vocal pitch-matching for the target songs' patterns when deconstructed from the melodies during all three years. As relatively limited evidence on vocal pitch-matching accuracy for musical intervals in a song has been collected, this area requires more attention from researchers.

(ii) Sex, Sex*Country, Sex*Training and Age*Sex

The effect of age on vocal pitch-matching accuracy for musical intervals or patterns is greater than that of sex (Moore et al., 1995/1996). Generally, previous studies (Moore et al., 1995/1996; Pereira & Rodrigues, 2019; Welch et al., 1997) agreed that girls matched musical intervals and patterns more accurately than boys. This seems to be true in many countries. For instance, Moore et al. (1995/1996) reported that girls aged 6 and 9 tended to match 16 musical intervals more accurately than boys from the two age groups, based on data collected from Argentina, Poland, and Spain.

While previous studies suggest that girls tend to match musical intervals or patterns more accurately than boys of the same age, this tendency could be influenced by children's age and singing ability, teachers' encouragement and sample size.

The sex difference in the vocal pitch-matching accuracy for musical intervals or patterns generally was reported to be small for kindergarten children (Leighton & Lamont, 2006; Moore, 1994; Sinor, 1984; Welch et al., 1997), but it was greater for Primary school children.

In terms of children's singing ability, Moore (1994) reported that no significant difference by sex was found in the pitch-matching accuracy of musical intervals singing for older, talented Primary school singers aged 8 to 11. The talent of these participants was judged by their specialist music teachers. Consequently, these participants may not have been representative of ordinary Primary school students.

Furthermore, the American boys in Moore et al.'s (1995/1996) study also showed strong vocal pitch-matching accuracy in musical intervals thanks to regular encouragement from music teachers. This indicates that boys' vocal pitch-matching accuracy for musical intervals can be improved with encouragement and attention from music teachers.

Similarly, Cooper (1995) did not find a statistically significant sex difference in vocal pitch-matching accuracy for a five-note pattern for Grades 1 to 5 children in the United States, using an objective measurement (Visi-Pitch) was applied to measure the song data. Because of large standard errors, the no sex difference finding might be because of the small sample size by sex in each of the

School grades (numbering between 14 to 19 children). However, it should be noted that there was a tendency for girls to match pitches more accurately than boys when echoing in a group.

2.5.5 Accuracy for Pitches

For the vocal pitch-matching accuracy of single pitches, previous studies paid attention to the influence of a starting pitch, pitch height, location, and the number of pitches. These elements are explained as follows.

In terms of the starting pitch, when Flowers and Dunne-Sousa (1990) asked *N* = 93 kindergarten children to sing a target song *The Little White Duck*, only about one-third of participants began on the given starting pitch. However, the characteristics of the starting pitches (sharp or flat) that most of the children used are unknown.

According to Trollinger's (2003) study, as pitch height increased, boys' vocal pitch-matching accuracy was substantially poorer, but that of girls was only slightly influenced. Her data was collected from *N* = 70 pupils aged 36 to 71 months who sang three patterns with the same intervals but different pitch heights.

Previous studies agree that the location of pitches has a significant impact on the vocal pitch-matching accuracy of pitches. For instance, Welch et al. (1996) and Feng et al. (2013) agreed that children tended to match pitches at the beginning of a sequence more accurately compared to pitches in the middle or at the end. This might be explained by diverse control of breath support.

Furthermore, previous studies (e.g., Sinor, 1984; Welch et al., 1996; Wolf, 2005) agreed that children matched pitches more accurately when there were fewer pitches. However, it should be noted that studies by Sinor (1984) and Welch et al. (1996) defined the number of pitches differently. Sinor (1984) meant the number of different pitches, while Welch et al. (1996) meant the total number of pitches. These different definitions could have a significant impact. For instance, if a target song (e.g., *Little Donkey,* see Figure 4.4) involved many unisons, its difficulty in vocal pitch-matching might be reduced due to the potential high pitch-matching accuracy for unisons reported above.

2.5.6 Reviewing Vocal Pitch-Matching Accuracy of Musical Elements in a Song

Compared with singing a song, echoing musical elements usually involves simpler rhythm and text (e.g., a neutral syllable in the study of Sinor (1994)), fewer pitches for each pattern, and more opportunity to breathe between a pair of singing tasks in previous studies. Echoing musical elements requires short-term memory, but singing a familiar song from memory requires more long-term memory. Consequently, it might be no surprise that previous studies have agreed that vocal pitch-matching accuracy for musical elements tended to be easier than vocally matching a song (e.g., Demorest et al., 2018; Nichols, 2016; Welch et al., 1996). Nichols (2016) reported that the correlations between musical elements (single pitch, interval, and pattern) and a song (*Jingle Bells*) were from $r = 0.55$-0.65, suggesting a relatively strong correlation. However, limited previous studies measured vocal pitch-matching accuracy for songs by an objective method because such research is very time-consuming. Two related studies are reviewed below.

Feng et al. (2013) explored the details of vocal pitch-matching accuracy for songs by asking $N =$ 130 kindergarten children aged 3-6 from Jiangsu Province, China to sing two composed songs. They reported that the total vocal pitch-matching accuracy across the two target songs was significantly correlated. However, the results were also different for each song, as children from each age group matched the second target song better than the first. Furthermore, based on their observations in regular music lessons, Feng et al. (2013) also reported that kindergarten children repeated their vocal pitch-matching errors when singing a familiar song, and also that they might sing one pattern in-tune and another out-of-tune.

Asztalos (2021) observed the singing of $N = 100$ Grades 4th and 5th students from ten Primary schools in Hungary for five continuous years. He reported that many students used the chest register to sing. This led to 'voice cracks in the area of register transition' (p. 48), and the poor vocal pitch-matching for high tones with limited dynamics. He also summarised three types of out-of-tune singers: (1) singing in a speech range; (2) 'false singing' (p. 51), which meant missed matching a big music interval with a smaller one; (3) singing with a low pitch range. After two years of exercises with specific strategies, he reported that all these singing problems were remediated.

Nichols et al. (2023) analysed vocal pitch-matching accuracy of each pitch and musical intervals in $N = 37$ adults' singing performances of *Happy Birthday*, as measured by deviation in cents (100 cents = one semitone in equal temperament). Vocal pitch-matching accuracy for each pitch was based on a key inferred by the starting pitch. If a pitch was matched within 50 cents from a target (equivalent to a quartertone), it was classified as accurately matched. The authors reported that 91.89% (34/37) of adults could not match the octave of the *Happy Birthday* target song accurately. However, vocal pitch-matching accuracy for pitches following the octave leap were not significantly influenced by the poor sung accuracy of the octave, although there was a tendency for pitches before the octave to be matched inaccurately. They also reported that vocal pitch-matching accuracy for four particular musical intervals (namely the 4th, 7th, 10th, and 13th intervals in sequence from the beginning – including two perfect 4th, one perfect 5th and a unison), could highly predict the vocal pitch-matching accuracy for the sung octave in *Happy Birthday*. The Nichols et al. (2023) study is the first to reveal details of vocal pitch-matching accuracy for musical intervals within a target song, and using a relatively objective measurement. However, their analyses did not include a complete picture of vocal pitch-matching accuracy for other categories of sung musical intervals within a target song.

Generally, because only a limited number of songs have been measured from what this thesis is terming a 'micro' perspective (song elements), relatively few details about vocal pitch-matching accuracy for songs have been revealed by previous studies. Although the process is extremely time-consuming, exploring the details of vocal pitch-matching accuracy for musical elements in songs may help to reveal more clearly the development of children's singing behaviour.

2.6 Summary

Previous studies revealed the production of fundamental frequency and the system of pitch control during singing from a physical perspective. Varied types of singing materials (e.g., musical intervals and patterns) were applied to explore vocal pitch-matching accuracy for musical elements. These usually were measured by a relatively objective approach. During the Primary school period, children's vocal range and comfortable singing range tend to expand with age. Girls' vocal range might be similar to that of boys. Children generally matched melodic contours accurately, although some authors report that contours, such as simple ascending or descending, tended to be more difficult to match accurately.

The existence and use of limited vocal range seems to lead to inconsistent key centres. Previous studies generally agreed that older children sang more consistently than younger children, although the age difference tended to disappear when matching the pitches of a difficult song. Vocal pitch-matching accuracy for different types of musical intervals was found to be varied. Smaller musical intervals tended to be easier to be matched more accurately than greater intervals. Previous studies suggested that older children tended to match musical elements more accurately than their younger peers. However, more studies are required to explore the age difference in vocal pitch-matching accuracy for the same type of musical interval. Previous studies also reported that girls tended to match musical elements more accurately than boys. However, the age and sex differences could be reduced if singing training and attention from music teachers were provided.

Most children were reported to not start with a given starting pitch to sing. However, researchers in previous studies did not specify the characteristics. According to previous studies, children tended to match pitches at the beginning of a sequence more accurately. They also generally matched more accurately for singing materials with fewer pitches. It should be noted that the definition of the number of pitches used in previous studies was varied.

Previous studies agreed that musical elements were easier to match accurately than when collected together in whole songs. Consequently, it is unsure whether the results of vocal pitch-matching accuracy for musical elements are suitable to be applied to songs, especially as the latter have the additional variable of lyrics (language). Limited studies paid attention to vocal pitch-matching accuracy for songs. According to existing studies, a child's vocal pitch-matching accuracy for two songs for a child could be both related and also different when the target songs had different complexities. Children were likely to repeat their vocal pitch-matching errors during a song singing. Primary school students mainly used the chest register (low pitches) to sing a song, which led to poor accuracy for high notes. However, these singing problems could be remediated by training.

Lastly, most of the findings above were received based on research in Western cultures. The country difference in vocal pitch-matching accuracy within Western cultures seems to be not significant, but more related to singing experience and appropriate singing pedagogy. Consequently, it is uncertain whether these general literature findings will be appropriate to describe the singing behaviour of children from another culture, such as Asia.

Chapter Three

Reviewing Singing Behaviour from a Whole Song – a Macro Perspective

The current chapter reviews two components of children's singing behaviour (vocal register use and vocal pitch-matching accuracy) for whole songs in order initially to understand development being influenced by variables, such as singing materials, measurements, age, sex, and socioeconomic status. Then, the results of the two singing perspectives as measured by the above variables were compared to understand how their combined scores demonstrate an overall singing ability for songs. Finally, a theory is discussed to explain children's singing behaviour.

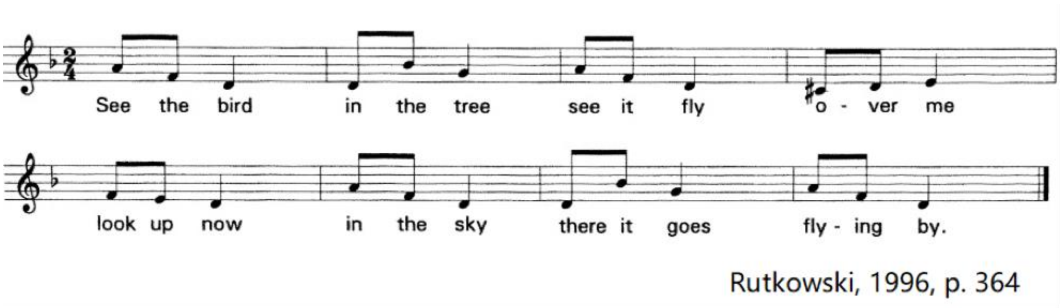
3.1 Reviewing Vocal Register Use for Song Singing

To better understand vocal register use, related singing materials and measurements were analysed by age, sex and socioeconomic status below.

3.1.1 Review of Selected Singing Materials

As transition from one vocal register to another can be influenced by the pitch range and musical intervals of the singing materials (Svec et al., 1999; Wurgler, 1990), Rutkowski (1996) composed a specific short song to test the application of vocal registers (see the song in Figure 3.1).

Figure 3.1 A Song Composed by Rutkowski (1996) to Test the Application of Sung Vocal Registers in Song Singing



The song's vocal range is a seventh with an upward minor sixth and an upward perfect fifth, which is a relatively difficult task (see more details of the complexity of singing tasks in section 3.2.1b). When the song was firstly used in America, then introduced to other countries, such as Portugal by Pereira and Rodrigues (2019), its melody was kept, but its lyrics were translated into Portuguese. When using Rutkowski's (1996) patterns to test vocal register use, previous studies (e.g., Pereira & Rodrigues, 2019; Rutkowski, 1997) usually asked young children to echo each bar phrase individually from an adult's model. This was thought to allow more air support from their lungs and more assistance from the researchers.

Differently, some studies (Mang, 2006; Welch et al., 2009a, 2009b) employed familiar songs, such as *Twinkle, Twinkle* and *Happy Birthday*, to test the application of vocal registers. Furthermore, using familiar songs as the singing tasks can also reduce the work for music teachers, as they can spend less time teaching these in their regular music classes.

3.1.2 Review of Measurement

The two most frequent measurements reported in the literature to test children's application of sung vocal registers in song singing are the original five-rating scale of Wurgler (1990) and the evolved nine-rating scale of Rutkowski (1996, see also 2015, 2018).

Figure 3.2 The Rating Scale of Wurgler (1990)

1	The student sings only in one register; range is severely limited.
2	The student independently chooses only one register, but can find a second register with help.
3	The student uses head and chest registers, but exhibits marked breaks in production.
4	The student can lighten production to increase range, but lacks a clear head tone.
5	The student makes smooth register transitions; there are no production problems related to range.

The content of Wurgler's (1990) rating scale is shown in Figure 3.2. Wurgler measured vocal registers by mainly distinguishing tone quality (including vocal timbre and volume). However, measurement by timbre is both objective and subjective as each timbre level is hard to define (McGlone & Brown, 1969). Consequently, the opinion of more than one judge is usually sought to improve the reliability of the results.

Whilst Wurgler's scale is essentially measured by sung timbre, the Singing Voice Development Measure (*SVDM*), suggested by Rutkowski (1990, 1996, 2015, 2018), uses perceived lift point and vocal range as the two key determinants (see Figure 3.3). The original 5-point continuous scale *SVDM* (Rutkowski, 1990) resulted from discussion between Rutkowski and several Primary school music teachers and drew on results from previous studies whose participants were mainly kindergarten children, including studies by Smith (1963), Gordon (1971), Joyner (1971), Young (1971), and Harkey (1978). These studies provided clues for register lift points (D4, F#4, Bb4) in the *SVDM* scale. The scale was designed to measure how children, especially children who were school Grade 1 or younger, used their voice to sing, and focused on the application of sung vocal registers.

Rutkowski believed that there were around three registers: low, middle, and upper, and divided her scale initially into a 5-point continuous scale based on singing range and lift points. After summarising the reported lift points for children and adults, Rutkowski (1990, 1996, 2015, 2018) used D4 (*cf*. Welch et al., 2009b), F4, and A#4 (A4 = 440 Hz) as the three critical lift points for her *SVDM* scale. Using these lift points, children's vocal register use could be divided from 'pre-singer' and 'speaking range singer' (a low register), 'limited range singer' and 'initial range singer' (a middle register), to 'singer' (an upper register) (Rutkowski, 2019, p. 7).

In 1996, another 4 ratings with a 0.5 point between each original two adjacent ratings were added to describe children who were inconsistent in using their singing voice between two vocal registers, which created a 9-rating scale (Rutkowski, 1996, p. 364). In 2016, she updated the 1996 version based on A4 = 440Hz (Rutkowski, 2015, 2018) (see Figure 3.3). Almost all previous vocal register studies before the current study for Primary school children use the *SVDM* scale – the nine-ratings scale, to measure the application of vocal registers in song singing.

Figure 3.3 The Singing Voice Development Measure (SVDM) Rating Scale of Rutkowski (2015, 2018)

1	***Presinger*** does not sing, but chants the song text.
1.5	***Inconsistent Speaking-Ranger Singer*** sometimes chants, sometimes sustains tones and exhibits some sentitivity 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).

The *SVDM* scale was the first scale to test primarily how children use their singing voice, unlike other existing scales, such as the rating scales of De Yarman (1971) and Ramsey (1983), which were designed to rate vocal pitch accuracy or combined pitch accuracy and use of the singing voice (Rutkowski, 1990).

Previous studies (e.g., Dansereau, 2005, Levinowitz et al., 1998; Runfola et al., 2012; Rutkowski, 2014; Rutkowski & Chen-Hafteck, 2001) reported a high reliability of the *SVDM* scale when they tested young children's vocal registers' use. However, its reliability seems lower for older children. Levinowitz et al. (1998) found that its reliability was reduced for Grade 6 students (around age 11+). This might be because puberty led to a decline in the pitch range and lift points for this oldest Primary school grade, and/or there is a tendency for older children to want to sound like adults.

When comparing the scales of Wurgler and Rutkowski to measure different aspects of vocal registers, the former scale focuses on the timbre of different vocal registers, while the latter one pays attention to lift points and vocal range, and has been more widely used. It seems that judgements based on more perspectives (lift points, change of timbres, and vocal range) for Rutkowski's scale are more seen as more objective than those based on timbre alone, as in Wurgler's scale. Furthermore, as the current author decided to compare the results of the current Chinese study with those of the English *Sing Up* study (see Welch et al., 2009b), it would be better to use the same measurement approaches – the *SVDM* rating scale (Rutkowski, 1996, 2015, 2018).

3.1.3 Review of the Influence of Age on Vocal Register Use

Most of previous studies agree that older Primary school children used statistically significantly more vocal registers than younger children. For instance, Mang (2006) reported that n = 56 nine-year-old (Cantonese monolinguals or English bilinguals) children used statistically significantly more vocal registers than n = 64 seven-year-old children in Hong Kong. In England, Welch et al. (2009a) reported that Year 6 children (age 10+) applied statistically significantly more vocal registers than Year 3 pupils (age 7+). In Portugal, Pereira and Rodrigues (2019) reported that 2nd Grade and 4th Grade children used significantly more vocal registers than kindergarten children. Although the three above studies included children from different areas (Hong Kong, England, and Portugal) and applied diverse singing materials (songs with diverse difficulties or patterns), each of them tested the age difference in vocal register use for children with a two-year age gap, as measured by the *SVDM* scale.

When the age gap was less than two-years-old, it seems more difficult to find an age difference in vocal register use, which could relate to similarities in physical size and development in adjacent age groups. For instance, Rutkowski (2015) did not find a statistically significant age difference in vocal register use between kindergarten children and 1st Grade children, although she applied the same patterns and measurements which were subsequently used by Pereira and Rodrigues (2019).

These previous studies primarily involved Western cultures, despite the study of Mang (2006) involving children from Hong Kong, and so it was uncertain whether Primary school children from a different culture, such as from the Chinese mainland, would show a similar trend for vocal register use by age.

Furthermore, the means of vocal register use for Primary school children varied across previous studies. While all the above previous studies measured vocal register use by Rutkowski's (1996) *SVDM* scale, it was evident that the mean scores for vocal register use reported by Pereira and Rodrigues (2019) (M = 3.60 to 4.55) and those reported by Mang (2006) (M = 4.15 to 4.43) were generally higher than the results reported by Welch et al. (2009a) (M = 3.26 to 3.60), albeit with the latter having a much larger number of participants. These varied means across the previous studies might be because of different degrees of related training, such as applied to all the children in the Pereira and Rodrigues (2019) study. Another potential explanation is due to the varied singing materials. For instance, while one of two target songs, *Twinkle, Twinkle* – whose pitch range is a major 6th, was applied in the study of Welch et al. (2009a), the song's relatively narrow vocal range might decrease the chance for children to apply more vocal registers. Overall, the relationship between vocal range of singing materials and vocal register use requires further research.

3.1.4 Review of the Influence of Sex on Vocal Register Use

Many previous studies agree that girls generally used more vocal registers than boys (e.g., Mang, 2006; Welch et al., 2009a), even for young pupils, such as three-year-old children in the pre-test of Dansereau's (2005) study. Children involved in the above three previous studies were aged from three to nine and were mainly educated in Western cultural settings. They were asked to sing a difficult target song (e.g., *Happy Birthday* in the study of Mang, 2006), a simple and a more complicated song (e.g., *Twinkle, Twinkle* and *Happy Birthday* in the study of Welch et al., 2009a), and to echo phrases of a criterion song (e.g., Dansereasu, 2005). Vocal register use of all the singing performances across the three previous studies was measured by the *SVDM* scale. The results suggested that the sex difference in vocal register use, weighted towards girls, generally was not influenced by the complexity of singing materials and age group. However, it was unclear whether the same conclusion could be reached for older Primary school children from another culture, such as school Grades 5 and 6 children from the Chinese mainland.

On the other hand, it emerged from the research that the sex difference in vocal register use could be reduced if children were trained in a rich singing environment. For instance, the sex difference in vocal register use was not found in the post-test of the study of Dansereau (2005) after *N* = 46 three-year-old participants experienced research-based musical engagement for 12 weeks. Similarly, Pereira and Rodrigues (2019) did not find a statistically significant sex difference in the vocal register use for *N* = 137 aged 4 to 9 children. According to information derived from personal contact by the current author during the fieldwork period with the Portuguese authors, the children involved in their study received professional training in sung vocal register extension.

3.1.5 Review of the Influence of Training or Socioeconomic Status on Vocal Register Use

A few previous studies have focused on possible socioeconomic influences on children's vocal register use. The study of Pereira and Rodrigues (2019) seems to support the positive impact of training on vocal register use. In their study, *N* = 137 students aged 4 to 9 from a private school in Lisbon, Portugal, were recruited. The students were provided with professional training on sung vocal register use. Their range of raw scores of the application of vocal registers measured by the *SVDM* scale was from 3.5 to 4.5, which was generally higher than those of other studies, such as School Years 3 and 6 children across England with varied socioeconomic backgrounds who did not receive the extra singing training (*M* = 3.26 to 3.60) in the study of Welch et al. (2009a). It is unsure whether higher mean scores reported by Pereira and Rodrigues (2019) were because of singing training, or higher social-economic status, or both. Their possible relationship could be that the opportunity of singing training from a music specialist was more likely to be provided under the high financial support from parents.

Levinowitz et al. (1998) suggested that singing training could help children use more vocal registers. There are differences in the use of the head register between untrained and trained singers. For all singers, only the edges of the vocal folds usually vibrate when using the upper register to sing (Titze, 1994), and the position of the larynx is higher than when singing in the chest register, which produces a smaller resonating space (Welch, 2022). The head register's resonating space is around 60% - 70% of that for the chest register (Bernardoni et al., 2014). The volume is usually low as the air support available might be not sufficient to vibrate tighter vocal folds, especially for high pitches. However, a trained singer may be able to keep the larynx in a relatively low position when singing high pitches and apply the head register by adjusting tongue and jaw actions (Welch, 2022). Trained singers usually have better breathing skills, which would likely enhance the beauty of the timbre and vocal volume when singing in the head register (Shipp, 1975).

3.2 Reviewing Vocal Pitch-Matching Accuracy for Song Singing

The structure of this section which reviews vocal pitch-matching accuracy for songs is the same as that for reviewing vocal register use (3.1 above). This opens with a review of the influence of singing materials and measurement approaches on scores of vocal pitch-matching accuracy for songs. Then, the influence of age, sex, and socioeconomic status on vocal pitch-matching accuracy for songs are reviewed subsequently.

3.2.1 The Influence of Singing Materials

Although previous studies agreed that there was a positive relationship for pitch-matching accuracy between accurate musical elements and song singing (Demorest & Pfordresher, 2015; Demorest et al., 2018; Flowers & Dunne-Sousa, 1990; Nichols, 2016; Roberts & Davies, 1975; Welch et al., 1996), they also agreed that children tended to match musical elements on their own more accurately than those of a song where music and lyrics are combined (Welch et al., 1996; 1997; 1998). The current section (a) lists a range of songs used to test vocal pitch-matching accuracy for songs in previous studies, and (b) analyses these subsequently from four following perspectives: starting pitch, vocal range, melodic contour and musical intervals. There are n = 19 songs used in previous studies listed in Table 3.1.

Table 3.1 A List of Songs Applied by Previous Studies to Test Vocal Pitch-Matching Accuracy of Children

Name of the Song	**Reference(s)**
America	Guerrini, 2006; Smith, 1973
Are You Sleeping?	Mizener, 1993
Bow Wow Wow	Green, 1994
Great Big House	Persellin & Bateman, 2009
Happy Birthday	Demorest & Pfordresher, 2015; Mang, 2006; Welch et al., 2009a, 2009b
In the Sea	Levinowitz et al., 1998; Pereira & Rodrigues, 2019
L'il Liza Jane	Persellin & Bateman, 2009
Jingle Bells	Apfelstadt, 1984; Mizener, 1993; Nichols, 2016
Jumbo Elephant	Levinowitz, 1987
Little Sir Echo	Western, 2002
Path to the Moon	Guerrini, 2006
Take a Trip (a composed song)	Miller, 2020
The British National Anthem	Joyner, 1969
The Jack in the Box	Levinowitz, 1987
The Little White Duck	Flowers & Dunne-Sousa, 1990
This Old Man	Mizener, 1993
Twinkle, Twinkle	Demorest & Pfordresher, 2015; Mizener, 1993; Welch et al., 2009a, 2009b; Western, 2002
Row Row Row Your Boat	Levinowitz et al., 1998
Yankee Doodle	Mizener, 1993
Did not report the name of a song	Leighton & Lamont, 2006; Ramsey, 1983; Smale, 1987; Welch et al., 1996, 1997, 1998
Self-selected Song	Flowers & Dunne-Sousa, 1990; Mizener, 1993; Robert & Davies, 1975
Judged by teachers directly based on their impression of students' regular singing	Davies & Robert, 1975

(a) A List of Songs Used in Previous Studies

As the table shows, most previous studies asked children to sing specific song(s). Three familiar songs, *Twinkle, Twinkle, Happy Birthday* and *Jingle Bells,* have been often selected in previous studies. Furthermore, many previous studies used a not previously known song taught by the music teacher. Some studies allowed children to self-select a song, or asked a music teacher to judge children's vocal pitch-matching accuracy for songs based on their daily observations.

For the self-selected songs, no agreement about the choice of songs was reached. For instance, Apfelstadt (1984) did not limit the options of songs for kindergarten children, and most children selected *Jingle Bells*. On the other hand, Mizener (1993) provided a list of songs from *Twinkle, Twinkle*, *This Old Man*, *Yankee Doodle*, to *Are You Sleeping?* for Third to Sixth graders. The list of songs offered was not reported in the studies of Roberts and Davies (1975). As a result, it seems that the difficulty level of songs selected by the children themselves was more likely to be outside the researchers' control than songs which they themselves had selected. Furthermore, their results could be more difficult to explain. For instance, Flowers and Dunne-Sousa (1990) reported that kindergarten children used a narrower singing range to sing their self-selected song than to sing a song set for them by the researchers. Three years later, Mizener (1993) reported that no significant age difference in vocal pitch-matching accuracy for a variety of songs selected by the children themselves was found, but when the children sang a specific song chosen by the researchers a significant difference by age was revealed. Overall, it seems that self-selected songs are not an ideal way to test Primary school children's vocal pitch-matching accuracy if the aim is to generate a comparison between children, and also the results from a variety of songs may make it more difficult to draw conclusions across different studies.

(b) Analysing the Complexity of the Criterion Songs

Song complexity appears to have the biggest impact on children' vocal pitch-matching accuracy for songs (Svec, 2018). In this chapter, a review of the literature found *N* = 24 target songs, including five unnamed taught songs from the studies of Leighton and Lamont (2006) and Smale (1987), that had been used in *N* = 27 previous studies[45] to test children's vocal pitch-matching accuracy for songs.

[45]Some criterion songs were applied in more than one previous study.

Table 3.2 Musical Characteristics of N = 24 Songs Applied by Previous Studies (↘ means a downward musical interval, and ↗ means an upward musical interval.)

Song	The starting pitch	Vocal range	No. of pitches	No. of musical intervals	The number of musical intervals by type and direction																						
					Unison	↘ Minor 2nd	↗ Minor 2nd	↘ Major 2nd	↗ Major 2nd	↘ Minor 3rd	↗ Minor 3rd	↘ Major 3rd	↗ Major 3rd	↘ Perfect 4th	↗ Perfect 4th	↘ Perfect 5th	↗ Perfect 5th	↘ Minor 6th	↗ Minor 6th	↘ Major 6th	↗ Major 6th	↘ Minor 7th	↗ Minor 7th	↘ Major 7th	↗ Major 7th	↘ Octave	↗ Octave
Twinkle, Twinkle	Tonic	A major 6th	42	41	18	4	-	13	2	-	-	-	-	-	1	-	3	-	-	-	-	-	-	-	-	-	-
A Song Used by Smale, 1987	Dominant	A perfect 5th	14	13	3	1	-	5	-	3	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	
America	Tonic	A minor 7th	41	40	8	6	6	9	7	1	-	1	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-
Are You Sleeping?	Tonic	A major 6th	32	31	4	2	2	4	8	1	-	4	1	2	2	-	1		-	-	-	-	-	-	-	-	-
Bow Wow Wow	Tonic	A major 6th	17	16	5	-	-	3	1	2	2	1	2	-	-	-	-		-	-	-	-	-	-	-	-	-
Great Big House	Mediant	A major 6th	25	24	4	-	1	3	5	3	5	-	-	3	-	-	-	-	-	-	-	-	-	-	-		-
Happy Birthday	Dominant	An octave	25	24	4	3	-	5	3	1	-	3	-	1	1	-	1	-	1	-	-	-	-	-	-	-	1
L'il Liza Jane	Mediant	A perfect 5th	23	22	6	-	-	7	1	3	3	-	2	-	-	-	-	-	-	-	-	-	-	-	-		-
In the Sea	Mediant	A minor 6th	24	23	2	2	2	2	2	6		4					1	1	1	-	-	-	-	-	-	-	-
Jingle Bells	Median	A perfect 5th	51	50	29	2	1	4	5	1	4	1	-	-	1	2	-	-	-	-	-	-	-	-	-	-	-
Jumbo Elephant	Tonic	An octave	33	32	5	6	5	5	7	-	-	1	-	1	1	1	-	-	-	-	-	-	-	-	-		-
Little Sir Echo	Tonic	A minor 6th	26	25	6	2	-	7	2	-	1	-	-	-	5	-	-	2	-	-	-	-	-	-	-	-	

Song	The starting pitch	Vocal range	No. of pitches	No. of musical intervals	The number of musical intervals by type and direction																						
					Unison	↘ Minor 2nd	↗ Minor 2nd	↘ Major 2nd	↗ Major 2nd	↘ Minor 3rd	↗ Minor 3rd	↘ Major 3rd	↗ Major 3rd	↘ Perfect 4th	↗ Perfect 4th	↘ Perfect 5th	↗ Perfect 5th	↘ Minor 6th	↗ Minor 6th	↘ Major 6th	↗ Major 6th	↘ Minor 7th	↗ Minor 7th	↘ Major 7th	↗ Major 7th	↘ Octave	↗ Octave
Path to the Moon	Dominant	An octave	56	55	4	3	6	9	15	4	1	1	-	-	5	2	-	2	-	1	-	-	-	-	-	-	1
Row Row Row Your Boat	Tonic	An Octave	27	26	11	1	1	4	4	1	-	1	-	1	1	-	1		-	-	-	-	-	-	-	-	-
This Old Man	Dominant	A major 6th	31	30	6	3	3	5	6	2	3	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-
Yankee Doodle	Tonic	A minor 7th	57	56	6	5	4	10	17	1	3	4	1	2	2	1		-	-	-	-	-	-	-	-	-	-
The 1st Song of Leighton & Lamont, 2006	Tonic	A major 6th	33	32	12	-	3	2	2	2		3	5	1	1	1	-	-	-	-	-	-	-	-	-	-	
The 2nd Song of Leighton & Lamont, 2006	Mediant	A major 6th	36	35	10	-	-	11	6	2	1	-	-	2	1	-	2	-	-	-	-	-	-	-	-	-	
The 3rd Song of Leighton & Lamont, 2006	Dominant	An octave	32	31	-	2	4	3	5	5	2	3	1	-	3	1	-	1	-	-	-	1	-	-	-	-	
The 4th Song of Leighton & Lamont, 2006	Mediant	A major 6th	47	46	16	1	2	7	1	8	5	-	2		3	1	8	-	-	-	-	-	-	-	-	-	
Take a Trip	Tonic	An octave	32	31	17	-	3	2	4	3	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-		1
The British National Anthem	Tonic	A minor 7th	42	41	8	6	6	11	7	1	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	
The Jack in the Box	Dominant	A major 9th	30	29	12	-	1	1	5	-	1	-	1	-	3	1	1	-	-	-	-	-	-	-	-	1	-
The Little White Duck	Tonic	A minor 7th	36	35	9	2	3	4	3	3	3	3	5	-	-	-	-		-	-	-	-	-	-	-	-	-

These have been analysed in Table 3.2 (above). The analyses for a song were based on the six following characteristics: the starting pitch; pitch range; the number of its pitches and musical intervals, and the type and number of each musical interval involved.

(i) Scale Degree of the Starting Pitch

As Table 3.2 shows, the scale degree of a starting pitch of all $N = 24$ criterion songs was either the song key's tonic (I) (52%), mediant (III) (24%) or dominant (V) (24%).

(ii) Vocal Range

As Table 3.2 shows, only the pitch range of *The Jack in the Box* used by Levinowitz (1987) was a major 9th, while those of the other criterion songs was an octave, or less than an octave. The pitch range of songs used in previous studies also varied and so, consequently, was the implied level of difficulty of the songs, especially if vocal register use is also taken into account (see the pitch range of songs in Apfelstadt, 1984; Demorest & Pfordresher, 2015; Guerrini, 2006; Joyner, 1969; Levinowitz et al., 1998; Mang, 2006; Mizener, 1993; Nichols, 2016; Welch et al., 2009a, 2009b; Western, 2002). For instance, while testing older Primary school children's (age ≥ 7) vocal pitch-matching accuracy for songs, the pitch range of *Jingle Bells* applied by Nichols (2016) was a perfect fifth, while that of *Happy Birthday* used by Mang (2006) was an octave. The songs with different vocal range increase the difficulty in comparing their results.

When considering the influence of a tone language on vocal range, Chen-Hafteck (1999b) did not find a difference in vocal range between Cantonese and English children, suggesting that sung vocal range may not be influenced by a tone language.

(iii) Melodic Contour

The melodic contours of songs, defined by the direction of relative musical intervals, were also diverse in previous studies. The types of contours varied between consistent (perfect unison), upward only (↗), downward only (↘), from upward to downward (Λ), from downward to upward (V), and more mixtures of direction (e.g., ~).

The author concludes that the size of musical intervals influences the shape and fluctuation in melodic contour. As reviewed in Chapter 2, section 2.5.4a, vocal pitch-matching accuracy for musical intervals tends to decrease with increasing size of intervals (*cf.* Jones, 1971; Moore et al., 1995/1996; Sinor, 1984; Wolf, 2005). Consequently, the 23 musical intervals from a unison to an octave with downward and upward directions as shown in Table 3.3 were divided into the three following groups:

(1) unison to major 3rd; (2) perfect 4th and perfect 5th, (3) minor 6th to octave. The percentage of each of the three groups as constituents in a criterion song is summarised in Table 3.3.

Table 3.3 The Percentage of Each of the Three Groupings Defined by Size of Musical Intervals for Each of Songs as Applied in Previous Studies

Song	Unison to Major 3rd	Perfect 4th and perfect 5th	Minor 6th to Octave
Bow Wow Wow	100.00%	0%	0%
*L'il Liza Jane**	100.00%	0%	0%
The Little White Duck	100.00%	0%	0%
The British National Anthem	97.56%	2.44%	0%
America	97.50%	2.50%	0%
Jingle Bells	94.00%	6.00%	0%
Take a Trip	93.55%	3.23%	3.23%
This Old Man	93.33%	6.67%	0%
A Song Used by Smale, 1987	92.31%	7.69%	0%
The 1st Song of Leighton & Lamont, 2006	91.60%	9.40%	0%
The 4th Song of Leighton & Lamont, 2006	91.30%	8.70%	0%
Yankee Doodle	91.07%	8.93%	0%
Jumbo Elephant	90.62%	9.38%	0%
Twinkle, Twinkle	90.24%	9.76%	0%
Row Row Row Your Boat	88.46%	11.54%	0%
Great Big House	87.50%	12.50%	0%
In the Sea	86.96%	4.35%	8.70%
The 2nd Song of Leighton & Lamont, 2006	85.70%	14.3%	0%
Are You Sleeping?	83.87%	16.13%	0%
The 3rd Song of Leighton & Lamont, 2006	80.70%	12.80%	6.50%
Path to the Moon	80.00%	12.73%	7.27%
Happy Birthday	79.17%	12.50%	8.33%
The Jack in the Box	72.41%	17.24%	6.90%

Little Sir Echo	72.00%	20.00%	8.00%

Note. * Only the first half of *L'il Liza Jane* was used.

The melodic contours of these songs can be separated into the following three groups based on the percentage of a group of musical intervals within a song.

A. When 90% of musical intervals of a criterion song were perfect unisons to major thirds (including minor seconds), the melodic contour of the song was relatively smooth (see the list of songs in the yellow section in Table 3.3).

B. When 80% – 90% of musical intervals of a song were perfect unisons to major thirds, and 10% – 20% of them were perfect fourths and perfect fifths, the melodic contour of the song indicated a slightly bigger movement compared with group 1 (see the list of songs in the blue section in Table 3.3).

C. When 70% – 79% of musical intervals of a song were perfect unisons to major thirds, 10% – 20% of musical intervals were perfect fourths and perfect fifths, and 5% – 10% of them were from minor sixths to an octave, the melodic contour of the song showed a more extensive and varied movement compared with that of songs in the second group (see the list of songs in the green section in Table 3.3).

According to the means of the raw scores measured by rating scales reported by previous studies (e.g., Chen-Hafteck, 1999b; Demorest & Pfordresher, 2015; Mang, 2006; Nichols, 2016; Welch et al., 2009a, 2009b), most Primary school students could sing the general contour of melody contours accurately. For instance, Welch et al. (2009a) used the 4-point *VPMD* rating scale (see Figure 3.4) to measure the vocal pitch-matching accuracy of song for N = 3,472 students aged seven to ten. The content of the 2nd rating of the *VPMD* scale is that 'Sung melodic outline begins to follow the general (macro) contours of the target melody or key constituent phrases', and that of the 3rd rating scale is that 'Melodic shape and intervals are mostly accurate' (*Welch*, 1998, p. 35). They reported that means for male and female students were M = 2.81 and M = 3.09, respectively, which were greater than the 2nd point on the *VPMD* rating scale and focusing on sung accuracy of the melodic contour. Furthermore, Chen-Hafteck (1999b) reported that the ability to match melodic direction increased with age. She also reported that the melodic direction increased faster than the musical intervals. Her findings were based on an objective measurement, which semitone errors were calculated based on the differences in fundamental frequency.

Figure 3.4 A Model of Vocal Pitch-Matching Development (Welch, 1998, p. 35)

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 intervals 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 singer's musical culture.

(iv) Musical Intervals

According to Table 3.3, most musical intervals (72% to 100%) of the 24 criterion songs listed in the table were perfect unisons to major thirds. The second most popular musical intervals involved in these songs were perfect 4th and perfect 5th (0% to 20%), while the third group consisting of greater musical intervals (from minor 6th to octave) had the least (0% to 8.7%).

Although a minor second was considered to be a more difficult interval to match accurately (Moore et al., 1995/1996), the analyses suggest that it regularly occurred in songs (see Table 3.2), such as *America* (Guerrini, 2006), and the British National Anthem (Joyner, 1969).

According to results reported by previous studies which used rating scale measures involving descriptions of musical intervals (e.g., Demorest & Pfordresher, 2015; Mang, 2006; Nichols, 2016), many Primary school students do not sing all the musical intervals within a song accurately. For instance, the third level category of the *VPMD* scale (Welch, 1998) suggested that most musical intervals of a song were sung accurately, but with some inaccurate ones. In contrast, the fourth category of the 4-point scale of the *VPMD* indicated 'No significant melodic or pitch errors in relation to relatively simple songs from the singer's musical culture' (Welch, 1998, p. 35). Welch et al. (2009a) used the *VPMD* scale in their large-scale national study and reported M = 2.81 and M = 3.09 (both around the third *VPMD* level rating) for both girls and boys among N = 3,472 pupils aged seven to ten in England. The data suggested that these participants could match most of the musical intervals of songs accurately, but that collectively across this age range they could not match all the musical intervals accurately.

In addition, researchers have varied in their choice of the number of songs used to test Primary school students' pitch-matching accuracy in song singing. For instance, only one song was used in studies by Smith (1973), Geringer (1983), Green (1994), and Nichols (2016), whilst two songs were employed in studies by Mizener (1993), Levinowitz et al. (1998), Guerrini (2006), Welch et al. (2009a, 2009b), and Demorest and Pfordresher (2015).

The current author agrees with the latter approach, namely that songs used to test vocal pitch-matching accuracy for children should involve both a relatively simple song (such as one with a small pitch range) and one or two more complex song(s). The simple song may also relax children who may believe that they cannot match pitches well, or who have little or no confidence in their singing ability, while the more complex song can help to distinguish between children's current singing abilities. Furthermore, if songs in future studies could be selected from a particular list, it would enable comparison across datasets. This list could involve common songs with varied difficulties, such as *Twinkle, Twinkle (or Jingle Bells)*, and *Happy Birthday*.

Finally, singing with or singing without a starting pitch could affect singing scores recorded by the rating (e.g., *VPMD*) scale measuring vocal pitch-matching accuracy of song singing, as a lower key and a lower pitch range were more likely to be used to sing if there is no starting pitch (e.g., Sergeant, 1992). Previous studies suggested that a person tended to match pitches of songs more accurately when they sang in a lower key (e.g., Joyner, 1969; Welch, 1979b), which implies that a higher score would be achieved as measured by the *VPMD* scale.

3.2.2 Measurement

Two dominant types of measurements were used to assess vocal pitch-matching accuracy in previous studies: (a) rating scales (e.g., Smith, 1973; Welch, 1998; Wise & Sloboda, 2008) and (b) software which was originally designed for the speaking voice (e.g., Western, 2002). Another type of measurement, (c) utilising software to investigate pitch height directly, was not commonly employed in previous studies. These three types of measurements are reviewed below as follows.

(a) Rating Scales Measuring Pitch-Matching Accuracy of Song

A rating scale is the most frequent method used to assess vocal pitch-matching accuracy of a song in previous studies. A wide range of rating scales were used in previous studies, but most of them were not named. It seems that there were four common perspectives describing vocal pitch-matching development of song, from singing like speaking, accuracy of melodic contour, key consistency, to accuracy of musical intervals. They are reviewed and analysed successively.

(i) Singing like speaking

Speaking or chant behaviour has been placed in the first rating of many rating scales to describe a chant-like singing behaviour (see scales of Atterbury & Silcox, 1993; Davies & Roberts, 1975; Guilbault, 2004; Mizener, 1993; Welch, 1998; Wise & Sloboda, 2008; Young, 1971). Besides the description 'chant-like', some scales also contain other descriptions, such as 'restricted variation of pitches' and 'a limited pitch range', to complement their explanation of speaking behaviour in song singing (Welch, 1998; Wise & Sloboda, 2008).

(ii) Melodic Contour

Melodic contour (melodic direction, or melodic line) is also commonly used in rating scales which aim to define vocal pitch-matching accuracy of song singing (Davies & Roberts, 1975; De Yarman, 1971; Dittmore, 1968; Guilbault, 2004; Joyner, 1969; Smith, 1973; Welch, 1998; Wise & Sloboda, 2008; Young, 1971).

Most scales employed two of their internal rating scales to describe aspects of sung melodic contour (Davies & Roberts, 1975; Guilbault, 2004; Joyner, 1969; Welch, 1979a; 1986; 1998; Wise & Sloboda, 2008; Young, 1971). The most popular positions for ratings of sung melodic contour in a scale were the second and third, i.e., after chant-like singing (e.g., Guilbault, 2004; Welch, 1998; Wise & Sloboda, 2008). Other scalic positions were also reported, such as the first and second ratings in the scale of Young's (1971) study, the third and fourth in that of Davies and Roberts (1975), and the first and third in Joyner's (1969) study.

Also, sometimes more than two ratings of melodic contour are used in a scale, such as the first to third ratings in Dittmore's (1968) study, or the first to fourth ratings in De Yarman's (1971) study. It seems unnecessary to apply more than three ratings to explain the development of melodic contour in song singing for children, both because its relative shape can be sung accurately by most pupils (e.g., Welch et al., 2009a), but also because too many ratings for melodic contour can distract attention from other important elements, such as musical intervals.

Furthermore, a significant leap between two adjacent ratings is revealed in some scales describing the development of melodic contour. For instance, in the scale of Davies and Roberts (1975), rating B is 'Does not follow the melodic line at all', while rating C is 'Can sing the melodic line correctly…' (p. 24). Consequently, it seems that the gap in description of melodic contour development between these two ratings is too large, at least when compared to other scales in the literature (e.g., Welch, 1998).

(iii) Key

The development of the music's key in song singing is described in many rating scales (e.g., Dittmore, 1968; Flowers & Dunne-Sousa, 1990; Ramsey, 1981; Smith, 1973; Welch, 1998; Wise & Sloboda, 2008). 'Tonality', 'tone', 'intonation', and 'key', were words frequently used to describe key in rating scales of pitch-matching accuracy of song (De Yarman, 1971; Welch, 1998; Wise & Sloboda, 2008; Young, 1971).

However, there is no agreement in previous studies on how many ratings should be used to describe the development of a sung key centre. For instance, while the *VPMD* scale summarised by Welch (1998) used two ratings to describe the development of key during song singing, its updated application in the scale by Wise and Sloboda (2008) extended it from two to five ratings to describe various stages of key consistency for adults. The later scale has a more specific set of definitions to define categories of singing in-tune than the former scale (Welch, 1998). It should be noted that the *VPMD* scale is mainly used to measure children's vocal pitch-matching accuracy, while the updated, latter version was designed to be used with adults.

This raises a critical question of how to define singing in one musical key, or in other words, how to define a tonality centre, or how many consistent accurate melodic intervals are needed to define a particular key centre.

Previous studies defined the consistency of key centres using various approaches. In earlier studies, De Yarman (1971, p.23) used words such as 'Poor' 'Fair', 'Good', 'Very good', 'Excellent' to describe five levels of pitch-matching accuracy. These descriptors without further elaboration and, by today's expectations, seem to need more clarity. A few years later, a scale created by Davies and Roberts (1975) measured the accuracy of key for a song sung by participants based on the key of a model. In this case, an appropriate starting pitch for the model would be vital to enhance the reliability of its results. Then, some rating scales (e.g., scales employed by Dittmore, 1968; Flowers & Dunne-Sousa, 1990; Ramsey, 1981) used the number of implied keys in the participant's singing of a target song to define their level of pitch-matching accuracy. Scales using this criterion typically counted tonal centres up to three, which means that a singing performance sung with more than three tonal centres would be marked as having the lowest rating of vocal pitch accuracy. A different measurement was employed by Mizener (1993) who focused on whether tones were the same at the beginning and at the end of a sung song. However, these types of scales both seem to underestimate the effects of the relatively internal complexity level of musical intervals within the song. These scales were not designed to show the relative accuracy for matched pitches. In a more nuanced rating scale which relates to such an approach, Guibault (2004) defined an accurate sung pitch of the song by the vocal pitch accuracy of the tonic, dominant, subdominant or cadential pattern in a song.

(iv) Musical Intervals

Although some rating scales have a final, most advanced, level of vocal pitch accuracy with a specific focus on the sung key, they do not contain a description of the development of sung musical intervals or pitches (e.g., rating scales applied by Atterbury & Silcox, 1993; Flowers & Dunne-Sousa, 1990; Guilbault, 2004; Ransey, 1981; Roberts & Davies, 1975). In contrast, other rating scales do seek to account for sung intervals (e.g., scales composed by Joyner, 1969; Mizener, 1993; Smith, 1973; Welch, 1998; Wise & Sloboda, 2008; Young, 1971). According to these latter previous studies, the description of musical intervals was usually placed developmentally higher than that of key consistency. If a sung key was mostly constant in the song singing, the developmental ability to match musical intervals with minimal pitch errors was described in the last one or two rating(s) on such scales. The number of interval or pitch errors at this relative stage of development was observed to be relatively small, and would not damage a sense of a prevailing key centre.

The rating scale used by Smith (1973) (see Figure 3.5) had a clear description of the relationship between a consistent key centre and the number of accurate melodic intervals being reproduced. The scale reveals that a longer sung pattern of consistently accurate intervals represents a more stable key centre, meaning that the integrity of sung intervals is crucial to the level of key centre preservation.

Overall, previous studies described the development of a sung key centre more thoroughly in some rating scales than in others. The description of key and musical intervals needs to consider the clarity of vocabulary and the relationship between a key's consistency and the accuracy of musical intervals.

Figure 3.5 A Seven-Rating Scale of Smith (1973)

1	Out-of-tune – No part of the singing is recognizable as outlining the intervals of the melody; pitch inflection may be in the proper direction, but the size of the intervals is incorrect; no indication of a tonal centre (Estimated 0 per cent to 9 per cent in-tune)
2	Mostly out-of-tune—Brief sections of the melody may be sung with the proper intervals; short sections of the melody may be sung with proper intervals but with a different tonal centre; (when the tonal centre is shifted while singing with an accompaniment, the result is melodic intervals parallel with the accompaniment) (Estimated 10 per cent to 29 per cent in-tune)
3	Out-of-tune the majority of the time—Several sections may be sung with proper melodic intervals, but the majority of melody is out-of-tune; several phrases of melody may be sung with the proper intervals but with a different tonal centre; (when the tonal centre is shifted while singing with an accompaniment, the result is melodic intervals parallel with the accompaniment) (Estimated 30 per cent to 44 per cent in-tune)
4	In-tune one-half of the time—Sections of the melody are in-tune and other sections are out-of-tune; loss of tonal centre may occur in out-of-tune sections (Estimated 45 per cent to 59 per cent in-tune)
5	In-tune the majority of the time—The melodic intervals of the melody may include short sections which are not recognizable as part of the melody; the tonal centre may be sharped or flatted in some sections or there may be a brief loss of the tonal centre (Estimated 60 per cent to 74 per cent in-tune)
6	Mostly in-tune—Several small errors in the melodic intervals may occur or slight sharping or flatting of the tonal centre may occur (Estimated 75 per cent to 89 per cent in-tune)
7	In-tune—Just noticeable pitch errors which do not distract from the overall performance may occur (Estimated 90 per cent to 100 per cent in-tune)

Arguably, the most successful rating scales measuring vocal pitch-matching accuracy of song singing reported in previous studies were usually those which took as their foci speaking, melodic contour, key, intervals, and their development. Among these musical elements, more ratings were typically employed to describe the development of key centre, followed by that of melodic contour and intervals, while predominantly using their speaking voice was frequently represented by a level one rating. These rating elements tended to be designed to develop gradually in an overlapped manner across ratings, except for speaking which was normally in the first rating only.

Finally, it should be noted that when measuring vocal pitch-matching accuracy in song singing, most previous studies did not focus on the accuracy of the song's rhythm, although some did measure rhythm accuracy (e.g., De Yarman, 1971; Dittmore, 1968; Welch, 1979a). These studies suggest that children could match rhythm accurately if the observer was required to notice this within the scalic descriptions. This was supported by Welch (1994) and Etopio (2009) who both suggest that children can match rhythm accurately earlier than they can match pitches.

(v) A List of Rating Scales Used in Previous Studies

Previous studies applying a rating scale are summarised in Table 3.4. They are analysed here based on the number of different perspectives describing vocal pitch-matching development of song, such as singing like speaking, accuracy of melodic contour, key consistency, accuracy of pitches or musical intervals, to singing in-tune.

As Table 3.4 shows, only the rating scale used by Wise and Sloboda (2008) involved all of the five perspectives. It should be noted that their rating scale was updated from the *VPMD* rating scale of Welch (1998), following his advice (see the content of the *VPMD* scale in Figure 3.4) to describe adults' vocal pitch-matching accuracy. The difference between the two ratings is that the scale of Wise and Sloboda (2008) contains 5 elements (leading to 8 levels of ratings), in order to describe in more detail participants' vocal pitch-matching accuracy in song singing than the original 4-rating scale *VPMD* summarised by Welch (1998), especially when describing the development of key consistency[46]. The current author is of the opinion that a rating scale with more ratings tends to be more useful than one with fewer ratings. However, having more rating levels may make it difficult to distinguish between adjacent ratings because of smaller differences between them, as all such rating scales are relatively subjective.

As Table 3.4 shows, as well as the *VPMD* scale summarised by Welch (1998), two other rating scales composed by Young (1971) and Mizener (1993) also involved four common perspectives of description of vocal pitch-matching development of song. As with the *VPMD* scale (Welch, 1998), the scale of Young (1971) does not involve a description of singing in-tune completely. This might be because his participants were very young, from kindergarten to first Grade children. On the other hand, the scale of Mizener (1993) does not describe melodic contour, but focuses on whether the initial and final tonalities were the same.

[46] Welch (1983) originally used a five-point scale in his doctoral thesis and in his invited 1986 BJME article which was based on this, and then simplified the scale to four categories for his 1998 article.

Table 3.4 A Checklist of Five Perspectives of a Rating Scale Used to Measure Vocal Pitch-Matching Accuracy of Song in Each Previous Study

Reference	Singing like speaking	Accuracy of melodic contour	Key consistency	Accuracy of pitches or musical intervals	Singing completely in-tune	No. of applied perspectives
Wise & Sloboda, 2008	X	X	X	X	X	5
Welch, 1998	X	X	X	X	-	4
Young, 1971	X	X	X	X		4
Mizener, 1993	X	–	X	X	X	4
Joyner, 1969	-	X	X	X	-	3
Smith, 1973	-	X	X	X	-	3
Davies & Roberts, 1975	X	X	X	-	-	3
Guilbault, 2004	X	X	X	-	-	3
Dittmore, 1968		X	X	-	X	3
Atterbury & Silcox, 1993	X	-	X	-	X	3
De Yarman, 1971	-	X	X	-	-	2
Ramsey, 1981	-	-	X	-	-	1
Flowers & Dunne-Sousa, 1990	-	-	X	-	-	1

The other rating scales listed in Table 3.4 involve less than four categories describing vocal pitch-matching development in song singing. When a scale with less ratings is used, the gap between each level can be relatively large. For instance, a considerable leap is evidenced from singing inaccurately to accurately among two adjacent ratings is evidenced in the scales used by Joyner (1969) and Atterbury and Silcox (1993). Joyner (1969) was the first researcher to report formally that children's vocal pitch accuracy was affected by the pitch range of the target melody, noting that some of his participants were much more in-tune when the target melody was transposed downwards into more of a spoken pitch range.

(b) Software Used to Test Vocal Pitch-Matching Accuracy for Songs

Only a limited number of studies have analysed vocal pitch-matching accuracy for songs with IT software. Although the application of IT could provide a more objective measurement than a perceptually-based rating scale, the process could be more time-consuming and the amount of emergent data would still need to be clustered in some way to be pragmatically useful. Furthermore, one ongoing challenge is that, although an IT-based analysis of singing can provide rich acoustic data in fine detail, this does not easily equate to the perception of musical behaviour where the brain tidies up the micro detail into more culturally shaped sonic elements (Welch, 1994).

Previous studies which used software to analyse vocal pitch-matching accuracy in song singing included the studies of Green (1994), Chen-Hafteck (1999b), Elmer and Elmer (2000), Western (2002), and Hedden and Baker (2010). It seems that the measurement of Elmer and Elmer (2000) was more musical in design than that used in the other four studies. They used a system of pitch approximation and showed the direction of minor mistuning, such as an upward or downward glissando. Their system focused on the status of stability of pitch (stable/unstable, with/without glissando), but it could not show the degree of the stability of vocal pitch.

An earlier study conducted by Green (1994) applied a distinct measurement based on the sum in vocal pitch accuracy of every sung pitch (a total of 17 pitches) and every interval (a total of 16 intervals) of a song *Bow, Wow, Wow* – a total maximum score was 33 points (one point for each accurate pitch and interval). If the song was sung in a different key from that of a model, it would be given 16 points. A high reliability $r = .94$ was reported. Compared with rating scales showing a general vocal pitch-matching accuracy for a song, this measurement seems more objective, i.e., open to replication by others, as it showed both the absolute and relative pitch accuracies in the sung product. However, it is unclear how the accuracy of vocal pitches and intervals was measured.

Chen-Hafteck (1999b) analysed kindergarten children's vocal pitch-matching accuracy of song by calculating semitone errors of musical intervals of a criterion song and assigning a score of melodic direction. Her semitone errors of musical intervals were taken by calculating semitone error between a sung pitch based on MIDI values and its standard pitch. This method allowed her to explore more detailed vocal pitch-matching accuracy of a song from semitone error of each musical interval involved in a criterion song, the overall semitone errors of musical intervals, to the overall accuracy level of melodic contour.

Similarly, Western (2002) used the Real-Time Pitch programme by CSL (4100) software-hardware systems (https://www.pentaxmedical.com/pentax/en/99/1/Computerized-Speech-Lab-CSL) to analyse mean fundamental frequency of pitches in one of two tested songs, but reported that this was enormously time-consuming to operate, as it could only report on a short piece of singing at one time. A similar software package to CSL is Praat (https://www.fon.hum.uva.nl/praat/). Another software package, Sing&See (https://www.singandsee.com) has been designed specifically for providing visual feedback of vocally matched pitches. This has a more pedagogically focused design, rather than the speech science applications of CSL and Praat, and has been used in some studies with adults (Wilson, 2006), but there appears to be no research literature on its use with children, although Welch (private correspondence) reports that Sing&See is being used regularly in a London mainstream Primary school with deaf and hearing impaired children in their singing lessons. Overall, software that automatically analyses each sung pitch's fundamental frequency of a whole song singing is not available.

Although some commercial applications have been designed for the assessment of song singing, such as a free online Chinese singing application Quanmin K ge[47] (All people sing), and is reported to evaluate vocal pitch-matching accuracy of song automatically, the measurement of vocal pitch-matching accuracy by software can be substantially influenced by accuracies of rhythm, key, and consistency. For instance, if someone sings a song slower than that of the accompaniment, their rating of vocal pitch-matching accuracy for the song as measured by the software would be relatively lower than that if measured by a human. One pitch can only be recognised as matched accurately by the software when its rhythm and pitch height are both precisely matched against a standard one. Because of the confidential nature of the company's information, it is not known how these pitches are processed and what format of software is used for them.

Similarly, Arora et al. (2023) designed a software program that could detect and analysis singing mistakes during the learning of Indian songs. Firstly, a standard audio recording was recorded by a teacher. Then, its time and key centre could be shifted by a learner based on their convenience. The learner sang the song with the shifted version of the teacher's recording. Four types of errors, including frequency error, were analysed[48]. Pitches of a matched song were calculated by Praat, shown as related to measures of sung fundamental frequency. The gaps in fundamental frequency (pitch) between the teacher and the learner were calculated, taking the absolute values to analysis. This system involved many elements used to measure vocal pitch-matching accuracy, such as Praat, fundamental frequency, adjusted time and key centre of a model. It could be efficient in its application if children sing a song continuously. However, it is unsure whether this will work

[47]全民 K 歌 (qq.com)

[48] Other three types of errors were amplitude, pronunciation, and timing errors.

effectively if a child sings a song with a few pauses, as the melodic contour will not match the standard version.

Overall, these previous studies provided rich clues in exploring the details of vocal pitch-matching accuracy for songs.

(c) Other Measurements

Other measurements have been employed in some former studies to assess vocal pitch-matching accuracy of song. For instance, a continuing scale from 0 to 15 points, used by Ramsey (1983), divided participants into higher and lower ability singers by the middle score of 7. Another measurement, using the marks '+' and '-', was employed by Brophy (1997) to quickly score each pupil's accurate and inaccurate vocal pitch-matching of pattern singing in a class setting. Errors of lyrics, melody, stops, and a starting pitch were not counted in his system. Although these two methods are subjective, they could help to quickly measure someone's vocal pitch-matching accuracy for songs. However, as a detailed description of each rating is missing, different judges might have a different interpretation of each rating.

The study of Roberts and Davies (1975) did not use any of the above tools, but relied on music teachers' personal judgement. Although this method was subjective, and teachers might have varied understanding of what they termed as a 'poor pitch singer' (p.29), having data from a huge number of participants (*N* = 18,902) boosted the reliability of their results.

In conclusion, various approaches to vocal pitch matching accuracy and development in song singing have applied in previous studies. One dominant type of measurement within this literature is the use of differentiated rating scales, whilst a few previous studies have sought to apply a more objective measurement through the analysis of more acoustically-based data, such as by using IT. A perceptual rating scale can be applied to assess the pitch-matching accuracy of song much quicker but more subjectively than a software program analysing fundamental frequency and translating this to pitch. In addition, software that has been originally designed for a speech signal is likely to be more appropriate for short singing materials, such as the musical elements/phrases of a song, because it is time-consuming to use. Although it might be possible to obtain greater detail from musical software designed for displaying visual feedback of a sung pitch (e.g., Sing&See), the reliability of the pitches shown in the software would have to be checked carefully if the intention is to use the data to generate detailed statistical analyses. Although the latest commercial software is reported to be able to calculate automatically the accuracy of vocally matched pitches of a song singing, its judgement involves other elements, such as rhythm and timing in song singing, so it cannot yet assess vocal pitch-matching accuracy of a song independently. When a singing performance is not consistent, this may not be recognised by the software.

Based on diverse song repertoire and measurements applied in previous studies, their results of vocal pitch-matching accuracy for songs are now reviewed below by the variables of age, sex, and social-economic status, which correspond to key elements that are the focus in the current study.

3.2.3 Review of the Influence of Age on Vocal Pitch-Matching Accuracy for Songs

Many previous studies paid attention to vocal pitch-matching accuracy for songs by age. The general trend is summarised below, followed by analyses of elements influencing the results of vocal pitch-matching accuracy for songs by age.

(a) A General Trend

In previous studies, age was typically illustrated by school Grade (e.g., Gould, 1969; Green, 1994; Pereira & Rodrigues, 2019), age in whole years (e.g., Davies & Roberts, 1975), or a six-month cluster (e.g., Welch et al., 2009b). As a Grade can cover several age groups, data analysed by age in whole year, or a six-month cluster is likely to describe the impact of age on vocal pitch-matching accuracy more accurately.

Commonly, most previous studies agree that age has a significant impact on vocal pitch-matching accuracy in song during the Primary school period, with older Primary school students tending to match pitches of songs more accurately (e.g., Davies & Roberts, 1975; Gould, 1969; Green, 1994; Mizener, 1993; Pereira & Rodrigues, 2019; Welch et al., 2009a). The age of Primary school students in most of these studies (e.g., Davies & Roberts, 1975; Green, 1994; Pereira & Rodrigues, 2019; Welch et al., 2009a) was ≤10, which might have been designed to reduce the interference of puberty in their data analyses. These studies collected data in a Western cultural setting either from a large number of music teachers' opinions by filling in a questionnaire (e.g., Davies & Roberts, 1975; Gould, 1969), or assessments of children's singing performances (e.g., Pereira & Rodrigues, 2019; Welch et al., 2009a) in varied conditions, such as singing individually or within a group (e.g., Green, 1994). While the assessment for vocal register use for patterns or songs has mainly used the *SVDM* scale (see Section 3.1.2), that for vocal pitch-matching accuracy for songs has been more varied (see Section 3.2.2). Although these diverse measurements increased the difficulty in comparing results across previous studies, the diversity can be seen as improving the reliability of anyone finding about increasing vocal pitch-matching ability with growing age for Primary school children.

Although there is evidence that some girls are entering puberty during their Primary school years, this has been observed to be a minority currently. According to the National Health Service in England (NHS), the average age of puberty is 11 for girls and 12 for boys, implying that singing data for Primary schools which uses large numbers of participants is unlikely to be effected statistically by the onset of puberty.

The previous studies that reported a statistically significant age difference in vocal pitch-matching accuracy for songs (e.g., Green, 1994; Leighton & Lamont, 2006; Welch et al., 2009a) shared two characteristics. The age group of children in these studies had at least a two-year age gap and the children were asked to match a specific criterion song(s) (e.g., Green, 1994; Welch et al., 2009a), instead of matching a self-selected song (e.g., Mizener, 1993). However, it is unclear whether the development of vocal pitch-matching accuracy with increasing age reported for Primary school children in Western culture will necessarily be applicable to children in other cultures, such as Asia, not least because tuning systems in traditional music are not identical.

It should be noted that the data in most previous studies (e.g., Davies & Roberts, 1975; Gould, 1969; Guilbault, 2004; Hornbach & Taggart, 2005; Mizener, 1993; Pereira & Rodrigues, 2019; Welch, 1998; Welch et al., 1998) have suggested that the development of vocal pitch accuracy of songs by age tends to be nonlinear. In other words, children's development of vocal pitch-matching accuracy for songs might improve faster in some periods, but may be more at a consistent level at other times.

Using data collected from music teachers by questionnaire, Gould (1969) and Davies and Roberts (1975) agreed on a trend for vocal pitch-matching accuracy in song during the Primary school period, from rapid development at age around 5 to 8, to relative stasis at age around 8 to 12. This was also evidenced across larger numbers of participants by Welch et al. (2009b) who asked children to sing individually, and partially supported by Green (1994). The study of Green (1994) pointed out that older Primary school children's (Grade 5) vocal pitch-matching accuracy could continue to improve at a fast pace in a group singing, implying that where children experience a nurturing environment, it is possible for their singing competency in pitching to continue to improve as a group. Again, it is unsure whether the same result will be evidenced in another culture, such as in Asia.

Compared with other Grades in the Primary school period, a limited number of previous studies have involved the oldest Primary school year group, Sixth Graders. This might be because of lower interest at that stage in singing children's songs (Welch et al., 2009b) and/or the imminent onset of puberty, especially for boys (Martin, 2006).

In terms of puberty, Davies and Roberts (1975) reported that—at that time—for most boys, their voice started changing at age eleven; for 20% of them it changed at age thirteen, and for almost all of them the change was complete by the age of fifteen. As a result, it is assumed that the picture of the singing voice for Grade Six students might differ from that of younger Primary school students (*cf*. Levinowitz et al., 1998). An additional issue for further studies of Sixth Graders, is that this age group can be disrupted by the educational context in Primary schools. For instance, School Year 6^{th} or Grade 6^{th} students may have an essential entrance examination for some secondary schools in many countries, such as the UK and China. The exam may cause school leaders to focus their attention on the main curriculum subjects, such as literacy and mathematics, with less attention paid to music classes for this oldest Grade in Primary schools.

Means of vocal pitch-matching accuracy for songs were varied in previous studies. For instance, studies of Mang (2006) and Welch et al. (2009a) both applied the *VPMD* scale to measure older Primary school students' vocal pitch-matching accuracy in song. They applied the same target song, *Happy Birthday,* while another simpler target song, *Twinkle, Twinkle*, was also applied in Welch et al.'s (2009a) study. Scores of Mang (2006) (Age 7 M = 3.35, Age 9 M = 3.47) were generally higher than those of Welch et al. (2019a) (Year 3[49] M = 2.81, Year 6 M = 3.07). The different results between the two studies might be because of varied understandings of each of the four ratings in the *VPMD* scale, as the scale was relatively subjective.

Means of vocal pitch-matching accuracy for songs reported by Welch et al.(2009a) tended to be in line with the Chinese national study (MCCEMEC, 2020). The latter reported that the vocal pitch-matching accuracy of most (65.9%) Grade 4 children fell into the second or third rating of their four-rating scale. Their data were based on a self-selected song from the children's musical textbook. While vocal pitch-matching accuracy for children from a middle school Grade (Grade 4) of the Primary school period was assessed in the Chinese national study to give an overall view of Primary school children's vocal pitch-matching accuracy for songs, arguably children from other school Grades in Primary schools need to be assessed in order to show the development of vocal pitch-matching ability for songs in the Chinese context.

[49] School Year 3 and Year 6 are around age 7+ and age 10+ in the UK.

(b) Elements Influencing the Results of Development of Vocal Pitch-Matching Accuracy for Songs by Age

Although previous studies agree that age growth generally improves Primary school children's pitch-matching accuracy in songs, we can still notice that this finding could be influenced by some other variables, including (i) two internal elements: children's interest in singing a song and native language, and (ii) four external elements: complexity of a song, selection of a measurement tool, organization during the process of a data collection, and training.

(i) Internal Elements

Firstly, low interest in a target song might influence participants' vocal pitch-matching accuracy for the song. Mizener (1993) reported differences when participants were invited to sing *Jingle Bells*, which Sixth Grade students sang significantly less well than Fourth and Fifth Graders. *N* = 542 students from Third to Sixth Graders were involved in her study, and their song performances were measured by a seven-point rating scale focusing on whether the initial and end keys were the same or different. Mizener's unusual finding might result from a lack of interest on the part of these Sixth Graders in singing *Jingle Bells*. This assumption is supported by a higher vocal pitch-matching score for this Grade when they sang a self-selected song. Singing materials should, therefore, consider children's potential attitudes towards the choice of a target song.

Furthermore, it is unsure whether maternal language has an influence on Primary school children's vocal pitch-matching accuracy. Mang (2006) reported that Cantonese (a tone language) monolingual children tended to show higher vocal pitch-matching accuracy for *Happy Birthday* – sung in English – than English bilingual children (English and another language). However, it is uncertain whether the significant difference in vocal pitch-matching ability for songs was because of the different language systems (with or without a tone) or because of unequal language learning experience (learning one or two languages). Learning two languages seems more difficult than learning one language. Welch et al. (2009b) reported that White and Black children in England generally matched *Happy Birthday* and *Twinkle, Twinkle* more accurately than participant Asian children. Compared with the White and Black children whose mother tongue was English, it seems that Asian children in England would be more likely to speak two languages at home. Consequently, the influence of the tone language or culture on Primary school children's vocal pitch-matching accuracy for songs require further study.

(ii) External Elements

For the complexity of songs, a song with large intervals (typically over a perfect 5th) tended to be a harder singing task in terms of pitch for all Primary school students to match accurately (e.g., Chen-Hafteck, 1999b), so any age difference in vocal pitch-matching accuracy for songs would be more likely to be smaller in this case. Chen-Hafteck (1999b) reported that children could match small musical intervals earlier than bigger musical intervals. *Happy Birthday*, with its large intervals, was applied only in Mang's (2006) study. She reported finding no significant age difference between seven- and nine-year-old participants in vocal pitch accuracy of the song, although there was a two-year difference in age.

Pereira and Rodrigues (2019) reported that pitch-matching accuracy in song for Kindergarten and Grade One pupils was significantly poorer than that of Second to Fourth Graders, while a significant improvement was reported between Grades One and Two. The slow development in reproducing the songs accurately for the two youngest Grades might be because the singing task was challenging. The task was a song with eight three-pitch patterns (see Figure 3.1) with two ascending minor sixths, one ascending perfect fifth and one chromatic pitch C#4 as composed by Rutkowski (1996) initially to test vocal register usage. This relatively difficult vocal task might also explain why there is a relative stasis of vocal pitch-matching accuracy for songs as reported in Pereira and Rodrigues' (2019) study which was one to two years earlier compared with that of Gould (1969), Davies and Roberts (1975) and Welch et al. (2009a).

A particular methodological challenge is to identify if the complexity level of target songs for each age group's singing task is comparatively different in a study. This is illustrated by Demorest and Pfordresher's (2015) study, in which Kindergarten pupils sang *Twinkle, Twinkle*, but Sixth Graders sang *Happy Birthday*. It should be noted that the data of these two age groups was taken from two separate previous studies.

Furthermore, results between sung self-selected songs and required songs by age were also varied. For instance, while Fifth Graders sang pitches of *Jingle Bells* more accurately than Sixth Graders in Mizener's (1993) study, they sang target pitches less accurately in a self-selected song than the older students. Leighton and Lamont (2006) reported that no significant difference by age in singing a self-selected song was found among four- to eight-year-old pupils, but the difference was revealed in the same participants when they had to sing required songs. The lack of difference by age in self-selected song singing might be down to difficulties in the songs chosen by young participants; some of them might even choose a song that was relatively too advanced for them. In general, it seems that the complexity level of self-selected songs was unpredictable, which increases the uncertainty of results in terms of comparison across studies.

Using different types of measurements with the same singing materials, can influence the results of vocal pitch-matching accuracy of song analysed by age. For instance, in Leighton and Lamont's (2006) longitudinal study, *N* = 28 4-6-year-old children were asked to sing two target songs initially and sing another two target songs two years later. They reported that data measured by a seven-point rating scale showed a significant age difference, biased towards children when they were older. However, when the same singing data were analysed by the researchers using *Praat* 4.0.5 that provides mean fundamental frequency for a pitch(es), no significant difference was revealed by age. Furthermore, the complexity of the four target songs was not significantly different measured by their rating scale, but one of the songs was statistically significantly different from the others when assessed by *Praat*.

Although the results of the two measurements in Leighton and Lamont's (2006) study seem contradictory, each type of measurement evaluated vocal pitch accuracy in songs from a different perspective. For example, the software program Praat can display vocal frequency from which absolute pitch accuracy can be inferred, but a perceptual rating scale's measures are more likely to be a form of judgement for relative pitch accuracy. Praat is not designed to recognise if a pattern of pitches in a song was sung relatively accurately but in a different key, this would need to be inferred by the user to check an individual's sung replication of the original pitch model. However, such similarity in melodic patterns could be noted by an appropriate rating scale where the measurement was based on a general sense of stability of key, even though the scale may not illustrate the detail of pitch errors in a song's reproduction. Consequently, these two types of measurement seem to assess vocal pitch-matching accuracy of song from two distinct perspectives, i.e., possible absolute and relative pitch errors. While there appears to be only one study which applied both a subjective and a more objective measurement to assess children's vocal pitch-matching ability for songs (Leighton & Lamont, 2006), it seems appropriate to research how these two approaches can be reconciled to allow acoustic and perceptual measures to be undertaken together.

Results of vocal pitch-matching accuracy of song by age could vary according to how the participants were organised. For instance, in terms of the influence of singing a song individually or in a group, Green (1994) reported that Fifth Grade students sang a simple song *Bow Wow Wow* more accurately within a group than they did individually, but First to Third Graders matched the same song with a smaller difference in their vocal pitch-matching. Consequently, results may be better related to the setting in which participants sing, individual or group.

Finally, many previous studies have reported that training could change a regular trend based on age for vocal pitch-matching accuracy in songs (Boardman, 1964; Guilbault, 2004), especially for younger Primary school children. In the study by Welch et al. (2009b), children who attended the *Sing Up* training programme tended to develop their vocal pitch-matching ability for songs two years earlier than their peers who were not provided with the same raft of training opportunities. Furthermore, younger children with the training experience developed even three years earlier than their peers without the experience. Similarly, when Jacobi-Karna (1996) found that four-year-old pupils matched pitches of a child's song more accurately than five-year-old pupils in the second posttest measure by using six-rating scale, she assumed that this might be because the younger group improved faster under the same strength of training. This suggests that younger children developed vocal pitch-matching ability for songs faster than older children when the same rich singing environment is provided.

Previous studies also pointed out that the impact of age development and that of singing training on vocal pitch-matching accuracy of song might be merged. As a child gets older, they could collect more singing experience in their daily life. Demorest and Pfordresher (2015) investigated the different impacts between age and training. Originally, they thought that increasing singing experience was the main reason for the improvement of children's vocal pitch-matching accuracy in song. However, after conducting further fieldwork, they updated their earlier view and now hold that age and singing training 'correlate independently with singing accuracy' (Pfordresher & Demorest, 2020, p. 1).

3.2.4 Review of the Influence of Sex on Vocal Pitch-Matching Accuracy for Songs

Overall, previous studies (e.g., Bentley,1969; Gould, 1969; Green, 1994; Mang, 2006; Welch et al., 1997; 2008; 2009b) agree that the vocal pitch-matching accuracy in song differed between boys and girls, with girls song singing – as a group – tending to be significantly more accurate than boys when assessed in a school setting. The data come from diverse socioeconomic groups in the United Kingdom (Bentley, 1969; Green, 1994; Welch et al., 1997, 2008, 2009b), the USA (Gould, 1969), and Hong Kong, China (Mang, 2006). Participants were asked to sing diverse types of songs, such as taught songs (Green, 1994; Welch et al., 1997) or familiar songs (Mang, 2006; Welch et al., 2008, 2009b), with diverse internal complexity levels.

On the other hand, studies by Chen-Hafteck (1999b) and Leighton and Lamont (2006) did not find a significant sex difference in vocal pitch-matching accuracy in song singing. However, participants of both studies were young children (age ≤ 6+), and both studies measured children's singing performances using software. It is uncertain whether the absence of sex difference in vocal pitch-matching of song was because the children were young or because the measurement was more objective, or—as in the case of the young children in the Welch et al. longitudinal study (1997)—whether the task is to reproduce both lyrics and melody together (=less accurate) or the melody alone (=more accurate).

Previous studies remind us that singing in a group could impact on the results of vocal pitch-matching accuracy by sex. When singing within a group, Leighton and Lamont (2006), and Goetze and Horii (1989) agreed that their young participant boys vocally matched pitches of songs apparently much less successfully, while young girls were disrupted less (Goetze & Horii, 1989), or even matched more accurately in group singing (Leighton & Lamont, 2006). This sex difference within group singing might be a product of the differing social effects of public singing in a group on boys and on girls. Welch et al. (2009b) reported girls generally had a more positive attitude towards singing than boys. It seems that singing is more likely to be perceived as a girls' activity than that of boys, particularly in a school context which is where most research studies take place.

*3.2.5 Review of the Influence of the Age * Sex Interaction on Vocal Pitch-Matching Accuracy for Songs*

Many previous studies have involved several different age groups in the Primary school phase, which have allowed a review of more details of sex difference by age. In terms of the literature reviewed here, it seems that an age and sex interaction effect is likely to be influenced by the age group. The literature suggests that there are three phases of vocal pitch-matching development in song related to an age and sex interaction:

- five- to seven-year-old children might demonstrate no statistically significant sex differences by group in vocal pitch-matching accuracy of a song (Davies & Roberts, 1975; Welch et al., 1997, 2009b); but
- data analyses from eight- to ten-year-old students tend to demonstrate a significant sex difference (Davies & Roberts, 1975; Welch et al., 2009b). This may be because girls' vocal pitch-matching accuracy generally develops faster than that of boys around this age (Gould, 1969; Welch et al., 2009b);

- data analyses for ten to twelve-year-olds tend to suggest a greater sex difference, although this might decrease slightly because of the relatively strong improvement in older Primary school boys (Welch et al., 2009b). According to Gould (1969), the ratio of 'poor' (Welch, 1979b) pitch singers between boys and girls in Grade Six was 12:1. Overall, according to previous studies, the changing sex difference trends across the three Primary school periods suggest a significant interaction impact between age and sex on vocal pitch-matching accuracy of song.

On the other hand, many previous studies (e.g., Leighton & Lamont, 2006; Smale, 1987) did not find a significant interaction difference between age and sex in vocal pitch-matching accuracy of song. They include Smale's (1987) study whose participants were *N* = 106 four- to five-year-old children singing a simple song; Green's (1994) study where *N* = 241 First, Second, Third, and Fifth Graders sang a simple song *Bow Wow Wow*; the longitudinal study of Leighton and Lamont (2006) whose *N* = 28 participants aged four to eight sang four songs; and Welch et al.'s (2008) study whose data came from *N* = 3,510 children aged Year 2 to Year 6 in the opening, baseline phase of their research who sang two songs with different complexity levels. These previous studies using complete songs are supported by some studies which used musical elements' singing as tasks (e.g., Goetze & Horii, 1989).

The conflicting conclusions on 'with' to 'without' an interaction influence between age and sex on singing behaviour might be explained by the studies' different sample sizes. For instance, Welch et al. (2008) reported no significant interaction effect based on *N* = 3,510 participants, while a significant interaction effect was reported from *N* = 8,162 participants in their later study using the same methods (Welch et al., 2009b). Consequently, it seems that a study with a large sample size is more likely to find the statistically significant interaction effect between age and sex. Furthermore, when participants were young, it may be more difficult to find a statistically significant interaction because their vocal structures and biographies are more similar.

3.3 Reviewing Singing Behaviour based on Vocal Register Use and Vocal Pitch-Matching Accuracy for Songs

When singing behaviour was defined as being based on both vocal register use and vocal pitch-matching accuracy in songs, only Mang (2006) and Welch et al. (2009b) combined scores of the two singing perspectives to describe singing behaviour of Primary school children. The study of Mang (2006) involved *N* = 120 aged 7 and 9 children who were Cantonese monolinguals or English bilinguals in Hong Kong, China, who were asked to sing *Happy Birthday* in English. The study of Welch et al. (2009b) involved *N* = 8,799 singing performances of *Twinkle, Twinkle* and *Happy Birthday* provided by Year 1 to Year 6 Primary school children from diverse socioeconomic backgrounds across England. At this point in their research, most (66.88%) of participants had received some form of enhanced singing experience under the English national singing programme *Sing Up*. Both Mang and Welch et al. studies applied Rutkowski's *SVDM* scale (Rutkowski, 1996, 2015, 2018) to measure vocal register use, and used Welch's *VPMD* scale (Welch, 1998) to measure vocal pitch-matching accuracy for song(s). Mang and Welch et al. agreed that there was a positively significant correlation between vocal register use and vocal pitch-matching accuracy for songs. Their results, based on the combined scores between vocal register use and vocal pitch-matching accuracy for song(s), are compared below.

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. On the other hand, age had a statistically significant impact on English Primary school children's singing behaviour, with the mean scores of older children being statistically significantly higher than that of younger children in the study of Welch et al. (2009b). It is uncertain whether the different results by age between the two above studies was because of the different complexity level of target song(s), or the wide gap in sample size, or the diverse ranges of age groups. (In terms of the current PhD thesis, perhaps this uncertainty can be resolved by conducting another study based on the context of China, including a similar complexity of target songs with the study of Welch et al. (2009b), and relatively great sample size and a wide range of age groups.)

Secondly, in terms of the sex variable influence on Primary school children's singing behaviour, the studies of Mang (2006) and Welch et al. (2009b) agreed that girls' singing behaviour generally was better than that of boys. When boys and girls were provided with professional singing training in the study of Welch et al. (2009b), the sex difference in their singing behaviour still existed, with a continued developmental bias towards girls (Welch et al, 2010).

Thirdly, in terms of 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 Primary school children. It was not sure whether the difference results reported in the two above studies was because of different complexities of target songs, sample size or age groups. Only two age groups (ages 7 and 9) were included in Mang's study, while Years 1 to 6 were included in Welch et al.'s (2009b) study.

Fourthly, both of the above studies agreed that Primary school children's singing behaviour varied statistically significantly by culture. Mang (2006) reported that n = 60 Cantonese monolingual children showed statistically significantly better singing behaviour (including vocal register use and vocal pitch-matching accuracy) than n = 60 English bilingual children. This suggested that children speaking a tonal language seem to have an advantage in singing behaviour. Welch et al. (2009b) reported that the overall singing behaviour for Black (n = 526) and White (n = 6,550) children was statistically significantly higher than that for Asian children (n = 1,223) (including Indian, Pakistani, Bangladeshi, other). Mang (2006) suggested that the reported cultural difference might be because of the different number of languages used daily. In England, it can be conjectured that Black and English children were likely to speak only English at home and school, whilst Asian children might speak English in school, but speak another language, such as their parents' mother tongue, at home. Furthermore, it is uncertain whether the singing behaviour of these Asian children living in England would be representative of children living in their parents' original culture. For instance, the singing behaviour of Chinese children living in China might be different from that of Chinese children living in the United Kingdom, not least as the aims of the experienced music curriculum in Primary schools in the two countries are likely to be different.

Furthermore, Welch et al. (2009b) also reported that English Primary school children (including Asian children) who received *Sing Up* experience generally showed better singing ability (including vocal register use and vocal pitch-matching accuracy) than peers who did not receive such an enrichment programme. While a significant age difference was found between Year 2 and Year 3 in the baseline of the initial *Sing Up* data (Welch et al., 2009a), it disappeared for the same year groups in the subsequent research data collected one year later. Although the second data included a smaller sample size (n = 324) than the baseline (n = 3,510), it was believed that the non-age difference in the second data collection was contributed by the *Sing Up* training, which for Year 2 children dramatically improved their vocal pitch-matching accuracy of song. This suggested a positive correlation between positive singing experience and an improvement of rated singing ability. This conclusion is supported by the Pereira and Rodrigues (2019) findings. Overall, these previous studies also suggest that Primary school students either in areas of a higher socioeconomic status or with local and regional support for music, tend to be more likely to access professional singing training.

The conclusion is that effective professional singing training, financially supported, is likely to lead to more advanced singing ability in students, including wider vocal register use.

Lastly, Welch et al. (2009b) also reported that the highest and the lowest 40 schools for ranking of singing behaviour were located across England, irrespective of the socioeconomic differences between their localities. The participant schools were primarily chosen from inner city areas across England and nominated by the Local Music Education Hubs as providing examples of both positive singing cultures and also schools where children had not yet received specialist support. For English children who came from higher indices of multiple deprivation (IMD)[50] areas with *Sing Up* experience, their singing behaviour generally was more advanced than children from lower IMD areas without *Sing Up* experience. This suggests that socioeconomic status would not be a barrier if singing support is provided. 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, the English data based on socioeconomic status may not be representative for the singing data in another country with big rural and urban income differences, such as China. Understanding the correlation between singing behaviour and socioeconomic status in a country with a large economic difference between locations, such as China, could help us understand Primary school children's singing behaviour under its context better.

3.4 Summary of Chapter 3

According to previous studies, both complete songs and melodic patterns have been used to test vocal register use. Rutkowski's *SVDM* has been the most common rating scale used to assess vocal register use in children, especially in young children. Previous studies tend to agree that, in terms of age, children were more likely to use more vocal registers as they got older, and—in terms of sex—girls were more likely than boys to use more vocal registers. Socio-cultural and economic factors may also play a part, such as allowing access to high-quality singing support and resources, irrespective of location. Previous studies suggest that the application of vocal registers in singing can be improved by effective education.

Songs with diverse internal complexities have been used to test vocal pitch-matching accuracy

[50] English indices of deprivation 2019 - GOV.UK (www.gov.uk)

in songs. Children's singing performances have been mainly measured by using specially designed rating scales, whilst a small number used voice-acoustics related software. Previous studies found that Primary school children's vocal pitch-matching accuracy for songs generally improved with increasing age. However, this tendency might be subject to other variables, termed as internal (e.g., personal interest in a target song, and spoken language) and external (e.g., song choices, measurement tools, methods organization, and training opportunities). In general, girls are reported to tend to match pitches vocally more accurately than boys. The age and sex interaction effect on vocal pitch-matching accuracy for songs was more likely to be seen as significant when number of participants was large. These studies were, however, mostly undertaken from a Western cultural perspective.

Combining data of vocal register use and vocal pitch-matching accuracy in songs can show an overall fundamental singing behaviour. This type of singing behaviour tends to increase with age during the Primary school period, although the age difference can be reduced when matching a difficult song. Using the combined measures, girls' singing behaviour is generally better than that of boys. Previous studies did not agree on cultural or language influences on children's singing behaviour. Although socioeconomic status might have limited influence on the singing behaviour in one country, it might have a different impact in another country with a different socioeconomic structure.

3.5 Overall Summary of Chapters 2 and 3

3.5.1 Summary findings of Singing Behaviour Reviewed in Chapters 2 and 3

Table 3.5 shows a summary of six musical elements related to vocal pitch-matching accuracy for songs: vocal range, comfortable singing range, melodic contour, consistency of sung key centres, musical intervals, and individual pitches. Understanding these elements can help to explain vocal pitch-matching accuracy for songs. Overall, vocal range, comfortable singing range, and consistency in sung key centres tend to improve with increasing age for children. The vocal range of boys and girls tends to be similar, or that of girls might be slightly larger. Children's vocal range and accuracy in reproducing musical intervals can vary according to the singing materials used. The starting pitch, quantity and location of pitches can have an influence on vocal pitch-matching accuracy for single pitches. Furthermore, relative pitch patterns, i.e., open to transposition by the singer, are generally easier to be matched accurately than absolute pitch patterns.

Furthermore, previous studies showed that vocal range, consistency of sung key centres, vocal register use, and individual pitch accuracy are related. When the pitch height of the singing materials was above children's customary vocal range, they were more likely to change their tonality (Jones, 1979). It could be that children changed their reproduction of sung key centres because of their poor skill in transferring between vocal registers, which limits their available sung vocal range (Jones, 1979).

Overall, previous studies (e.g., Welch et al., 1996) agreed that vocally matching the pitches of a subset of musical elements tend to be easier than matching those of a song. When vocally matching a song, children tend to repeat their vocal pitch-matching errors when asked to match two similar melodies in a song (Feng et al., 2013). When children matched one target song accurately, they tended to match another target song accurately (Feng et al., 2013).

Table 3.5 A Summary of Musical Elements Related to Vocal Pitch-Matching Accuracy for Songs

Element	Independent variable	Findings summarised from previous studies
Vocal range	Age	Previous studies (e.g., Fieldhouse, 1937; Jersild & Bienstock, 1934; Paulsen, 1895; Wilson, 1970) tend to agree that Primary school children's vocal range expands with increasing age in a non-linear manner.
	Sex	The vocal range between boys and girls tends to be similar (e.g., Fieldhouse, 1937) or that of girls might be slightly larger (Jersild & Bienstock, 1934; Paulsen, 1895).
	Singing materials	Children tended to use a wider vocal range when matching a pitch pattern than a complete song (Flowers & Dunne-Sousa, 1990).
Comfortable singing range	Age	Previous studies tended to agree that children's comfortable singing range expands with increasing age (Welch, 1979b).
Melodic contour	A general tendency	Previous studies suggested children tended to match melodic contour more accurately than the constituent pitch elements (e.g., Demorest et al., 2018; Leighton & Lamont, 2006; Nichols, 2016; Welch et al., 1996; Wolf, 2005).
Consistency of key centre	Age	The ability to maintain a key centre when singing musical elements or songs generally improved with increasing age (e.g., Flowers & Dunne-Sousa, 1990; Welch et al., 1996).
Musical intervals	Type	Previous studies tended to agree that smaller musical intervals were easier to match accurately than bigger musical intervals (e.g., Chen-Hafteck, 1999; Moore et al., 1995/1996; Tillotson, 1972), except the minor second (Moore et al., 1995/1996).
Pitches	A starting pitch	Children, especially young children (Flowers & Dunne-Sousa, 1990), may find that it is difficult to start a song with a given target pitch.
	Quantity	Children tend to match a short singing material more accurately (e.g., Sinor, 1984; Welch et al., 1996; Wolf, 2005).
	Location	Children tend to match pitches at the beginning of a phrase more accurately (Feng et al., 2013; Welch et al., 1996).

	Relative pitches vs Absolute pitches	Relative pitches' patterns tend to be easier to match accurately than absolute pitches' patterns (Flowers & Dunne-Sousa, 1990; Moore et al., 1995/1996).

Table 3.6 compares the results of vocal pitch-matching accuracy for musical elements and songs, and vocal register use by key independent variables. Note: The empty spaces in Table 3.6 relate to gaps in the data from previous studies. As the table illustrates, singing materials and measurements applied to test singing behaviours have been diverse. The choice of such materials is likely to have an impact on the reports of children's vocal pitch-matching accuracy for musical elements and songs. Children tend to show better vocal pitch-matching accuracy when singing a favourite song. However, when findings in previous studies are analysed by the variables age and sex, they generally agreed that each of the three singing perspectives (melodic elements, whole songs, vocal register use) improved with increasing age for Primary school children. As groups, girls tend to show better scores for each of these compared to boys. Nevertheless, vocal pitch-matching accuracy for musical elements and songs and vocal register use can be developed by effective singing education. Such education could then lead to a smaller age and sex difference across a range of measured singing behaviours. The singing environment can also have an impact on children's vocal pitch-matching accuracy in songs. Older children, especially girls, appear to get more benefit from group singing. The influence of home language on children's vocal pitch-matching accuracy for songs requires further studies.

Furthermore, most of the previous studies were undertaken in Western cultural settings. Whether these findings are able to be replicated in another culture, such as Asia, is not certain.

Table 3.6 Comparing Findings on Vocal Pitch-Matching Accuracy for Musical Elements and Songs and Those on Vocal Register Use for Songs by Key Variables

Variable	**Vocal pitch-matching accuracy for musical elements**	**Vocal pitch-matching accuracy for songs**	**Vocal register use for songs**
Singing material	Single pitches, musical intervals, patterns (e.g., Demorest & Pfordresher, 2015; Demorest et al., 2018; Moore et al., 1995/1996; Nichols, 2016; Wolf, 2005)	Familiar or taught songs, self-selected songs or teachers' judge (see Table 3.1)	Rutkowski's (1996) patterns or familiar songs
Measurement	Mainly software showing fundamental frequency (e.g., Alberston, 1979; Cooper, 1995; Flowers & Dunne-Sousa, 1990; Goetze, 1985; Howard & Welch, 1993)	A range of rating scales or more objective measurements (see Table 3.4)	Wurgler's (1990) scale or Rutkowski's (1996, 2015, 2018) *SVDM* scale
Age	Previous studies (Cooper, 1995; Feng et al., 2013; Moore et al., 1995/1996; Sinor, 1984) generally agreed that vocal pitch-matching accuracy for musical elements improved with increasing age.	Most previous studies found that older children matched pitches of songs more accurately than younger children (Davies & Roberts, 1975; Gould, 1969; Green, 1994; Mizener, 1993; Pereira & Rodrigues, 2019; Welch et al., 2009a).	Most previous studies agreed that older Primary school children applied more vocal registers than younger children (e.g., Mang, 2006; Pereira & Rodrigues, 2019; Welch, et al., 2009a). However, the age difference in vocal register use might be small for young children, such as kindergarten children and 1st Grade pupils in Rutkowski's (2015) study.

Variable	Vocal pitch-matching accuracy for musical elements	Vocal pitch-matching accuracy for songs	Vocal register use for songs
Sex	Previous studies tended to agree that girls matched pitches of musical elements more accurately than boys (Moore et al., 1995/1996; Pereira & Rodrigues, 2019; Welch et al., 1997).	Previous studies (e.g., Bentley,1969; Gould, 1969; Green, 1994; Mang, 2006; Welch et al., 1997; 2008; 2009b) agreed that girls generally matched pitches of songs more accurately than boys.	Most previous studies found that girls generally used more vocal registers than boys (e.g., Mang, 2006; Welch et al., 2009a).
Training	-	Welch et al. (2009b) reported that English children who received *Sing Up* training tended to show better singing ability, including vocal pitch-matching for songs.	Children who were trained were more likely to use more vocal registers (Levinowitz et al., 1998; Pereira & Rodrigues, 2019; Welch et al., 2009b; Wilson, 1970).
Socioeconomic status	-	Welch et al. (2009b) reported that children from less advanced socioeconomic status tended to show better singing behaviour, including vocal pitch-matching accuracy for songs.	It is assumed that children who received singing training were more likely from higher socioeconomic status.
Language	-	Mang (2006) suggested that children who spoke a tonal language generally showed better singing ability. However, the influence of language on children's singing behaviour was uncertain in the study of Welch et al. (2009b).	-

Variable	Vocal pitch-matching accuracy for musical elements	Vocal pitch-matching accuracy for songs	Vocal register use for songs
Age * Sex	The sex difference in vocal pitch-matching accuracy for musical elements tended to be smaller for younger children than older children (Leighton & Lamont, 2006; Moore, 1994; Sinor, 1984; Welch et al., 1997).	The age * sex interaction effect on vocal pitch-matching accuracy for songs might be significant. The sex difference might increase with increasing age (Davies & Roberts, 1975; Gould, 1969; Leighton & Lamont, 2006; Welch, 1979a; Welch et al., 1997; 2009b). The sex difference could be smaller or disappear for young children (age ⩽ 6+) (e.g., Chen-Hafteck, 1999b; Leighton & Lamont, 2006; Smale, 1987) or when sample size was not big enough, such as comparing studies between Welch et al. (2008) and Welch et al. (2009b).	-
Age * Singing Material	-	The age difference in vocal pitch-matching accuracy for songs was more likely reduced when Primary school students matched a difficult target song (Mang, 2006; Rutkowski, 1996) or when children self-selected their songs (Mizener, 1993).	-
Age * Measurement	-	The result of the age difference in vocal pitch-matching accuracy for song might be different when measured by a rating scale or a more objective method (e.g., Leighton & Lamont, 2006).	-

Variable	Vocal pitch-matching accuracy for musical elements	Vocal pitch-matching accuracy for songs	Vocal register use for songs
Age * Training	The age difference in musical intervals tended to be smaller when singing training was provided (Moore et al., 1995/1996).	The age difference in vocal pitch-matching accuracy for songs might be reduced if training is provided (Boardman, 1964; Guibault, 2004; Welch et al., 2009b). The improvement in vocal pitch-matching accuracy for songs was generally faster for young than older children (Welch et al., 2009b; Jacobi-Karna, 1996).	-
Age * Singing alone/in a group	-	The age difference might be greater when Primary school children sang in a group than alone, especially for older children (e.g., Green, 1994).	-
Age * Interest	-	Older children might match a song poorly if they did not like the target song (Mizener, 1993).	-
Sex * Singing materials	The increasing pitch height had greater negative impact on boys' vocal pitch-matching accuracy than that of girls (Trollinger, 2003).	-	-
Sex * Training	Boys could match pitches of musical elements well when they were encouraged by music teachers regularly (Moore et al., 1995/1996).	-	The sex difference in vocal register use might be reduced if boys received singing training (Dansereau, 2005).

Variable	Vocal pitch-matching accuracy for musical elements	Vocal pitch-matching accuracy for songs	Vocal register use for songs
Sex * Singing individually/in a group	-	Singing in a group had greater negative impact on young boys' vocal pitch-matching accuracy for songs than that of girls (Goetze & Horii, 1989; Leighton & Lamont, 2006).	-
Age * Sex * Talent	It might be difficult to find the sex difference in vocal pitch-matching accuracy for musical elements for older talented singers in a Primary school (Moore, 1994).	-	-

Note. If grids are marked with the same colour, it means that their contents reached a similar summary based on an independent variable or any of their interaction.

3.5.2 Using Bronfenbrenner's Ecological Theory to Explain the Summary of Section 3.5.1

The rationale for applying Bronfenbrenner's ecological theory is that there is considerable literature concerning the impact of different environmental contexts on child development. In terms of children's musical development, this has been reported by Wu (2018), Wu and Welch (2022) and Welch (2022).

Bronfenbrenner's (1977, 1979) ecological theory of human development involves five levels, named microsystem, mesosystem, exosystem, macrosystem, and chronosystem. The microsystem means the interaction between a developing person and a setting that has an immediate effect on their activity. The mesosystem focuses on the interaction effect between two or more settings at a particular point in time for a developing person. The exosystem does not directly influence a developing person, but influences settings in the mesosystem. The macrosystem refers to a particular culture or sub-culture bias for the mesosystem. The chronosystem consists of all the environmental changes that occur during a lifetime and their influence on a developing person.

Figure 3.6 shows a summary of children's singing behaviour (related to vocal register use and vocal pitch-matching accuracy) suggested by previous studies (Tables 3.5 and 3.6) by using the structure of Bronfenbrenner's ecological theory. As the figure illustrates, the microsystem includes four settings: children's physical development (see Sections 3.1.3, 3.14, 3.2.3 and 3.2.4), research design (see Sections 3.1.1, 3.1.2, 3.2.1 and 3.2.2), school (see Section 3.1.5) and home. Most of their immediate influence on Primary school children's singing behaviour is reviewed in the above sections. Compared with school, the setting of the home seems to receive less attention to its effect on Primary school children's singing behaviour, although some studies have paid attention to it (e.g., Welch et al., 2009b; see Williams et al, 2015 for pre-school family influences). This might be because children in this phase of their lives spent more time in school than at home. The power of singing activities at home seems to be more obvious for younger children. Previous studies (Aherne, 2011; Kirkpatrick, 1962; Michaud, 2014; Shelton, 1965) agreed that young pupils who came from families with a wider range of musical activities tended to match pitches of song singing more accurately, compared with their peers who had less musical experience at home as infants.

The mesosystem focuses on the interaction between two or more settings within the microsystem. Based on reviewing previous studies, five interactions between two settings of the microsystem: (1) singing material * bio-development (measured by age or sex), (2) measurement * bio-development, (3) singing training in schools * bio-development, (4) music teachers' encouragement on boys' vocal pitch-matching ability, and (5) settings (e.g., singing individually or within a group) * bio-development, are summarised. Here is an example illustrating their interaction effects.

Firstly, for singing material * bio-development, due to the growing length of the vocal folds by increasing age during the Primary school period (see a review in Chapter 2, Section 2.2.1), children's vocal range and comfortable range tends to expand with age (see a review in Chapter 2, Section 2.5.1). This wider vocal range increases the opportunity to sing a song with a wider pitch range more accurately.

Secondly, in terms of measurement * bio-development, the study of Leighton and Lamont (2006) reported that neither software nor the choice of rating scale found a difference between the two target songs, when participants were young (aged 4 to 6). However, they had different results when the children were older. When they were aged 6 to 8, using software (*Praat*) to show fundamental frequency, the researchers found a significant difference in vocal pitch-matching accuracy between two target songs. Nevertheless, when measured by a rating scale, these children did not register a difference between the two songs.

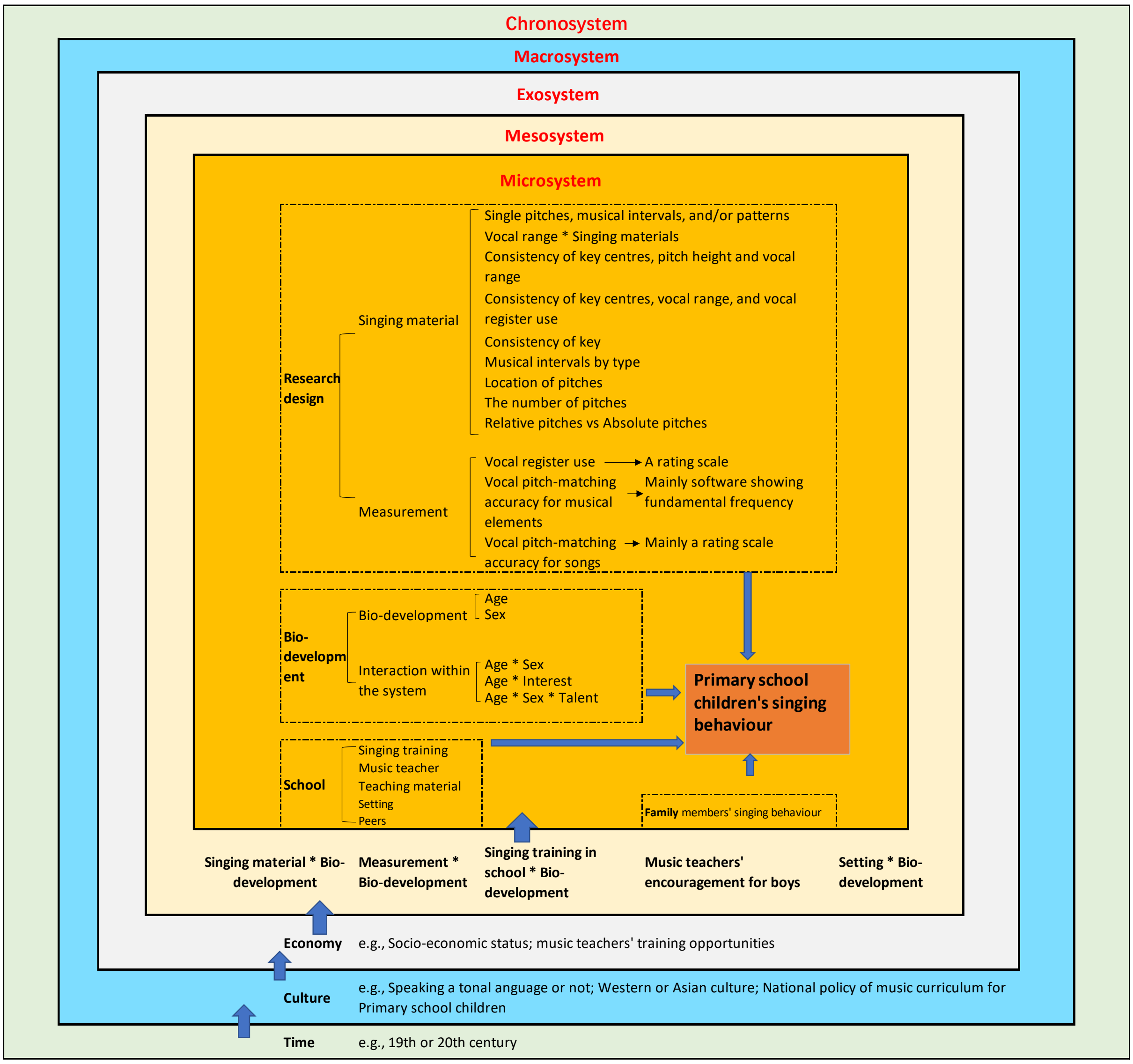
For the interaction between singing training in school and bio-development, Welch et al. (2009b) and Papageorgi et al. (2022) showed that younger Primary school children appeared to benefit more from singing training than older Primary school children to improve their singing ability.

Lastly, for setting * bio-development, previous studies (Goetze & Horii, 1989; Leighton & Lamont, 2006) agreed that singing in a group had a greater positive impact on young girls' vocal pitch-matching accuracy for songs than on that of boys. Older children appear to benefit more from group singing than younger children (Green, 1994).

For an example in terms of the exosystem, while Portuguese young children from a private school showed a high vocal register use in the study of Pereira and Rodrigues' (2019) study, this might be because of training provided by a music specialist, in the context of a high socioeconomic group. On the other hand, Welch et al. (2009b) reported that English children from less advantaged groups did not necessarily have less developed singing behaviour, especially if their school leaders had positive attitudes toward singing. The research team suggested that school leaders' attitudes to singing seem to have greater influence than a disadvantaged socioeconomic status (with a higher IMD[51] score) on children's singing behaviour. While most previous studies were undertaken in Western culture (the macrosystem), it is not known what the singing behaviour of children from other cultures looks like in the same (21st) century (the chronosystem).

[51] English indices of deprivation 2019 - GOV.UK (www.gov.uk)

Figure 3.6 A Summary of Previous Findings about Children's Singing Behaviour by Bronfenbrenner's Ecological Theory



Chapter Four

Methodology

The literature on children's singing behaviour and development is primarily based in Western cultures, with very little published empirical research having been undertaken in mainland China. Consequently, there was a need to discover if the evidence from elsewhere in the world might apply in a Chinese mainland context. The evidence from Hong Kong suggested that this might be applicable. The available literature tended to focus on age and sex as key variables, with little published study of geographical location and social class which is why these two variables were included in the thesis design.

Under the guidance of the three research questions (see Chapter 1, Section 1.7), the current study is essentially descriptive research, i.e., 'Descriptive research is an appropriate choice when the research aim is to identify characteristics, frequencies, trends, and categories' (Scribbr, 2023 – see https://www.scribbr.co.uk/research-methods/descriptive-research-design/), although inferential statistics are also used to determine aspects of potential statistical significance within the emergent dataset. The research seeks to illustrate Chinese Primary school children's singing behaviour from both macro (whole song) and micro (song elements) perspectives. The current Chapter reports methodological choices as applied to children's singing behaviour in a participant population from China. It starts with the process and findings of the pilot study, and this is followed by the details of the main phase of data collection.

4.1 The Pilot Study

A pilot study was undertaken in China with appropriate ethical approval, in order to explore the validity and reliability of the proposed main research study methodology.

4.1.1 Participants

During the pilot singing assessment, five adults (three relatives and two music teachers) helped to collect data from 15 participant children, drawn from five Primary schools in the northeast of Hunan Province, China.

The pilot study was undertaken from August to September 2016 in two phases. All the pilot participants were seven-year-olds from the 2nd Grade in Primary school. The age group was the second youngest age group proposed for the main fieldwork and was chosen on the hypothesis that, if the pilot procedure worked with them, it would probably also be appropriate for a slightly wider, including older, age range of participants in the main study. In the pilot study, the prime focus was in ensuring that the research protocol would work appropriately. Therefore, a small number of children aged 7y+ were included and these were chosen by their teachers.

There were children of both sexes (10 boys and 5 girls) in the pilot study to explore any possible sex differences in children's understanding of the pilot study tasks. Participants from two Primary schools were selected by their music teachers.

All participants had Han ethnicity, which makes up 91.6% of the population of China (State Council Census Office & NBSC, 2011)[52] with a ratio of 2:3 pilot children drawn from rural areas. These schools were chosen as a convenience sample through personal contacts.

---

[52] index (stats.gov.cn)

None of the child participants had received any specialised singing training and all were volunteers (with appropriate informed consent and ethical approval). All participants undertook an individual singing assessment (see below). Under the ethical approval, all children and adults were permitted to withdraw from the study at any time for any or no reason, in line with the British Educational Research Association guidelines (BERA, 2018)[53]. Personal information, including any school and family details, was anonymised in the subsequent reporting.

4.1.2 Singing Assessment Protocol

The singing behaviours of three songs were assessed. The three songs embraced one with a relatively simple pitch range (*Twinkle, Twinkle*), one that had a more complex pitch range (*Happy Birthday*), and a Chinese nursery song (*Little Donkey*) (see more details of the analysis of the relative complexity of each criterion song in Section 4.2.2).

Mang (2006) suggested that singing a whole song is the most frequent activity in a Chinese music class. The research protocols for the singing assessment were based on the assessment criteria used in the evaluation of the UK *Sing Up* programme (see Welch et al., 2009a, 2009b, 2011) (see Figure 4.1). The three criterion songs were modelled on a recording by an adult female and were sent to relatives and music teachers by email or social media.

The music teachers were told that the purpose of the pilot study was to gather information on Chinese Primary school children's singing behaviour and whether the proposed research protocol was effective. They were asked to play the model recordings to the participant child several times across a one- to two-week period and to note how many times these were played before the children became familiar with the songs. Two weeks later, individual child song singing data were collected, either by the researcher's volunteer relatives using video at the child's home (school 1) or by the music teachers (schools 1 and 2) who recorded individual children's singing in a quiet place within their schools using audio recording only.

No starting pitch was given in any of the tasks so that children could start with their own choice

---

[53] https://www.bera.ac.uk/researchers-resources/publications/ethicalguidelines-for-educational-research-2018

of comfortable singing pitch.

The three songs were measured against the criteria listed in the Vocal Pitch Matching Development (*VPMD*) and Singing Voice Development Measure (*SVDM*) rating scales (as in Figure 4.1) (see more analyses of the two rating scales in Chapter 4, Section 4.2.3).

Figure 4.1 The Singing Voice Development Measure (SVDM) and the Vocal Pitch-Matching Accuracy (VPMD), Based on Rutkowski (1996, 2015) and Welch (1998)

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 intervals 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.

The *SVDM and VPMD* ratings (derived from the framework in Figure 4.1) were used together to assess participant Primary pupils' singing behaviour. The *SVDM* (Rutkowski, 1996, 2015) scale was designed to measure children's use of voice registers in singing. Register boundaries in *SVDM* were reported to be C4, F4, and Bb4, and these created five basic rating categories, some of which were then subdivided. The *VPMD* rating was summarised by Welch (1998) as measuring participants' vocal pitch accuracy in song singing.

Mang (2006) used the two rating systems (*SVDM* and *VPMD*) in combination to evaluate the singing competence of 60 Cantonese-speaking children and 60 English-speaking children in Hong Kong. A high correlation between the two scales was reported ($r = .821$, $N = 120$, $p = .05$), based on the ratings of two judges.

Subsequently, Mang's approach was adopted in the evaluation of the UK Government's National Singing Programme *Sing Up* in England (Welch et al., 2009a, 2009b). This UK research also used the two rating scales, both singly and in combination, initially to measure the singing behaviour of $N =$ 8,799 children aged from 67 months to 144 months in England. In addition, ratings from the two measures were combined to build a single assessment of a child's song singing behaviour – a 'normalised singing score' (NSS) (Welch et al, 2009b).

4.1.3 Data Analyses

For the data synthesis of the singing assessment in the pilot study, a small selection of the video or audio recordings of the three criterion songs were reviewed by two senior judges to check the reliability of the researcher's initial assessments.

The *SVDM* and *VPMD* scores were combined into one final normalised singing score (NSS), using an overall mean of the mean scores across the three focus songs as measured by the two rating scales, and calculated out of 100. For the pilot study, the independent variable was sex.

Sex differences in the singing of the three songs was calculated by an Independent – samples t-test, using a significance value at $p < .05$.

4.1.4 Results

It was noted that participants seemed more relaxed when just being asked to make an audio recording of their singing voice than when giving a singing performance in front of a camera.

Fifteen participants (*n* = 5 girls, *n* = 10 boys) sang the three criterion songs, except for one participant who did not sing *Little Donkey* as the adult forgot to ask her.

In terms of **sung vocal register use**, as assessed by Rutkowski's *SVDM* scale for each song, the ratings of the three songs were significantly correlated (see Table 4.1). This suggests that participants sang the three target songs using similar vocal registers. In terms of sex differences, the mean of vocal register use for the three target songs was not statistically different significantly between boys (*M* = 3.37, *SD* = 0.67) and girls (*M* = 3.80, *SD* = 0.42), (*t*(12) = 1.30, *p* = .219).

Looking at the data in terms of the **vocal pitch matching** according to the *VPMD* scale (see Table 4.2), the inter-correlations between the three songs were significantly positive. This suggests that a participant's vocal pitch matching for one song would be statistically similar to that for the other songs. In terms of any sex difference, girls (*M* = 3.33, *SD* = 0.24) were rated as matching pitches of the three target songs significantly better than boys (*M* = 2.43, *SD* = 0.57), *t*(13) = 3.35, *p* = .005.

When the two separate sets of ratings vocal register use (*SVDM*) and vocal pitch matching (*VPMD*) were combined into a normalised singing score (NSS), the data for the participant seven-year-olds indicated that the mean of girls' singing competence (*M* = 79.67, *SD* = 5.42) was statistically significantly greater than that of boys (*M* = 64.72, *SD* = 12.77), *t*(12) = 2.46, *p* = .03.

Overall, the pilot study suggested that the singing materials and measurements would be appropriate for the main fieldwork (see more reports on the pilot study from Lu et al., 2019).

Table 4.1 Mean Values, Standard Deviations, and Intercorrelations of SVDM Ratings of the Three Criterion Songs

Song	*M*	*SD*	1	2	3
Twinkle, Twinkle	3.43	0.73	-	-	-
Little Donkey	3.54	0.69	0.766**	-	-
Happy Birthday	3.61	0.59	0.814**	0.598*	-

Note. *. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4.2 Mean Values, Standard Deviations, and Intercorrelations of VPMD Ratings of Three Criterion Songs

Song	*M*	*SD*	1	2	3
Twinkle, Twinkle	2.67	0.90	-	-	-
Little Donkey	2.87	0.74	0.677**	-	-
Happy Birthday	2.67	0.62	0.557**	0.519*	-

Note. *. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

4.2 The Main Study

One long-term applied research purpose of the study is to help music teachers at Primary schools in China to understand more about students' singing behaviour and development. Consequently, the research aim was to build a rounded picture of students' singing behaviour and development at different ages, taking into account variables related to age, sex, geographic location (rural/urban), and inferred parental income.

The singing tasks were the same as those applied in the pilot study (see Section 4.1.2), in which participants were asked to sing three familiar songs. In the following section, more details about participants and context, singing tasks, the process of data collection, measuring singing data, data storage, and analyses are shown.

4.2.1 Participants

(a) Schools and Location across Two Visiting Years (2017 – 2019)

All participants came from six schools located in Hunan Province, China. Four schools were in four rural areas – one school in each, while two schools were in an urban area. The reason why the number of rural schools was greater than that of urban schools is that the size of an urban Primary school is usually greater[54]. Two rural schools were in villages, while two of them were near local town centres. Participants in one urban school were from the town, while children in another urban school were originally from rural areas. 'Different types of schools were selected to ensure a diverse range of school singing cultures were accessed' (Welch et al., 2009b, p.19). All participants were from the Han ethnic group, and they tended to speak Mandarin in schools. Furthermore, some participating schools were near each other, such as Schools C and D, and Schools E and F. This proximity may reduce any cultural or linguistic disparities between them.

With regards to the tonal structure of the three target songs, these were chosen because of their cultural familiarity and the subsequent analyses using the new assessment model investigated the overall sung tonality of each song and also any shifts in tonality. Note was made of the relative accuracy of particular intervals within the songs (see Section 5.3). There is no evidence that intervals from a pentatonic scale were sung much more accurately than other intervals.

Music textbooks used across the six participating schools were the same, which were published by the Hunan Literature and Art Publishing House. As an example, Appendix E provides analyses of the curriculum of the first academic term of a music textbook for fourth-grade students published by Hunan Literature and Art Publishing House. The textbook includes n = 18 songs for learning to sing. 14 of the 18 songs are Chinese. The melodies of two further songs are from Western countries

[54] 中央编办 教育部 财政部关于统一城乡中小学教职工编制标准的通知 - 中华人民共和国教育部政府门户网站 (moe.gov.cn)

(America and Italy), but their lyrics are translated into Chinese. Furthermore, either melodies or lyrics of another two songs are from Japan. Each of the songs involves one tonality only, but the choice of key varies across the songs. The pitch ranges of 16 of the 18 songs are ≥ an octave. Consequently, these songs required the use of chest-to-head registers. More analyses are needed to explore further in future research.

All participants in the six research schools that were visited in 2017-2018 – the first year of the main study data collection – were invited to continue in the second year 2018-2019 in order to explore any difference in singing behaviour across two visiting years. In the second visiting year, more participants from the six schools were invited in order to create an overview of Primary school children's singing behaviour across age groups.

Figure 4.2 Location and the Number of Participating Schools of Each of the Provincial Cities in Hunan Province, China

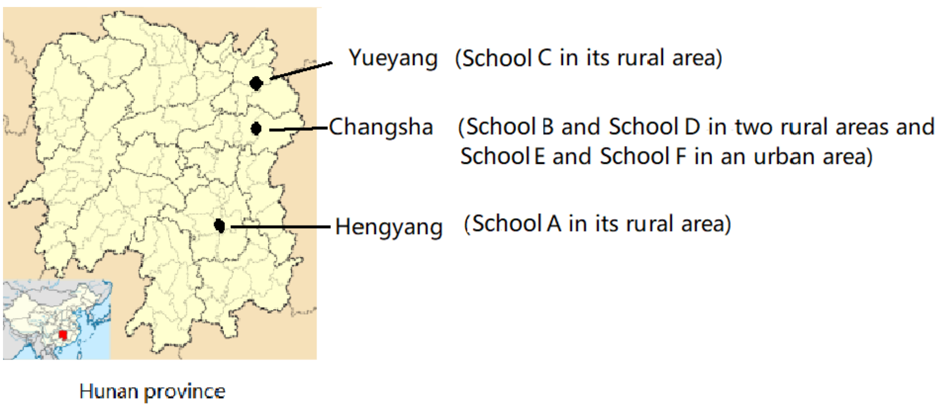
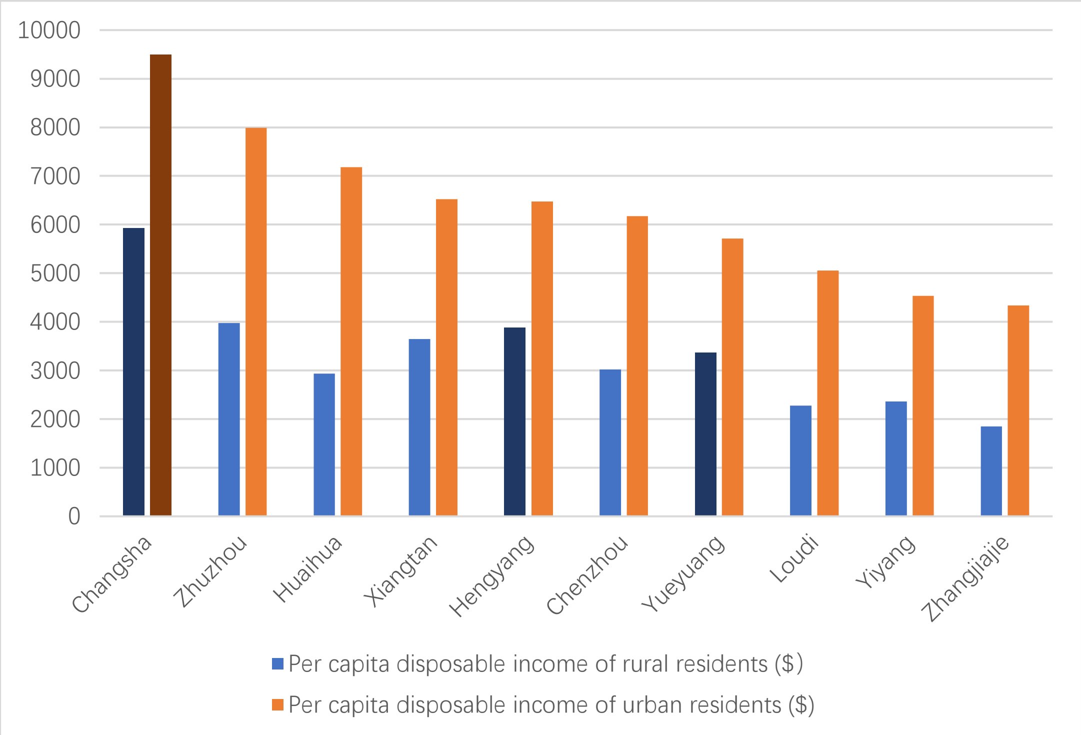


Figure 4.3 A Line Figure Showing Per Capita Disposable Income between Rural and Urban Residents among Ten Cities of Hunan Province[55]



Note. Four bars are marked with darker colours to show per capita disposable income for residents from participating areas.

[55] Data for the figure were collected from each of the ten cities' official websites that reported the latest per capita disposable income for the ten cities in Figure 4.3 in 2021 or 2022.

Table 4.3 School Size and Resources of Music Teachers of Each of the Participating Schools in Hunan Province, China

School	School Size from Grades 1 to 6	No. of Music Specialists	No. of Music Specialists Teaching Music Only
A	≈120	1	0
B	≈250	2	0
C	≈260	2	0
D	≈500	2	1
E	≈600	3	3
F	≈700	4	4

The six participating schools came from three sub-administrative regions of Hunan Province, including rural (Schools B & D) and urban areas (Schools E & F) of Changsha, a rural area of Yueyang (School C), and Hengyang (School A) (see locations in Figure 4.2). Per capita disposable incomes between rural and urban residents in the three cities are reported in Figure 4.3. As the figure illustrates, the rural and urban per capita disposable incomes of Changsha (the capital city of Hunan Province) were the highest among other cities in Hunan Province. Per capita disposable incomes in rural areas of Hengyang and Yueyang were closer to that of most provincial cities. Furthermore, some participating schools were located near each other, such as Schools C and D, and Schools E and F. These factors are believed to have reduced the cultural or linguistic disparities among them.

The choice of schools was based on seeking to contrast those in rural and urban areas, drawing initially on personal contacts, then checking how they might match against the focus variables related to location, likely parental occupations and their interest in participating in the research. 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. Students at these schools came from local areas. It was 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.

As Table 4.3 shows, although each of the four rural participating schools (Schools A to D) employed music specialists, six of the seven teachers were not able to focus solely on teaching the music curriculum because they were expected to cover other subjects as well. In contrast, all the music teachers in the two urban participating schools (Schools E and F) were music specialists who could teach music curriculum only. Also, compared with the four rural participating schools, the music classrooms in the two urban schools were used regularly. One difference in the music environment between the two urban schools was that only School F had an orchestra. Furthermore, School F was also famous for its advanced academic performance in the province. Overall, the unequal opportunity to teach the music curriculum only between rural and urban areas across the six schools provides an opportunity to explore the difference in children's singing behaviour by geographic location.

Furthermore, the impact of unequal socioeconomic development within a geographic area on children's singing behaviour in China was explored by relative analysis of parents' income. The families' income was inferred, based on the parents' occupation. 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, they were categorised as a group of rural parents with relatively lower income[56] (*cf.* children at Schools A and B). If rural parents owned a shop in a local town centre, they were categorised as another group of rural parents with relatively higher income (*cf.* children at Schools C and D). In urban areas, if parents worked in a factory or ran a small shop in a city, they were categorised as a group of urban parents with relatively lower income (*cf.* children at School E). If urban parents were professionals, such as doctors, they were more likely to be recognised as higher-income parents (*cf.* children at School F). Although the description might be not suitable for every participant within a school, informal conversations with teachers and pupils suggested that these inferences were appropriate. The above unequal socioeconomic development was partly caused by unequal payment among varied occupations in society (*cf.* Hong, 2022).

[56] According to Hong (2022), farmers' income in rural areas is the lowest compared with those working in other occupations.

(b) The Number of Participants and Singing Performances across the Two Years (2017-2019)

From 2017-2018 to 2018-2019, *N* = 1,193 Primary school students from six participating schools in Hunan Province, China were involved[57]. In the first year of data collection (2017-2018), *n* = 453 children in Grade 2 (age 7+) and Grade 4 (age 9+) were assessed. In the second year (2018-2019), *n* = 346 (76%) of these (now in Grade 3 and Grade 5) could be assessed again, thus providing a longitudinal perspective of their singing development. In addition, in 2018-2019, another *n* = 740 participants from Grades 1, 2, 4, and 6, were assessed to build a picture of children's singing behaviour across all Primary school grades. Across the two years, *N* = 1,539 singing performances were recorded from the *N* = 1,193 participants. Overall, this large number of participants allowed for a smaller standard error and ensured that data from a certain number of participants were available when exploring independent variables and their relationships, related to age, sex, geographic location, inferred parental income, and school.

Across the whole Chinese dataset, the Primary school participants were drawn from Grades 1 to 6, mainly aged between 6+ and 11+. The number of participants and singing performances for each age group (by School year) in each visiting year are shown in Table 4.4. It is recognised that using whole years has the drawback of having some children relatively very young and some older within a Year group, but this whole year approach was adopted here as it is conventional in such research internationally. Due to a comparatively small number of 12-year-olds (*n* = 41), their singing performances were combined with those of 11-year-olds to reduce the standard error for the oldest participants.

In the current study, there were *n* = 655 boys (*n* = 780 singing performances) and *n* = 538 girls (*n* = 759 singing performances) participating. Of these, *n* = 810 participants (*n* = 1,015 singing performances) came from rural areas, and *n* = 383 participants (*n* = 524 singing performances) were from urban areas.

---

[57] Using a 95% confidence interval, 5% margin of error, 50% of population proportion, and *N* = 5,116,600 child population size in Hunan Province[57], the minimum sample size was suggested as *N* = 384 participants[57].

Table 4.4 The Number of Participants and Singing Performances by Age in Whole Years across Two Visiting Years (2017-2019)

Age in whole years	Participants in 2017-2018		Participants in 2018-2019		Performances in 2017-2018	Performances in 2018-2019	Performances Total
	Tested in 2017-2018 only	Longitudinal data in 2017-2018	Longitudinal data in 2018-2019	Tested in 2018-2019 only			
6+	-	-	-	76	-	76	76
7+	25	74	-	158	99	158	257
8+	29	95	74	158	124	232	356
9+	27	96	95	144	123	239	362
10+	25	80	96	116	105	212	317
11+	1	1	81	88	2	169	171
Total	107	346	346	740	453	1086	1539

Also, *n* = 610 (*n* = 711 singing performances) and *n* = 583 participants (*n* = 828 singing performances) were from lower- and higher-income families, respectively. The reason that the number of singing performances was greater than that of participants was that a small number (*n* = 346) of participants were tested twice.

Before analysing all (*N* = 1,539) singing performances provided by all (*N* = 1,193) participants, one singing performance for each participant was analysed (reported in Chapter 6, Section 6.1) – without any longitudinal perspective – by taking the singing data of participants who were tested once only, either in 2017-2018 or in 2018-2019. This dataset includes all participants with one singing assessment for each participant. This could avoid the error of statistical analyses that mixed singing data provided by participants tested both once and twice. The related number of singing performances based on the dataset was shown in Table 4.5 (below).

Table 4.5 The Number of Participants Tested Only in 2017-2018 and All Participants Tested in 2018-2019 (N = 1,193)

Variable	Sub-group	Age in Whole Years						Total
		6+	7+	8+	9+	10+	11+	
Sex	Boys	29	111	134	155	137	89	655
	Girls	47	69	125	115	102	80	538
Geographic	Rural	76	137	166	142	147	142	810
Location	Urban	-	43	93	128	92	27	383
Inferred	Lower	35	90	122	139	137	87	610
Parental Income	Higher	41	90	137	131	102	82	583
School	A	14	29	28	27	28	31	157
	B	21	39	40	51	49	40	240
	C	21	40	46	41	38	54	240
	D	20	29	53	23	32	17	174
	E	-	22	53	61	60	16	212
	F	-	21	39	67	32	11	170
Total		76	180	259	270	239	169	1193

Then, in order to maximise the utilisation of singing data, all singing performances (*N* = 1,539), including those of *n* = 346 children tested twice across two visiting years, were analysed (reported in Chapter 6, Section 6.2). The number of singing performances by age-related interactions under the dataset is shown in Table 4.6 (below). Furthermore, the number of participants tested twice using age-related interactions is shown in Table 4.7 (the results were reported in Chapter 6, Section 6.4).

There was no consideration of absolute pitch ability as such in the participant profiles because this is reported to be relatively rare in the general population, i.e., less than 1:10,000 people (Deutsch, 2019). Although some participant children may have had extra musical experiences and thus developed a more acute sense of pitch, the demographic data suggest that any such children were likely to be in a very small minority across the whole dataset.

Table 4.6 The Number of All Singing Performances by Age Related Interactions (N = 1,539)

Variable	Sub-group	Age in Whole Years						Total
		6+	7+	8+	9+	10+	11+	
Sex	Boys	29	134	174	185	167	91*	780
	Girls	47	123	182	177	150	80	759
Geographic Location	Rural	76	183	221	193	198	144	1015
	Urban	-	74	135	169	119	27	524
Inferred Parental Income	Lower	35	110	156	164	158	88	711
	Higher	41	147	200	198	159	83	828
School	A	14	34	32	38	31	31	180
	B	21	44	53	56	62	41	277
	C	21	59	71	59	63	54	327
	D	20	46	65	40	42	18	231
	E	-	32	71	70	65	16	254
	F	-	42	64	99	54	11	270
Total		76	257	356	362	317	171	1539

Note: * This number of male participants aged 11 does not include those whose voices were changing at the time of assessment due to the onset of puberty because their vocal ranges and registers would be different to the rest of the male participants with unchanged voices (*cf.* Cooksey & Welch, 1998).

Table 4.7 The Number of Participants Tested Twice Across Two Visiting Years by Age in 2017-2018, Sex, Geographic Location and Income (n = 346)

Variable	Sub-group	Age in Whole Years in 2017-2018					Total
		7+	8+	9+	10+	11+	
Sex	Boys	36	49	46	40	1	172
	Girls	41	43	52	38	-	174
Geographic Location	Rural	39	54	54	54	1	202
	Urban	38	38	44	24	-	144
Inferred Parental Income	Lower	39	47	49	41	-	176
	Higher	38	45	49	47	1	170
School	A	12	9	13	6	-	40
	B	11	23	13	23	-	70
	C	15	21	12	17	-	65
	D	1	1	16	8	1	27
	E	16	16	23	12	-	67
	F	22	22	21	12	-	77
Total		77	92	98	78	1	346

4.2.2 Singing Tasks

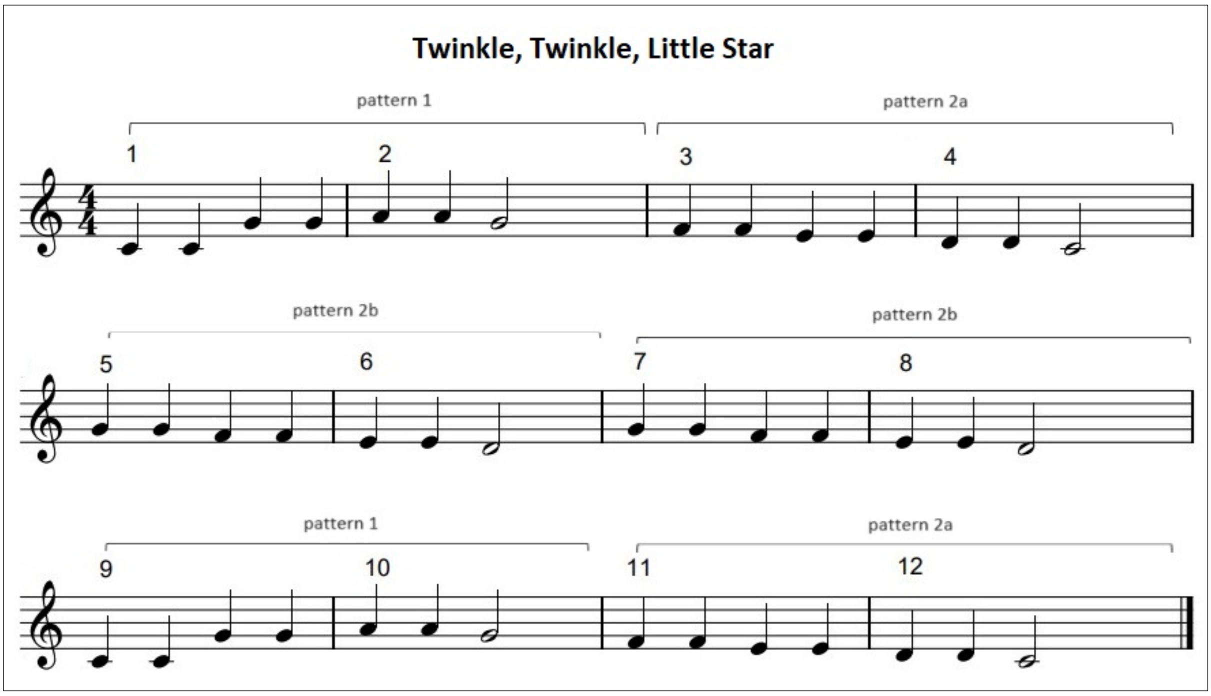
In order to test Primary school children's ability in song singing, participants were asked to sing three well-known criterion songs, *Twinkle, Twinkle*, *Little Star* (called *Twinkle, Twinkle* hereafter), *Little Donkey* (a typical Chinese nursery song) (see the tune in Figure 4.4), and *Happy Birthday,* with lyrics provided.

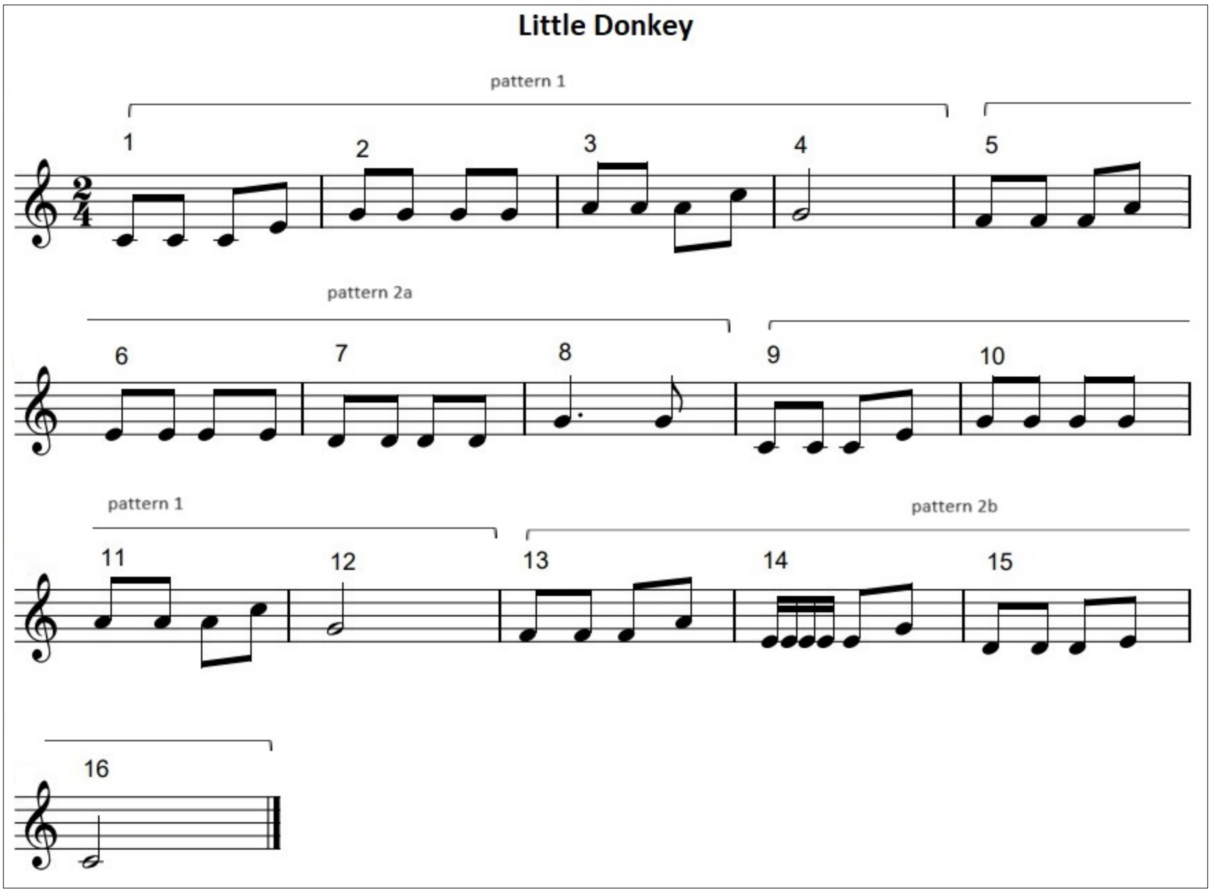
Twinkle, Twinkle, and *Happy Birthday*, thought to be two typical Western children's songs but also well-known in China, were selected to be able to compare their singing data between Chinese participants in the current study and English participants evaluated in the *Sing Up* programme. It should be noted that both *Twinkle, Twinkle,* and *Happy Birthday* were introduced to China from overseas during the 20th century. Their melodies were kept, but their lyrics were translated into Chinese. It seems that the text-melody relationship has a limited impact on children's vocal pitch-matching accuracy of song singing (Chen-Hafteck, 1999b). Furthermore, both *Twinkle, Twinkle,* and *Happy Birthday* are popular with children in China according to participants' comments. They might learn these two songs at kindergarten and/or from their parents. Furthermore, *Happy Birthday* is a typical song sung at a Chinese birthday party. The traditional Chinese song *Little Donkey* was chosen to explore any possible difference in singing when the place of origin of a song is different.

The three criterion songs are defined as short familiar songs for participants. The conventional pitches of the three songs with the bar numbers of each of the target songs presented in Figure 4.4 (below). The starting pitch of *Twinkle, Twinkle,* and *Little Donkey* is the tonic, while that of *Happy Birthday* is the dominant.

Furthermore, thirteen characteristics of the three target songs are summarised in Table 4.8 to illustrate the diverse complexities of the three target songs. As the table illustrates, the pitch range of *Little Donkey* and *Happy Birthday* is wider than that of *Twinkle, Twinkle*. Only the melodies of *Twinkle, Twinkle,* and *Little Donkey* include repeated patterns. The number of notes, types of musical intervals and rhythm, and lyrics across the three target songs vary. As previous studies reported that children generally showed high accuracy when matching rhythm and lyrics (e.g., Welch, 1994; Welch et al., 1996), these two musical elements were not involved in the current analyses of children's singing behaviour. Instead, the current study focuses on pitch-related singing behaviour.

Figure 4.4 Melodic Contour and Pitches with Patterns of Each of the Three Target Songs





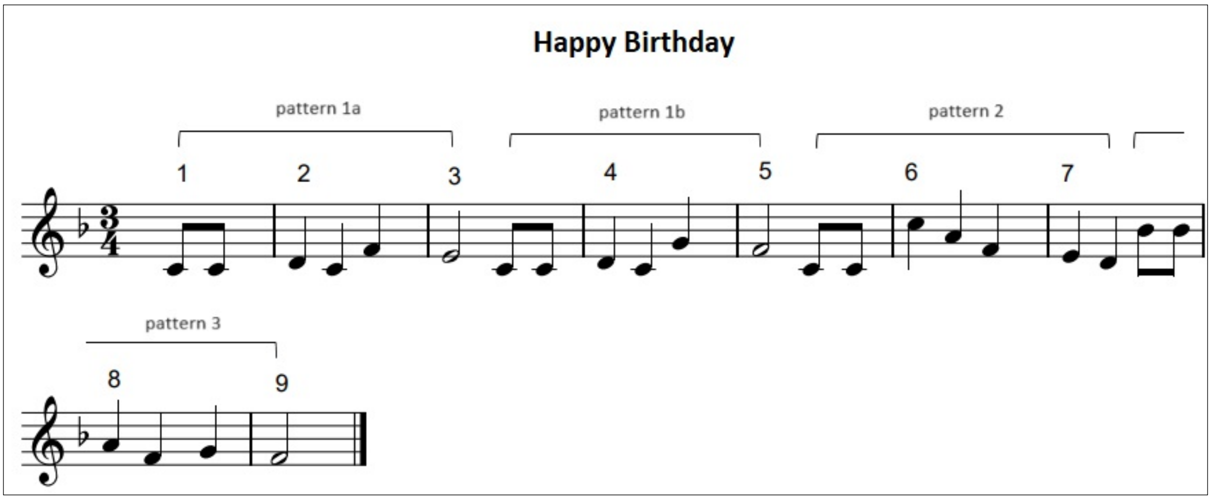


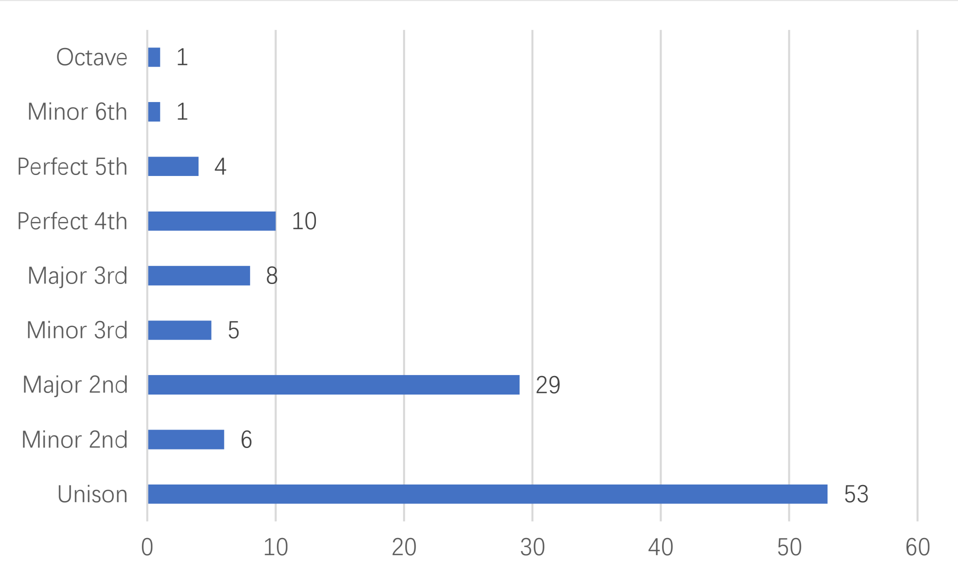
Table 4.8 Characteristics of the Three Target Songs

	Range	No. of patterns	Repeated patterns (P)	No. of notes	No. of different intervals	Types of intervals	No. of each interval	The interval direction		Time signature	Rhythm: upbeat or not	Types of rhythms	Lyrics same or not
								No. of rising interval	No. of falling interval				
Twinkle, Twinkle, Little Star	major sixth	12	P1-P9	42	5	Perfect unison	18	-	-	4/4	No	♩ 𝅗𝅥	Lyrics of patterns 1 to 4 are same with that of patterns 9 to 12. Lyrics of patterns 5 to 8 are different from that of other patterns.
			P2-P10			Minor second	4	0	4				
			P3-P11			Major second	15	2	13				
			P4-P12			Perfect fourth	1	1	0				
			P5-P7			Perfect fifth	3	3	0				
			P6-08										
Little Donkey	octave	4	P1-P9	55	6	Perfect unison	31	-	-	2/4	No	𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯 ♫ ♩. ♪ 𝅗𝅥	Lyric of each pattern is different.
			P2-P10			Major second	5	3	3				
			P3-P11			Minor third	5	5	0				
			P4-P12			Major third	5	4	1				
			P5-P13			Perfect fourth	6	1	5				
			P6-P14(rhythms are different)			Perfect fifth	1	0	1				
			P7-P15 (one note is different)										
Happy Birthday	octave	4	No	25	9	Perfect unison	4	-	-	3/4	Yes	♫ ♩ 𝅗𝅥	Lyric of each pattern is same.
						Minor second	3	0	3				
						Major second	8	3	5				
						Minor third	1	0	1				
						Major third	3	0	3				
						Perfect fourth	2	1	1				
						Perfect fifth	1	1	0				
						Minor sixth	1	1	0				
						Octave	1	1	0				

Note. The number of intervals includes the intervals within two patterns.

In terms of musical intervals among the three target songs, unison and major 2nd are the two most common musical intervals (see Table 4.8). The three target songs also have seven other types of musical intervals, minor 2nd, minor 3rd, major 3rd, perfect 4th, perfect 5th, minor 6th, and octave. The number of each type of musical interval involved in the three target songs is shown in Figure 4.5. Such diversity of musical intervals allowed comparison of children's vocal pitch-matching accuracy when singing a song. Furthermore, the different combinations of musical intervals across the three target songs also provided an opportunity to explore their influence on participants' vocal pitch-matching accuracy when singing the three target songs. Aspects of this are explored in Chapter 5.

Figure 4.5 The Number of Each Type of Musical Interval Including in the Three Target Songs



When combining musical elements for a target song, *Twinkle, Twinkle* was thought to be the easiest song to match, based on its narrower vocal range (a major sixth) and small musical intervals (≤ perfect 5th). The complexity of *Little Donkey* was judged to be at a medium level across the three target songs due to its wider vocal range (an octave), but small musical intervals (≤ perfect 5th). On the other hand, *Happy Birthday* was thought of as the most difficult song to sing because of its wide vocal range (an octave) and generally wider and more numerous musical intervals, although it has the fewest notes. The different complexities of the three target songs helped explore the impact of complexity on participant children's vocal pitch-matching accuracy for songs.

The choice of singing tasks to include both melody and lyrics was to gather data from a large number of participant children which could be seen as culturally conventional in terms of what might count as singing in schools. The possible use of melody without lyrics (syllables, or single vowels) was not a focus for this particular study.

4.2.3 Measurement

All singing performances were assessed against two existing rating scales: (a) Rutkowski's (1990, 1996, 2015) *SVDM* scale measures vocal register use for the three target songs; and (b) Welch's (1998) *VPMD* scale assesses vocal pitch-matching accuracy for songs (see Figure 4.1). In addition, (c) a new measurement system was created, termed the *Melodic Analysis of Pitch-Matching* (*MAPM*), to explore the details of vocal pitch-matching accuracy for each song's constituent pitches within the target melodies themselves. Details and reliability of each of the three measurements are described below.

(a) Rutkowski's *SVDM* Scale Applied to Measure the Application of Sung Vocal Registers in Songs

Participant children's vocal register use for target songs was measured by Rutkowski's (1990, 1996, 2015) Singing Voice Development Measure (*SVDM*) scale. As Figure 4.1 shows, the latest version of the *SVDM* scale (Rutkowski, 1996, 2015, 2018) used specific lift points and pitch ranges to define a particular type of vocal register used among its nine-rating scoring system.

The *SVDM* scale has been applied in many previous studies to test children's application of vocal registers in song singing, mainly in Western cultures, such as America (Guerrini, 2006; Levinowitz et al., 1998; Rutkowski, 1990, 2004; Rutkowski & Chen-Hafteck, 2001; Rutkowski et al., 2002, 2007; Rutkowski & Miller, 2003) and England (Welch et al., 2009a). However, in Mang's (2006) study, half of the Primary school participants were Cantonese monolinguals, suggesting that the *SVDM* scale could be applied to other cultures, including Asia. Using the same measurement to assess participants' vocal register use as in previous studies increased the possibility of comparing results with previous studies.

High consistency and interrater reliability coefficients were stated for the *SVDM* scale (see Rutkowski, 1990, p. 90; 1996, p. 358; 2018; Guerrini, 2006). However, Levinowitz et al. (1998) reported lower reliability for Grade 6 students. The same problem was also found in the current study when using the *SVDM* scale to measure pubertal boys whose voice changed, as their common vocal ranges and lift points did not match those suggested in Rutkowski's (1996) *SVDM* scale.

Consequently, to improve the reliability of results, as noted above, the singing performances of participant pubertal boys whose voice were changing at the time of assessment are not included in the current study.

(b) Welch's *VPMD* Scale Used to Measure Vocal Pitch-Matching Accuracy in Songs

The *VPMD* scale was summarised by Welch (1998) based on children's vocal pitch-matching behaviour reported by previous studies. As reviewed in Chapter 3, Section 3.2.2, the *VPMD* scale involves all elements defining vocal pitch-matching accuracy for songs suggested by previous studies, arranged with a hierarchy order from (i) singing like speaking, (ii) the accuracy of melodic contour, (iii) the ability to maintain a key centre, (iv) the accuracy of pitches, and (v) singing completely in tune (see the content of the *VPMD* scale in Figure 4.1).

The *VPMD* scale has been used by many previous studies to test the vocal pitch-matching accuracy of song singing for children from diverse cultures, such as Asia (e.g., Mang, 2006) and Western culture (Welch et al., 2009a, 2009b). It was also updated to an eight-point rating scale by Wise and Sloboda (2008) to assess adults' vocal pitch-matching ability for songs. Compared with Welch's *VPMD* scale, the eight-rating scale of Wise and Sloboda (2008) used more levels (4/8) to describe the development of consistency of key centres for songs. This updated rating scale has been applied by Demorest and Pfordresher (2015) to assess children's and adults' vocal pitch-matching accuracy for songs. However, it seems that Welch's *VPMD* scale would be sufficient to measure children's vocal pitch-matching accuracy for songs. Furthermore, as the *VPMD* scale has been applied by Welch et al.'s (2009b) large-scale study of children's singing behaviour and development in England, using the same measurement would allow the current author to compare singing data with Welch et al.'s (2009b) study more directly.

However, it must be admitted that using a rating scale (including the *SVDM* and *VPMD* rating scales) to measure singing behaviour is relatively subjective. This is because the boundary between two neighbouring scale ratings can be fuzzy. Consequently, to improve the reliability of raw singing data, the singing performances of 20 participants who were randomly selected were measured by one experienced English judge and the current author together, using the *SVDM* and *VPMD* scales. Kendall's tau statistics, which are designed to test the level of agreement between judges for ranking data (Kinnear & Gray, 2009), were used to analyse the agreement between the two judges. As Table 4.9 shows, the agreements on raw singing scores measured by each of the two rating scales for each target song was relatively high. Overall, the test suggested that the judgment of the current author was reliable.

Table 4.9 Agreement of Raw Singing Scores Measured by Two Existing Rating Scales

Type of rating scale	Song	Kendall's tau
Rutkowski's *SVDM*	*Twinkle, Twinkle*	0.801
	Little Donkey	0.921
	Happy Birthday	0.736
Welch's *VPMD*	*Twinkle, Twinkle*	0.743
	Little Donkey	0.606
	Happy Birthday	0.855

To further improve the reliability of the raw singing data, the author measured all singing performances by both existing scales three times to ensure more consistency of judgment[58]. Scores gained from the third judgment were applied to analyses in the current study as it was seen as more consistent than the other two earlier assessments.

(c) The Melodic Analysis of Pitch-Matching (*MAPM*) System to Measure Vocal Pitch-Matching Accuracy in Songs

In addition, to deepen understanding of the details of vocal pitch-matching accuracy for songs, a new measurement, named the *Melodic Analysis of Pitch-Matching* (*MAPM*) System, was designed by the current author and Graham F. Welch to analyse a small part of singing performances from n = 134

[58]Because of the large numbers of participants, measuring all singing performances once usually took around six weeks when the author measured alone. Her understanding about each of ratings of the scales could change slightly from the beginning to the end of the measurement. However, this uncertainty seemed to be much reduced by the time of the third round of judgements.

participants who were selected on a systematic sampling basis by choosing every 10th child from the dataset of the whole Chinese singing performances, including participants who were tested twice. Overall, the *MAPM* system had two parts. Each sung pitch was assigned a number on the basis of its semitone difference from the target within the melody – the numbering approach, and with degrees of pitch error shown by differences in colour labelling – the colour characteristics.

The current study used these two components to show (i) the degree of vocal pitch-matching accuracy for each sung pitch of a target song, (ii) a similar accuracy rating of each musical interval of a target song, and (iii) the ability to maintain key centre during singing of a target song. Details of the three above points are described separately below.

(i). Showing the Degree of Vocal Pitch-Matching Accuracy for a Sung Pitch of a Target Song by the MAPM System

To measure the accuracy of each of the vocally matched pitches of the three target songs, the following six steps with their sub-steps in Figure 4.6 were designed. An example for each step is provided in Figure 4.7.

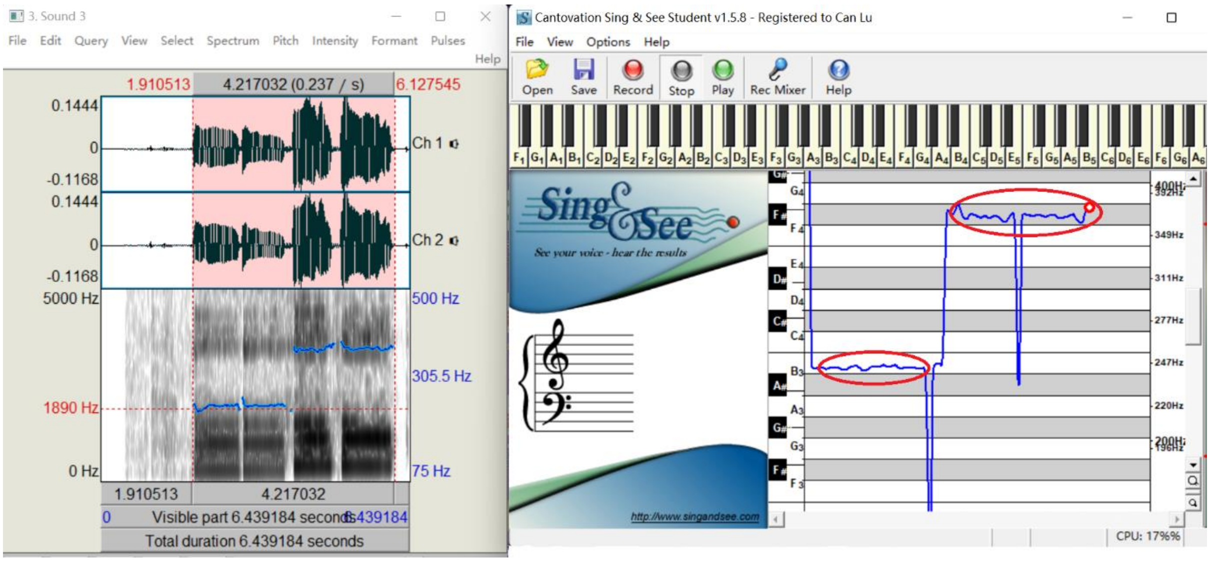
Figure 4.6 Six Steps to Collect the Accuracy of Each of the Vocally Matched Pitches of the Three Target Songs

Step 1. Noting vocally matched pitches
1.1 An audio recording of a song singing was uploaded into *Praat*.
1.2 A short piece of audio recording was selected and then played in Praat, while these audio voices were also inputted into the software program Sing&See, which visually displayed movement of the height of pitches of the short pattern automatically.
1.3 Then, all matched pitches shown by Sing&See were noted in an Excel file.
Step 2. Calculating vocal pitch-matching errors of sung pitches of a target song based on the starting pitch and showing the semitone errors by colours
2.1 Measuring vocal pitch-matching semitone error of each of sung pitches of a target song (except the starting pitch) by calculating the number of semitone gaps between a matched pitch and a related standard pitch of a target song whose key was based on the chosen starting pitch. A temporary key was applied based on the first pitch sung in each song by each participant to provide a foundation for the underlying prevailing key that was illustrated later.
2.2 The semitone errors were shown using the colour characteristic (see Table 4.11) of the *MAPM* system.
Step 3. Recalculating semitone errors of vocally matched pitches of a target song for a participant using an inferred prevailing key and updating the colours given in step 2.
3.1 After colours were inserted for related sung pitches, one colour for most pitches of a song could be seen. A key that was inferred by having a predominant colour was selected as a prevailing key.
3.2 Semitone errors of each vocally matched pitch of a target song for each of participants were then recalculated based on the new, inferred key, noting the semitone errors by the first numbering system (-8 to 8). '0' meant no semitone error, while '-8' or '8' meant matching a pitch with 8 or more than 8 semitones flat or sharp. The recalculation of semitone errors could be based on a gap of semitone difference between an original and a new key.
3.3 Then, the colour characteristics were updated based on the revised semitone errors.

Step 4. Marking semitone errors of vocally matched pitches of a target song for a participant using the second numbering system (0 to 8). '0' meant no semitone errors, while '8' meant matching a pitch with 8 or more than 8 semitone errors.
4.1 Using the replace function of an Excel file transferred all negative numbers (from -8 to -1) into positive (from 8 to 1). This avoided having to calculate with positive and negative numbers to be offset.
Step 5. Using the third numbering approach (1 to 9) to represent semitone errors of vocally matched pitches of a target song. '1' meant with 8 or more than 8 semitone errors, while '9' meant matching a pitch with no semitone errors.
5.1 By using the 1 to 9 numbering system, a greater sum of these numbers meant matching a target song more accurately. This step could be achieved efficiently by using the replace function of Excel, cooperating with the colour characteristic for each of the vocally matched pitches.
Step 6. Sum up all pitches' numbers based on the third numbering system (1 to 9) initially, and using the sum divided to a full score of a target song to show a degree of vocal pitch-matching accuracy of a target song for a participant
6.1 Adding all numbers together using the third numbering system (1 to 9). 6.2 Using the sum of the scores divided by the overall maximum possible score for the target number of pitches within a target song, and using the data out of 100 for a computer analysis.

Figure 4.7 Examples Showing the Six Steps to Calculate the Accuracy of Each of the Vocally Matched Pitches of a Target Song in the MAPM System

Step 1: Praat to Sing&See



Step 2

Sub-step: 2.1 - Finding semitone errors of vocally matched pitches based on a key set by a starting pitch							
Standard pitch based on a temporary key set by a starting pitch	A3	A3	E4	E4	F#4	F#4	E4
Vocally matched pitch	A3	A3	D4	D4	E4	E4	D4
Semitone error(s)	0	0	-2	-2	-2	-2	-2
Sub-step: 2.2 - Marking semitone errors with colours							
Standard pitch based on a temporary key set by a starting pitch	A3	A3	E4	E4	F#4	F#4	E4
Vocally matched pitch	A3	A3	D4	D4	E4	E4	D4
Semitone error(s)	0	0	-2	-2	-2	-2	-2

Note. The melody was *do-do-so-so-la-la-so* from the first two patterns of *Twinkle, Twinkle*.

Step 3

Sub-step: 3.1- Detecting a prevailing key centre

Twinkle, Twinkle								
Patterns	Type	Pitch						
1 - 2	Original standard pitch	A3	A3	E4	E4	F#4	F#4	E4
	Vocally matched pitch	A3	A3	D4	D4	E4	E4	D4
	Original semitone error(s)	0	0	-2	-2	-2	-2	-2
3 - 4	Original standard pitch	D4	D4	C#4	C#4	B3	B3	A3
	Vocally matched pitch	C#4	C#4	B3	B3	A3	A3	G3
	Original semitone error(s)	-1	-1	-2	-2	-2	-2	-2
5 - 6	Original standard pitch	E4	E4	D4	D4	C#4	C#4	B3
	Vocally matched pitch	D4	D4	C4	C4	B3	B3	A3
	Original semitone error(s)	-2	-2	-2	-2	-2	-2	-2
7 - 8	Original standard pitch	E4	E4	D4	D4	C#4	C#4	B3
	Vocally matched pitch	D4	D4	C4	C4	A#3	A#3	G#3
	Original semitone error(s)	-2	-2	-2	-2	-3	-3	-3
9 - 10	Original standard pitch	A3	A3	E4	E4	F#4	F#4	E4
	Vocally matched pitch	G3	G3	D4	D4	E4	E4	D4
	Original semitone error(s)	-2	-2	-2	-2	-2	-2	-2
11 - 12	Original standard pitch	D4	D4	C#4	C#4	B3	B3	A3
	Vocally matched pitch	C4	C4	B3	B3	A3	A3	G3
	Original semitone error(s)	-2	-2	-2	-2	-2	-2	-2

Sub-steps: 3.2 to 3.3 - Recalculating semitone errors based on a new key centre, updating semitone errors and colours

Twinkle, Twinkle								
Patterns	Type	Pitch						
1 - 2	New standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	A3	A3	D4	D4	E4	E4	D4
	New semitone error(s)	-2	-2	0	0	0	0	0
3 - 4	New standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C#4	C#4	B3	B3	A3	A3	G3
	New semitone error(s)	1	1	0	0	0	0	0
5 - 6	New standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	B3	B3	A3
	New semitone error(s)	0	0	0	0	0	0	0
7 - 8	New standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	A#3	A#3	G#3
	New semitone error(s)	0	0	0	0	-1	-1	-1
9 - 10	New standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	G3	G3	D4	D4	E4	E4	D4
	New semitone error(s)	0	0	0	0	0	0	0
11 - 12	New standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C4	C4	B3	B3	A3	A3	G3
	New semitone error(s)	0	0	0	0	0	0	0

Step 4

Sub-step: 4.1 - Transferring the first numbering system (-8 to 8) into the second system (0 to 8)

Twinkle, Twinkle								
Patterns	Types	Pitch						
1 - 2	The updated standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	A3	A3	D4	D4	E4	E4	D4
	The updated semitone error(s)	2	2	0	0	0	0	0
3 - 4	The updated standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C#4	C#4	B3	B3	A3	A3	G3
	The updated semitone error(s)	1	1	0	0	0	0	0
5 - 6	The updated standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	B3	B3	A3
	The updated semitone error(s)	0	0	0	0	0	0	0
7 - 8	The updated standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	A#3	A#3	G#3
	The updated semitone error(s)	0	0	0	0	1	1	1
9 - 10	The updated standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	G3	G3	D4	D4	E4	E4	D4
	The updated semitone error(s)	0	0	0	0	0	0	0
11 - 12	The updated standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C4	C4	B3	B3	A3	A3	G3
	The updated semitone error(s)	0	0	0	0	0	0	0

Step 5

Sub-step: 5.1 - Transferring the second numbering system (0 to 8) into the third system (1 to 9)

Twinkle, Twinkle								
Patterns	Type	Pitch						
1 - 2	The updated standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	A3	A3	D4	D4	E4	E4	D4
	The updated semitone error(s)	7	7	9	9	9	9	9
3 - 4	The updated standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C#4	C#4	B3	B3	A3	A3	G3
	The updated semitone error(s)	8	8	9	9	9	9	9
5 - 6	The updated standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	B3	B3	A3
	The updated semitone error(s)	9	9	9	9	9	9	9
7 - 8	The updated standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	A#3	A#3	G#3
	The updated semitone error(s)	9	9	9	9	8	8	8
9 - 10	The updated standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	G3	G3	D4	D4	E4	E4	D4
	The updated semitone error(s)	9	9	9	9	9	9	9
11 - 12	The updated standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C4	C4	B3	B3	A3	A3	G3
	The updated semitone error(s)	9	9	9	9	9	9	9

Step 6

Sub-step: 6.1 - Adding all numbers together based on the third numbering system (1 to 9)	7+7+9+9+9+9+9+8+8+9+9+9+9+9+9+9+9+9+9+9+9+9+9+9+9+8+8+8+9+9+9+9+9+9+9+9+9+9+9+9+9+9=369
Sub-step: 6.2 - Using the sum divided by the total score	369/378=97.62%, noted as 98 for a computer analysis

Further information for each of the six steps is given below.

For the <u>first</u> step, details of two software programs, Praat and Sing&See, are described below. Sing&See[59] is a programme that is designed to provide real-time visual feedback on the vocal pitch in singing. It does this by using a standard musical correlate of the perceived fundamental frequency with conventional octave pitch labels (such as C4 for Middle C), displayed on a Western classical musical staff system.

Praat[60] is a speech analysis system that displays fine-grained moment-by-moment mean measures of fundamental frequency (F0, such as C4 as 260Hz). It[61] was chosen because it allowed the author to easily select a short piece of audio recording, such as an audio recording of three sung pitches that might take only one second to play. The short pattern of an audio recording of a song allowed every pitch of a short pattern included on one page of a screen of the Sing&See software (see Figure 4.7 for an example[62]), which was a technically convenient technique for the author when inputting pitch heights into an Excel file. Furthermore, selecting a short piece of audio recording accommodated playing them repeatedly by Praat to confirm the pitch height of played pitches on Sing&See, as some pitches' height might not be clear due to background noise or the low volume of a recorded singing voice. This demonstrated how a quiet environment was required to ensure the reliability of vocally matched pitches of the three target songs' singing caught by the Sing&See

[59] http://www.singandsee.com

60 http://www.fon.hum.uva.nl/praat/

[61] www.praat.org

[62] A sung pitch collected by Sing&See was presented by a relatively steadying line on the Sing&See screen (see red circles in Figure 4.7). The pitches collected by Sing&See were more reliable with a slow speed during a song singing.

software, so that sound caught by Sing&See would be only that played by Praat.

However, some singing could not be recognised clearly by Sing&See, which may have been because of extraneous noise or the low volume of a child's voice. The problem could be solved by copying pitches of the child's singing piece-by-piece with a louder volume by the researcher, so this could be recognised by Sing&See.

Some people might argue that the pitches displayed by Sing&See were unreliable as each pitch has a range on the screen of Sing&See (see a screen of Sing&See in Figure 4.7). Indeed, the pitch height of a sung pitch could be at the bottom or top of its range scale, which means that it could not typically represent the height of the pitch. In other words, if a sung pitch's height was in the middle of its range, its value would be more representative. However, it was found that if a musical interval was matched accurately, the location of its pitch heights would be displayed similarly on the screen of Sing&See. For instance, when the fundamental frequency of two neighbouring pitches was at the bottom of a range of fundamental frequencies for both their pitches on the screen of Sing&See, this suggested that they were matched relatively accurately.

In other words, Sing&See could illustrate relative pitch accuracy for a musical interval by displaying the vocal products visually against a musical framework, being the equivalent of an adjustable vertical piano keyboard. If a few continuing musical intervals were matched relatively accurately, as shown by a similar location of a scale of the Sing&See screen, these constituent pitches would create a temporarily consistent key centre. Consequently, it was believed that the pitch heights collected from Sing&See were reliable. By calculating semitone errors between a vocally matched pitch shown by Sing&See and a related standard pitch, the current system could reveal an absolute semitone error of each vocally matched pitch of a target song for a participant as well as an overall sense of tonality across the vocal products.

Furthermore, vocal sound usually vibrates naturally. One target pitch, especially one with a long duration, might be sung using two (or more) neighbouring pitches. This fluctuation would be revealed by the screen of Sing&See. To select a pitch among multiple adjacent pitches to represent its pitch height, a pitch that lasted longer based on evidence on the screen of Sing&See would be selected.

In order to further check the reliability of Praat and Sing&See, a test was undertaken to analyse participants' speaking fundamental frequency measured between Praat and Sing&See. The speaking fundamental frequency was tested by backward counting from 20 to 1. The speaking data of *n* = 102 participants were randomly selected from the whole Chinese dataset[63]. A Pearson correlation coefficient showed that the outcome measures of the two software packages were in general agreement (see Table 4.10). An online table[64] that demonstrated pitch to frequency and vice versa was matched, and agreed, upon with output from an Online Tone Generator[65].

For the second step, details of the colour characteristic of the *MAPM* system are described below. The colour characteristics of the *MAPM* system could be used to illustrate the size and type of semitone error for matching pitches of a target song. Table 4.11 shows 17 colours applied to show 17 types of semitone errors, including white which meant no semitone error. Except for white, eight colours with different degrees of red and yellow were selected to represent eight levels of sharpness, while another eight colours with diverse degrees of blue, green, and purple illustrated eight levels of flatness. The reason that more than one type of colour was applied to show sharpness or flatness was to distinguish different degrees of a type of semitone errors. A darker colour suggested a greater semitone error for a matched pitch compared with a related standard pitch from the inferred key centre. If two or more than two neighbouring pitches were marked with the same colour, this meant that the singer was using the same temporary key centre (see an example in Figure 4.7, step 2). If colours of two neighbouring pitches were different, this indicated that they matched with two different key centres.

The current study used the colours of the *MAPM* system to analyse a range of characteristics of vocal pitch-matching accuracy for a target song, such as the accuracy of musical intervals and the number of tonalities used to sing a target song (see reports in Chapter 5).

---

[63] This was another task that all participants undertook. Its results are not reported in the current PhD study.

[64] https://pages.mtu.edu/~suits/notefreqs.html

[65] http://www.szynalski.com/tone-generator/

Table 4.10 A Pearson Correlation Coefficient within Sing&See and Praat for Speaking Fundamental Frequency

Measurement	Sing&See and Praat
no. of participants	102
Pearson Correlation	.755
p	< .001

Table 4.11 Illustration of Colours for Each Type and Size of Semitone Errors

Flat (b)			Sharp (#)		
colour	no. of semitone error	type and size of semitone error	colour	no. of semitone error	type and size of semitone error
	-	0		-	0
	1b	-1		1#	1
	2b	-2		2#	2
	3b	-3		3#	3
	4b	-4		4#	4
	5b	-5		5#	5
	6b	-6		6#	6
	7b	-7		7#	7
	8b/8b+	-8		8#/8#+	8

The first numbering system, '-8' to '8' in the third step is explained below. Under this system, no semitone error was noted with '0'. A sharply matched pitch was noted with its positive semitone errors. The maximum sharp semitone error was '8', representing matching a pitch ≥ 8 semitones sharp compared with a related standard pitch whose key was based on a starting pitch. Similarly, if a pitch was matched flat based on a key centre set by a starting pitch, it was noted with a negative number. ≤ -8 meant matching a pitch of 8 or more than 8 semitones flat. Consequently, it shows the number and type of the semitone errors of a matched pitch. Using the numbering system allowed for later statistical analysis.

The reason why a prevailing key was finally applied was that it was found that most participants changed their key after singing the starting pitch. For instance, 69.18% (110/159) of participants changed their original key after singing the first two pitches of *Twinkle, Twinkle*.

For the fourth step, it should be noted that the second system conceals details of flatness and sharpness of vocally matched pitches. A reason why the first number system ('-8' to '8') needed to be transferred into the second system ('0' to '8') was to avoid offsetting negative and positive numbers when adding them up, to show a general view of the vocal pitch-matching accuracy of a target song for a participant. For the second numbering system ('0' to '8'), '0' meant no semitone error of a vocally matched pitch of a target song, while 8 meant a vocally matched pitch with ≥ 8 semitones sharp or flat.

In terms of the fifth step, to positively compare data measured by the *VPMD* scale (Welch, 1998), in which a higher rating means vocally matched more accurately, the second numbering system (0 to 8) was transferred to the third numbering system ('1' to '9'). '9' means no semitone errors, while '1' means ≥ 8 semitone errors.

Finally, for the sixth step, data were summed up according to the third numbering system ('1' to '9') initially. A larger sum indicates that the participant vocally matched the three target songs more accurately. The sum was divided into a sum based on standard pitches to show the degree of vocal pitch-matching accuracy for a song for a participant. The data out of 100 also allowed more direct comparisons of vocal pitch-matching scores among all three songs which have different full scores because of the varied numbers of total pitches within each song.

(ii) Using the MAPM System to Calculate Semitone Errors of Vocally Matched Musical Intervals of the Three Target Songs

Semitone errors of vocally matched musical intervals of the three target songs could be calculated by the following four steps (see Figure 4.8). An example was provided for each step and was shown in Figure 4.9. Semitone errors of vocally matched musical intervals could be negative, 0, or positive. A negative number meant that the second pitch of a musical interval was matched flat compared with a target one, while a positive number indicated that it was matched sharp. 0 meant that a musical interval was matched accurately.

Figure 4.8 Four Steps Using the MAPM System to Calculate Semitone Errors of Vocally Matched Musical Intervals of the Three Target Songs

Step 1. Using the colour characteristic of the *MAPM* system to find every musical interval that was accurately matched, marked as '0'
1.1 Using the Excel file created from Step 3 in Figure 4.6. The file showed the updated semitone errors and colour of each of vocally matched musical intervals of a target song for a participant based on a prevailing key. 1.2 If a musical interval was matched relatively accurately, the colour of two pitches of the musical interval would be same. For any of the musical intervals marked with a same colour, the second pitch of a musical interval was marked as '0' to show no semitone error for the musical interval.
Step 2. Calculating semitone errors for other musical intervals that were not vocally matched relatively
2.1 If semitone errors of two neighbouring pitches (a musical interval) were different from related standard pitches based on a prevailing key, these two pitches would be marked with different colours. To show a degree of errors in the musical interval, a gap of semitones from the second colour to the first one was calculated, based on the colour characteristics in Table 4.11.
Step 3. Transferring negative numbers into positive numbers
3.1 All negative numbers were then transferred into positive numbers to avoid being offsetting when adding all numbers up.
Step 4. Creating an SPSS file to analyse the accuracy of vocally matched musical intervals by different variables

Figure 4.9 An Example Showing the Four Steps for Calculating Semitone Errors of Vocally Matched Musical Intervals of the Three Target Songs

Sub-steps: 1.2 - Identifying any accurately matched musical interval and marking as '0'

Twinkle, Twinkle								
Patterns	Type	Pitch						
1 - 2	Original standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	A3	A3	D4	D4	E4	E4	D4
	Semitone error(s)		0		0	0	0	0
3 - 4	Original standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C#4	C#4	B3	B3	A3	A3	G3
	Semitone error(s)		0		0	0	0	0
5 - 6	Original standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	B3	B3	A3
	Semitone error(s)	0	0	0	0	0	0	0
7 - 8	Original standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	A#3	A#3	G#3
	Semitone error(s)	0	0	0	0		0	0
9 - 10	Original standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	G3	G3	D4	D4	E4	E4	D4
	Semitone error(s)		0	0	0	0	0	0
11 - 12	Original standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C4	C4	B3	B3	A3	A3	G3
	Semitone error(s)	0	0	0	0	0	0	0

Sub-steps: 2.1 - Identifying any inaccurately matched musical interval and marking semitone errors of musical intervals

Twinkle, Twinkle								
Patterns	Type	Pitch						
1 - 2	Original standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	A3	A3	D4	D4	E4	E4	D4
	Semitone error(s)		0	2	0	0	0	0
3 - 4	Original standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C#4	C#4	B3	B3	A3	A3	G3
	Semitone error(s)	1	0	-1	0	0	0	0
5 - 6	Original standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	B3	B3	A3
	Semitone error(s)	0	0	0	0	0	0	0
7 - 8	Original standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	A#3	A#3	G#3
	Semitone error(s)	0	0	0	0	-1	0	0
9 - 10	Original standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	G3	G3	D4	D4	E4	E4	D4
	Semitone error(s)	1	0	0	0	0	0	0
11 - 12	Original standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C4	C4	B3	B3	A3	A3	G3
	Semitone error(s)	0	0	0	0	0	0	0

Sub-step:3.1 - Transferring negative semitone errors into positive semitone errors

Twinkle, Twinkle								
Patterns	Type	Pitch						
1 - 2	Original standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	A3	A3	D4	D4	E4	E4	D4
	Semitone error(s)		0	2	0	0	0	0
3 - 4	Original standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C#4	C#4	B3	B3	A3	A3	G3
	Semitone error(s)	1	0	1	0	0	0	0
5 - 6	Original standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	B3	B3	A3
	Semitone error(s)	0	0	0	0	0	0	0
7 - 8	Original standard pitch	D4	D4	C4	C4	B3	B3	A3
	Vocally matched pitch	D4	D4	C4	C4	A#3	A#3	G#3
	Semitone error(s)	0	0	0	0	1	0	0
9 - 10	Original standard pitch	G3	G3	D4	D4	E4	E4	D4
	Vocally matched pitch	G3	G3	D4	D4	E4	E4	D4
	Semitone error(s)	1	0	0	0	0	0	0
11 - 12	Original standard pitch	C4	C4	B3	B3	A3	A3	G3
	Vocally matched pitch	C4	C4	B3	B3	A3	A3	G3
	Semitone error(s)	0	0	0	0	0	0	0

Figure 4.10 Two Steps to Calculate the Number of Tonalities Used to Sing a Target Song

Step 1. Counting the number of colours applied to sing a target song for each of the participants
Step 2. Creating an SPSS file to analyse the consistency of key centres

(iii) Using the MAPM System to Calculate the Number of Keys Used to Sing a Target Song for a Participant

This task could be undertaken using the two steps shown in Figure 4.8. The first step is illustrated by an example shown in Figure 4.9. As the figure shows, four colours, including white, pink, light blue, and darker blue, were used to show vocally matched accuracy for each of the pitches of a target song for a participant. The four colours indicate that the participant used four keys to sing the song, as each of the colours represents a temporary key for a target song of a participant.

Overall, participants' vocal pitch-matching accuracy of a song was measured generally (macro) by Welch's (1998) Vocal Pitch-Matching Development (*VPMD)* scale and in a more detailed (micro) approach by the Melodic Analysis of Pitch-Matching (*MAPM)* system. When these two measurements are applied together (macro + micro), it allows the creation of a more complete picture of participants' vocal pitch-matching of a song.

4.2.4 Process

The process of data collection in the current thesis can be divided into three periods of time, (a) before the first visit to a school, (b) the initial visit to a school, and (c) the second visit for collecting singing data.

(a) Before the First Visit to a School (Before February 2017)

Before visiting a participating school, four preparations were undertaken, (i) recording models of the three target songs; (ii) preparing an information letter and an ethics form for music teachers and participant children's parents (see an English Version in Appendix A); (iii) printing materials for participants, and (iv) contacting music teachers.

Firstly, audio recording models for the three target songs were prepared to ensure that each of the six participating schools had the same basic learning materials. Three target songs were sung by a Chinese adult female with a normal speed and limited vibrato. The pitch range of the model for the three target songs was from A3 to C5. No accompaniment was provided for the audio recording models (see the influence of comfortable singing range, and accompaniment on vocal pitch-matching accuracy in Chapter 3, Section 3.2.3).

Then, an information letter and ethics form were prepared under the policy guidance of University College London (UCL). They were written initially in English and then approved by a committee from a music education team in the Culture, Communication, and Media department at UCL, Institute of Education. After approval, they were translated into Chinese for Chinese music teachers, parents, and participant children to read. These documents were printed before visiting a participating school.

A form designed to collect participant children's information was also printed. It contained sections for a participant's name, school, grade, date of birth, gender, weight, height[66], whether or not they had taken private music training (if yes, whether it was for singing)[67], and their father's job (see the Chinese model of the form in Appendix B). Most of the information helped the researcher analyse participant children's singing behaviour afterward. For studies undertaken in China (e.g., Li & Qiu, 2016; Wu, 2013), the father's occupation, instead of that of the mother, was usually used as an independent variable to test a family's socioeconomic status. This was guided by the different roles in a family in traditional Chinese culture, in which the man goes out to work, while the women look after the family (Liu, 2008).

Furthermore, two new audio recording machines, an iPhone 6s (model: iOS 14.8) and a Philips voice tracer (model: VTR5000), were prepared. As revealed by the pilot study, participants seemed to be more relaxed when an audio recording was taken compared with a video recording.

Finally, one music teacher from each of the six participating schools was contacted. They were found through various sources. Four of them, one each from the rural schools (A to D) were introduced by friends and relatives, while two of them, from schools E and F, were suggested by a senior researcher in an education council in Changsha, Hunan Province. The six schools were

---

[66] *Note*. The reason for including weight and height in the personal information sheet was to explore the influence of physical development on participants' singing performances. However, many participants did not know their weight and height, so they could ignore these questions.

[67] This was designed to collect more information about music training for participant children. A review of the demographic data suggested that, where the information was provided, only a small number of participant children had private music lessons outside their schools, i.e., less than n = 50 overall out of N = 1193 (<5%). The possible impact of this variable on children's singing needs to be explored systematically in future research.

characterised by diverse geographic locations (rural/urban) and family income contexts (lower/higher parents’ income based on occupation), and they had students from Grade 1 to Grade 6, with a mixed-sex intake.

Information about the research aim and singing tasks was then sent to each of the contacted music teachers by email after the initial introduction. Audio recording models were sent to them around two weeks before an initial visit to each school to ensure the same basic teaching materials were provided. Each teacher was asked to rehearse the three target songs with the participant students a few times during the coming two weeks to help the children become more familiar with the target songs.

(b) The Initial Visit to a School (February to April 2017 & February to June 2018)

During the initial visit to a participating school, three tasks were undertaken, (i) distributing information letters and ethics forms to the music teachers; (ii) discussing the arrangements with the music teachers; and (iii) having initial contact with the participant children.

Each music teacher was given a letter and two identical teachers' ethics forms. The letter was used to introduce the aim and singing tasks of the author's study, while the ethics forms allowed the relevant teachers to express their agreement with each detail of data collection. After a music teacher agreed on each part of the ethics form, she/he signed both forms, then kept one and returned the other to the author. They could ask any questions during the process. Furthermore, time and space for testing were discussed with music teachers. The author suggested that a quiet but familiar room within a school would be an ideal space to ensure a good quality audio recording and to reduce the stress on participants when giving a singing performance in front of a group of classmates.

To relax the participant children, the author met them in their classroom during recess and said ‘hello’ to them. For the main study, all children in the focus classes were invited 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). Participants were also told that they could withdraw from the test for any reason or no reason. Participants were told that they would be asked to sing three familiar children’s songs

(*Twinkle, Twinkle, Little Donkey,* and *Happy Birthday*). The author checked with them whether these songs had been rehearsed by their teacher. If the target songs had not been rehearsed, the author would use the recording models to lead them in singing the three target songs a few times.

Although ethics forms for parents were printed, they were only distributed to participants of one class in one participating school. The author was told that this was because the process of collecting ethics forms from guardians was too time-consuming, and that teachers were allowed to make such a decision on behalf of guardians for this type of school task.

(c) The Second Visit to a Participating School (February to April 2017 & February to June 2018)

The second visit to a participating school can be divided into three time periods, (i) before, (ii) during, and (iii) after a singing assessment.

Before a singing assessment, a room arranged to test singing performances was checked by the author. At least two chairs were allocated for the researcher (the author) and a participant to sit at a similar horizontal level, which was good for eye contact between the researcher and the participant. Eye contact could comfort and encourage participants during the singing test. On a desk arranged between the two chairs, was a piece of A4 paper with the lyrics of the three target songs written on it, two items of recording equipment, and a few pens.

Before a formal singing test, participant children's information sheets were distributed to save time during the test itself[68]. Then, around five to seven participants in a group went to a testing room. The group size was designed to relax participants and prevent the testing room from becoming overcrowded. Participant children were told that they could see the lyrics of the songs written on a sheet of paper if they needed. This helped avoid the negative influence of forgotten lyrics on the scores for song singing. Furthermore, singing melodies with lyrics seems to be a more natural way for people when they vocally matched a familiar song. They were encouraged to sing the three target songs at a normal speed with no starting pitch to provide a better opportunity for them to sing

[68] Most participants only had two music classes of 45 minutes per class per week, so the author tried to minimise disruption to participants' music classes by finishing some preparation before the singing tests, such as filling in a personal information form.

within a comfortable singing range. They were also encouraged to sing individually, but were allowed to sing with one of their friends if they felt more comfortable that way. When a participant was singing, 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 individual child was asked to sing for approximately five minutes, this being the overall time needed for the three target songs. There was no evidence of task fatigue. Children attended in small groups because this had been reported as a useful protocol to allow each child to understand the nature of the task and to share the experience with friends. There was no evidence of competition between children because of the time involved in which each child was given a preliminary task of counting backwards from twenty – in part because this was a more cognitively focused vocal task and different from singing and also because it allowed the researcher to get some sense of the child's natural speaking pitch, although this data was not needed for the thesis.

If they asked to sing privately, other participants were asked to leave the room. If a participant child did not know how to sing a particular phrase, the author helped them by teaching the phrase, which helped to create a relaxing and supportive environment. After the singing tests for a group were completed, the children were thanked and asked to go back to their regular classroom and call the next group to come.

The participant teachers were provided with a recording of the three songs prior to the data collection visits. They were asked to familiarise the children with the songs, although they had been selected because of the likely widespread knowledge of the songs within the general population. The children's singing was assessed in a quiet and familiar place within the school. Children attended in small groups so that they could share the experience – in line with the research protocol adopted in England for the *Sing Up* programme and as suggested by Plumridge (1972). They were not provided with a starting pitch, but were asked to sing from memory without accompaniment. All vocal products were recorded digitally.

All participants were expected to sing from memory. Only a few participants, approximately five out of N = 1,193 were unsure about a particular part of a target song, and in such cases the children were reminded of this part a few times before the formal singing performance. It was assumed that their singing ability would not be altered because the short term of assistance was brief (Welch et al., 1998) and also a judgement could be made of their singing of the remainder of the song which they had remembered. In addition, a copy of the lyrics was provided to prompt participants' memory. This protocol was adopted in order to reduce the threat of any researcher and/or practice effect in the data. 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*. The researcher checked with the children whether these songs had been formally rehearsed by their teacher. If the target songs had not been rehearsed, the researcher would use the recorded models to lead them in singing the three target songs as a group a few times to remind them. This only needed to be applied to one class, as all the other classes had sung the songs with their teacher. Looking at the resultant data, there was no evidence that this class singing profile was any different to others in the same school.

4.2.5 Data Analyses

In the current section, (a) the coding system for data, (b) types of data, (c) the structure of data analyses and related statistical tests, and (d) calculation for data measured by the *MAPM* system are described successively.

(a) The Coding System

For coding used in the current study, boys were represented by '0', while girls were shown by '1'. In the geographic location column, rural participants were identified by '1', while urban participants were identified by '2'. In the income column, participants from lower-income families were illustrated by '1', while their peers from higher-income families were shown by '2'. The name of a participating school was replaced by a capital letter from A to F to protect the privacy of the participating schools. For the coding for songs, *Twinkle, Twinkle* was noted as '1', *Little Donkey* and *Happy Birthday* were marked as '2' and '3', respectively.

(b) Types of Data

In the SPSS file, age was set as a scale value. Data on sex, geographic location, income, school, and song were set as a nominal value. Raw scores measured by the *SVDM* and *VPMD* scales were numerical ratings. Analysing ratings has been 'a grey area' (Kinnear & Gray, 2009, p. 2). Previous studies do not agree on whether to treat ratings as scale data or ordinal data (Kinnear & Gray, 2009). Kinnear and Gray (2009) suggest that the decision about which statistics to use should be based on 'the distribution of data and the number of points on the rating scale' (p. 2). In the current study, all raw data measured by both rating scales was normally distributed (see Appendix C). Raw scores measured by the *SVDM* scale had nine points, while those measured by the *VPMD* scale had four points. Consequently, scores measured by the two ratings were treated as scale data (*cf*. Mang, 2006; Welch et al., 2009b). Also, as data measured by the *MAPM* system were continuous, they are treated as scale data.

(c) Calculation for Data based on the *MAPM* System

The number of participants whose singing performances were measured by the *MAPM* system is shown in Table 4.12. The related analyses are shown in Table 4.13. Calculation of overall vocal pitch-matching accuracy for a target song measured by the *MAPM* system, semitone errors of musical intervals, and the number of key centres are described below.

Firstly, in terms of an overall score of vocal pitch-matching accuracy for a target song measured by the *MAPM* system, this was an accuracy rate, calculated out of 100 (see an example in Figure 4.7, the 6^{th} step). It used a sum of the vocal pitch-matching score of all vocally matched pitches in a target song divided by a full score based on a standard target song. The full score of a vocally matched pitch was 9, while the smallest score of a vocally matched pitch was 1. For instance, if a participant matched all (n = 42) pitches of *Twinkle, Twinkle* accurately, his overall vocal pitch-matching score for *Twinkle, Twinkle* would be (9x42)/(9x42) = 1, analysed using 100. If he matched the song using a speaking manner, his overall score would be (1x42)/(9x42) = 0.11, analysed using 11 for statistical analyses.

Table 4.12 The Number of Participants Whose Singing Performances were Tested by the MAPM System

Variable	Sub-group	Age in Whole Years						Total
		6+	7+	8+	9+	10+	11+	
Sex	Boys	3	13	18	14	13	7	68
	Girls	6	12	11	12	13	12	66
Geographic Location	Rural	9	19	15	17	13	13	86
	Urban	-	6	14	9	13	6	48
Inferred Parental Income	Lower	4	10	13	14	15	11	67
	Higher	5	15	16	12	11	8	67
Total		9	25	29	26	26	19	134

Secondly, in the application of the new system, semitone errors are calculated in terms of the local tonality as set by the individual singer. An ‘error’ in this instance is in comparison with the expected pitch from the previous sung pitches within the chosen tonality. Deviations were noted in semitones rather than anything smaller. All judgements were in musical terms, rather than in terms of acoustic frequency analyses. In addition to the researcher, two expert academic colleagues undertook their own analyses of 10% of the vocal products to validate the original judgements.

For the accuracy of musical intervals, the absolute value of semitone errors of vocally matched musical intervals of the three target songs was calculated based on the colour characteristics of the *MAPM* system. For instance, if two neighbouring pitches were vocally matched relatively accurately, their colour was the same, suggesting the absolute value of semitone error of the musical interval was ‘0’. If the first pitch of a musical interval was in line with a prevailing key (noted with the colour white), and the second pitch was vocally matched one semitone sharper compared to a target pitch, it would be coloured pink (see the colour characteristics of the *MAPM* system in Table 4.11). This suggests that the absolute value of the semitone error of the musical interval was ‘1’.

Similarly, if the first pitch of a musical interval was in line with a prevailing key (noted with white), and the second pitch was matched one semitone flatter compared with a target pitch, it would be coloured light blue, suggesting the semitone error of the musical interval was '-1', and the absolute value of the semitone error would be '1'. An absolute value was used as raw semitone errors of matched musical intervals would be offset when summed up. Overall, an absolute value of semitone errors for a vocally matched musical interval of each target songs could be achieved by calculating the absolute value of semitone difference between two colours for two pitches of a musical interval.

In the current analyses, the possibility of categorical perception has not been formally considered as this is outside the scope of the thesis. However, this is an area for further exploration. The challenge is that categorical perception may not apply in real-time during the process of singing in which the behaviour may not be open to real-time manipulation.

As n = 2 of n = 134 participants sang some phrases of the three target songs using a speaking manner which made it difficult to define semitone errors for related musical intervals, all singing data of these two participants were not included when analysing the semitone errors of musical intervals of the three target songs. Consequently, the absolute value of semitone errors of vocally matched musical intervals of the three target songs was analysed based on n = 132 participants' singing performances.

The three standard target songs in total included n = 119 musical intervals. The number for each type of musical interval across the three target songs is shown in Figure 4.5. However, one unison, one major 2nd, and one minor 3rd were not involved in the current analyses for vocal pitch-matching accuracy of musical intervals due to a series of missing data. In more detail, n = 26 of n = 132 (19.70%) participants did not sing a unison (the fifth pitch of Pattern 14 of *Little Donkey*, *mi*) (See Figure 4.4). n = 20 of participants (15.15%) did not sing a minor 3rd (the sixth pitch of Pattern 14 of *Little Donkey, so*). n = 29 of participants (21.97%) missed a major 2nd (the second pitch of Pattern 7 of *Happy Birthday*, *la*). These common missing data seem to be because of either learning or remembering mistakes. Consequently, semitone errors of the other n = 115 musical intervals were involved in the analyses of vocal pitch-matching accuracy for musical intervals of the three target songs. All the absolute values of semitone errors of each matched musical interval of the three target songs were analysed.

Thirdly, the number of key centres used to sing a target song was analysed by calculating the number of colours shown for a vocally matched song. This is because each colour meant a type of semitone error of a vocally matched pitch when singing a target song (see the colour characteristic in Table 4.11). If a series of neighbouring pitches were vocally matched relatively accurately, they were noted with the same type of colour. If a series of neighbouring pitches were not vocally matched accurately, they were given different colours based on real semitone distance from vocally matched pitches to a standard pitch of a target song.

Lastly, common characteristics of vocal pitch-matching accuracy for the three target songs measured by the colour characteristic of the *MAPM* system were also analysed. These perspectives included (i) a percentage of participants selecting an appropriate starting pitch for a target song; (ii) common characteristics when matching two similar melodies; (iii) showing common types of pitch errors before vocally matching an upward wide musical interval; (iv) summarising common vocal pitch-matching errors when singing a series of upward or downward pitches toward a lift point; (v) showing the influence of the size of musical intervals on vocal pitch-matching accuracy of a high pitch; (vi) illustrating common errors when vocally matching a series of different musical intervals during singing a song. These were based on observing n = 134 participants' vocal pitch-matching accuracy for each of the three target songs measured by the colour characteristics of the *MAPM* system. Each perspective was illustrated with an example in Chapter 5, Section 5.4.

Table 4.13 Research Questions and Related Statistical Analyses for Three Datasets Used in Chapter 5

Dataset	Question	Test
Vocal Pitch-matching accuracy for each pitch	The relationship between scores measured by the *MAPM* system and those measured by the *VPMD* scale	A Pearson correlation coefficient
Semitone errors of vocally matched musical intervals	Comparing different types of musical intervals' semitone errors	One-way ANOVA
	Comparing semitone errors between minor 2^{nd} and major 2^{nd}	Independent-samples t-test
	Comparing semitone errors between minor 2^{nd} and minor 3^{rd}	Independent-samples t-test
	Semitone errors of vocally matched musical intervals by age * type of musical intervals	Two-way ANOVA
	Semitone errors of vocally matched musical intervals by sex * type of musical intervals	Two-way ANOVA
	Semitone errors of vocally matched musical intervals by geographic location * type of musical intervals	Two-way ANOVA
	Semitone errors of vocally matched musical intervals by income * type of musical intervals	Two-way ANOVA
	Semitone errors of vocally matched musical intervals by song * type of musical intervals	Two-way ANOVA
	Comparing vocal pitch-matching accuracy of one type of musical interval between two target songs	Independent-samples t-test
Consistency of key centres	The relationship between the number of key centres applied between two target songs	A Pearson correlation coefficient
	The relationship between the number of key centres used to	A Pearson

sing a target song and age	correlation coefficient
The number of key centres used to sing a target song by age	One-way ANOVA
The number of key centres used to sing a target song by age * sex	Two-way ANOVA
The number of key centres used to sing a target song by age * geographic location	Two-way ANOVA
The number of key centres used to sing a target song by age * income	Two-way ANOVA
The number of key centres used to sing a target song by age * song	Two-way ANOVA
The relationship between the number of key centres used to sing a target song and scores measured by the *SVDM* scale	A Pearson correlation coefficient

(d) The Structure of Data Analyses with Statistical Tests Applied

Statistical tests used to analyse diverse research questions for results in Chapter 6 are summarised independently in Table 4.14.

As the table shows, the relationship between age and raw scores for vocal register use and age and those of vocal pitch-matching accuracy for songs were explored initially in the first three Chinese datasets (see the left-hand column of Table 4.14). Then, raw scores measured by the two existing scales (*SVDM* and *VPMD*) were combined and normalised out of 100. The normalised singing scores were then analysed by age and age-related interaction effects (age * sex, age * geographic location, age * income, and age * song). The data analyses for semitone errors of musical intervals and the number of key centres applied to sing a target song measured by the *MAPM* system also shared a similar strategy (see Table 4.13). The analyses for the longitudinal dataset were based on changes in singing behaviour from the first to the second visiting years. For all statistical tests, $p < .05$ was set as a significant value.

The calculation of a normalised singing score (NSS) shown in Table 4.14 took a mean score of vocal register use and a mean score of vocal pitch-matching for the three target songs initially, calculated out of 100. 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 0.8(4/5), 0.9(4.5/5) and 0.8(4/5). In this case, her normalised score of vocal registers' application of the three criterion songs was around 0.83 [(0.8+0.9+0.8)/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(4/4), 1(4/4), and 0.75(3/4), and her normalised score of vocal pitch-matching accuracy of song singing was around 0.92 [(1+1+0.75)/3]. Her final normalised score was 0.88 [(0.83+0.92)/2], noted with 88 for computer calculation.

Table 4.14 Research Questions and Related Statistical Analyses for Four Datasets Used in Chapter 6

Dataset	Question	Test
Chinese children tested once, either in 2017 or 2018	The relationship between age in years and vocal register use	A Pearson correlation coefficient
	The relationship between age and the vocal pitch-matching ability	A Pearson correlation coefficient
	Vocal register use by age	One-way ANOVA
	Vocal pitch-matching accuracy by age	One-way ANOVA
	Vocal register use by age * song	Two-way ANOVA
	Vocal pitch-matching accuracy by age * song	Two-way ANOVA
	The relationship between scores of vocal register use and scores of vocal pitch-matching accuracies	A Pearson correlation coefficient
	Normalised singing scores (NSS), combined scores of vocal register use and vocal pitch-matching accuracy, analysed by age	One-way ANOVA
	NSS by age * sex	Two-way ANOVA
	NSS by age * geographic location	Two-way ANOVA
	NSS by age * income	Two-way ANOVA
	NSS by age * school	Two-way ANOVA
All Chinese children's singing performances	Ditto	Ditto
Chinese longitudinal dataset	Compare the mean improvement in vocal register use and that of vocal pitch-matching accuracy	An independent samples t-test
	Change of vocal register use by age	One-way ANOVA
	Change of vocal pitch-matching by age	One-way ANOVA

	Change of vocal register use by age * song	Two-way ANOVA
	Change of vocal pitch-matching by age * song	Two-way ANOVA
	Change of NSS by age	One-way ANOVA
	Change of NSS by age * sex	Two-way ANOVA
	Change of NSS by age * geographic location	Two-way ANOVA
	Change of NSS by age * income	Two-way ANOVA
	Change of NSS by age * school	Two-way ANOVA
Chinese children and English children without *Sing Up* experience	The relationship between Chinese NSS and age	A Pearson correlation coefficient
	The relationship between English NSS and age	A Pearson correlation coefficient
	NSS analysed by age * country	Two-way ANOVA
	NSS analysed by age * sex * country	Three-way ANOVA
	The relationship between the overall NSS and age	A Pearson correlation coefficient
	The overall NSS by age	One-way ANOVA
	The overall NSS by age * sex	Two-way ANOVA

Table 4.15 The Number of Singing Performances by Age and Sex Interaction Based on All Chinese Singing Performances and English Students Without Sing Up Experience

Group	Sex	Age in Whole Years						Total
		6+	7+	8+	9+	10+	11+	
Chinese	Boys	29	134	174	185	167	91	780
	Girls	47	123	182	177	150	80	759
English	Boys	112	616	156	137	720	106	1847
	Girls	96	569	140	113	649	84	1651
Total		284	1442	652	612	1686	361	5037

When analysing the relationship between the Chinese and English datasets, only age and the age * sex interaction effect were tested, as these two independent variables were only two shared variables between the current and Welch et al.'s (2009b) studies. There were *n* = 3,498 singing performances collected from English students without *Sing Up* training in the national Singing Program *Sing Up*[69] from 2007 to 2011 (Welch et al., 2009a; 2009b). Together with all singing performances collected in the current study (*n* = 1,539), this created a dataset of *N* = 5,037 singing performances across the two countries. The number of singing performances for Chinese and English participants by age and sex interaction is shown in Table 4.15.

The reason that singing performances across the two countries could be compared was that they used the same singing materials and the same measurements (the *SVDM* and *VPMD* rating scales). Both studies used *Twinkle, Twinkle,* and *Happy Birthday*, although the current Chinese study

[69] English participants could be divided into two groups based on whether they had *Sing Up* experience or not. Most of the English participants (*n* = 7,873, 69.24%) had joined the *Sing Up* programme which provided extra professional support for their singing skills, while another group of participants who were not involved in the Sing Up programme (*n* = 3,498, 30.76%) were also tested to evaluate the effectiveness of the *Sing Up* programme. As none of the Chinese participants involved in the current study had any extra singing support from the researcher, their singing behaviour was best compared with that of the English participants without *Sing Up* experience as neither group had extra singing support.

also used another Chinese nursery song, *Little Donkey*, to explore any cultural differences. As previous studies (e.g., Svec, 2018) suggested that singing tasks had a great impact on children's song accuracy, only the data of the same two target songs (*Twinkle, Twinkle* and *Happy Birthday*) applied in Chinese and English *Sing Up* studies were used for analysis in the current comparative study.

4.2.6 Data Storage and Sharing

All data were inputted into Excel files initially and uploaded into SPSS files later. Audio data was saved with permission from the music teachers. All this information is protected in an online drive by password requirement, and the data will be used only for academic purposes. Only the two supervisors were invited to access *n* = 20 audio recordings, to check the reliability of the researcher's judgments.

4.2.7 Summary

Overall, the study involved a large number of participants in order to explore the singing behaviour of participant children in Hunan Province, China by singing three target songs. When analysing the details of singing behaviour, the vocal pitch-matching accuracy for a target song was analysed by a new measurement protocol, called the *MAPM* system. When analysing the general singing behaviour, two established singing perspectives were applied related to vocal register use and vocal pitch-matching accuracy for the three target songs. Using these three measurements provides a general view of participants' singing ability and also explores more individual details of singing behaviour. The process of data collection was undertaken using the suggestions of previous studies and guidance on ethical issues from UCL.

Chapter Five

Results based on the Melodic Analysis of Pitch Matching *(MAPM)* System

The Melodic Analysis of Pitch Matching *(MAPM)* system (see Section 4.2.3) is a new method proposed in the current thesis for the analysis of an individual's vocal pitch matching when singing an existing target melody. In the current doctoral study, the *MAPM* system was applied to analyse the correlation for scores measured between the *VPMD* scale and the *MAPM* system (see Section 5.1), ability to maintain a key centre in a target song (Section 5.2), semitone errors related to nine types of musical intervals of the three criterion songs (Section 5.3) and summarising common characteristics during the three target songs' singing (Section 5.4).

Data in the current chapter are based on a sample of n = 134 participants who were quasi-randomly selected by choosing every tenth child from the whole Chinese dataset.

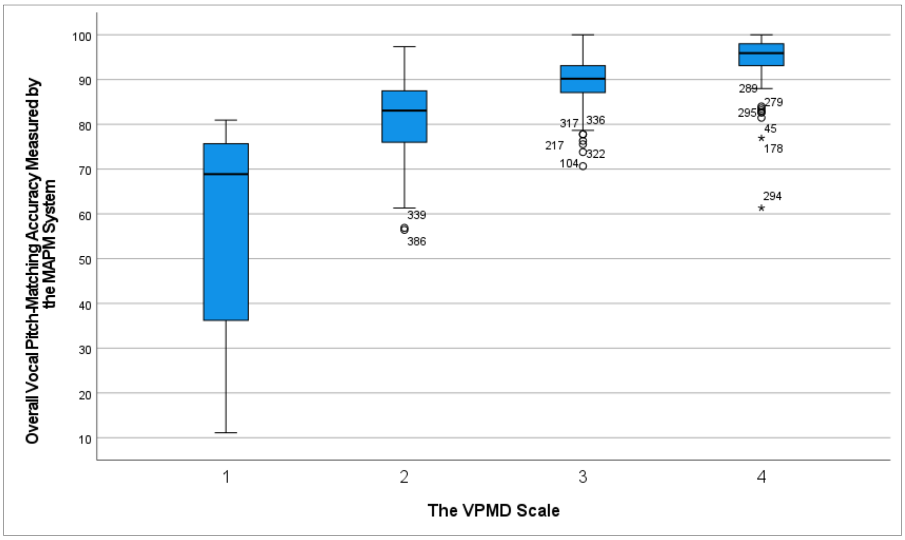
5.1 Correlation for Scores Measured between the *MAPM* System and the *VPMD* Scale

In the current section, scores for vocal pitch-matching accuracy for each pitch of a criterion target song were assigned based on the accuracy of matching each individual pitch within the target melody. A score of 9 meant no pitch error compared to the original target melody pitch, while 1 meant ≥ 8 semitone errors from the original target pitch – irrespective of whether the child is singing sharp or flat compared to the original. When all the individual pitch ratings are added together, the sum provides a measure of how accurate the child was overall in their pitch matching. A higher score implies a better vocal pitch-matching ability for a target melody. Then, each sum for a criterion melody for each participant was divided by a total number of pitches being within the focus song. For example, if a song had 10 constituent pitches and each was sung accurately by the child, the vocal pitch accuracy rating scores would be 9,9,9,9,9,9,9,9,9,9. A child who was less accurate in their vocal pitch matching for the same song might have achieved scores of 7,5,6,3,4,5,4,3,5,4. Consequently, dividing the sum of the individual scores by the ideal matched total score (90 in this example) would generate an overall rating of 1 for the first child (actual 90/90 ideal = 1) or 100%, computed as 100 in the subsequent application of inferential statistics. The second example child would be assigned an overall rating of 0.51 (46/90 = 0.51) or 51%, computed as 51 in the inferential statistics. Finally, an overall mean of the three vocal pitch accuracy ratings for the three criterion songs was computed for each child.

A Pearson correlation coefficient was computed to assess the relationship for scores of vocal pitch-matching accuracies for the three target songs measured between the numbering approach of the *MAPM* system and the *VPMD* scale (*cf.* Welch, 1998, as reported in the findings in Chapter 5). There was a statistically significant positive correlation between the two measures, $r = .642$, $n = 402$[70], $p < .001$ (see Figure 5.1).

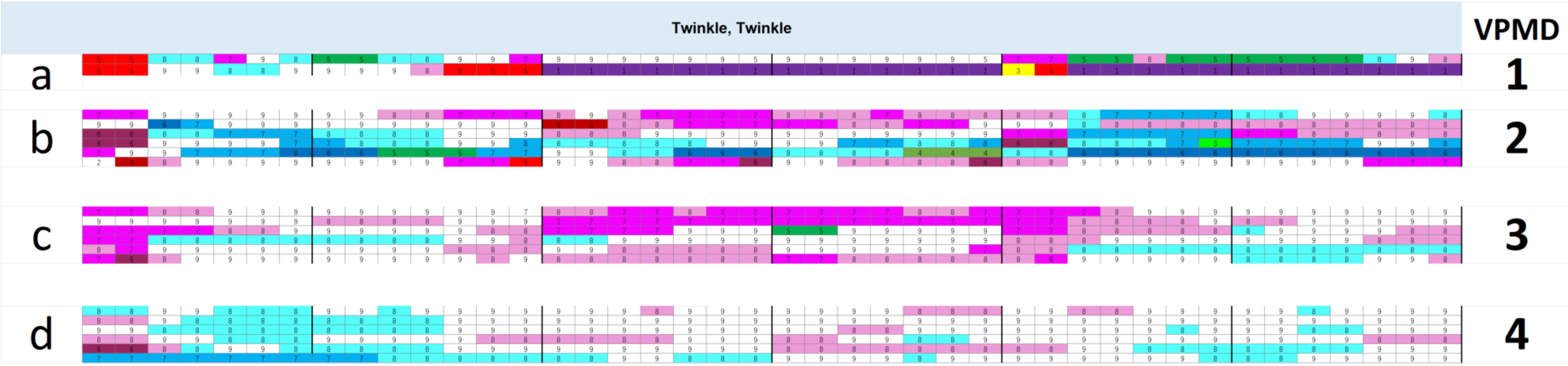
[70] $n = 402$ (134 * 3) indicates that the data analyses were based on each of the three target songs' vocal pitch-matching for the $n = 134$ participants.

Figure 5.1 Correlation for Participants Vocal Pitch Matching Accuracy for the Three Target Songs as Measured by Comparing the Ratings of the New MAPM System and the VPMD Scale (Welch, 1998)



In order to visualise the relative levels of vocal pitch matching accuracy for each individual on each song, a colour coding system was used (see Table 4.11). Lighter colour shading implies greater vocal pitch accuracy, with the colour white indicating an accuracy rating of 9 out of 9. Each of the sung pitches of the three songs for each participant were rated and assigned a colour code. These *MAPM* visual ratings were then clustered according to their approximate equivalence to the *VPMD* rating scale. An example is provided in Figure 5.2 and the equivalence in terms of rating criteria are summarised in Table 5.1.

Figure 5.2 An Example of Vocal Pitch Matching Accuracy for the Target song 'Twinkle, Twinkle' as Measured by the MAPM System in Colour and the VPMD Scale (Welch, 1998, numbered)



Note. Sharp semitone errors were noted with varied degrees of red and yellow, while flat semitone errors were marked with diverse degrees of blue, green and purple.

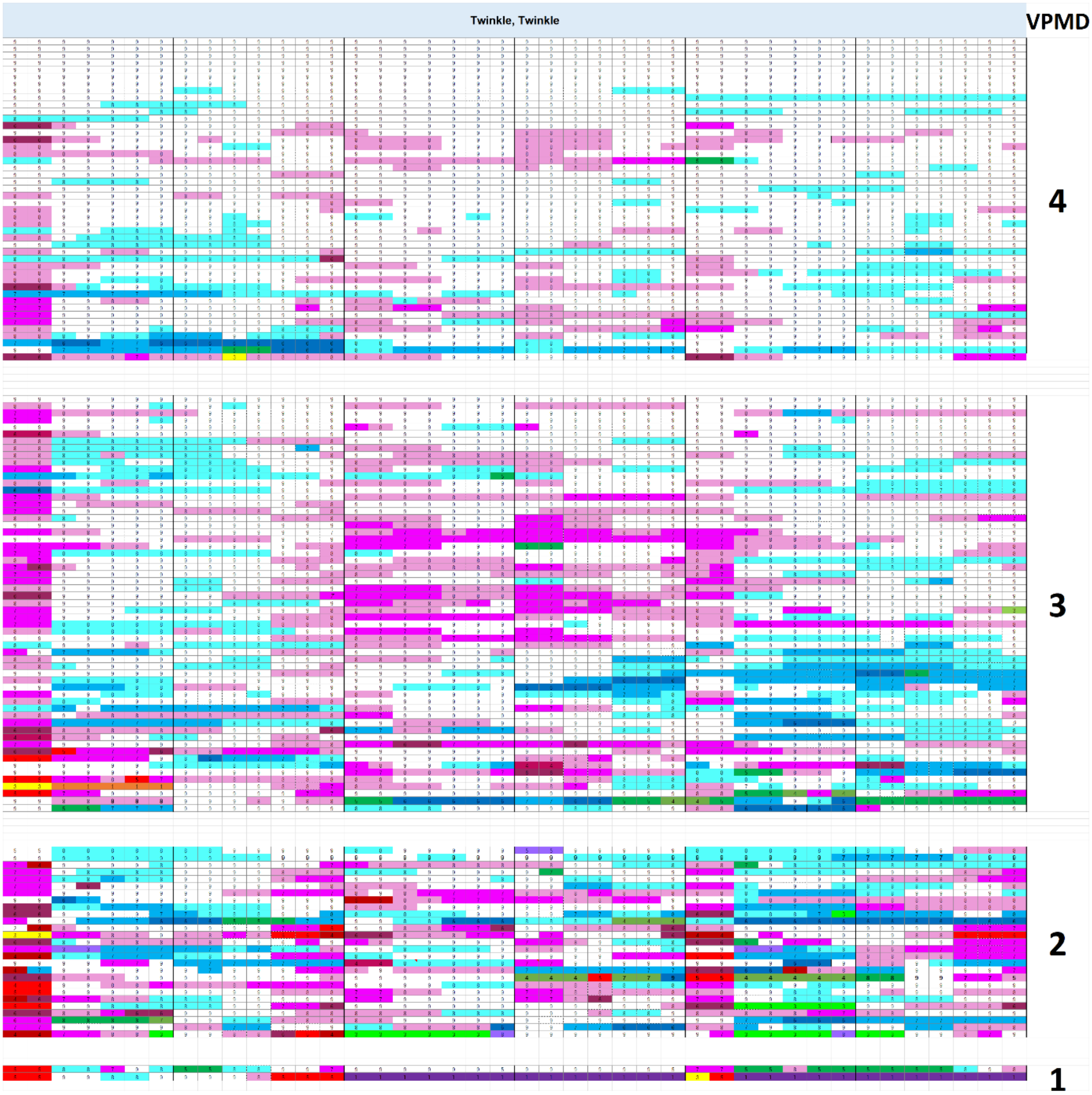
Table 5.1 Matching Each Rating of the VPMD Scale (Welch, 1998) to Descriptions of Song Singing Measured by the MAPM System

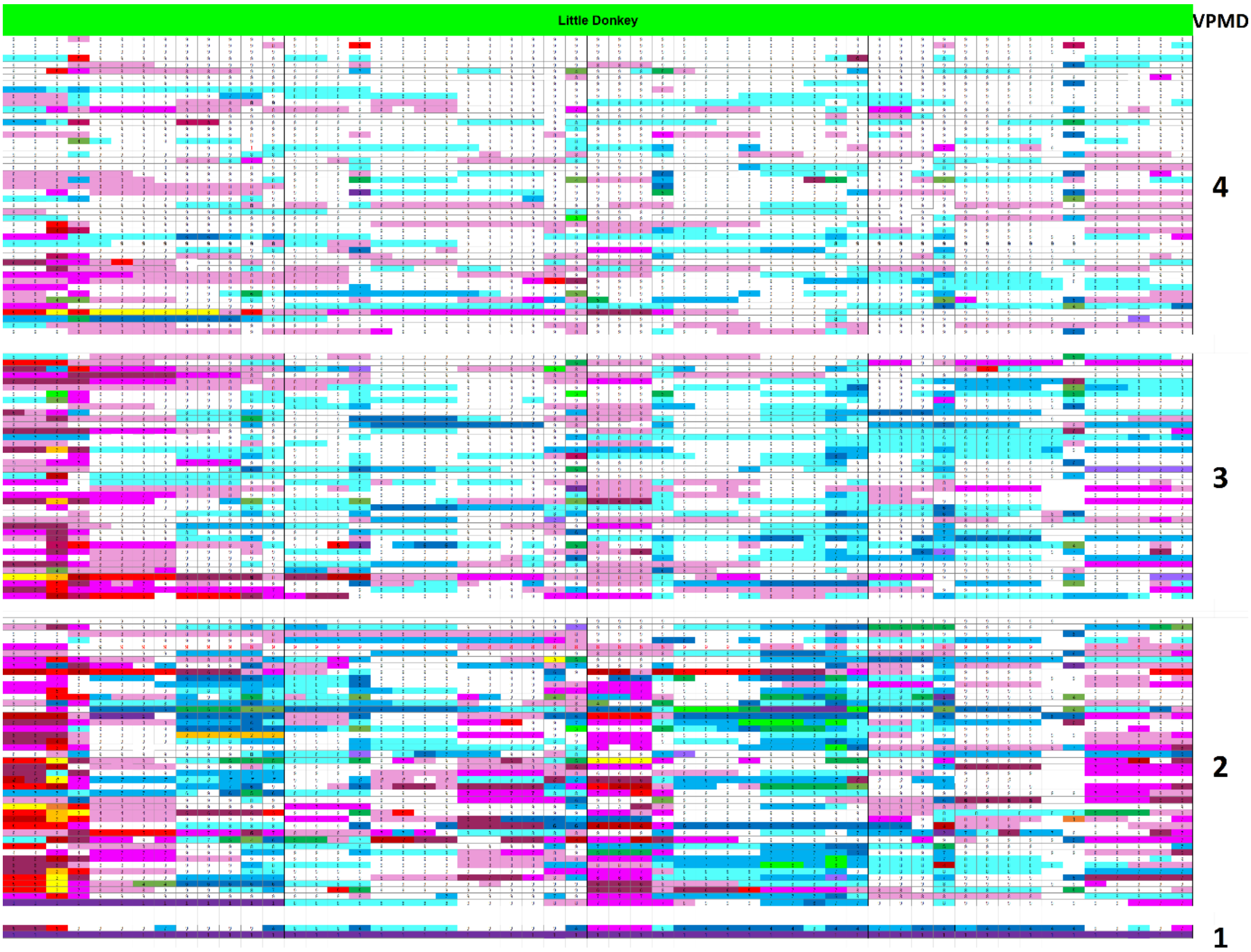
Content of each rating of the *VPMD* scale	Summary of participants' singing based on the *MAPM* system	Example
1. The words of the song appear to be the initial centre of interest rather than the melody, singing is often described as 'chant-like', employing a restricted pitch range and melodic phrases. In infant vocal pitch exploration, descending patterns predominate.	When singing a song, some melodies or a whole song were sung in a spoken manner. A temporary tonality often lasted only a short time, and a clear prevailing tonality could not be recognised. The fluctuation range of tonalities was large, such as within ≤ 14 semitones from the sharpest tonality to the flattest tonality when compared with that of the pitches in the melody of the target song.	See Figure 5.2a
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. The vocal pitch range used in 'song' singing expands.	A song was sung with a generally accurate melodic contour. However, a temporary tonality usually lasted only a short while, and a prevailing key centre was not clear. A gap from the sharpest to a flattest tonality was within ≤ 10 semitones. Compared to a standard target song, absolute semitone errors of most of best matched pitches of the song were ≤ 6 semitones.	See Figure 5.2b
3. Melodic shapes and intervals are mostly accurate, but some changes in tonality may occur, perhaps linked to	Fewer tonalities were applied to sing a target song, and an established tonality tended to last longer. The range from the sharpest to the flattest	See Figure

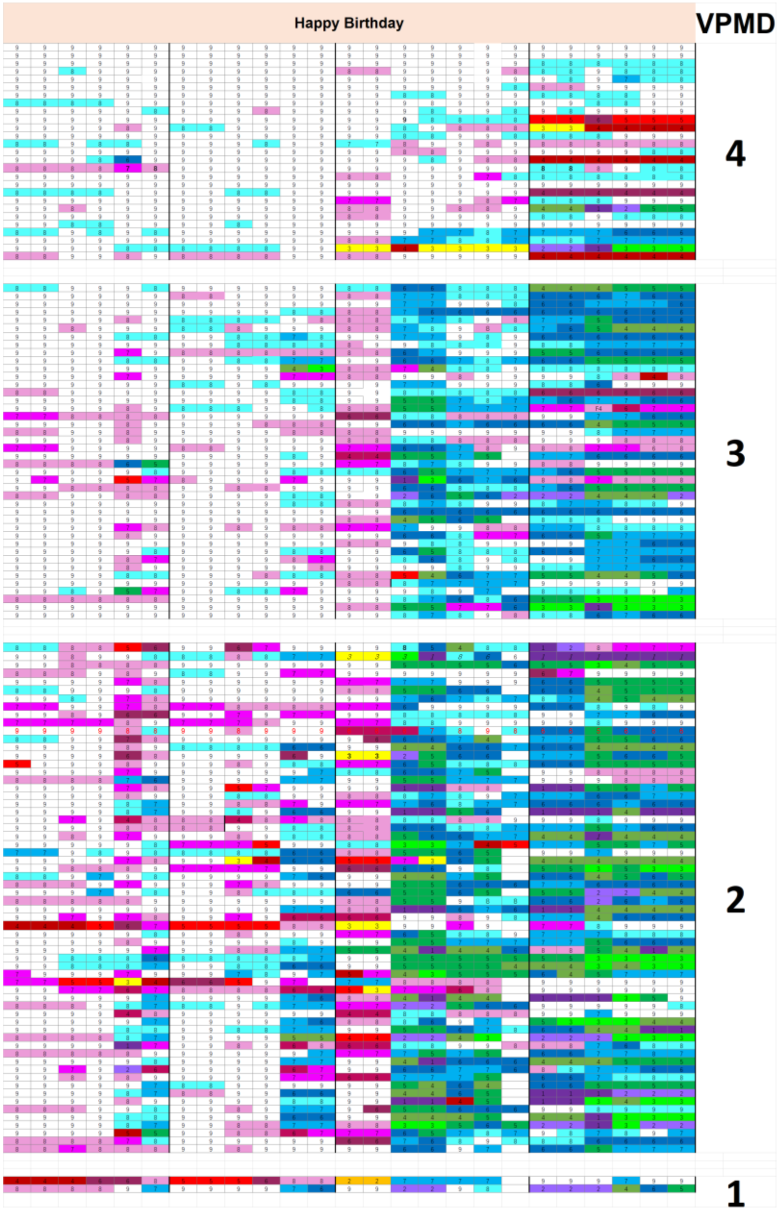
	inappropriate register usage. Overall, however, the number of different reference pitches is much reduced.	tonality was often within 4 semitones. Compared with a standard target song, absolute semitone errors of most matched pitches of a song were ≤ 2.	5.2c
4.	No significant melodic or pitch errors to relatively simple songs from the singer's musical culture.	Most or all of the pitches of a target song were sung with a stable prevailing key, although a few sets of pitches might be sung temporarily with another tonality. However, the variation of tonalities was much reduced. A gap from the sharpest to the flattest tonality was usually within 3 semitones. Compared with a standard target song, semitone errors of most matched pitches of a song were ≤ 1.	See Figure 5.2d

As Table 5.1 illustrates, for each of the four example summaries of vocal pitch-matching accuracy criteria as specified within the *MAPM* system tended to be distinct from each other by (a) the fluctuating degree of tonalities within the sung product and (b) absolute semitone errors compared with the standard target song. When comparing the content of each rating of the *VPMD* scale and summaries based on the *MAPM* system, the content of the two systems generally supports each other. Consequently, the *MAPM* data analyses suggest that the overall summative perceptual judgments based on the *VPMD* scale (ratings of 1 – 4, being least to most accurate in terms of vocal pitch matching accuracy, Welch, 1998) in the current study (as reported in Chapter 6) were reliable (see Figure 5.3).

Figure 5.3 Matching the Colour Characteristics of the MAPM System with Scores Measured by the VPMD Scale for Each of the Three Target Songs for n = 134 Participants




Little Donkey
VPMD
4
3
2
1



Note. Greater vocal pitch matching accuracy is indicated by lighter colours/white as each melody unfolds from left to right in this visual representation.

5.2 Describing the Consistency of Key Centres Applied for the Three Target Songs' Singing

A Pearson Correlation coefficient was computed to assess the linear relationship between the number of key centres applied when matching each two (as pairs) of the three target songs. The correlation between each of the three pairs of target songs was statistically significantly positive:

- *Twinkle, Twinkle* (M = 4.16, SD = 1.60) and *Little Donkey* (M = 5.69, SD = 1.92), r = 0.67, n = 134, $p < .001$;
- *Twinkle, Twinkle* (M = 4.16, SD = 1.60) and *Happy Birthday* (M = 5.69, SD = 1.96), r = 0.56, n = 134, $p < .001$; and
- *Little Donkey* (M = 5.69, SD = 1.92) and *Happy Birthday* (M = 5.69, SD = 1.96), r = 0.53, n = 134, $p < .001$.

These significantly positive correlations suggest that the key centres applied to sing one target song by participants can predict those used to sing another. For instance, when a participant changed many keys to sing *Twinkle, Twinkle*, they were more likely to apply many keys to sing *Little Donkey*.

A Pearson correlation coefficient was computed to assess the relationship between the number of tonalities used to sing a target song and age in whole years. There was a significantly negative correlation between the two variables (r = -.183, n = 402[71], $p < .001$), as illustrated in Figure 5.4, with a significant decrease in the number of sung keys with increasing age $F(5, 395) = 4.53$, $p < .001$.

The decreasing trend in the number of keys used to sing the three target songs with increasing age was not influenced by the other four key independent variables, sex, geographic location, income, and song, as the interaction effect between age in whole years and each of the four variables was not statistically significant, for age in a whole year and sex $F(5, 389) = 0.93$, $p = .459$; for age in the whole years and geographic location $F(4, 390) = 1.32$, $p = .264$; for age in whole years and income $F(5, 389) = 2.00$, $p = .078$; for age in whole years and song $F(10, 383) = 0.51$, $p = .884$.

[71] n = 402 was because the correlation analysis was based on data from each of the three target songs.

Figure 5.4 The Number of Key Centres Used to Sing a Target Song by Participant Age in Whole Years

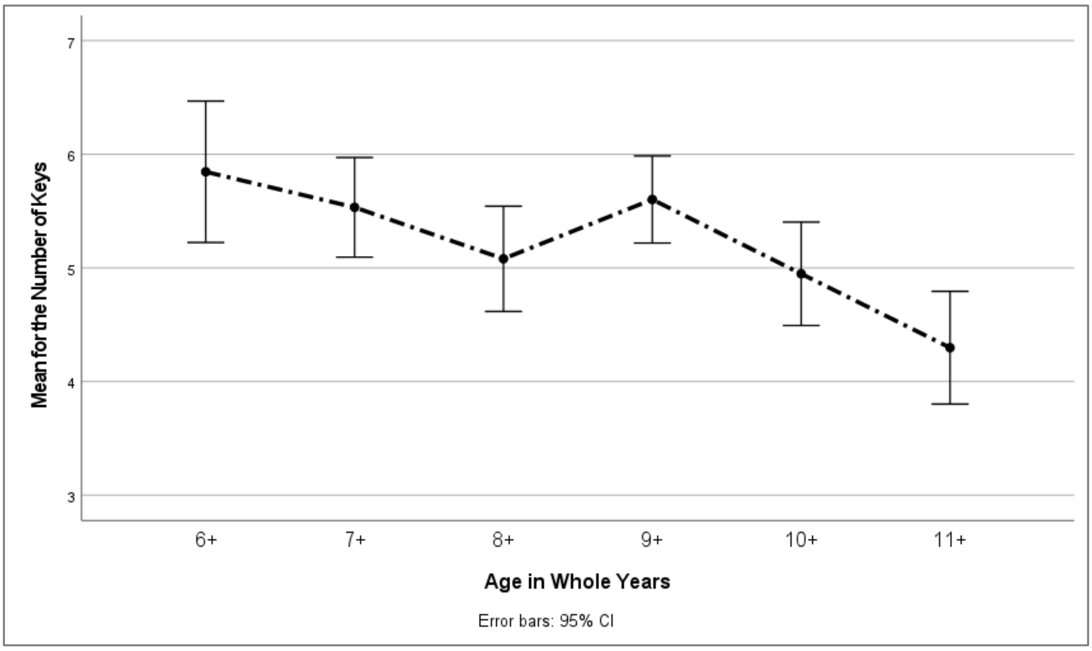
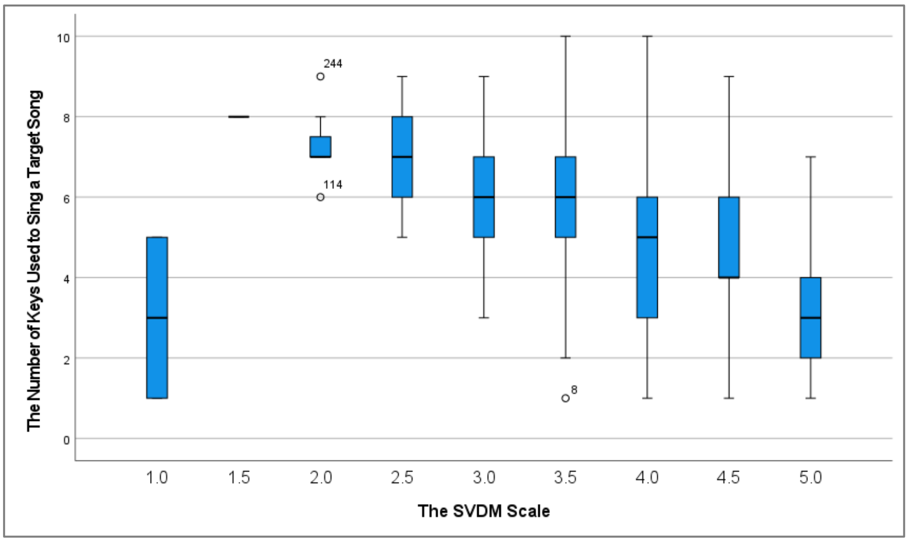


Figure 5.5 Correlation between the Number of Key Centres Used by Participants to Sing the Three Target Songs as Measured in Comparison with the SVDM Scale (Rutkowski, 1996)



It was hypothesised that changing key centres when singing a target song was perhaps an indication of changes in vocal registers.

Consequently, a Pearson correlation coefficient was computed to assess the linear relationship between the number of key centres used to sing the three target songs and the scores as measured by the *SVDM* scale (*cf.* Rutkowski, 1996, 2015, 2018). There was a statistically significant negative correlation between these two variables, $r = -0.57$, $n = 401$, $p < .001$. As Figure 5.5 demonstrates, when participant children received a higher score as measured by the *SVDM* scale, they tended to change fewer key centres in their song singing.

5.3 Describing Semitone Errors of Vocally Matched Musical Intervals of the Three Target Songs

Vocal pitch matching data for two participants are not included in the analyses in this section on musical intervals because they tended to speak the target melodies rather than sing them. Consequently, statistical analyses on vocal pitch-matching accuracy for each of the musical intervals from the three target songs included here are based on sung data from $n = 132$ participants. Overall, there were $N = 15{,}708$ musical intervals were analysed from the three target songs sung by $n = 132$ participants in the current section.

In terms of noting the accuracy of a vocally matched musical interval, all semitone errors were noted with their absolute values, as an equivalent positive or a negative value were treated the same degree of error. A rating of '0' meant no semitone error for the matched musical interval compared to a related target. A larger assigned rating meant poorer vocal pitch-matching accuracy for a musical interval.

The mean accuracy for each type of musical interval was calculated. Participants' vocal matching accuracy for musical intervals was statistically significantly influenced by the type of musical intervals $F(8, 15377) = 303.05$, $p < .001$. As Figure 5.6 illustrates, the absolute mean error was generally raised with expanding the size of musical intervals, except that the mean error for minor 2nd was greater than that for major 2nd.

Figure 5.6 The Means for Vocal Pitch Matching Accuracy of Nine Types of Musical Intervals within the Three Target Songs (n = 132 participants and 15,385 individual sung intervals)

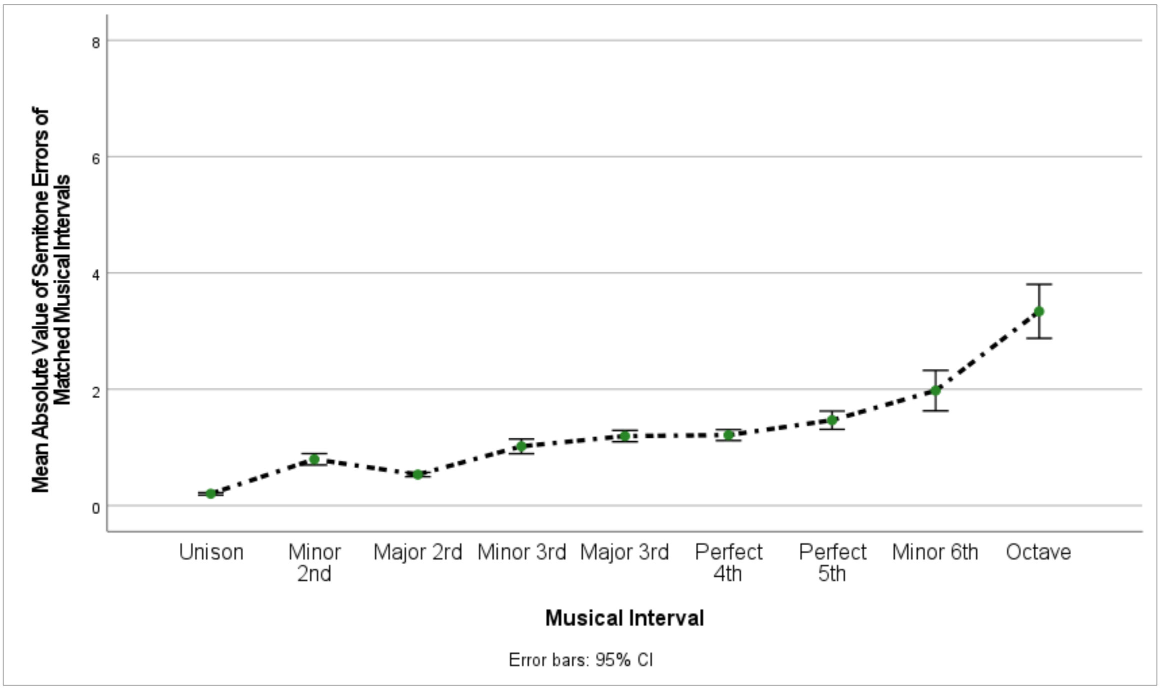


Figure 5.7 Relative Accuracy of Each Type of Sung Musical Interval for the Three Target Songs by Participant Age in Whole Years

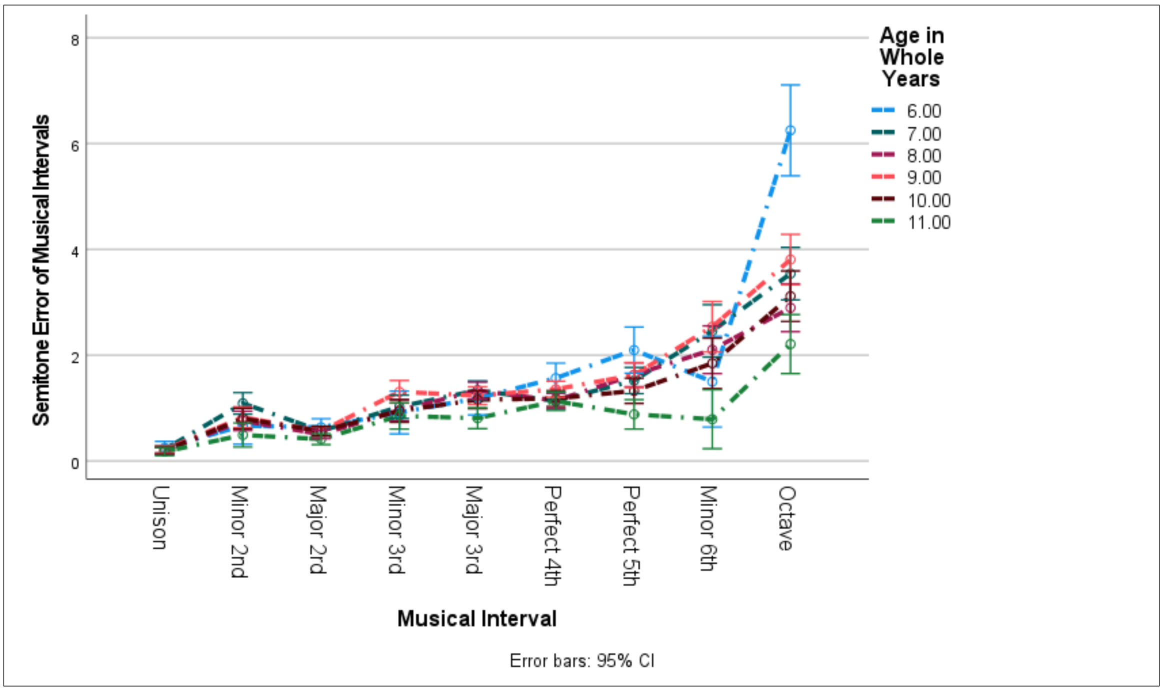
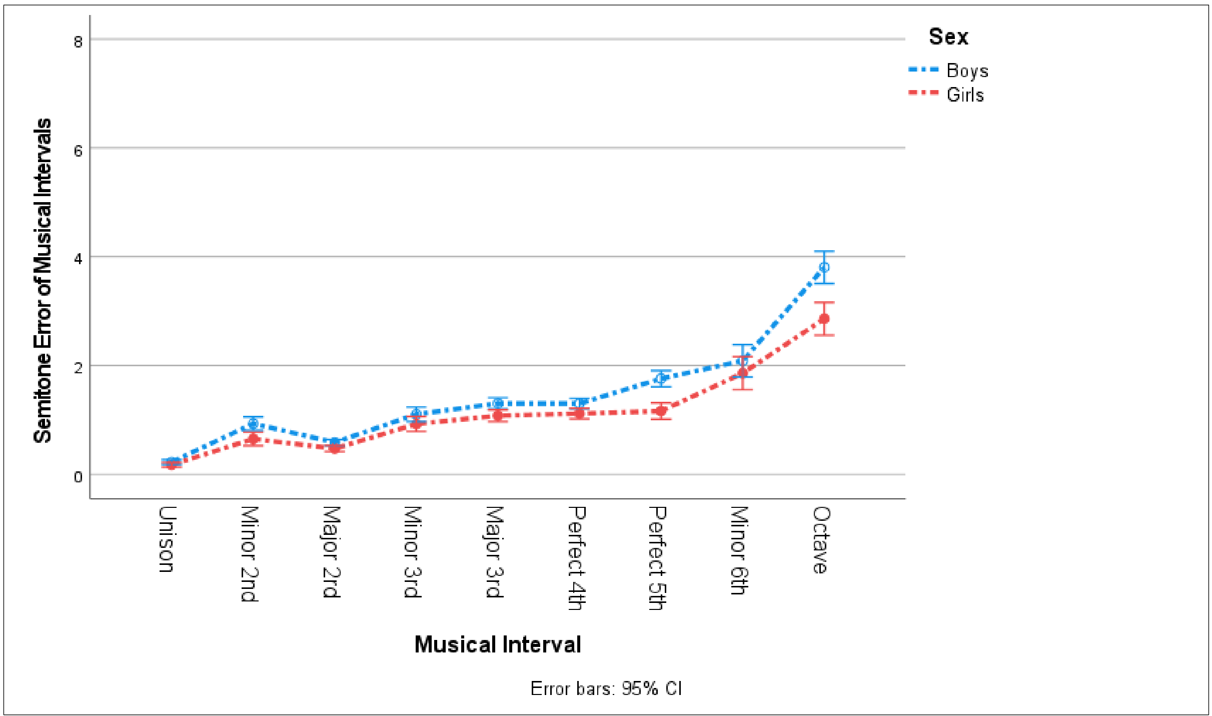


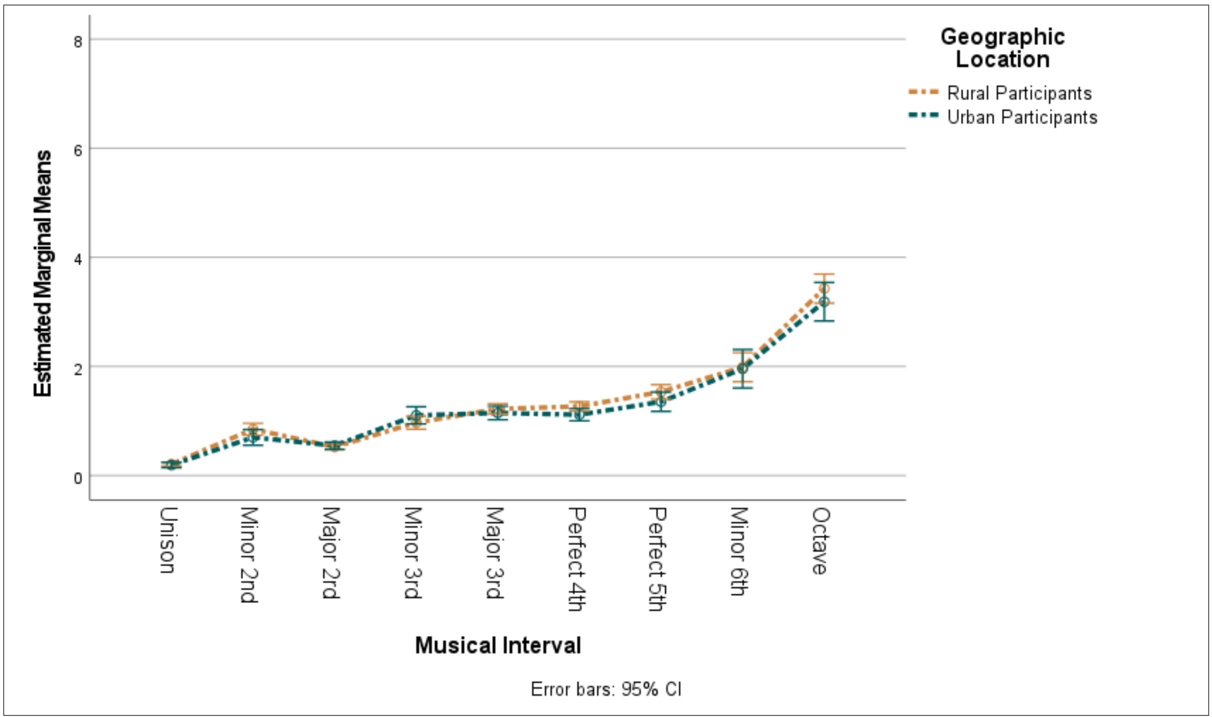
Figure 5.8 Relative Accuracy of Each Type of Sung Musical Interval for the Three Target Songs by Participant Sex



To explore the difficulty of matching minor 2nd, a univariate regression was conducted to compare the difference in vocal pitch matching difficulty between minor 2nd and major 2nd, and minor 2nd and minor 3rd, using a significant level at $p < .01$ to reduce the type one error. Two independent sample-t tests showed that semitone errors for matching major 2nd ($M = 0.53$, $SD = 1.17$) were statistically significantly smaller than those for minor 2nd ($M = 0.80$, $SD = 1.41$) $t(4612) = 5.53$, $p < .001$. Also, the mean of semitone errors for matching minor 2nd ($M = 0.80$, $SD = 1.41$) was statistically significantly smaller than that for minor 3rd ($M = 1.02$, $SD = 1.65$) $t(1446) = -2.78$, $p = .006$. Overall, these finding suggest that the matching difficulty for a minor 2nd in song singing was between a major 2nd and a minor 3rd (see Figure 5.6).

There was also a statistically significant impact of age on the relative accuracy of vocally matched musical intervals $F(40, 15332) = 3.80$, $p < .001$. As Figure 5.7 shows, the accuracy of smaller musical intervals (from unison to perfect 4th) seems to be relatively consistent across different age groups. However, when matching greater musical intervals (from perfect 5th to octave), older participants generally had smaller semitone errors.

Figure 5.9 Accuracy of Each Type of Sung Musical Interval for the Three Target Songs by Geographic Location



The interaction effect between the type of musical intervals and sex was also statistically significant $F(8, 15368) = 5.65$, $p < .001$. As Figure 5.8 shows, boys and girls show similar accuracy in their mean ratings when matching most types of musical intervals. However, there was a tendency for girls to match perfect 5^{th} and octave more accurately than boys.

The interaction effect between the type of musical intervals and geographic location (rural and urban) was not statistically significant $F(8, 15368) = 1.51$, $p = .148$, suggesting that rural and urban participant children showed similar relative levels of accuracy when matching different types of musical intervals (see Figure 5.9).

Figure 5.10 Accuracy of Each Type of Sung Musical Interval for the Three Target Songs by Inferred Income

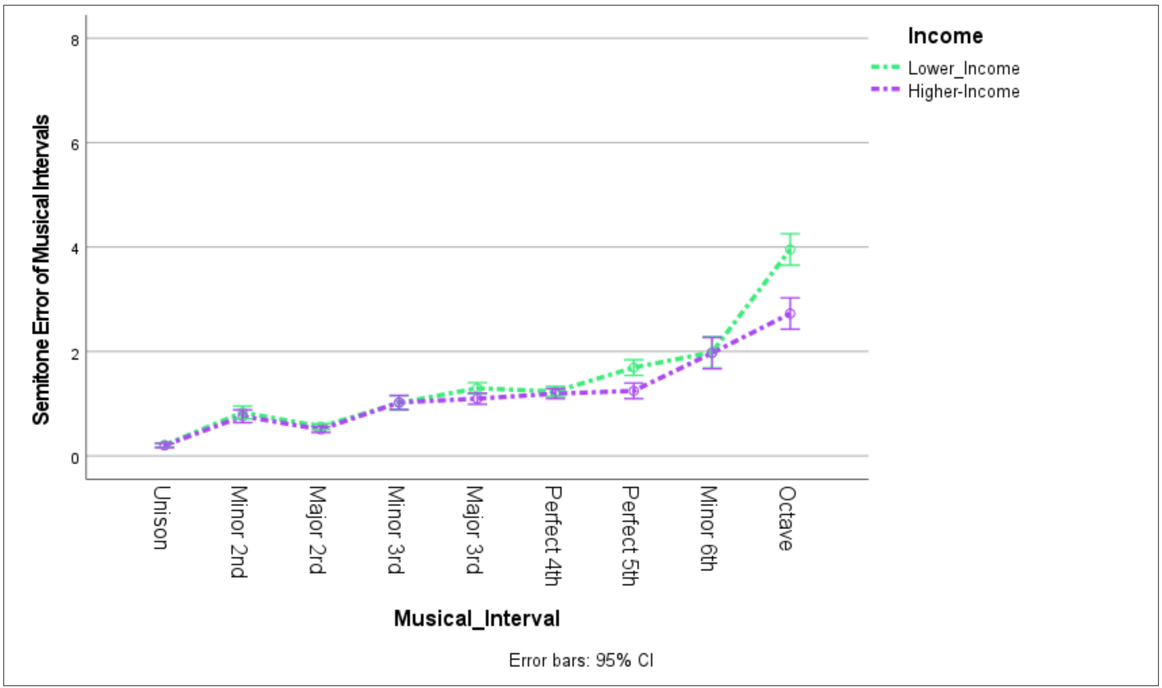
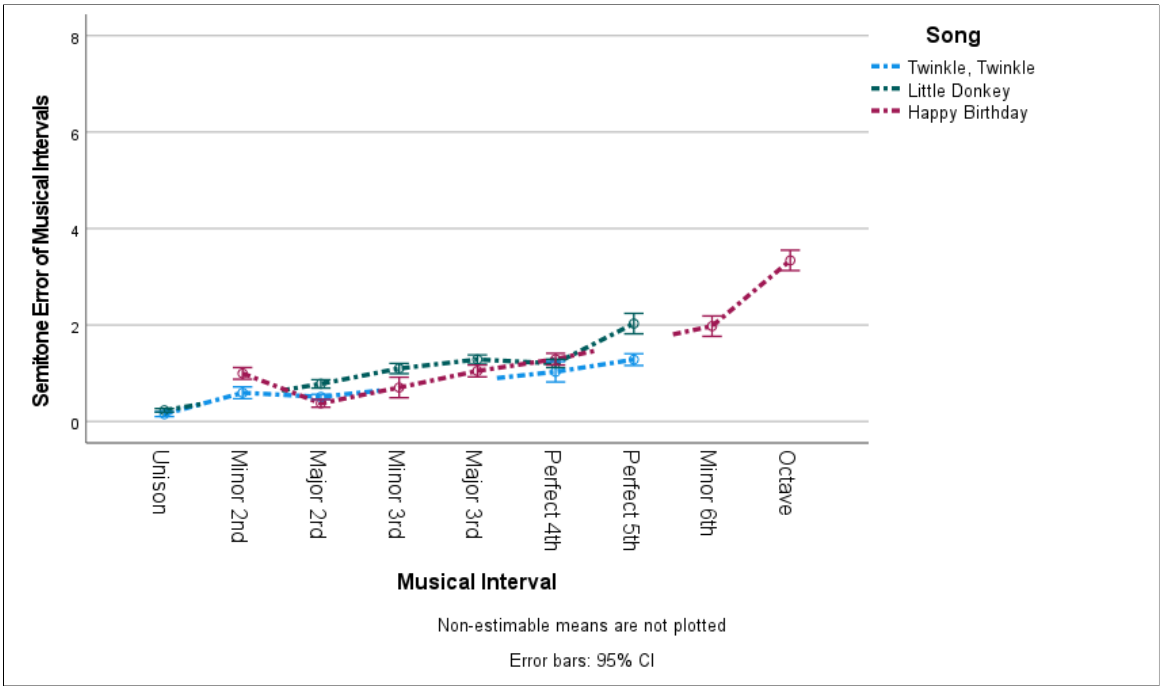


Figure 5.11 Accuracy of Each Type of Sung Intervals for the Three Target Songs by Song



On the other hand, the interaction between the type of sung musical interval and inferred income had a statistically significant impact on the accuracy of vocally matched musical intervals included in the three target songs $F(8, 15368) = 6.06$, $p < .001$. As Figure 5.10 shows, participant children from lower and higher-income families generally show similar vocal pitch accuracy when matching most types of musical intervals, except for the perfect 5th and octave. Participants from higher-income families matched these two types of musical intervals more accurately than peers from lower-income families.

Finally, the interaction effect between the types of musical intervals and the song in which they appeared was also statistically significant $F(7, 15368) = 11.75$, $p < .001$. As Figure 5.11 shows, only major 2nd and perfect 4th are two shared musical intervals included in each of the three target songs. Generally, the accuracies by means of these two musical intervals across the three target songs were similar. However, the mean accuracy for other types of musical intervals varied statistically significantly. For instance, a minor 2nd was matched significantly more accurately in *Twinkle, Twinkle* ($M = 0.60$[72], $SD = 1.38$) than in *Happy Birthday* ($M = 1.00$, $SD = 1.40$) $t(790) = -4.06$, $p < .001$, and a perfect 5th was matched more accurately in *Twinkle, Twinkle* ($M = 1.28$, $SD = 1.56$) than *Little Donkey* ($M = 2.03$, $SD = 2.38$) $t(525) = -4.13$, $p < .001$.

5.4 Describing Common Characteristics of Vocal Pitch-Matching for the Three Target Songs

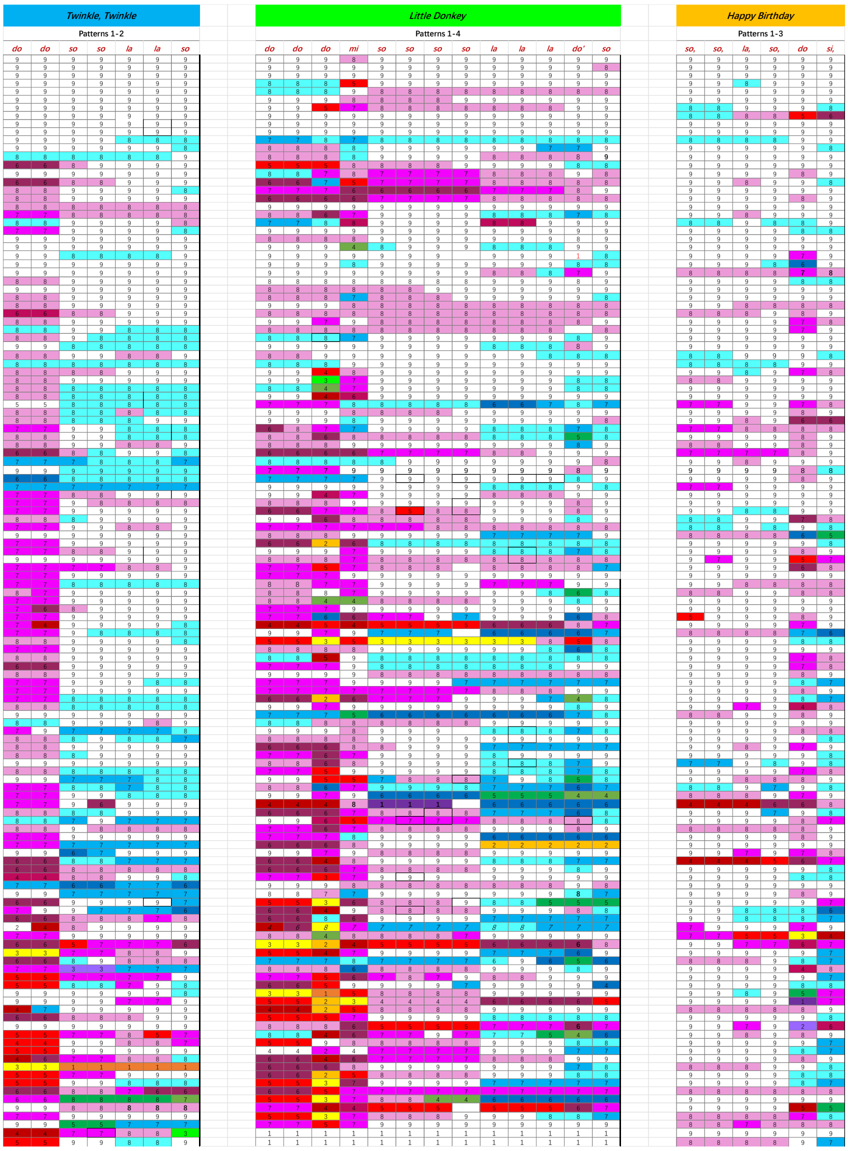
Based on the colour characteristic of the *MAPM* system, common vocal pitch-matching characteristics and types of semitone errors for each sung pitch across the three criterion songs are summarised below.

[72] A smaller '*M'* means matching a musical interval more accurately.

5.4.1 Common Errors in the Starting Pitch of a Target Song

Most of the *n* = 134 participants did not start in a key that they were able to sustain across the whole song when singing *Twinkle, Twinkle* (73.13%, *n* = 98), nor *Little Donkey* (70.90%, *n* = 95). Instead of starting in a key that they could sustain, more than half of participants (66.42%, *n* = 89 for *Twinkle, Twinkle*, and 58.21%, *n* =78 for *Little Donkey*) started in a relatively sharper key compared to their subsequent sung vocal pitches. The opening was usually one to two semitones relatively sharper (the colours represented by pink and purplish red) than subsequent pitches (with absolute accuracy represented by the white colour) – see colour of the first pitch(es) of each column in Figure 5.12.

Figure 5.12 Illustrating the Type of Pitch Error for a Starting Pitch of Each of the Target Songs as Measured by the Melodic Analysis of Pitch Matching (MAPM) System (n = 134 Participants)



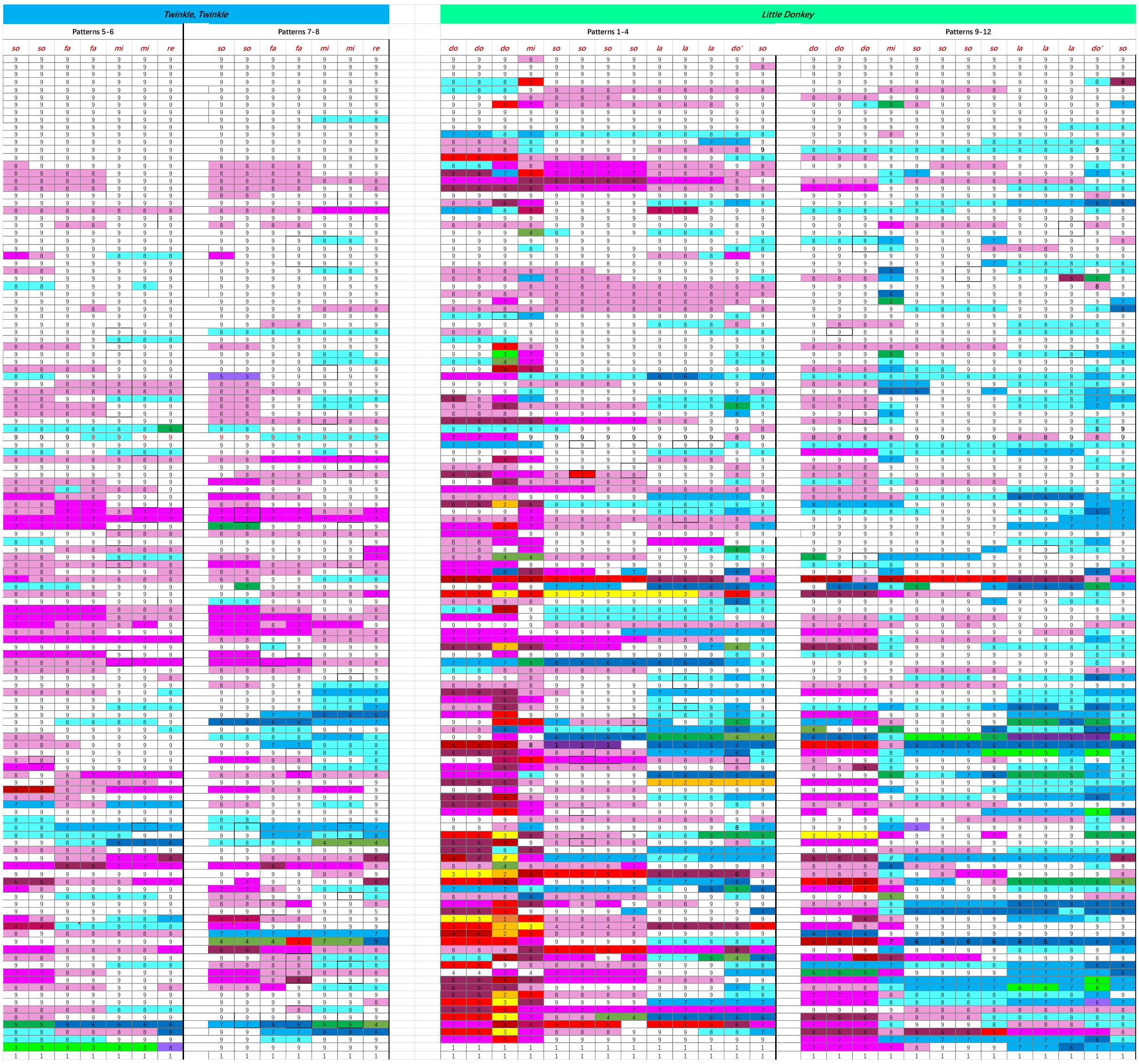
Note. The first pitch in each of the three columns above was the chosen participant starting pitch for each of the three target songs. Also note that, in terms of its melodic design, *Happy Birthday* begins on the dominant of the written key centre[73], whereas *Twinkle, Twinkle* and *Little Donkey* begin on their relative tonic.

73 As Patterns 5-9 of *Happy Birthday* usually were vocally matched with poor accuracy, participants' key centres usually were defined by the first 4 patterns of the song (see Figure 5.3).

5.4.2 Vocal Pitch-Matching Accuracy of Two Similar Melodies

Participants showed similar vocal pitch-matching accuracy across melodic fragments which were identical across the target songs. For instance, the type of vocal pitch-matching errors between two identical melodic fragments (Patterns 5-6 and Patterns 7-8 of *Twinkle, Twinkle*) with a sol-fah melodic representation of *so-so-fa-fa-mi-mi-re* was identical (see Figure 5.13, left columns). Furthermore, the relative sung vocal pitch-matching accuracy between Patterns 1-4 and Patterns 9-12 of *Little Donkey* shared the same melody (*do-do-do-mi-so-so-so-so-la-la-la-do'-so*) were also close (see Figure 5.13, right columns).

Figure 5.13 Example of Vocal Pitch-Matching Errors of Pairs of Melodies Measured by the Colour Characteristics of the Melodic Analysis of Pitch Matching (MAPM) System (n = 134 Participants)



Note.

1. The left two columns are for Patterns 5 to 8 of *Twinkle, Twinkle,* while the right two columns are for Patterns 1 - 4 and 9 -12 of *Little Donkey* (see the list of patterns of each criterion song in Figure 4.4).

2. Each horizontal line above represents the sung vocal pitch-matching accuracy of a criterion song for the same participant.

5.4.3 Pitch Error Caused by Unconscious Preparation Before Vocally Matching a Wide Upward Musical Interval

When vocally matching a target high pitch with an upward wide musical interval during a singing performance, many participants tended to sing the first pitch of the musical interval sharp and thus unconsciously to reduce the semitones interval to the next pitch. For instance, more than half (66.42%, n = 89) of participant children sang *do-do* relatively sharply before matching the third pitch *so* (an upward perfect fifth) in Patterns 1 and 9 in *Twinkle, Twinkle.* Similarly, nearly half (47.76%, n = 64) of participant children matched *so, so,* relatively sharply before matching the third pitch *so'* (an upward octave) in Patterns 5-6 in *Happy Birthday* (as illustrated by colour in Figure 5.14). However, they tended to match the second pitch of an upward wide musical interval flat (see Figure 5.14).

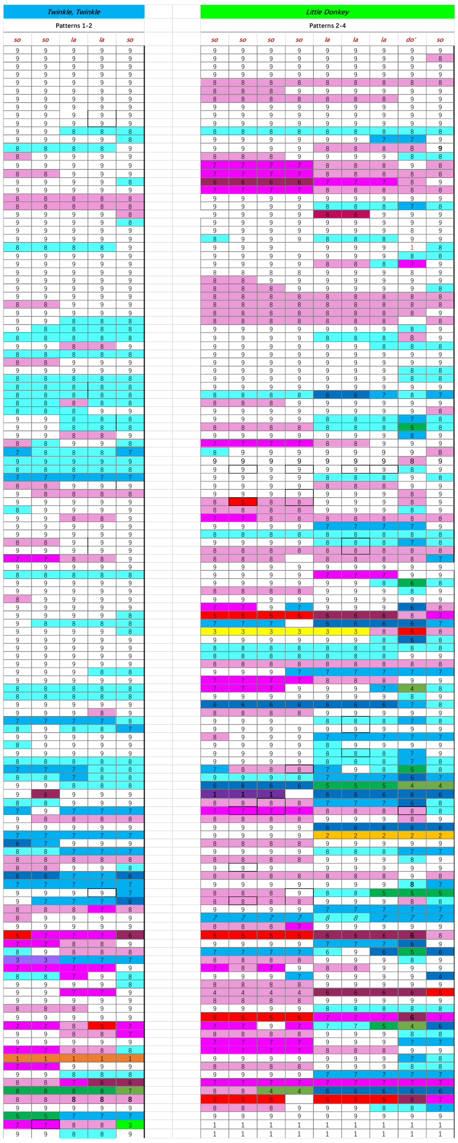
Figure 5.14 Examples of a Type of Sung Pitch Error when Matching an Upward Wide Musical Interval as Measured by the Colour Characteristics of the Melodic Analysis of Pitch Matching (MAPM) System (n = 134 Participants)

Note. The first two columns on the left represent the first four pitches (*do-do-so-so*) of Pattern 1 and Pattern 9 of *Twinkle, Twinkle*. The third column are the three pitches (*so,-so,-so*) of Patterns 5-6 of *Happy Birthday*.

5.4.4 The Influence of the Size of Musical Intervals on Vocal Pitch-Matching Accuracy of a High Pitch

When matching a high second pitch in a musical interval, most participants matched the high pitch poorly. For instance, when matching *so'* of a phrase *so,-so,-so'* in Patterns 5-6 of *Happy Birthday*, the musical interval between so and so' was an upward octave. Only 13.43% (*n* = 18) of participants matched *so'* relatively accurately (see Figure 5.14 above). However, when matching a high pitch that is led into by a small musical interval, most participants could vocally match these relatively accurately. For instance, 73.13% (*n* = 90) of participants matched *la-la* of *so-so-la-la* in Patterns 1-2 of *Twinkle, Twinkle* relatively accurately (noted with the same colour) (see Figure 5.15). Furthermore, half (50%, *n* = 67) of participant children showed a relatively good vocal pitch-matching accuracy (noted with the same colour) for *do'* of a melody *so-so-so-so-la-la-la-do'-so* in Patterns 2-4 of *Little Donkey* (see Figure 5.15).

Figure 5.15 Examples of Matching a Relatively High Pitch that is Prefaced by a Small Interval as Measured by the Colour Characteristics of the Melodic Analysis of Pitch Matching (MAPM) System (n = 134 Participants)



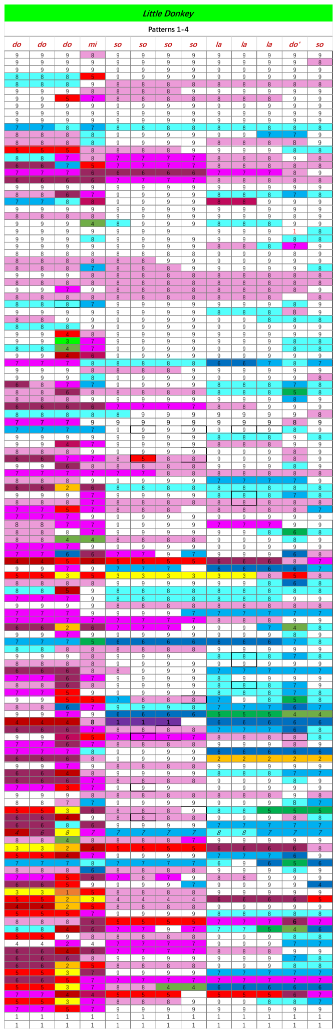
Note. The pitches in the left column were *so-so-la-la-so* from Patterns 1 to 2 of *Twinkle, Twinkle*, while those in the second column were *so-so-so-so-la-la-la-do'-so* from Patterns 2-4 of *Little Donkey*.

5.4.5 A Common Vocal Pitch-Matching Error Towards a Lift Point during Song Singing

When singing a continuous upward melody and closing at a lift point, many participant children sung these pitches flat. For instance, 41.04% (*n* = 55) of participants matched *la-la-la-do'-so* in Patterns 3-4 relatively flatly after singing a series of upward pitches (*do-do-do-mi-so-so-so-so*) in Patterns 1-2 of *Little Donkey* (as illustrated in Figure 5.16).

Similarly, when singing continuously downward pitches close to a lift point, many participants matched the relatively low pitches sharply. For instance, when matching consecutive pitches downwards towards *re* in Patterns 3-4 (*fa-fa-mi-mi-re-re-do*) in *Twinkle, Twinkle*, 44.78% (*n* = 60) of participants sang one to three pitches of *re-re-do* relatively sharply (see Figure 5.16). Similarly, 39.55% (*n* = 53) of participants matched *re-re-re-re* in Patterns 7-8 relatively sharply after *mi-mi-mi-mi* in Pattern 6 of *Little Donkey* (see Figure 5.17).

Figure 5.16 An Example of Vocal Pitch-Matching Accuracy of an Upward Melody Measured by the Colour Characteristics Used in the Melodic Analysis of Pitch Matching (MAPM) System (n = 134 Participants)



Note. This column included sung pitches of Patterns 1 to 4 of *Little Donkey* (*do-do-do-mi-so-so-so-so - la-la-la-do'-so)*.

Figure 5.17 Example of Vocally Matching Accuracy for a Downward Melody as Measured by the Colour Characteristics Used in the Melodic Analysis of Pitch Matching (MAPM) System (n = 134 Participants)

Note. The pitches in the left-hand column were *fa-fa-mi-mi-re-re-do* from Patterns 3 to 4 of *Twinkle, Twinkle*. Those in the second column were *fa-fa-fa-la-mi-mi-mi-mi-re-re-re-re-so-so* from Patterns 5 to 8 of *Little Donkey*.

5.5 Summary

This Chapter reports the application of a new rating system *MAPM* which seeks to identify the relative sung accuracy of the constituent pitches of a target melody. The system was applied to each sung vocal pitch compared to its model target, whilst allowing the participants (n = 134) to choose their own starting pitch and relative choice of sung key for each of three well-known songs. The numerical rating system was based on the relative vocal pitch accuracy, ranging from 1 to 9, with 9 indicating an absolute pitch match with the target pitch in the melody. The system was adapted to explore the participants' relative vocal pitch accuracy in matching intervals within the three target melodies and a lower number rating indicated greater accuracy in sung interval reproduction. A further adaptation was applied to explore the consistency of sung key centres, where a lower number rating indicated a greater internal consistency of key centre usage, linked to vocal register use.

There was a significant positive correlation for scores measured between the *MAPM* system and the *VPMD* scale (Welch, 1998) used in Chapter 5 to assess whole song vocal pitch matching. A selection of sung products from n = 134 participants drawn from the whole cohort were analysed in terms of the three applications of the *MAPM* system, i.e., by individual pitch accuracy, by relative consistency in sung key centres and by individual intervals.

In terms of the consistency of tonalities across the three target songs, the number of tonalities used to sing each of the two target songs was correlated statistically significantly. In other words, when a participant changed many keys when singing one target song, they would be more likely to use many keys to sing another target song. However, as participants got older, they tended to use fewer keys to sing a target song, suggesting a greater sense of key with increasing age. This development was not affected significantly by sex, geographic location, nor inferred income. There was a negative correlation between the number of keys used to sing a target song and related scores in vocal register use as measured by the *SVDM* scale (Rutkowski, 1996, 2015, 2018). When participants used more vocal registers to sing a target song, they generally used fewer keys to sing the song.

The three target songs included nine types of musical intervals. Unison and a Major 2nd were the two most common musical intervals. Generally, accuracy decreased significantly as the size of musical intervals increased, except for a minor 2nd. The accuracy for a minor 2nd was lower than for a major 2nd but higher than for a minor 3rd.

Participants across different age groups showed similar accuracy when matching musical intervals from unison to a perfect fourth. However, when matching larger musical intervals, older participants matched these more accurately than younger children. When matching most musical intervals, girls and boys showed similar accuracy. However, when matching a perfect 5th and an octave, the mean accuracy was higher for girls than boys. The inferred participant family income effect was similar to that of sex in that participants from higher-income families showed greater accuracy when matching a perfect 5th and an octave than peers from lower-income families, while they showed similar accuracy when matching other types of musical intervals. Furthermore, the relative sung accuracy for the same type of musical interval varied slightly across the three target songs.

Lastly, many common characteristics of participants' song singing were summarised. When the starting pitch was a tonic, many participants tended to start with a key that was around one to two semitones higher than an appropriate key as inferred by their subsequent sung vocalisations. When singing two similar melodies, participants tended to repeat pitches errors. Before matching a high pitch prefaced by a wide musical interval, many participants raised the first (lower) pitch of the musical interval, and their accuracy of the high pitch was also poor. However, when matching a high pitch which was approached more by interval step or smaller intervals, participants generally showed good vocal pitch-matching accuracy for the target high pitch. When matching a series of upward pitches towards a lift point, many participants matched these pitches flat. When matching a series of downward pitches towards a lift point, many of them matched the pitches sharp.

Overall, the new *MAPM* system appears to both complement and extend the existing models of children's singing behaviours as applied in Chapter 5. This is seen as a robust and novel approach to understanding children's sung vocal pitch accuracy within and across a target melody.

Chapter Six

Results of Singing Performances Measured from a Macro, Whole Song Perspective

This chapter starts by showing the results of singing performances for participants tested once, either in 2017-2018 or 2018-2019 (Section 6.1), followed by a presentation of the results of all singing performances for Chinese participants across the two years of data collection (Section 6.2). Then, the results based on the above two datasets are compared in Section 6.3. Results from the longitudinal dataset are reported in Section 6.4. Finally, a comparison of singing performances between all Chinese participants in the current study and English students with non-*Sing Up* experiences from the UK is made to explore any cultural influence on Primary school children's singing behaviour.

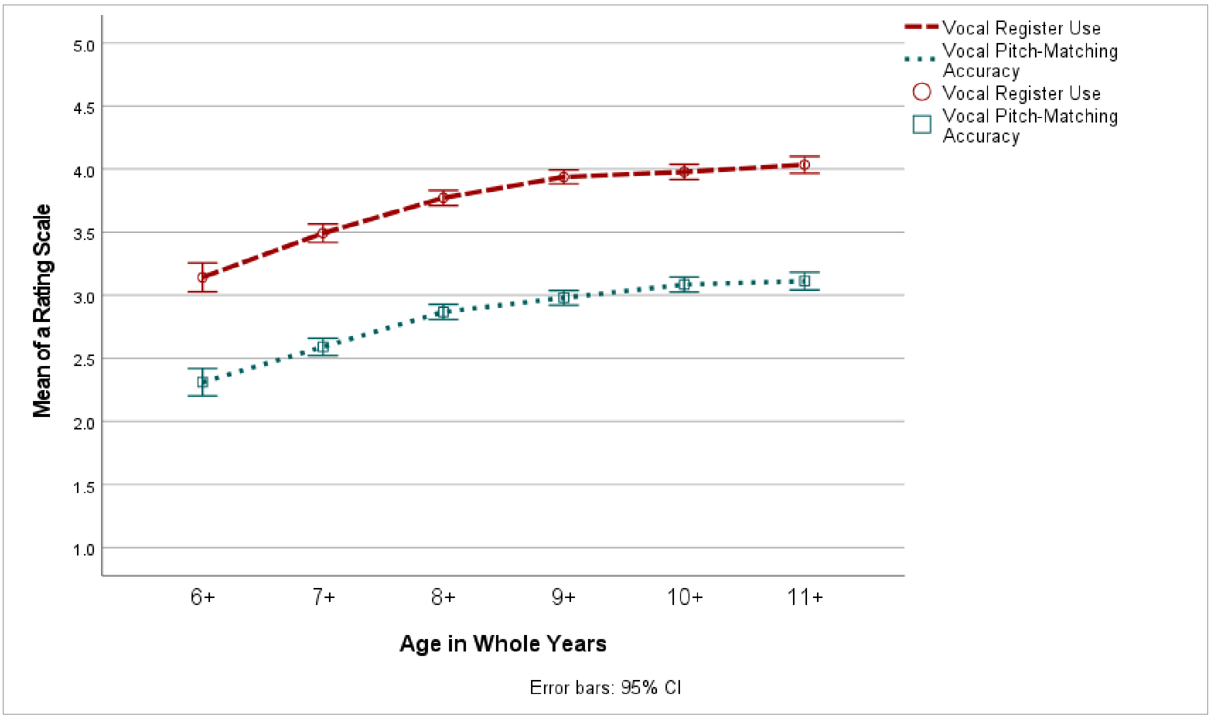
6.1 Chinese Children's Developing Singing Behaviour based on Participants Tested once, either in 2017-2018 or 2018-2019

Each participant sang three familiar target songs, *Twinkle, Twinkle*, *Little Donkey*, and *Happy Birthday*. The singing performances were measured by two existing rating scales, the Singing Voice Development Measure (*SVDM*) scale (Rutkowski, 1996, 2015, 2018), and the *Vocal Pitch-Matching Development* (*VPMD*) scale (Welch, 1998). The first scale measured the application of vocal registers, while the second scale measured vocal pitch-matching accuracy for song singing.

When analysing the two above singing perspectives for participants who were tested once only, either in 2017-2018 or in 2018-2019 by age in whole years (7y+, 8y+, etc), two Pearson correlations were computed to assess (a) the relationship between age in years and vocal register use and (b) the relationship between age and the vocal pitch-matching ability, all data arising from performances for the three target songs. There was a positive correlation between each of the two pairs: age and vocal register use (r = .262, n = 3579[74], p < .001), and age and vocal pitch-matching accuracy (r = .248, n = 3,579, p < .001). As illustrated in Figure 6.1, participants' scores for the two singing perspectives improved statistically significantly with increasing age: for vocal register use (F(5, 3573) = 63.28, p < .001), and for vocal pitch-matching accuracy (F(5, 3573) = 53.56, p < .001).

[74] N = 3579 relates to the cumulative number of individual ratings across participants for the three target songs. There were N = 3,579 singing performances of the three target songs for N = 1,193 participants.

Figure 6.1 Vocal Register Uses and Vocal Pitch-Matching Accuracy for the Three Target Songs and Children's Ages (in Whole Years) based on Participants Tested Once Only Either in 2017-2018 or in 2018-2019



As Figure 6.2 shows, percentages of singing performances gaining a relatively high score of vocal register use for the three target songs (rating ≥ 3 on the *SVDM* scale) increased from 69% at age 6 to 97% at age 11. Similarly, percentages of singing performances receiving a relatively high score of vocal pitch-matching accuracy for the three target songs (rating ≥ 3 on the *VPMD* scale) increased from 34% at age 6 to 76% at age 11.

The percentage of singing performances receiving low scores for vocal register use for the three target songs (rating < 3 on the *SVDM* scale) decreased from 31% at age 6 to 3% at age 11. Similarly, the percentage of singing performances matching the three target songs poorly (rating < 3 on the *VPMD* scale) decreased from 66% at age 6 to 25% at age 11.

Figure 6.2 Stacked Bars Showing Development of (a) Vocal Register Use and (b) Vocal Pitch-Matching Accuracy for the three Target Songs with Age in Whole Years for Participants Tested Once only, Either in 2017-2018 or in 2018-2019

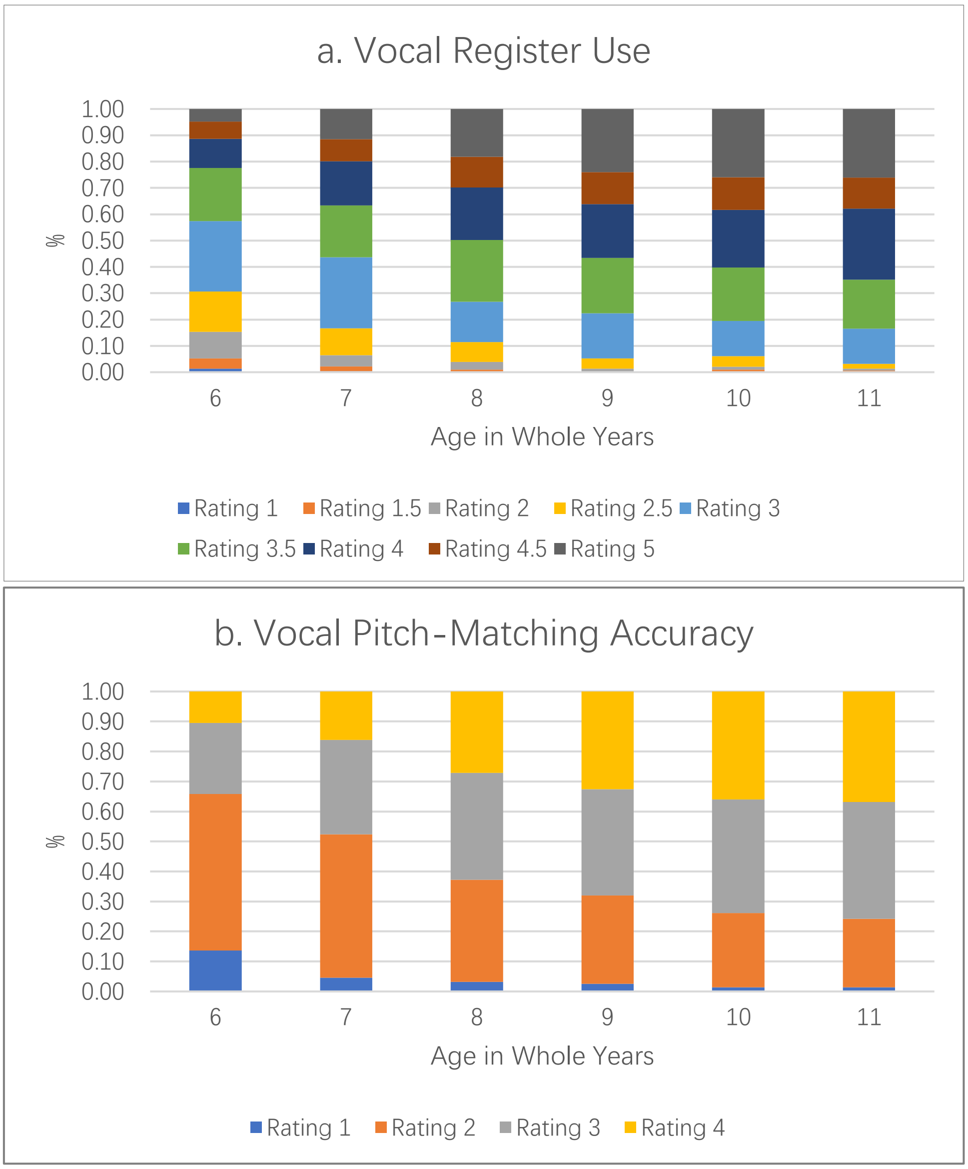
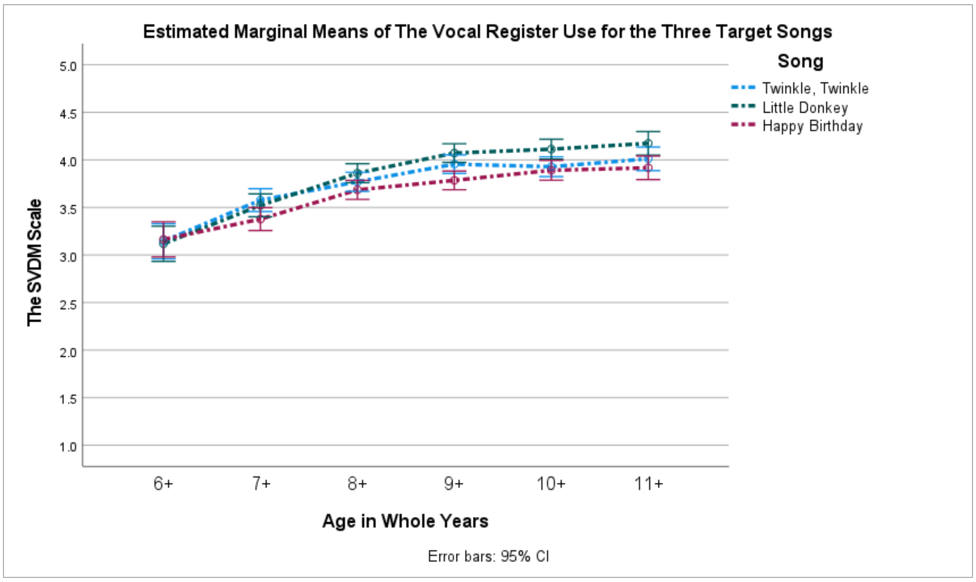
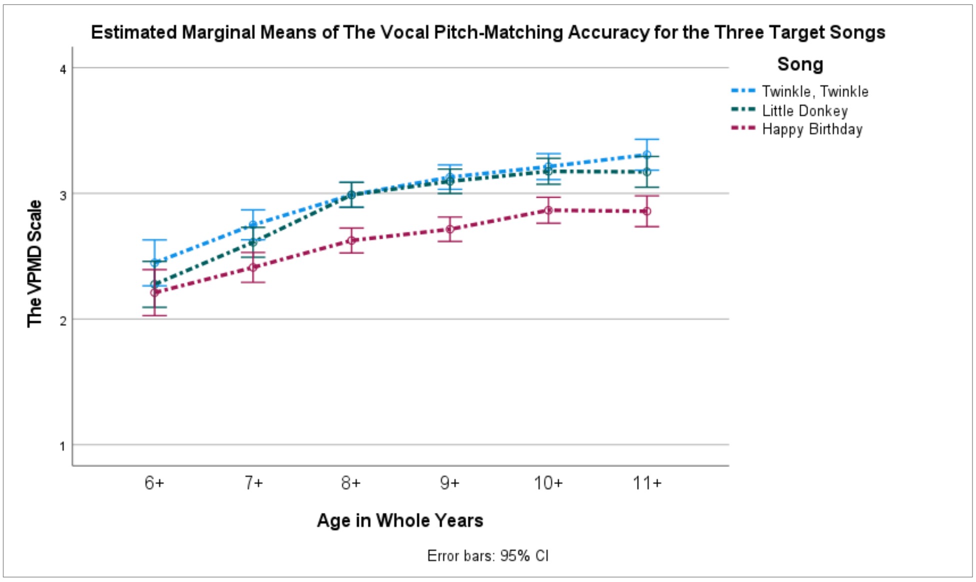


Figure 6.3 Development of Vocal Register Use and Vocal Pitch-Matching Accuracy for the Three Target Songs by Age in a Whole Year and Songs Based on the Dataset Including Participants Tested Once only, Either in 2017-2018 or in 2018-2019





The participants' age and individual song interaction effect were not statistically significant, neither for the application of vocal registers $F(10, 3561) = 1.04$, $p = .403$, nor vocal pitch-matching accuracy $F(10, 3561) = 0.82$, $p = .610$ for the three target songs. This suggests that the development of both singing variables by age was not different significantly between the three target songs.

As Figure 6.3 shows, participants' vocal register use and vocal pitch-matching accuracy for each of the three target songs generally improved with increasing age. For vocal register use, participants across ages showed similar vocal register uses statistically for each of the three target songs. All participants across ages also showed the poorest vocal pitch-matching accuracy when singing *Happy Birthday*, while their accuracy for the other two target songs seems similar.

Combined, the two above singing variable perspectives offer a holistic picture of participant children's singing behaviour. A Pearson correlation coefficient was conducted to assess the relationship for scores measured between the *SVDM* scale and the *VPMD* scale. There was a strong positive correlation between the two dependent variables (r = .834, n = 3579, p < .001). Consequently, the subsequent data analyses were based on the combined two rating scales and a normalised scale out of 100, taking the means of two existing scales for the three target songs.

Overall, participant children's normalised singing scores (*NSS*) increased statistically significantly with age F(5, 3573) = 63.20, p < .001 (see Figure 6.4).

There was also a statistically significant impact of the 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. As Figure 6.5 shows, the sex difference in the NSS was smaller when participants were younger (ages 6 and 7); this increased during the middle of the Primary school period (ages 8 and 10), but it decreased at the end of the Primary school period (age 11).

Similarly, the age and geographic location interaction effects 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 (Figure 6.6), and for the age and income interaction effect F(5, 3567) = 11.16, p < .001 (Figure 6.7). For each interaction, older children tended to receive higher normalised singing score ratings. However, the difference in NSS scores related to income appeared to disappear for the oldest participants (Figure 6.7).

Figure 6.4 Normalised Singing Scores for the Three Target Songs and Children's Ages (in Whole Years) based on Participants Tested Once only, Either in 2017-2018 or in 2018-2019

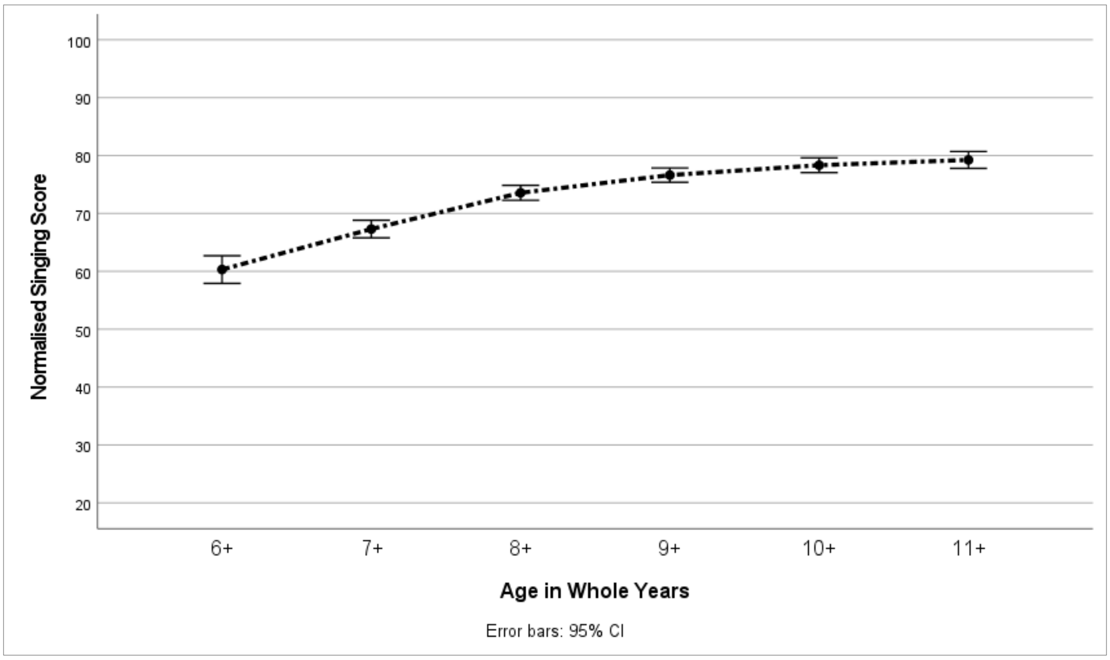


Figure 6.5 Normalised Singing Scores and Participant Children's Age (in Whole Years), Differentiated by Sex based on Participants Tested Once only, Either in 2017-2018 or in 2018-2019

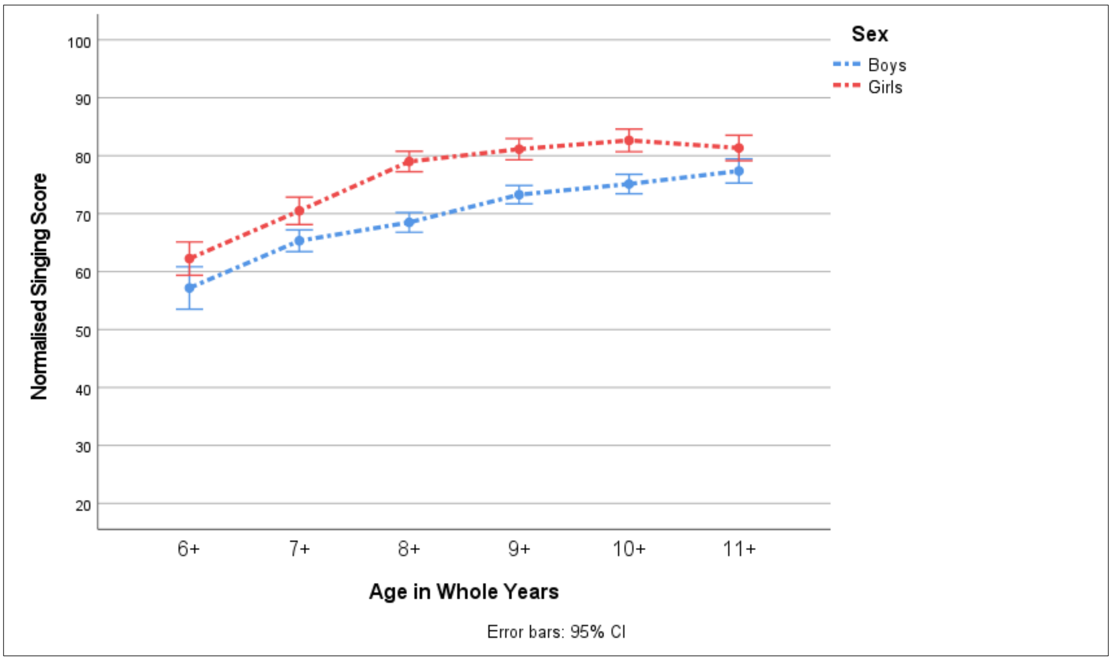


Figure 6.6 Normalised Singing Scores and Participant Children's Age (in Whole Years), Differentiated by Geographic Location based on Participants Tested Once only, Either in 2017-2018 or in 2018-2019

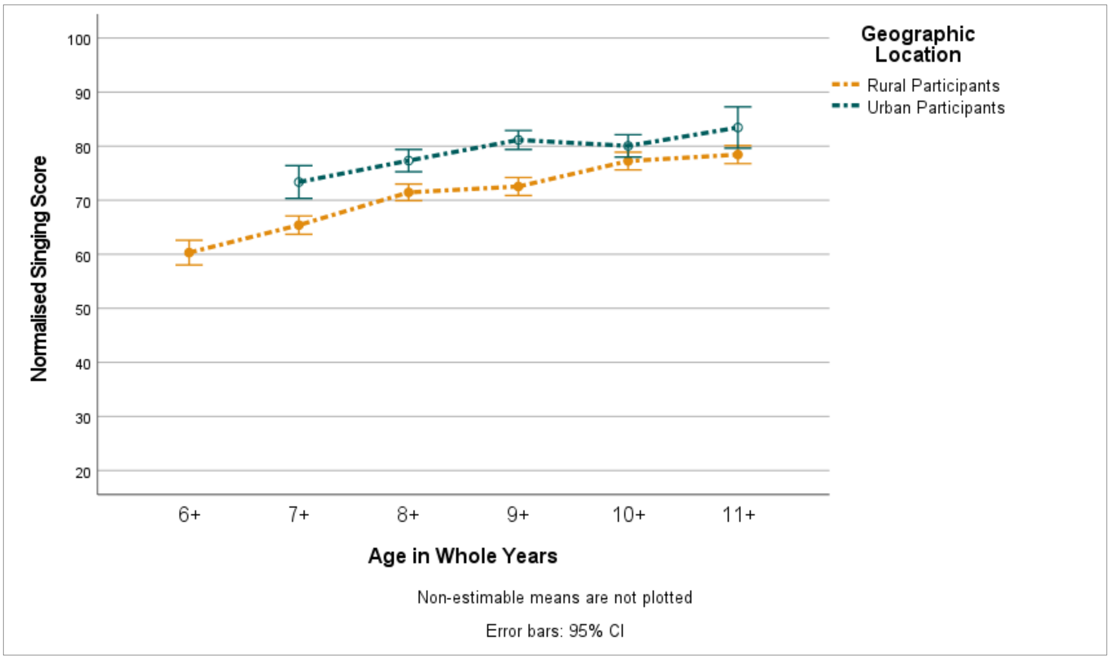


Figure 6.7 Normalised Singing Scores and Participant Children's Age (in Whole Years), Differentiated by Income based on Participants Tested Once only, Either in 2017-2018 or in 2018-2019

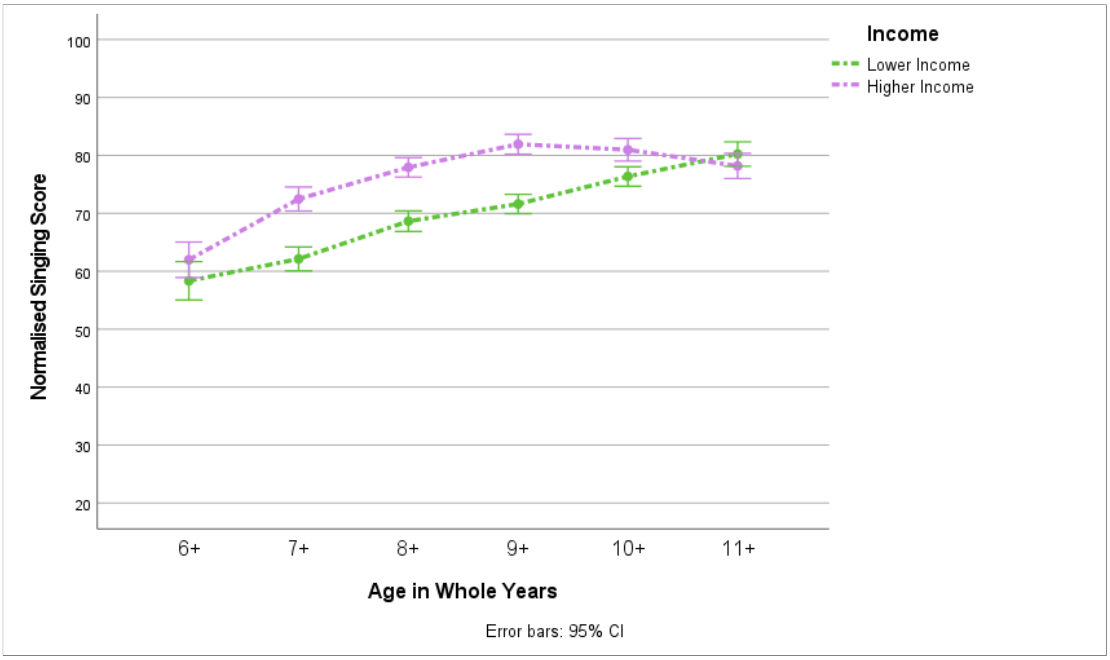


Figure 6.8 Normalised Singing Scores and Participant Children's Age (in Whole Years), Differentiated by School based on Participants Tested Once only, Either in 2017-2018 or in 2018-2019

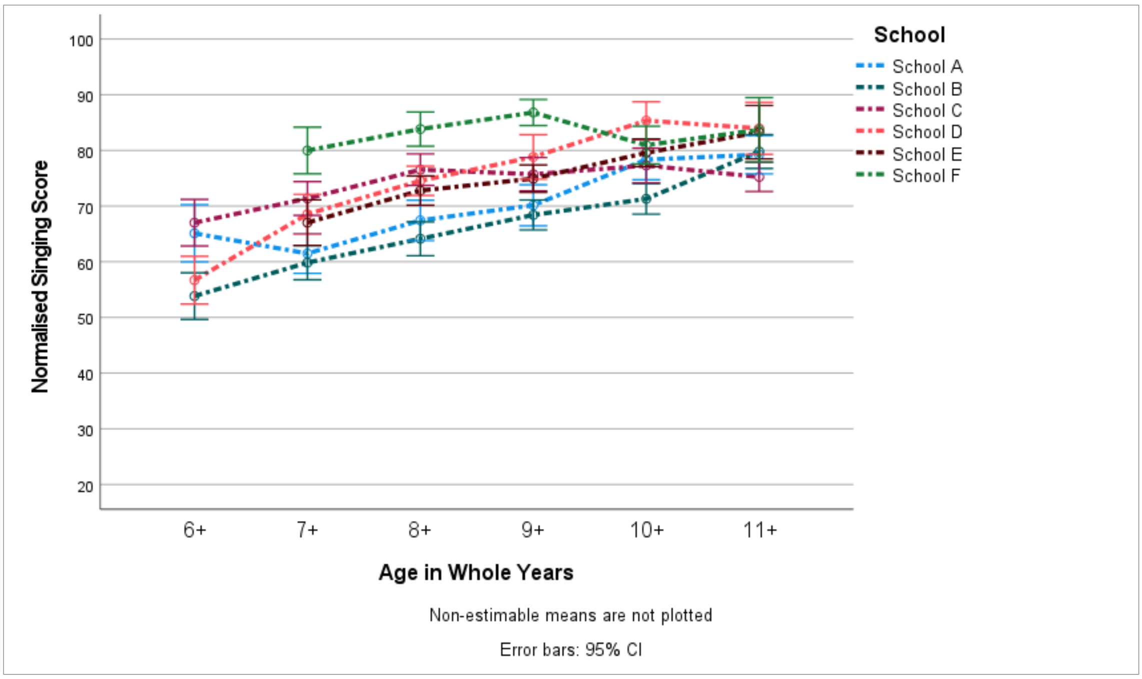
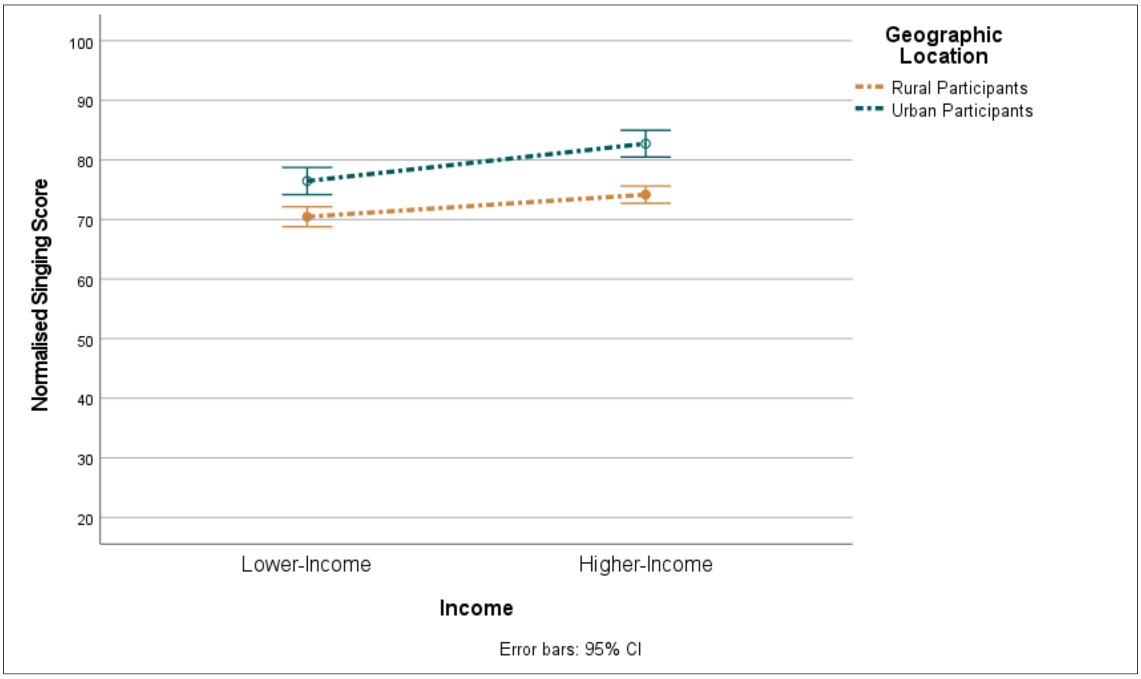


Figure 6.9 Normalised Singing Scores and Participant Children's Family Income, Differentiated by Geographic Location Tested Once only, Either in 2017-2018 or in 2018-2019



Furthermore, age and school interaction also had a statistically significant influence on the NSS $F(23, 3545) = 4.99$, $p < .001$. As Figure 6.8 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.

The interaction between geographic location and family income effect on participants' singing behaviour was explored. The test had no statistically significant interaction $F(1, 1189) = 1.68$, $p = .196$ (see Figure 6.9). In each geographic location, rural participants were less developed in their measured singing competency than their urban peers, irrespective of family income.

6.2 Chinese Children's Developing Singing Behaviour Based on All Singing Performances by Age across the Two Years of Data Collection, 2017-2018 and 2018-2019

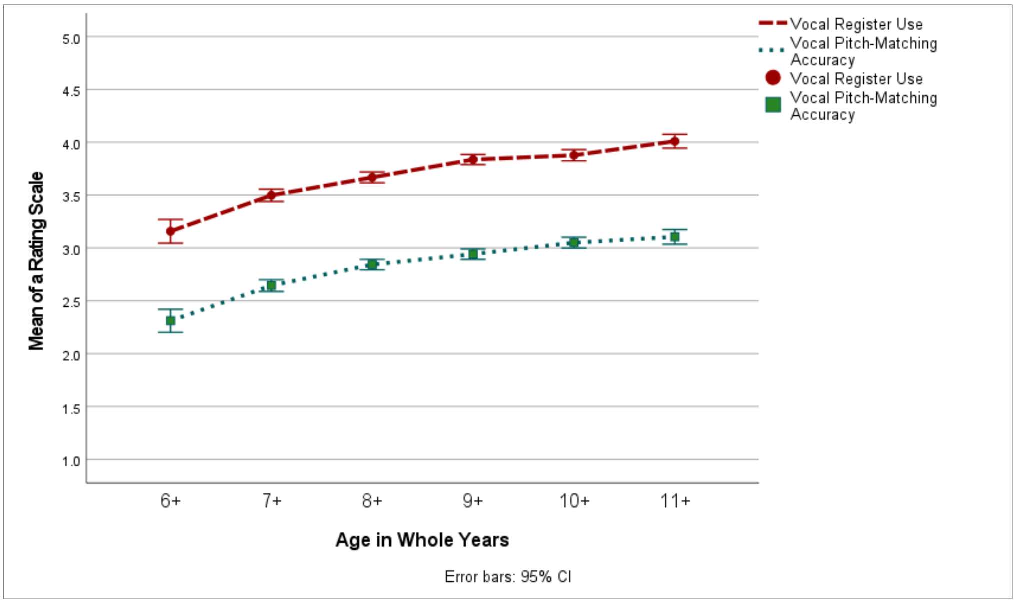
Analyses presented in the current section include data on all ($N = 1,539$) singing performances across the three target songs from all ($N = 1,193$) Chinese participants tested in both 2017-2018 and 2018-2019. Children who were tested twice, once each year, were excluded from Section 6.1 analyses. However, every singing assessment is included in this Section 6.2 to create the fullest picture of the assessment data. Overall, the results in the current Section 6.2 are similar to those reported in Section 6.1, as might be expected.

For the dataset including all of the Chinese participants' singing performances, two Pearson correlation coefficients were tested to explore the relationship between age clusters and each of the two singing perspectives (vocal register uses and vocal pitch-matching accuracy for the three target songs). There were positive correlations between age and vocal register use ($r = .232$, $n = 4617$[75], $p < .001$) and age and vocal pitch-matching accuracy ($r = .223$, $n = 4617$, $p < .001$) (see Figure 6.10).

[75] The reason for $n = 4,617$ is that each of the $N = 1,539$ singing performances included three target songs. The correlation between age and each of two singing perspectives was based on 1539*3=4617 singing performances of all of three target songs.

With increasing age, both of these singing perspectives of participant children improved statistically significantly, for vocal register use $F(5, 4611) = 57.41$, $p < .001$ and for vocal pitch-matching accuracy $F(5, 4611) = 53.66$, $p < .001$.

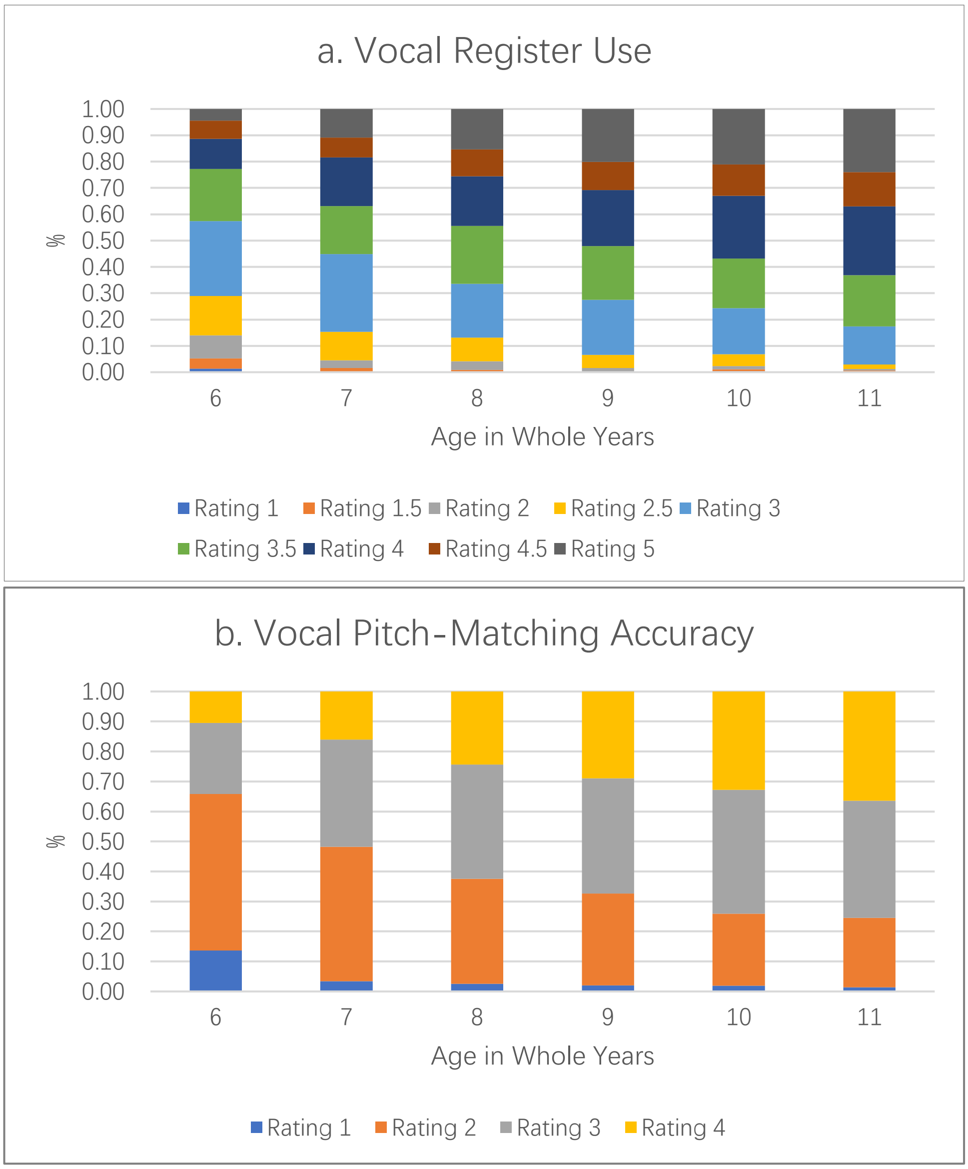
Figure 6.10 Vocal Register Use and Vocal Pitch-Matching Accuracy for the Three Target Songs and Children's Ages (in Whole Years) based on All Singing Performances Provided by Chinese Participants



Results based on all singing performances were similar to those based on the dataset of participants tested once only, either in 2017-2018 or 2018-2019. As Figure 6.11 shows, the percentage of singing performances gaining a relatively high score of vocal register use (rating ≥ 3 on the *SVDM* scale) increased from 71% at age 6 to 97% at age 11. The percentage of singing performances receiving a relatively high score of vocal pitch-matching accuracy (rating ≥ 3 on the *VPMD* scale) increased from 34% at age 6+ to 75% at age 11.

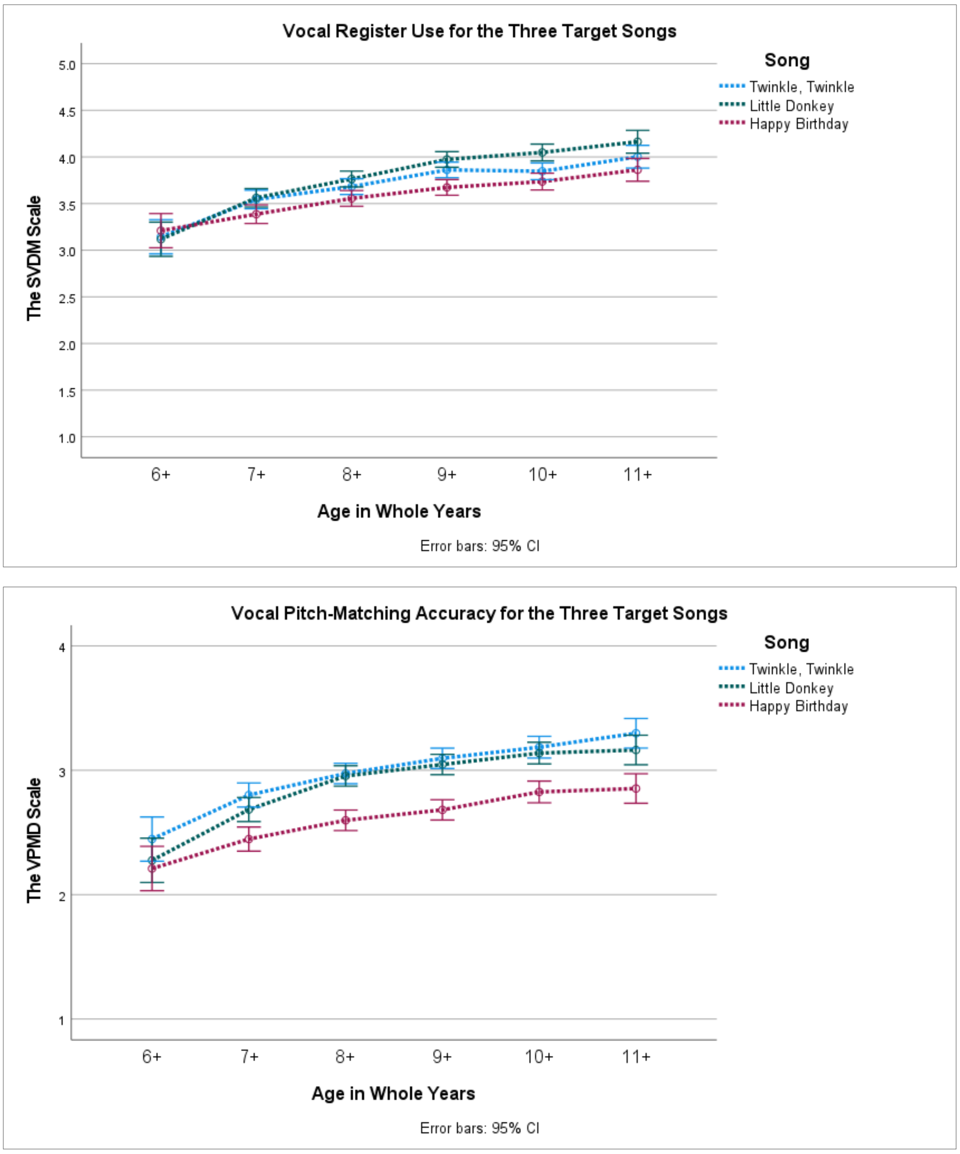
The percentage of singing performances receiving low scores for vocal register use for the three target songs (rating < 3 on the *SVDM* scale) decreased from 29% at age 6 to 3% at age 7. The percentage of singing performances matched poorly (rating < 3 on the *VPMD* scale) decreased from 66% at age 6 to 25% at age 11.

Figure 6.11 Stacked Bars Showing Development of (a) Vocal Register Use and (b) Vocal Pitch-Matching Accuracy for the Three Target Songs with Age in Whole Years for All Singing Performances Provided by Chinese Participants



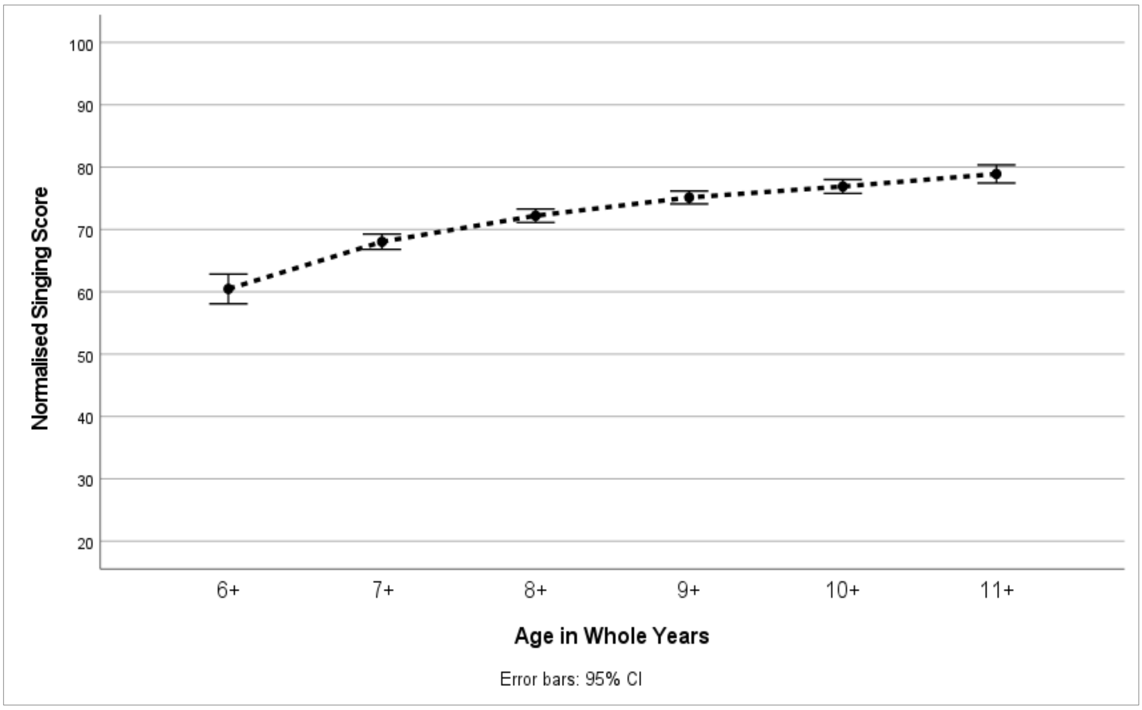
The age and song interaction effect was not statistically significant for vocal register use $F(10, 4599) = 1.25$, $p = .255$, and vocal pitch-matching accuracy $F(10, 4599) = 0.74$, $p = .684$. This suggests that the improvement tendency of vocal register use and vocal pitch-matching accuracy for the three target songs with increasing age was not changed by the diverse difficulties presented by the three target songs (see Figure 6.12). As the figure shows, participants across ages used similar vocal registers to sing each of the three target songs.

Figure 6.12 Development of Vocal Register Use and Vocal Pitch-Matching Accuracy for the Three Target Songs by Age in Whole Years and Songs Based on the Dataset Including All Singing Performances of Participants



In terms of vocal pitch-matching accuracy, there was a tendency that participants across ages matched *Happy Birthday* with the poorest accuracy, while they showed similar vocal pitch-matching accuracy when singing *Twinkle, Twinkle,* and *Little Donkey* (see Figure 6.12).

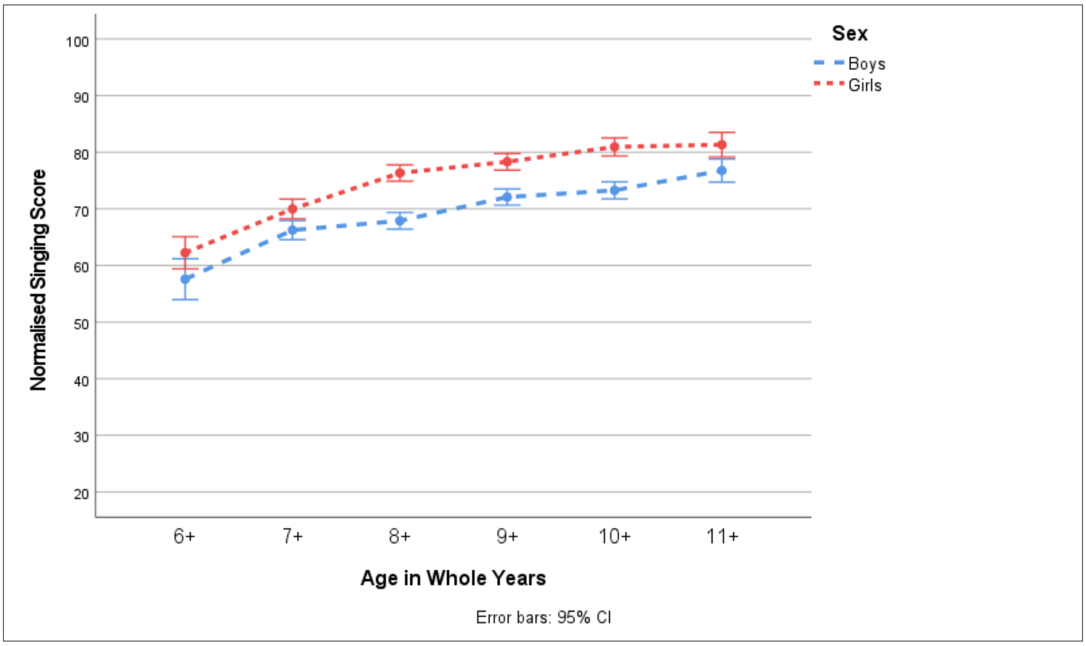
Figure 6.13 Normalised Singing Scores and Children's Age (in Whole Years) based on All Singing Performances Provided by Chinese Participants



A Pearson correlation coefficient was conducted to explore the relationship between vocal register use and vocal pitch-matching accuracy for the three target songs. There was a strong positive correlation between the two variables (r = .831, n = 4617, p < .001) based on all singing performances provided by Chinese participant children. Consequently, in the subsequent data analyses scores for the two above singing perspectives were combined and normalised out of 100 based on means of scores measured by the two existing scales. Participant children's normalised singing scores (NSS) improved statistically significantly with increasing age F(5, 4611) = 60.34, p < .001 (see Figure 6.13).

There was also a significant impact of the age and sex interaction effect on the NSS F(5, 4605) = 2.34, p = .039. As Figure 6.14 shows, the sex difference by means in the NSS for younger participants (ages 6 and 7) was smaller than that for older participants (ages ≥ 8), although there was a tendency for the sex difference for the oldest age group (age 11) to then decrease slightly.

Figure 6.14 Normalised Singing Scores and Participant Children's Age (in Whole Years), Differentiated by Sex based on All Singing Performances Provided by Chinese Participants



Similarly, the age and geographic location interaction and the age and income interaction had statistically significant effects on the NSS, for age * geographic location $F(4, 4606) = 3.01$, $p = .017$ and age * income $F(5, 4605) = 9.05$, $p < .001$. As Figures 6.15 and 6.16 show, the means NSS were higher for older children. However, the mean NSS differences related to geographic location and income were smaller for older participant children. Moreover, the mean differences of the NSS related to income almost disappeared for the oldest participants (ages 10-11).

Furthermore, the age and school interaction effect was also statistically significant $F(23, 4583) = 5.34$, $p < .001$. As Figure 6.17 shows, the mean scores of NSS for aged 7 to 9 participants in School F were higher than those of their peers in the other five schools. However, the means of NSS for the oldest age group (age 11) across the six participating schools were similar.

The interaction effect between geographic location and income was explored on all Chinese participants' singing behaviour. The test showed that the interaction effect was not statistically significant $F(1, 1535) = 0.20$, $p = .652$ (see Figure 6.18). Rural participants tended to be less developed in their singing competency than their urban peers, irrespective of family income.

Figure 6.15 Normalised Singing Scores and Participant Children's Ages (in Whole Years), Differentiated by Geographic Location based on All Singing Performances Provided by Chinese Participants

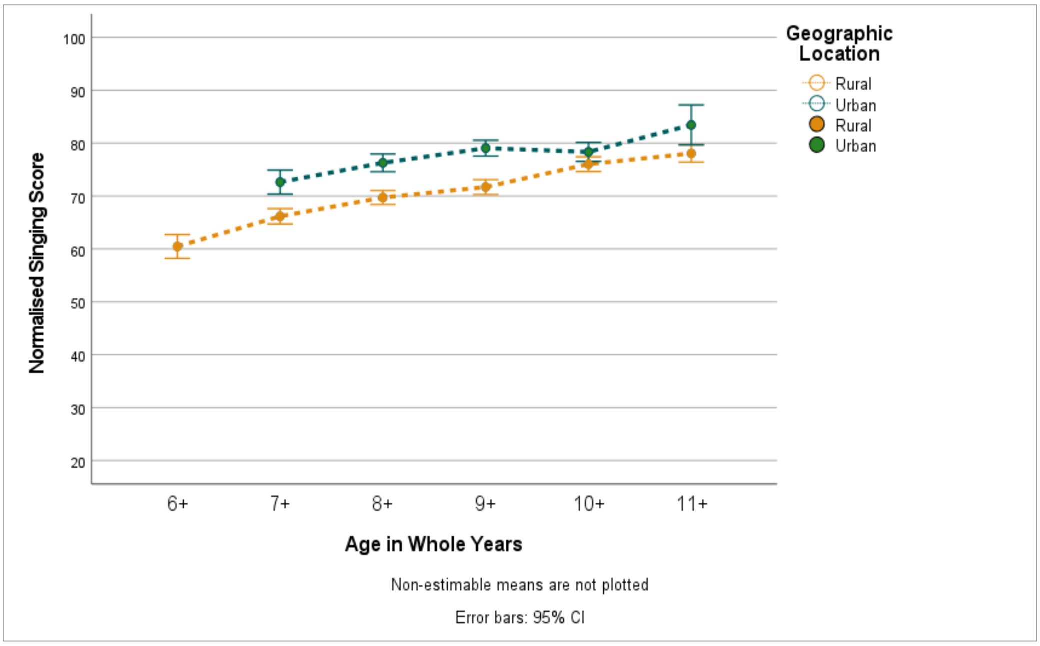


Figure 6.16 Normalised Singing Scores and Participant Children's Age (in Whole Years), Differentiated by Income based on All Singing Performances Provided by Chinese Participants

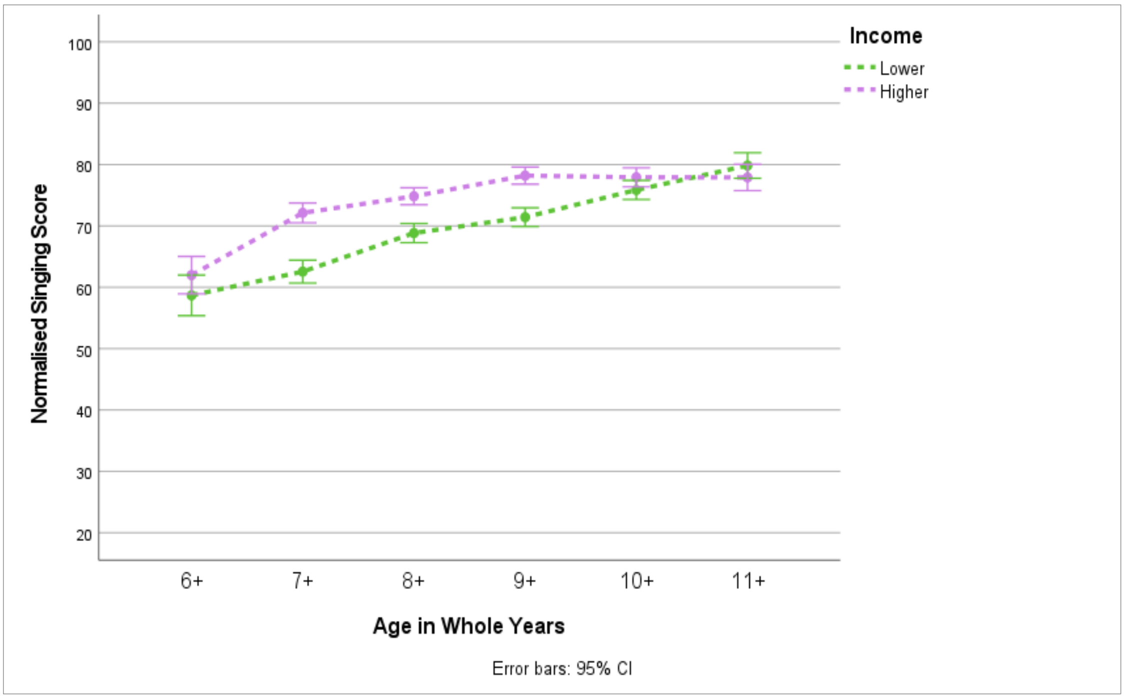


Figure 6.17 Normalised Singing Scores and Participant Children's Age (in Whole Years), Differentiated by School based on All Singing Performances Provided by Chinese Participants

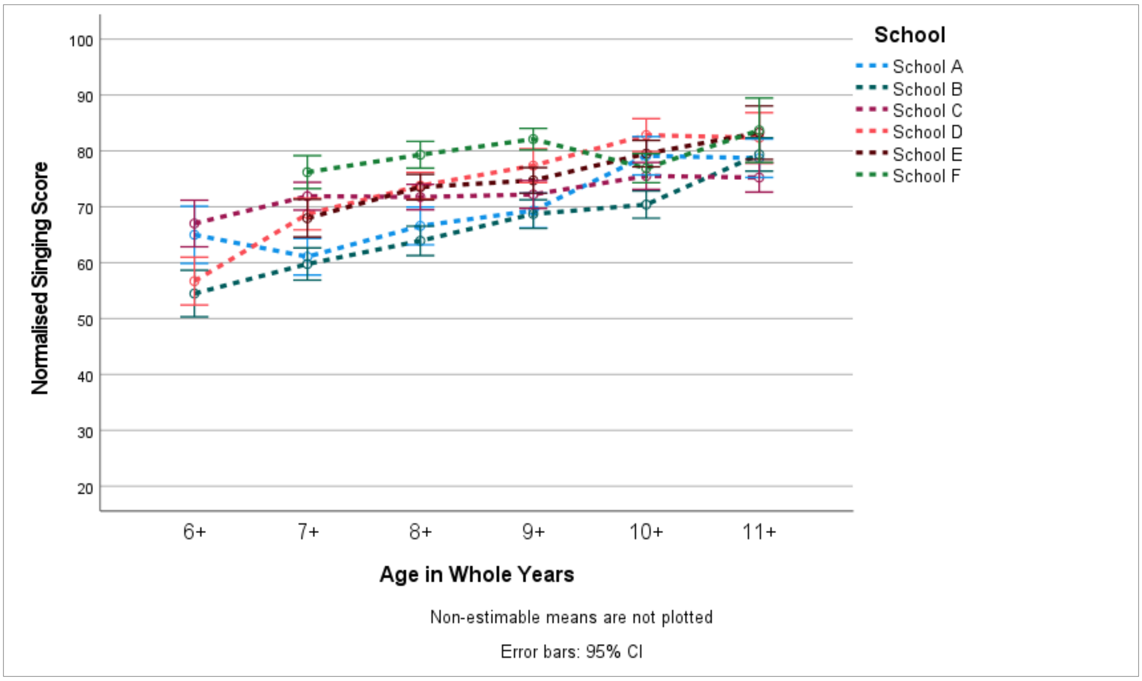
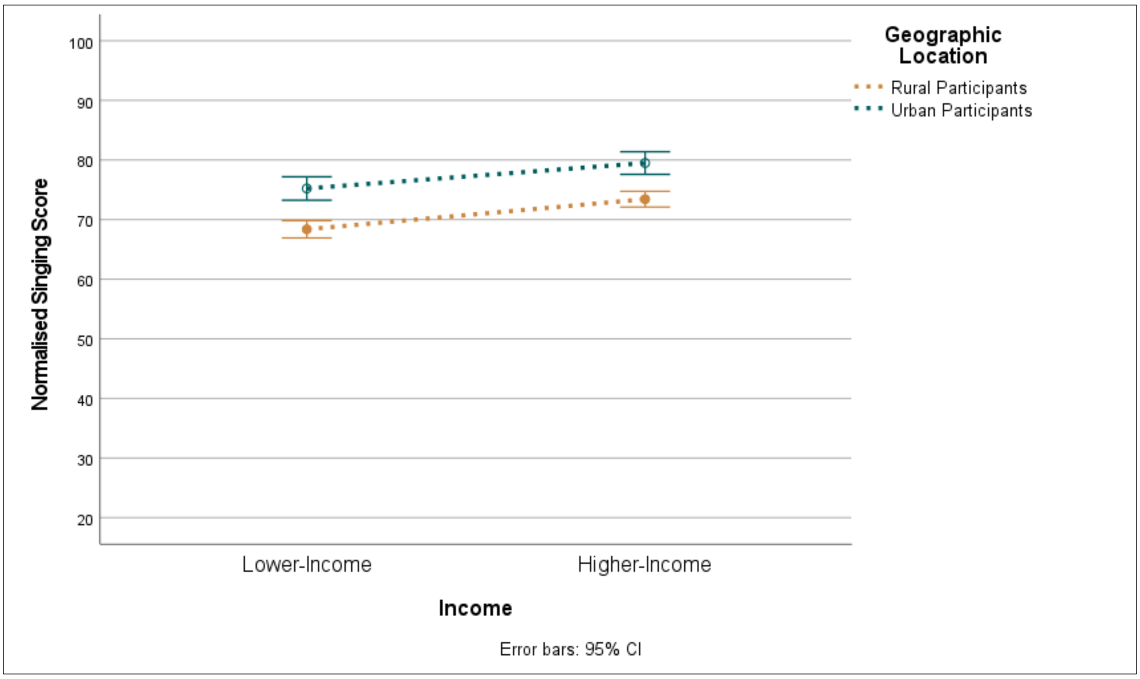


Figure 6.18 Normalised Singing Scores and Participant Children's Family Income, Differentiated by Geographic Location on All Singing Performances Provided by Chinese Participants



6.3 Comparison of Singing Data (Normalised Singing Scores) for Children Tested Once only in 2017-2018 or 2018-2019 (Section 6.1) and All Singing Data, Including Children Assessed Twice at Different Ages (Section 6.2)

The results of NSS based on children tested once only, either in 2017-2018 or in 2018-2019 were compared with those based on all singing data provided by Chinese participants, including children tested twice at different ages to check the reliability of the results based on the above second dataset (all Chinese singing performances). As Table 6.1 shows, the results of the NSS—based analyses of the two above datasets were essentially the same. This suggested that the results based on all singing performances (including children assessed twice at different ages) were similar to the results based on singing performances tested once only in 2017-2018 or 2018-2019.

Table 6.1 Comparing Results of the Normalised Singing Scores (NSS) between Participant Children Tested Once only, either in 2017-2018 or in 2018-2019, and All Chinese Singing Data, Including Children Tested Twice at Different Ages

Source	Children tested once only in 2017-2018 or 2018-2019	All Chinese Singing Data
NSS by age	NSS increased statistically significantly with age.	Ditto
NSS by age * Sex	NSS varied statistically significantly by age * sex. Girls > Boys. The sex difference was smaller for ages 6, 7, and 11, but it was greater for ages 8 to 10.	Ditto
NSS by age * Geographic location	NSS varied statistically significantly by age * geographic location. Urban > Rural. The location difference was smaller for ages 10 and 11 than for younger participants.	Ditto
NSS by age * Income	NSS varied statistically significantly by age * income. Higher-income> Lower-income. The income difference disappeared at age 11.	Ditto
NSS by age * School	NSS varied statistically significantly by age * school. The mean of the NSS was higher for ages 7 to 9 in School F than those of peers in the other five participating schools. However, the school difference was reduced for the older participants.	Ditto

6.4 Chinese Children's Developing Singing Behaviour Based on the Longitudinal Dataset

Out of a total of (N = 1,193) participants, n = 346 were tested twice to explore any longitudinal changes in their singing behaviour. Selected children were in Grade 2 and Grade 4 in 2017-2018, and subsequently in Grade 3 and Grade 5 in 2018-2019 (see Methodology Chapter, Section 4.2.1). In order to examine the information in more detail, the change in singing behaviour was analysed by age in whole years, instead of School Grade, as one Grade could include more than one age group. Singing performances for n = 3 participants who were aged 11 in 2017-2018 were not analysed because of the small numbers.

Raw scores of vocal register use and vocal pitch-matching accuracy were analysed initially. Then, the change of normalised singing scores (NSS) for each of the two singing behaviours was calculated and analysed.

Overall, means for vocal register use and vocal pitch-matching accuracy for the three target songs improved for longitudinal participants from 2017-2018 to 2018-2019 (see Figure 6.19). However, a paired samples t-test indicated that the mean improvement on vocal register use (M = 0.61, SD = 0.73) was statistically significantly greater than that for vocal pitch-matching ability (M = 0.35, SD = 0.71) $t(1031) = 11.63$, $p < .001$. For vocal register use, the improvement decreased statistically significantly with increasing age $F(3, 1031) = 6.11$, $p < .001$ (see Figure 6.19a). However, for vocal pitch-matching ability, the improvement did not vary statistically significantly with increasing age $F(3, 1028) = 0.83$, $p = .480$ (see Figure 6.19b). Overall, the changes in normalised singing scores across the two visiting years varied statistically significantly with age $F(3, 1028) = 2.81$, $p = .038$. As Figure 6.19c shows, the improvement in the normalised singing scores (NSS) was smaller for older participants than younger participants.

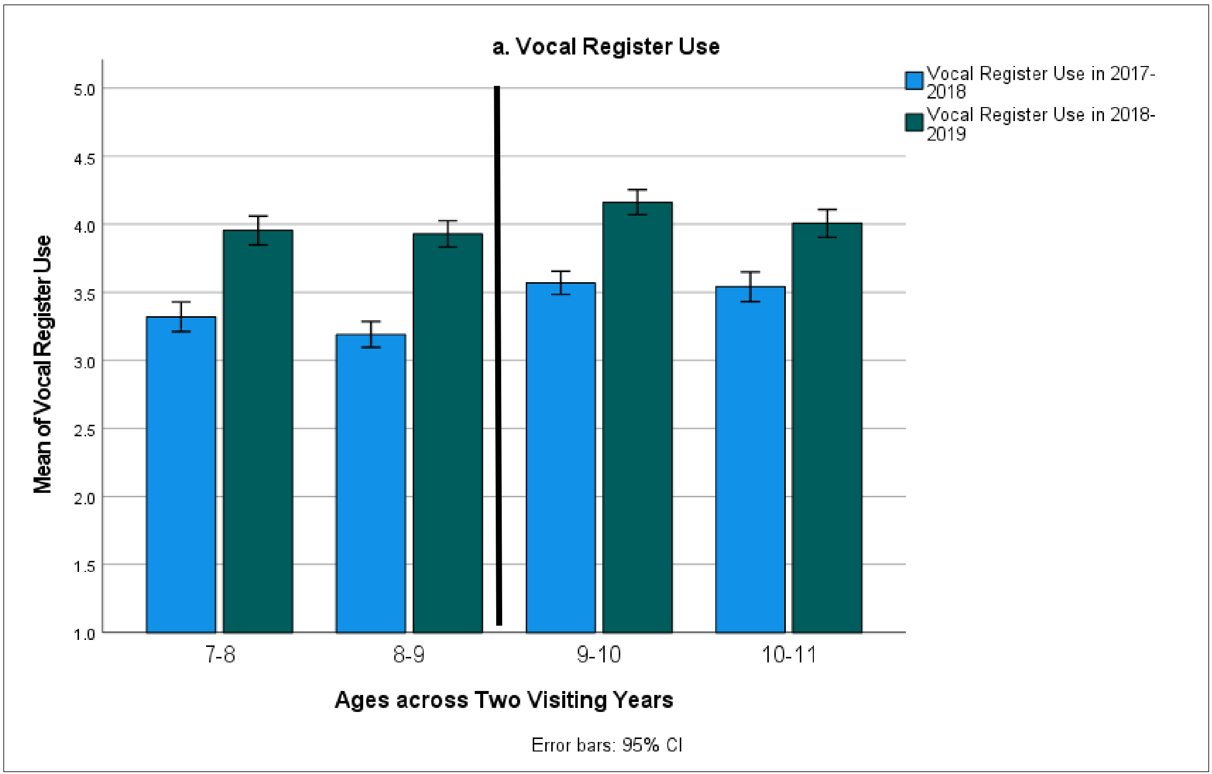
The interaction between age in whole years and the song was not statistically significant for vocal register use $F(6, 1023) = 0.92$, $p = .482$, nor vocal pitch-matching accuracy $F(6, 1020) = 0.30$, $p = .936$. This suggests that the change of both singing perspectives with increasing ages was not different significantly by the varied complexities of the three target songs. As Figure 6.20a shows, there was a tendency that younger participants had greater means of improvement on the vocal register use for the three target songs than older participants. In terms of vocal pitch-matching accuracy, the improvements for the three target songs for the oldest age group (ages 10-11) were generally smaller than those for the younger children (see Figure 6.20b).

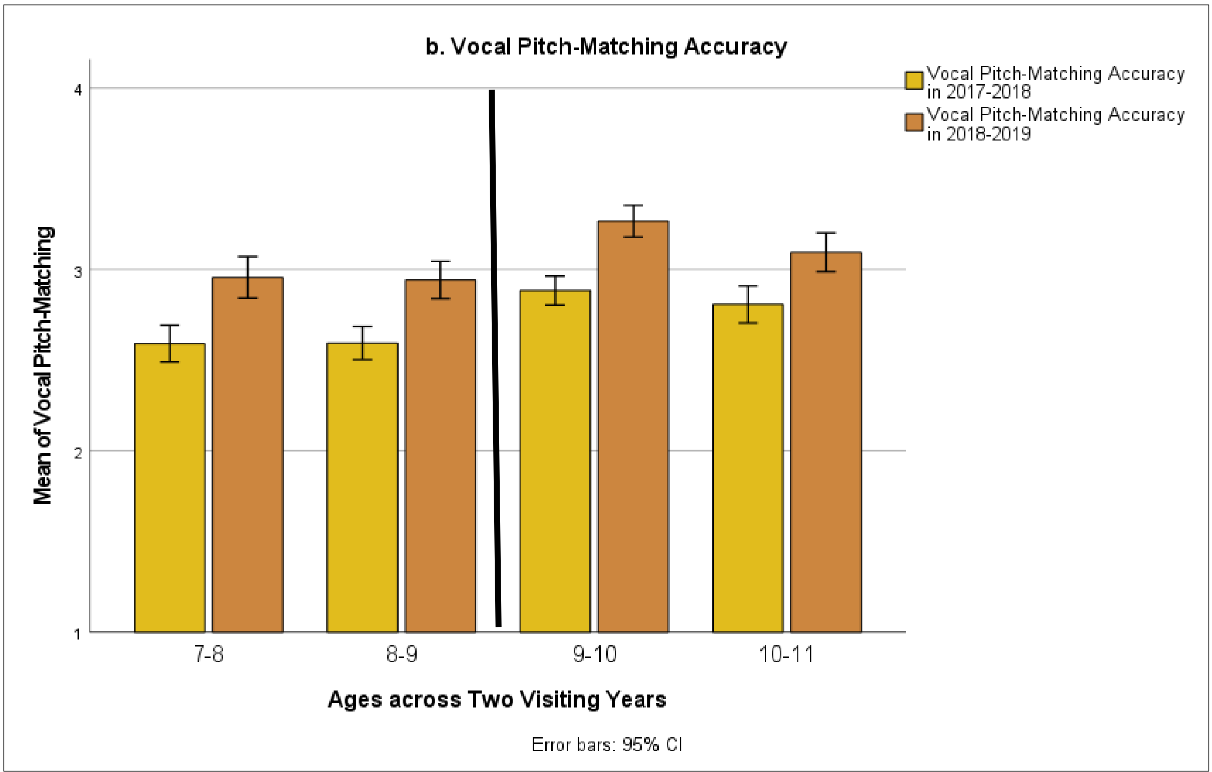
Then, mean changes of vocal register use and vocal pitch-matching accuracy for the three target songs were taken to offer a holistic perspective of participant children's complete change of singing behaviour across two visiting years. Overall, singing behaviour for each of the age groups in 2017-2018 showed improvement in 2018-2019 (see Figure 6.20c). However, the improvement was statistically significantly greater for younger participants than older children $F(3, 1031) = 2.84$, $p = .037$.

The interaction between age and sex did not reach a statistically significant effect on the changes in singing behaviour for participant children $F(3, 1027) = 2.15$, $p = .092$. Similarly, the other three interactions related to age had no statistically significant effect on the changes of the NSS, for participant children for age and geographic location $F(3, 1027) = 1.35$, $p = .256$; for age and income $F(3, 1027) = 0.24$, $p = .872$, and for age and school $F(15, 1011) = 1.01$, $p = .444$.

The interaction effect between geographic location and income was explored. The interaction effect was not statistically significant $F(1, 1034) = 1.36$, $p = .245$ (see Figure 6.21).

Figure 6.19 Change of Means for Vocal Register Use, Vocal Pitch-Matching Accuracy, Normalised Singing Score and Children's Age (in Whole Years)





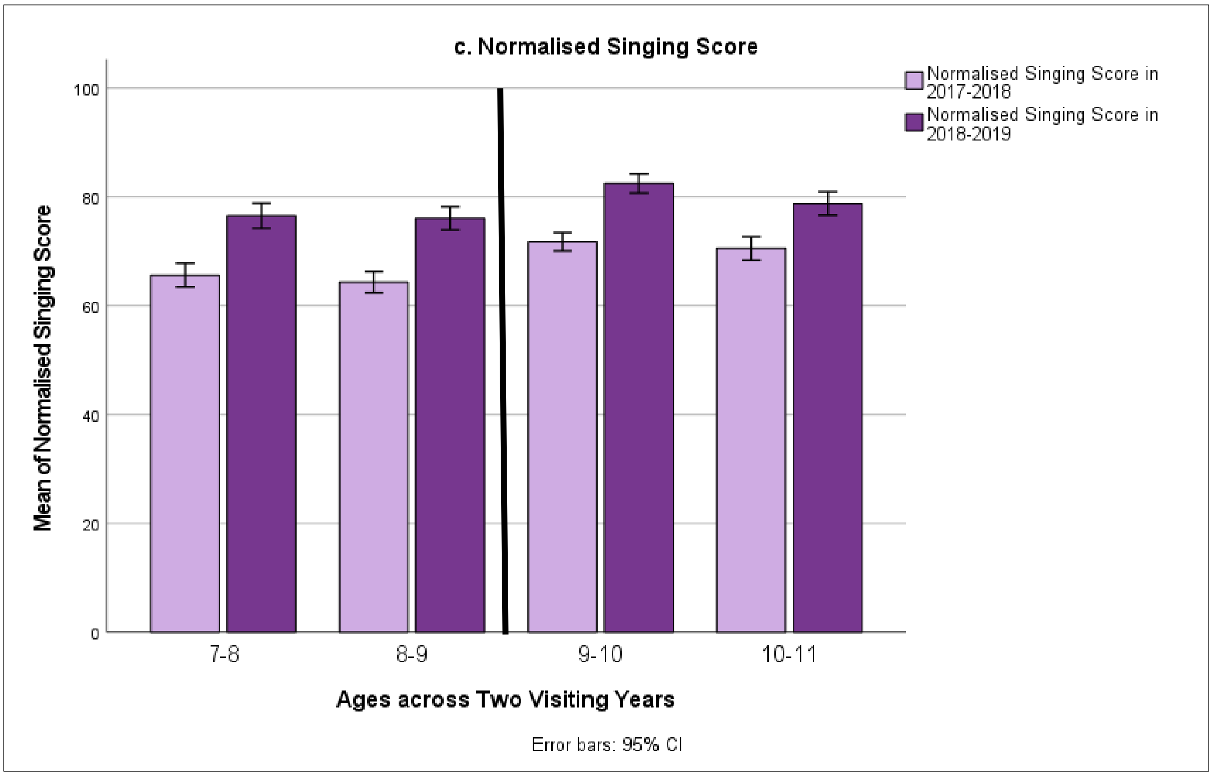
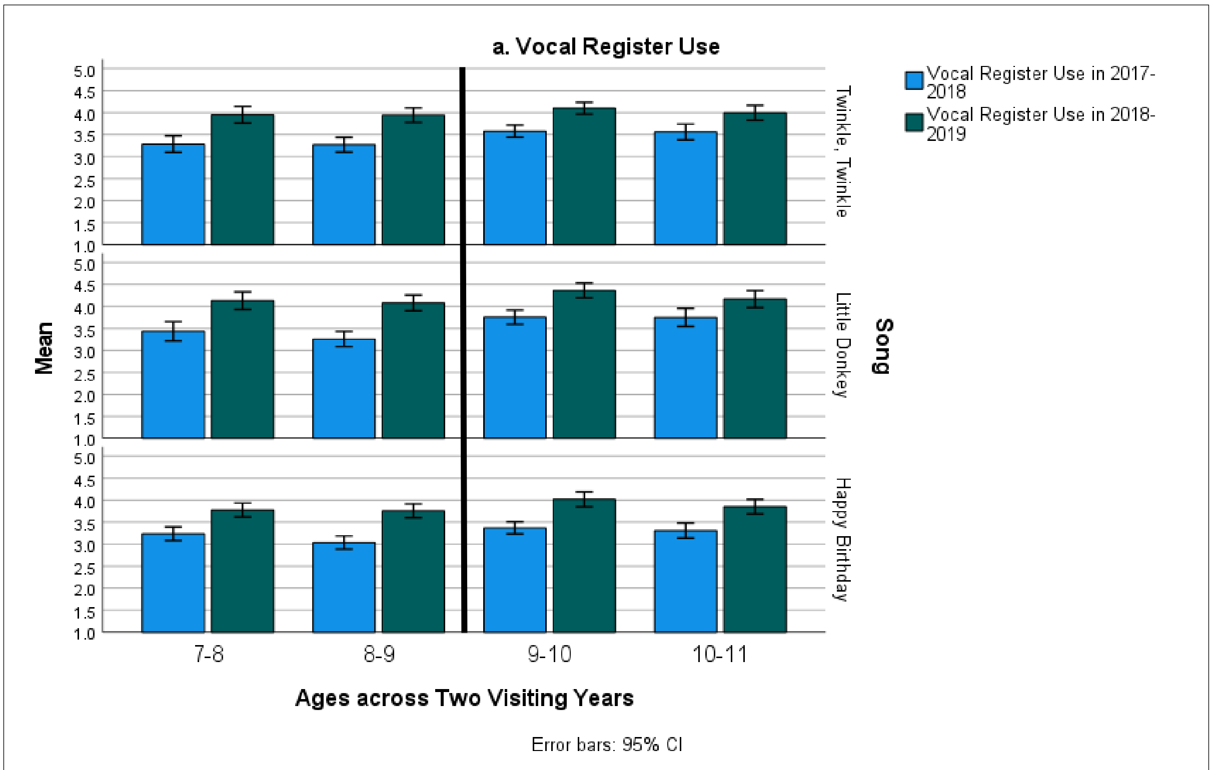
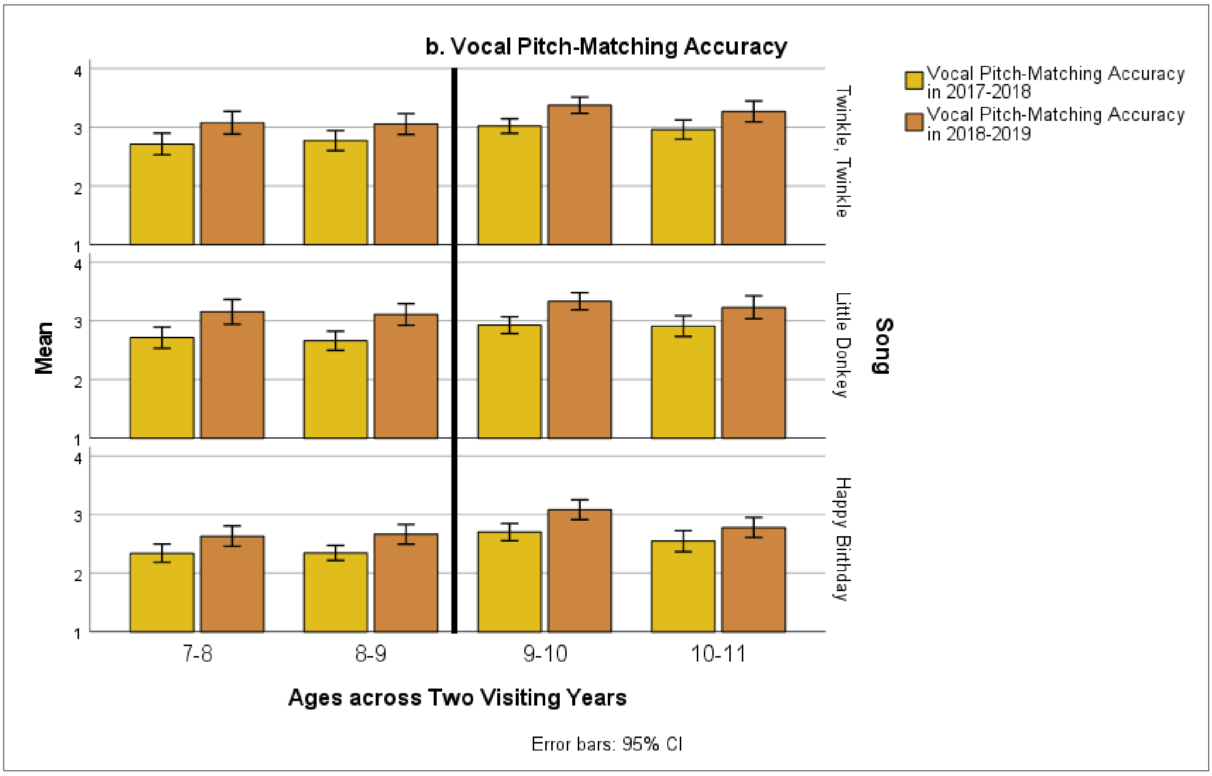


Figure 6.20 Change of Vocal Register Use, Vocal Pitch-Matching Accuracy, Normalised Singing Score and Age in Whole Years, Differentiated by Song




b. Vocal Pitch-Matching Accuracy
Vocal Pitch-Matching Accuracy in 2017-2018
Vocal Pitch-Matching Accuracy in 2018-2019
Mean
Twinkle, Twinkle
Little Donkey
Happy Birthday
Song
7-8
8-9
9-10
10-11
Ages across Two Visiting Years
Error bars: 95% CI

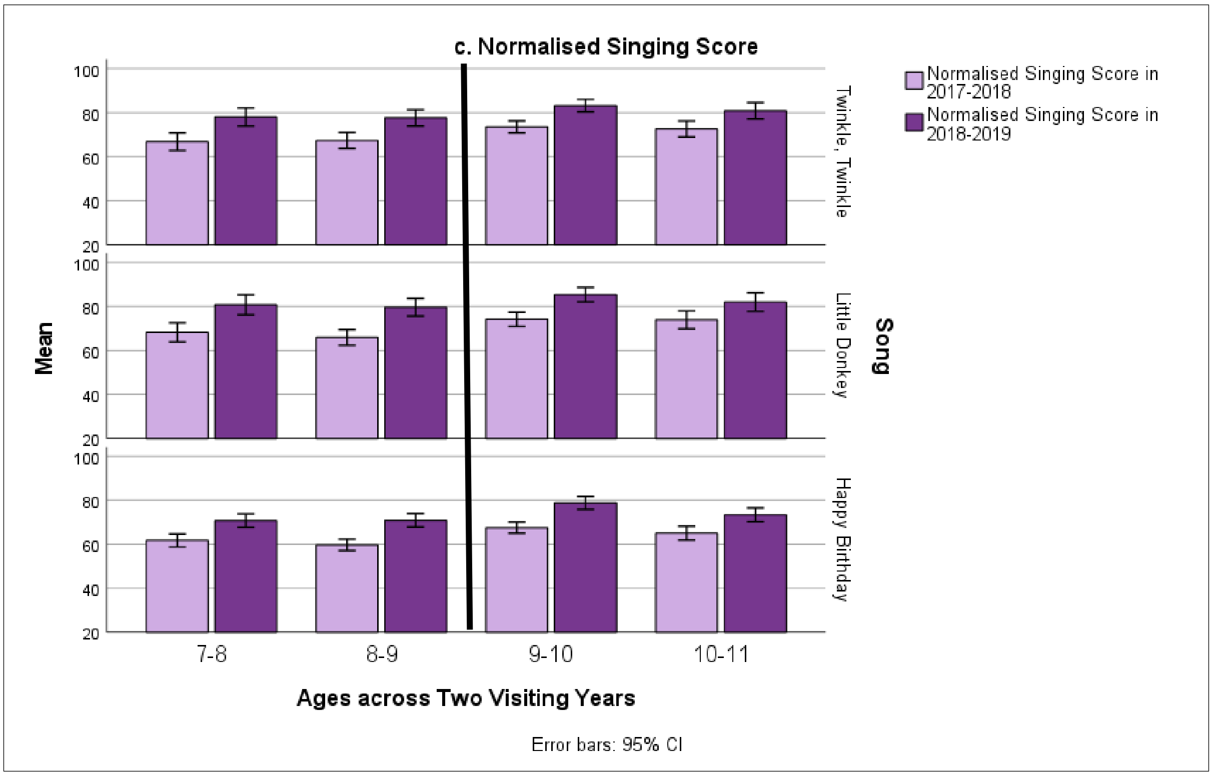
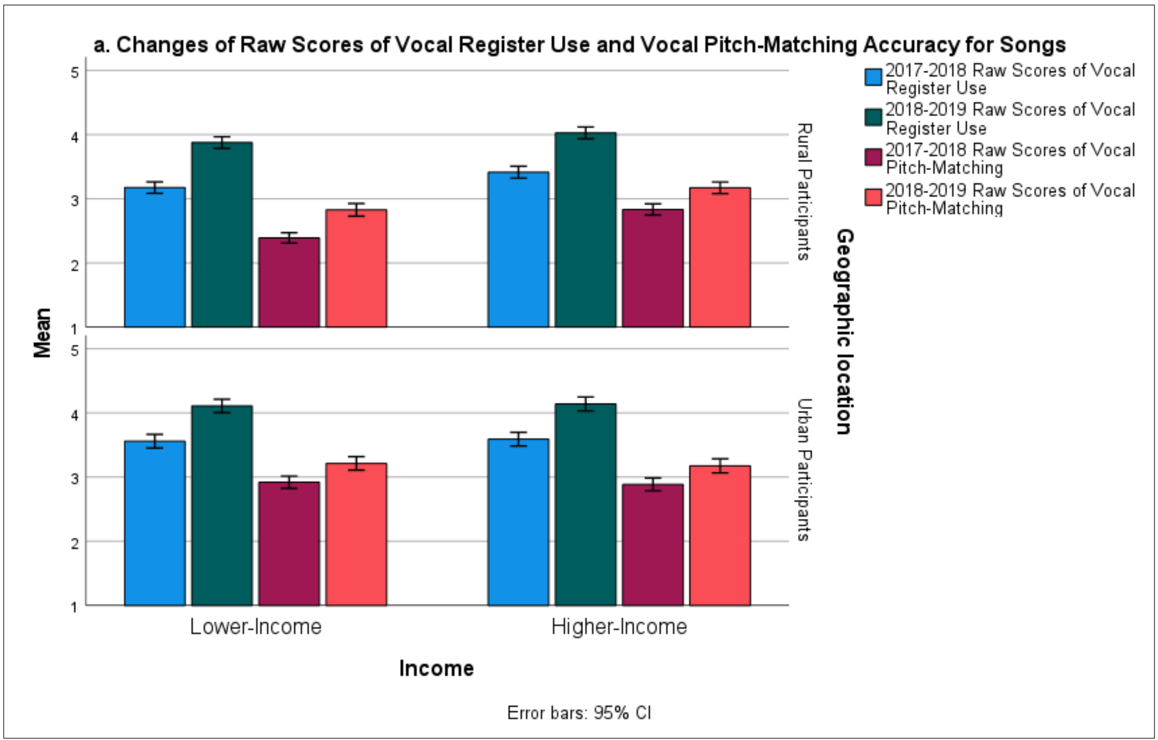
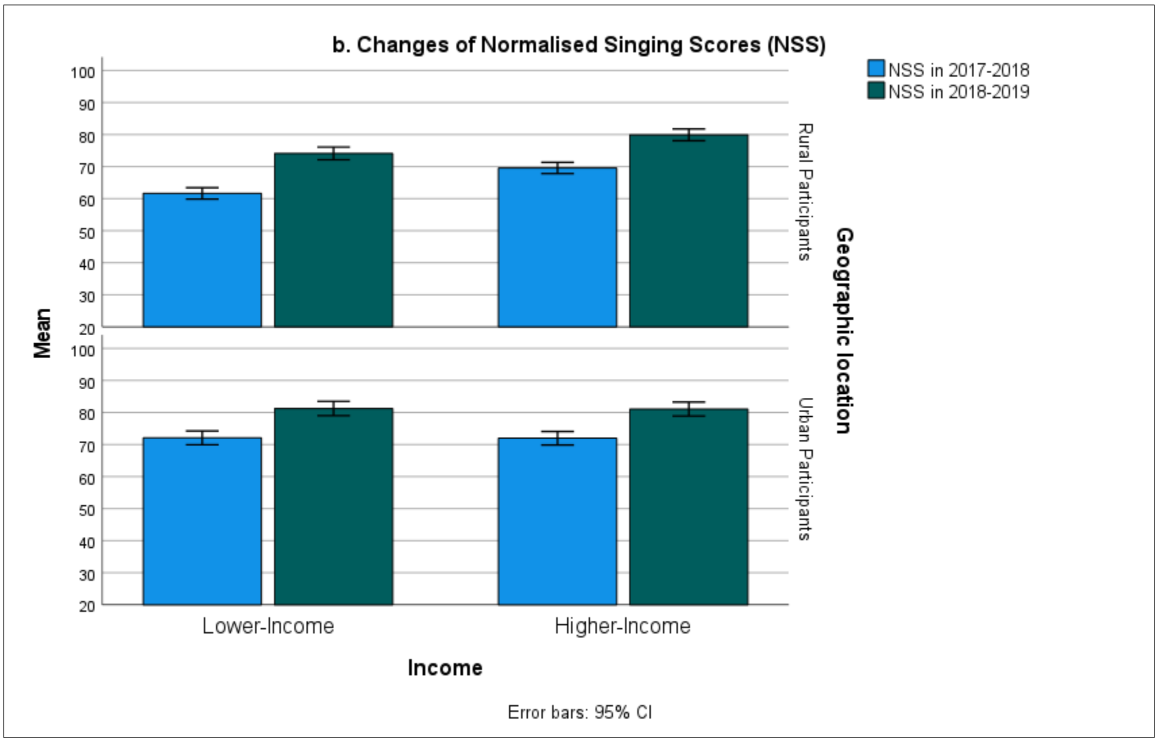

c. Normalised Singing Score
Normalised Singing Score in 2017-2018
Normalised Singing Score in 2018-2019
Mean
Twinkle, Twinkle
Little Donkey
Happy Birthday
Song
7-8
8-9
9-10
10-11
Ages across Two Visiting Years
Error bars: 95% CI

Figure 6.21 Change of Vocal Register Use, Vocal Pitch-Matching Accuracy, Normalised Singing Score and Inferred Family Income, Differentiated by Geographic Location





Note. The raw full score of vocal register use for songs was '5', which that of vocal pitch-matching accuracy for songs was '4'.

6.5 A Comparison of NSS for Chinese and English Children, Using Non-*Sing Up* Data from the UK

Data analyses in the current section are based on all the singing performances of Chinese participants (*N* = 1,539) collected by the current study and all the singing performances provided by English Primary school children without any *Sing Up* experience (*N* = 3,498) collected by the study of Welch et al. (2009b). These 'Non-Sing Up' children were treated as a control set in the English analyses to be compared to children of the same ages and backgrounds who had experience of the UK Government's National Singing Programme *Sing Up*, termed *Sing Up* participants in the original analyses.

In total, the two studies, Chinese and English, collected *N* = 5,037 singing performances from Primary school students who had no extra singing training other than in their regular music classes.

Both studies involved the same two target songs measured by the same rating scales. The comparisons of singing performances below were based on two shared target songs, *Twinkle, Twinkle* and *Happy Birthday*, measured by the *SVDM* scale (Rutkowski, 1996, 2015) and the *VPMD* scale (Welch, 1998) respectively for vocal register use and vocal pitch-matching accuracy for the two target songs, correspondingly. Normalised singing scores (NSS) are used to show collective singing behaviour. As Appendixes C and D illustrate, the singing scores of both Chinese and English participants were nearly normally distributed. Consequently, a series of parametric tests were undertaken to analyse the combined Chinese and English dataset.

Two Pearson coefficient correlations were conducted to explore the relationship between age in whole years and the NSS of all the Chinese singing performances and the relationship between age in a whole year and the NSS of all the English singing performances provided by English Primary school students without *Sing Up* experience. Both relationships were positively significant, for age and Chinese performances ($r = .252$, $n = 1{,}539$, $p < .001$), and for age and English performances ($r = .176$, $n = 3{,}498$, $p < .001$) (see Figure 6.22), suggesting that the NSS rose with increasing age for each group. Furthermore, as Figure 6.22 shows, the means NSS for each age group in the Primary school period were close between Chinese participants and English students without *Sing Up* training.

The dependent variable, the NSS was a continuous variable. A two-way ANOVA was conducted in order to explore the difference in the NSS by the increasing age during the Primary school phase between Chinese and English participants without *Sing Up* training. Overall, there was no statistically significant difference in the NSS by age in whole years between Chinese participants and English children without *Sing Up* experience $F(5, 5025) = 1.78$, $p = .112$.

Figure 6.22 Relationship between Age and the NSS of Each of the Groups

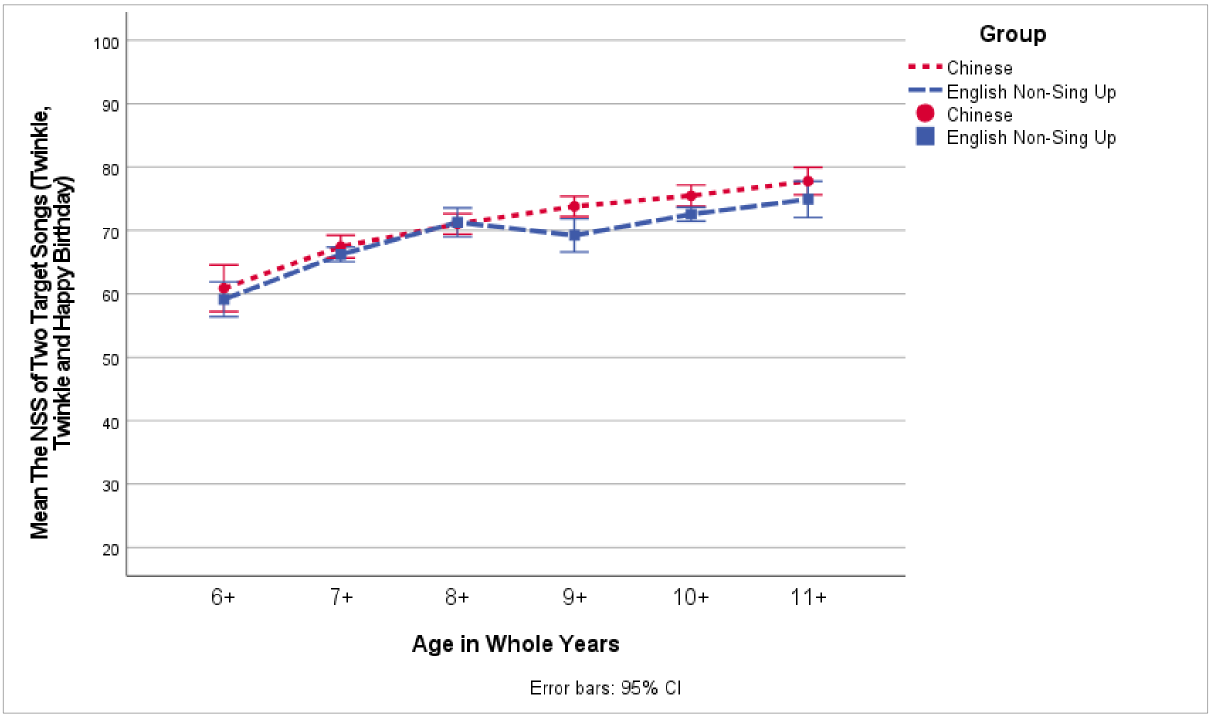
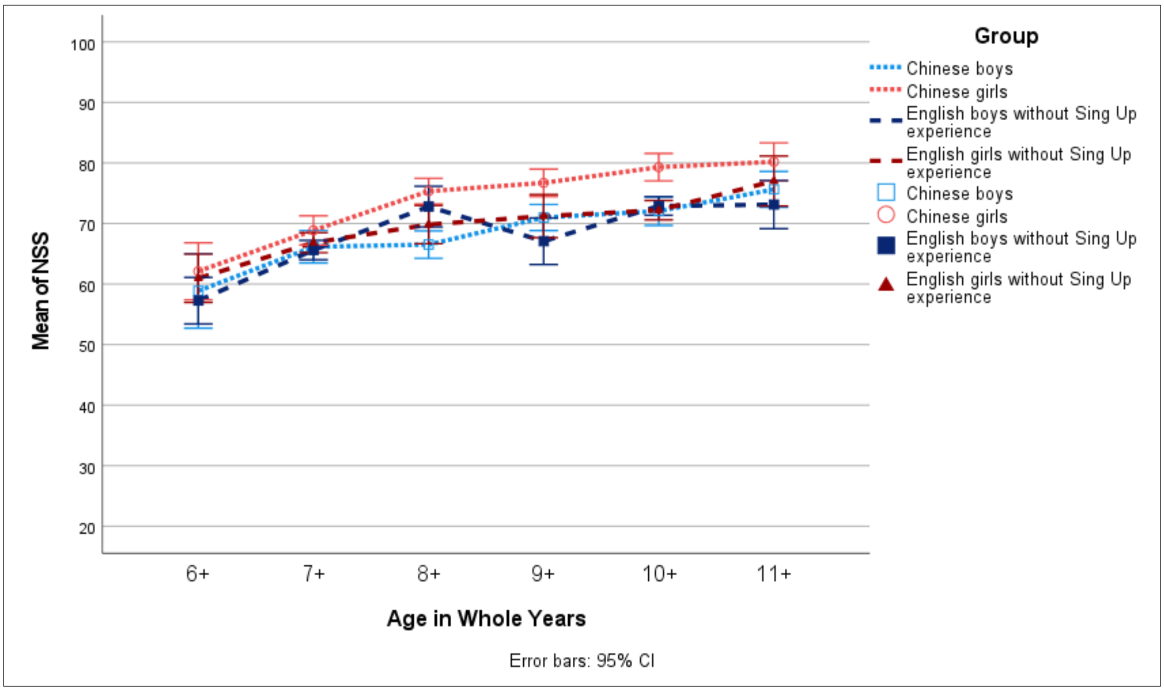


Figure 6.23 Normalised Singing Scores and Age in Whole Years, Differentiated by Nationality and Sex



When all children were divided into a younger (ages 6-8) and an older (ages 9-11) group, the age group and nationality interaction effect on the NSS was not statistically significant $F(1, 5033) = 0.82$, $p = .366$. These tests suggest that Primary school children from Chinese mainland and England shared similar development on the singing behaviour with increasing age.

A three-way ANOVA was conducted to compare the development of the NSS among Chinese boys and girls and English boys and girls without *Sing Up* training with increasing age during the Primary school period. The interaction effect of age in whole years*sex*country was not statistically significant $F(5, 5013) = 1.22$, $p = .299$. As Figure 6.23 suggests, the means NSS across ages in whole years were greater for girls than boys between Chinese participants and English students without *Sing Up* training.

Finally, singing performance data from all Chinese participants and English students without *Sing Up* experiences were combined to create an overview of singing behaviour for Primary school students across nationalities. A Pearson correlation coefficient was computed to assess the relationship between age clusters and means of the NSS. There was a positive correlation between the two variables (r = .195, n = 5037, p < .001). A one-way ANOVA was conducted to explore the development of the NSS changed by increasing age during the Primary school period. Overall, children's NSS increased statistically significantly with increasing age $F(5, 5031) = 45.43$, $p < .001$. As Figure 6.24 shows, the mean improvement for younger children (age ≤ 8) was greater than that for older children.

Finally, a two-way ANOVA was conducted to explore the change of the NSS with increasing age during the Primary school period between boys and girls. There was also a statistically significant difference by age and sex interaction effect $F(5, 5025)=2.65$, $p = .021$. As Figure 6.25 shows, the means of NSS were greater for girls than boys across ages. However, the mean sex difference for young participants (age ≤ 7) tended to be smaller than that for older children.

Figure 6.24 Normalised Singing Scores and Age in Whole Years based on Singing Performances of All Chinese Participants and English Students without Sing Up Experience

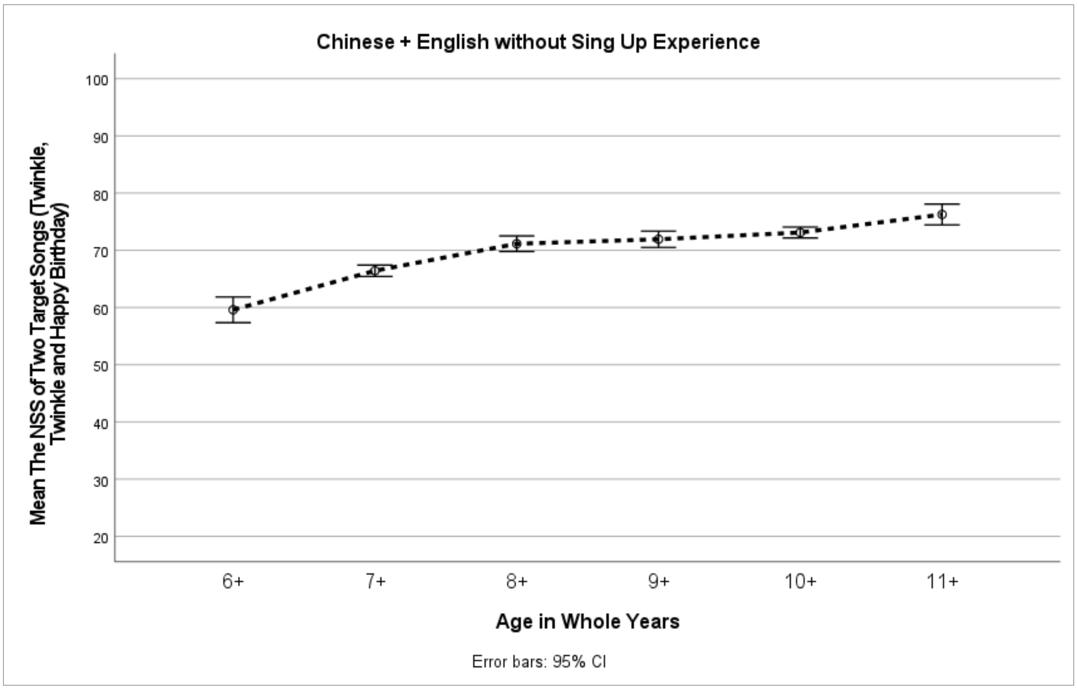
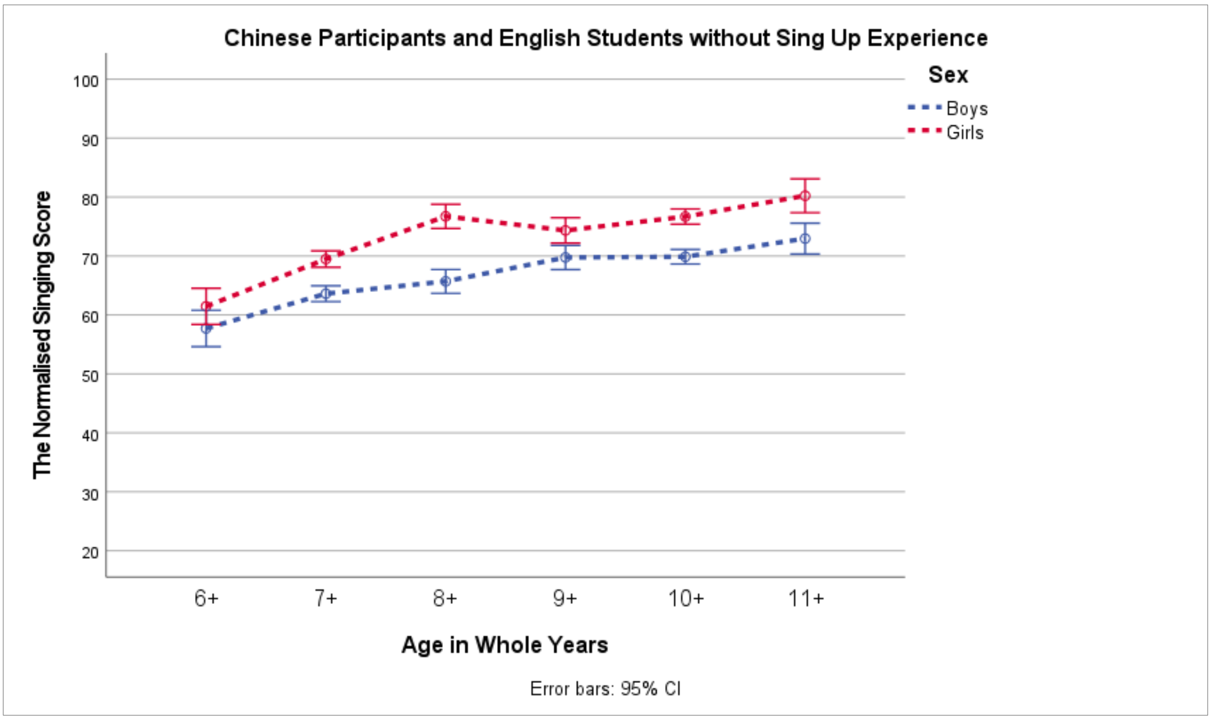


Figure 6.25 The Normalised Singing Scores Analysed by Age in Whole Years and Sex Based on Singing Performances of All Chinese Participants and English Students without Sing Up Experiences



6.6 Summary

There was a positive relationship between the application of vocal registers and vocal pitch-matching accuracy when singing the three target songs for Chinese children. This suggests that participants tended to match the pitches of the three target songs more accurately when they used more vocal registers. As participants' vocal register use and vocal pitch-matching accuracy for the three target songs increased statistically significantly with growing age in whole years, their singing behaviour as assessed by the two rating scales also increased significantly with growing age. When participants were six years old, around 70% of them could use the vocal registers equal or above 'limited range singer' (rating > 3 on the *SVDM* scale, Rutkowski, 1996), and only around 35% of participants could generally match the target songs well (rating > 3 on the *VPMD* scale). However, for participants who were 11 years old, nearly all of them (97%) were 'limited range singers' or above, while 75% of them could generally match the target songs accurately. 25% of participants still vocally matched pitches of the target songs poorly (rating < 3 on the *VPMD* scale).

Participants used similar vocal register to sing the three target songs in different age groups. In terms of vocal pitch-matching, they showed similar vocal pitch-matching accuracy when matching *Twinkle, Twinkle* and *Little Donkey*, across varied age groups. However, their vocal pitch-matching accuracy for *Happy Birthday* across age groups was poorer than that of two above songs.

Furthermore, the influence of sex, geographic location, and income is lower on the singing behaviour of the youngest and oldest Primary school participants. When participants were six to seven years old, the NSS difference related to sex or income difference was relatively small. For the middle age band of participants in Primary schools, girls, urban participants, and participants from higher-income families tended to have better singing competence. However, their singing ability developed slowly, while that of boys, rural participants, and participants from lower-income families continued to develop, which reduced the gap in the NSS between them and their peers. The interaction between geographic location and inferred family income had no statistically significant effect on participants' singing behaviour.

According to data collected from the longitudinal dataset, participants' vocal register use and vocal pitch-matching accuracy for the three target songs generally improved from 2017-2018 to

2018-2019. However, the growth rate for vocal register use was significantly greater than that for vocal pitch-matching accuracy when singing the three target songs. Furthermore, the growth rate for younger participants was greater than that for older participants. Overall, the improvement in singing ability was greater for younger than older participants and did not vary with sex, geographic location, or parents' income.

When comparing the NSS between Chinese participants and those English students without *Sing Up* experiences, the NSS for students from both groups increased statistically significantly with growing age. Furthermore, the means of the NSS for each age group across the two countries were similar. In both countries, the means of the NSS were greater for girls than boys from 6 to 11 years old.

Combining the data from the two countries shows that children's NSS improved statistically significantly with increasing age, especially for children aged ≤ 8. The means of the NSS were greater for girls than boys over time, but the sex difference was smaller when children were aged ≤ 7 years.

Chapter Seven

Discussion

The current chapter is organised into four sections. Each section discusses a contextual influence drawn from one aspect of Bronfenbrenner's ecological theory as applied to children's singing behaviour. It was noted that the microsystem of Bronfenbrenner's (1977, 1979) theory focuses on the interaction between a developing person and a setting that has an immediate influence on their developing person's behaviour. Their conceptualized mesosystem seeks to examines the influence of the interaction between two or more settings at a particular point in time on the developing person. The exosystem does not influence the developing person directly, but influences them by shaping the context of settings in the mesosystem. The macrosystem is related to a particular culture or sub-culture for the three above systems (see more details in Section 3.5.2). Furthermore, any interaction effects related to two different systems are analysed in a section of a later system to better present the interaction effects.

Using the theoretical constructs in the current study, the microsystem (Section 7.1) focuses on the influence of research design (including the selection of songs and measurements) and connected findings related to bio-development (age and sex) of Primary school children's singing behaviour. Families and schools are seen as central elements of the microsystem (Ettekal & Mahoney, 2017). The mesosytem (Section 7.2) discusses the effect of any interactions between two above elements of the microsystem on children's singing behaviour. The exosystem (Section 7.3) argues the impact of socioeconomic status (defined by geographic location and assumed family income) and the likely related interactions with the two former systems on children's singing behaviour. The macrosystem (Section 7.4) compares singing behaviour between Chinese and English Primary school children from the three following perspectives: bio-development, socioeconomic status and culture. It seems that the potential influence of the chronosystem may have only limited influence on children's singing behaviour in terms of the current study as the longitudinal aspects are short (two successive years of data collection) and the possible influences of any changes in macro-policy on these children's singing is unlikely to be evidenced in such a short period, although it is recognised as a feature of curricular policy related to singing in schools; consequently, it is not discussed further in the current chapter.

7.1 The Influence of the Microsystem on Children's Singing Behaviour

According to Bronfenbrenner's Ecological Theory, the microsystem includes elements that have a direct impact on children's development, such as families and schools. In terms of the current study and locating the new empirical data within the available literature, the current microsystem analysis focuses in particular on the influence of the choice of research design (including song task and its measurement) and of how this relates to what counts as bio-development in Primary school children's singing behaviour. To understand better participants' general singing behaviour, ratings of vocal register use and vocal pitch-matching accuracy for songs are also compared.

Although school has been a common element influencing children's development included in the microsystem (e.g., Bronfenbrenner, 1979), there is also the potential influence on development arising from a school's geographic location and variation in parental income which were allowed for as variables in the current study.

7.1.1 The Influence of the Research Design—Song on Children's Singing Behaviour

This section discusses the influence of choice of song in the current study on what counts as Primary school children's singing behaviour, including vocal register use and vocal pitch-matching accuracy. It starts by (a) distinguishing the influence of varied songs' complexities on vocal register use and vocal pitch-matching accuracy, followed by (b) discussing the difference in participants' singing behaviour between matching *Little Donkey* and matching other two target songs. Then, vocal pitch-matching of four musical elements during singing of the three target songs is noted from (c) the selection of a starting pitch, (d) accuracy of melodic contour, (e) consistency of key centres, to (f) accuracy of musical intervals are discussed.

(a) The General Influence of Song Choice on Primary School Children's Singing Behaviour

In the current study, the data analyses suggest that individual participants – evident in each age group – tended to use similar vocal registers to sing the three target songs. In terms of pitch matching, they tended to match two of the three songs (*Twinkle, Twinkle* and *Little Donkey*) more accurately than the third (*Happy Birthday*) (see Chapter 6, Figure 6.3). The hypothesised reasons for this are as follows. Although participants' potential vocal range (both comfortable range and maximum range) tends to become extended with age and experience due to the physical development of the voice source (see Chapter 2, Section 2.5.1) and its psychological management, individuals might not yet be able to control the voice source well, which requires cooperation between the vocal folds and other parts of vocal mechanism (see Chapter 2, Section 2.2.2). The data differences between two of the selected songs and the thirds imply that the varied internal musical complexity of the three target songs may have more impact on Primary school children's vocal pitch-matching accuracy than on their vocal register use.

(b) Differences in Children's Singing Behaviours between Matching *Little Donkey* and Other Two Target Songs

Participants' singing performance for the Chinese song *Little Donkey* is specifically compared with their performances for the other two target songs which have been evidenced more internationally in research into children's singing. *Little Donkey* is the only original Chinese children's song among the three target songs used in the current study, and its inclusion may help to shed light on the influence of local culture on children's singing behaviour (see more details about culture in Section 7.4).

It might be helpful at this point to review the degree of musical complexity of *Little Donkey.* Its internal complexity is considered to be around a medium level (see Chapter 3, Table 3.3 for the analyses of the complexity of songs), as 87.04% of its musical intervals are from unison to major 3rd, and 12.96% of them are a perfect 4th or 5th. Although the vocal range of both *Little Donkey* and *Happy Birthday* is an octave, the melody of *Little Donkey* includes a greater ratio of smaller musical intervals than those of *Happy Birthday* (see Chapter 4, Table 4.8). Furthermore, the melody of *Happy Birthday* includes two large musical intervals, a minor 6th and an octave. These make the internal musical complexity of *Little Donkey* less than that of *Happy Birthday*. Compared with the melody of *Twinkle, Twinkle*, the vocal range of *Little Donkey* is wider. However, both these target songs include mainly small musical intervals (see more details of the three target songs in Chapter 4, Section 4.2.2).

According to Figure 6.3 in Chapter 6, data analyses suggest that participants across the age range in this study generally used similar vocal registers to sing *Little Donkey* and *Happy Birthday*. However, there was a tendency for older children (age ≥ 8) to use more vocal registers to sing *Little Donkey* than *Happy Birthday* (see standard errors in Figure 6.3). This suggests that the large musical intervals contained within *Happy Birthday* (the minor 6th and the octave) did not necessarily encourage these participants to use more vocal registers. On the contrary, it is hypothesised that a series of upward small musical intervals in *Little Donkey* may have helped participants to expand their vocal range and, by implication, to use more vocal registers. Overall, it is suggested that these findings illustrate the potential influence of size and combination of musical intervals on children's vocal register use for a target song.

Participants generally matched *Little Donkey* and *Twinkle, Twinkle* with similar levels of vocal pitch-matching accuracies (see Chapter 6, Figure 6.3). This might be because the internal musical complexity of the two target songs is similar, in which both songs mainly included small musical intervals. Although the vocal range of *Little Donkey* is wider, this challenge seems to be reduced by the combination of the small musical intervals in its melodies. Generally, these findings suggest that the size and relative ratio of musical intervals can have an important impact in children's vocal pitch matching of a song.

Overall, it would seem that a study using these three particular target songs—*Twinkle, Twinkle*, *Little Donkey*, and *Happy Birthday* – was more likely to generate a slightly higher overall mean score of singing behaviour compared to a study that used *Twinkle, Twinkle* and *Happy Birthday* only. This is because of a generally higher rating in vocal register use and vocal pitch-matching accuracy for *Little Donkey* and *Twinkle,* Twinkle together will bias the resultant data positively when compared to *Happy Birthday*. Consequently, this implies that any comparison of singing behaviour data between these Chinese participants and English children without *Sing Up* training (Welch et al., 2009b) should be based on the scores of the two songs which they both sang, namely *Twinkle, Twinkle* and *Happy Birthday* only (see Section 7.4).

On the other hand, according to the results reported in Chapter 5, vocal pitch-matching accuracy across the target songs shared similar characteristics. For instance, the current study found that the number of key centres used to sing one target song had a positive correlation with those used to sing another target song.

Overall, when considering the subset of participants for whom more detailed note-by-note vocal pitch matching analyses were undertaken (Chapter 5), four common vocal pitch-matching errors were evidenced (see c to f below).

(c) Fitness for the Self-Selection of a Starting Pitch

The current study found that most participant children started with a key that was one or two semitones sharper than the most appropriate key for themselves, as judged by the overall key centre that emerged as their singing products unfolded (Figure 5.3). Similarly, Flowers and Dunne-Sousa (1990) had earlier reported that two third of *n* = 93 kindergarten children did not start with the target key when asked to sing a target song. Consequently, although Leighton and Lamont (2006) used 'the first stable pitch sung by the child' (p. 319) to define a standard key centre for each child, it seems that identifying the key centre used most when singing a song may be more appropriate measure. This could be achieved by applying the *MAPM* system (see Chapter 4, Figure 4.6), although it is recognised that this would be a practical challenge in the mainstream classroom. However, from a pedagogical perspective, the current data suggest that the likely variation in the development of children's sung key centres should be explored actively by vocal exercises across several different key centres in order to enable children to become more aware of tonality and sense of key.

(d) Discussing Vocal Pitch-Matching Accuracy for Melodic Contour of a Song

The current study reports that most participants could accurately match the melodic contour of the three target songs (see the percentage of participants across ages whose vocal pitch-matching scores were equal to or higher than rating 2 in Chapter 6, Figure 6.2, and also see the content of the *VPMD* rating scale in Chapter 3, Figure 3.4). This is in line with previous studies (e.g., Flowers & Dunne-Sousa, 1990; Moore et al., 1995/1996; Welch, 1994; Welch et al., 2009a). While some previous studies generated their findings by using musical intervals or melodic patterns, the current study reports this based on the use of three target songs. This illustrates a shared developmental characteristic when children seek to match musical elements in songs.

On the other hand, the current study found that participants were more likely to sing out-of-tune in continuously ascending or descending melodies (see Chapter 5, Section 5.4.5). This was reported earlier by Sinor (1984) in the finding that kindergarten children showed poorer accuracy for simple ascending and descending contours. The current findings suggest that micro-adjustments of the vocal tuning mechanism are less developed if transitioning across vocal registers

Furthermore, it was found that participant children in the current study usually repeated their pitch errors when matching identical melodic phrases within a song. This is in line with the observation of Feng et al. (2013) in regular music classes. This might be because of the muscle memory of the vocal mechanism as well as the related vocal pitch schema (Welch, 1985).

(e) Discussing the Consistency of Key Centres during Singing a Song Singing

The current study reports that participants used around a mean of four to five key centres when asked to sing one target song (see Chapter 5, Section 5.2). This was in line with the report of Flowers and Dunne-Sousa (1990) who found that most kindergarten children were either somewhat modulating (one to three key centres) or more modulating (more than three key centres) when asked to sing a self-selected song. However, the measurement of key centres in the current study differs from that used by Flowers and Dunne-Sousa's (1990). Flowers and Dunne-Sousa (1990) divided key centres into three categories, including 'more than three modulations', 'somewhat modulating', and 'none or one modulation' (p. 105). The method of analyses chosen in the current study could not only visually show the number of modulations used to sing a target song, but also show the length of a particular modulation, the type (sharp or flat), and difference between modulations compared with a target. This was achieved by applying the colour characteristics of the *MAPM* system in the current study, when two neighbouring pitches were noted with different colours if they were identified as being matched with different key centres.

(f) Discussing Vocal Pitch-Matching Accuracy for Different Types of Musical Intervals

The current study found that vocal pitch-matching accuracy for different musical intervals generally decreased with the increasing size of musical intervals (see Chapter 5, Section 5.3). This is in line with the results reported by previous studies (e.g., Chen-Hafteck, 1999; Moore et al., 1995/1996; Tillotson, 1972). However, compared with these previous studies, the current study found that a decreasing tendency for vocal pitch-matching accuracy was closer to a linear trend when the size of musical intervals increased (see Chapter 5, Figure 5.6). This linear trend was not influenced by different age groups, sexes, nor socioeconomic backgrounds. The more linear trend in the current study might be because of being able to draw on a large dataset, in which vocal pitch-matching accuracy for N = 15,708 musical intervals was analysed from three target songs sung by n = 132[76] quasi-randomly selected participants. Furthermore, most types of musical interval included varied

[76] Singing data of n = 2 participants were not included in the analysis of vocal pitch-matching accuracy for different types of musical intervals, as a part of their singing performances was in a spoken manner.

pitch heights among the three target songs (see Chapter 4, Figure 4.5). For instance, there were *n* = 29 major 2nds with diverse pitch levels across the three target songs.

When looking at the details for each type of musical interval in the current study, unison was seen to be the easiest to match compared with the other eight types of musical intervals. The current finding is supported by similar results reported by Moore et al. (1995/1996) and Feng et al. (2013). The study of Moore et al. (1995/1996) involved children aged 6 and 9 from four Western countries, while that of Feng et al. (2013) included aged 3 to 6 kindergarten children from Jiangsu Province, China. That the same results are evidenced across varied age groups and cultures suggests that unison is a relatively easy interval for children to match.

In terms of vocal pitch-matching accuracy for a minor 2nd in the current study, this interval was seen as more difficult to be matched accurately than a major 2nd, but was easier to be matched accurately than a minor 3rd. This finding is partly supported by Sinor (1984) and Hedden and Baker (2010). Sinor's (1984) study involved *N* = 96 American children aged 36 to 71 months who were asked to match 12 patterns that were randomly selected from 48 possible patterns. Each pattern involved four pitches. She found that half-steps (semitones) did not have a significant influence on the overall vocal pitch-matching accuracy. Similarly, according to a later study conducted by Hedden and Baker (2010), vocal pitch-matching accuracy for a minor 2nd was similar to that of unison, a major 2nd and a minor 3rd. Their study was based on *N* = 26 Second Grade children in America. The children were asked to sing a target song, *America*. However, only the first nine pitches of the song were analysed due to a large amount of data. These pitches were measured by a software programme MultiSpeech applied for acoustical analysis and also two judges' perceptual judgement. The fact that no significant difference between the four types of musical intervals was found in their study might be because these musical intervals were all relatively small and easy to match accurately (*cf.* vocal pitch-matching accuracy for different types of musical intervals as reported in Chapter 5, Figure 5.6 in the current study).

Similarly, the vocal pitch-matching accuracy for a minor 2nd reported in the current study seems also to be partly supported by studies of Moore et al. (1995/1996) and Feng et al. (2013), who reported that children generally matched a minor 2nd with poor vocal pitch-matching accuracy.

Moore et al. (1995/1996) reported that a minor 2nd was matched more poorly than most of larger musical intervals (except an octave), such as a major 3rd, a perfect 4th and a perfect 5th. The study of Moore et al. (1995/1996) included N = 480 children aged 6 and 9 from Argentina, Poland, Spain, and the USA, who were asked to echo 16 musical intervals, including two minor 2nd (Eb4-D4, G#4-A4). Vocal pitch-matching accuracy for the musical intervals was measured by two music teachers with 20 years of musical experience. The current study included 6 pairs of minor 2nds across the three target songs, measured by the *MAPM* system (see Chapter 4, Section 4.2.3c).

In Feng et al.'s (2013) study, the children were N = 130 kindergarten children (ages 3-6) who were asked to sing two specially composed songs. The authors reported that children matched a minor 2nd poorly compared to most other types of musical intervals, such as a major 6th. Although the vocal pitch-matching accuracies for the minor 2nd across the two target songs in Feng et al.'s (2013) study were relatively consistent, vocal pitch-matching accuracies for other types of musical intervals were not. For instance, children across different age groups matched a perfect 4th of the first target song with the poorest accuracy (accuracy rate: 0%-9.5%) among nine types of musical intervals, but they showed a good matching accuracy (accuracy rate: 58.1%-84.4%) for another perfect 4th in the second target song. It seems that other elements, such as pitches' height, might have a great impact on children's vocal pitch-matching accuracy for musical intervals.

It can be hypothesised that a minor second requires a subtle adjustment of the vocal mechanism (e.g., Titze, 1994) at the voice source and that this needs to be experienced perceptually as well as an appropriate vocal gesture. Being able to hear a minor second perceptually, does not imply that someone will be equally skilled in adjusting their voice to match this (Yank Porter, 1977).

A major 2nd was the second easiest interval to match in the current study, maybe because it was combined with unison within a melody. Although a major 2nd was the second largest group of musical intervals (n = 29) in the current study, most of these involved unisons in front of or behind (see Chapter 4, Figure 4.4). It is assumed that the high vocal pitch-matching accuracy for unisons might have contributed to that for a major 2nd. However, the assumption requires to be studied further.

In terms of a minor 3rd, it was not the easiest musical interval to match accurately in the current study, as participants matched it poorer than unisons, minor and major 2nds. However, Jones (1979) and Sinor (1984) agreed that it was the easiest type of musical interval for children to match accurately, particularly a falling minor third. The results of the two above studies were based on echoing musical intervals or patterns, rather than songs. Furthermore, the singing materials of both studies involved a limited number of unisons, although they involved a range of minor and major 2nds. As the results of the current study are based on singing three target songs, it is possible that the different results were obtained because of the different singing tasks (musical elements or songs) used between studies.

An octave was the hardest type of musical interval to match accurately in the current study. This is in line with the findings of previous studies (e.g., Feng et al., 2013; Moore et al., 1995/1996; Nichols et al., 2023; Tillotson, 1972). The poor matching accuracy for an octave is likely to be related to a limited current skill in transferring between vocal registers. According to Welch (1986, 2022 – see Chapter 2, Section 2.2.1), children can become out-of-tune when a melody reaches a certain pitch height because they do not adjust their singing register, but rather continue with the previous vocal gesture of trying to raise pitch solely by the action of increased tension in the *vocalis* muscles within the vocal folds rather than relaxing these muscles and letting the cricothyroid muscles assume more responsibility for stretching the vocal folds to achieve higher pitches.

In addition, the current study found that many participant children matched the first pitch of an ascending large musical interval sharp when matching a target song (see Chapter 5, Section 5.4.3), which has not been reported before. This pitch error might be because participants other-than-consciously prime the vocal mechanism to match a large ascending musical interval. They generally then matched the second higher pitch of the musical interval flat (see Chapter 5, Figure 5.14), which might be because of limited air support and/or a not yet developed transfer skill from a chest to a head register.

Lastly, a series of ascending musical intervals was found by the current study to help participant children to match a higher pitch more accurately (see Chapter 5, Section 5.4.4). This could be compared with generally poor vocal pitch-matching accuracy for a high pitch when included in a big upward musical interval reported above. It might be because a series of upward small intervals allowed children to adjust their vocal mechanism and air support slowly, priming the vocal system for upward steps and thus sing the target pitches more accurately.

Overall, the current study suggested that Primary school children matched different types of musical intervals with varied accuracy. Most of findings on vocal pitch-matching accuracy for different musical intervals agreed with previous studies, while some of them were newly found.

7.1.2 The Influence of Research Design—Measurement on Children's Singing Behaviour

(a) Reliability of the *SVDM* and *VPMD* Scales

The *SVDM* and *VPMD* scales had been used together in some previous studies, such the research by Mang (2006) – the first combined use of the two scales – and Welch et al. (2009b), with the aim to measure vocal register use and vocal pitch-matching accuracy for songs, respectively. Consequently, these two scales have been used in the current study. Each rating of the two scales represents a characteristic of a singing performance (see Chapter 3, Figures 3.3 and 3.4). Judges are expected to measure participants' singing performances against the content of the two rating scales.

The current study found that scores measured by the *SVDM* and *VPMD* scales were positively correlated at a high level (see Chapter 6, Sections 6.1 and 6.2), which is supported by the earlier findings from Mang (2006). This positive correlation between the two rating scales might be because of a shared concept— singing range between vocal register use and vocal pitch-matching accuracy for songs. In more detail, Rutkowski's *SVDM* scale (1996, 2015, 2018) distinguishes different vocal register uses by employing the concepts of vocal range, pitch height, and lift points. Meanwhile, when a participant child vocally matches the pitches of a target song accurately, it seems logical that they tend to have covered the vocal range of the song accurately as well.

However, some difficulties also arise when applying the two ratings. Two musical elements of Rutkowski' *SVDM* scale (1996, 2015, 2018) – lift points and vocal range – increase the difficulty of judging vocal registers' application. During the process of assessment for the singing performances, the current author found that it was hard to match the suggested lift points (e.g., D4 and A4) suggested in the *SVDM* scale for each individual participant with the obtained data. Instead, it emerged in that these hypothesised lift points with their associated vocal ranges were usually extended slightly at the individual participant level. Consequently, judgment based on the *SVDM* scale was relatively subjective. The *VPMD* scale, as a four-point rating scale, also shared the same psychological assessment feature. For instance, judges might be uncertain which one of two neighbouring ratings to choose, because of an unclear demarcation between the pair.

However, it should be noted that these drawbacks of the two above rating scales tended to be reduced by measuring singing performances multiple times. In the current study, all singing performances were measured three times. The final singing data were based on the third judgement to increase the consistency of the current author's decisions and a sample of these judgements was independently correlated with two other judges.

Furthermore, information on lift points, vocal range, and timbre were usually used simultaneously to give a final score for the vocal registers' application for a song for each participant in the current study. For instance, when a participant sang a song from D4 to G4 (a perfect fourth), their score might be measured as '3' of the Rutkowski *SVDM* scale based on the combined information of vocal range and timbre in the current study. If their score was measured as '3.5', this suggests that only vocal range with suggested pitch heights was considered, because the rating '3.5' of the *SVDM* scale meant that the participant was an 'inconsistent initial range singer' whose singing travelled between 'limited range singer' (usually pitch range from D4 to F4) and 'initial range singer' (usually pitch range from D4 to A4) unstably. This might be due to an individual current difficulty of transferring between vocal registers. However, these pitch heights (e.g., D4, F4, and A4) suggested by Rutkowski (2015, 2018) might or might not be lift points for a particular individual. This variation can increase the uncertainty of the judgement. According to Rutkowski's (2019) latest explanation, her *SVDM* scale does not measure vocal range, but measures the use of vocal registers only. Consequently, it seems that timbre characteristics of a singing voice should be considered during the judgement process.

Additionally, when measuring older participants' (e.g., age 11+) singing performances, the lift points and vocal range have been updated in the current study, as the suggested lift points did not match the older participants' singing. These children, most often boys, were usually in the process of changing voice, but had not finished the voice change process. For instance, in the current study, if an older boy's speaking voice range was centred around G3 to A3, and he sang a target song from B3 to D4 with a different timbre, his score of vocal register use was more likely to be '3' of the *SVDM* scale ('limited range singer'), instead of '2' ('speaking ranger singer'). This is because his speaking voice was lower than that suggested by Rutkowski (2015, 2018), usually between A3 to C4. While Levinowitz et al. (1998) also suggested a relatively lower reliability of the *SVDM* scale when measuring 6th Graders' vocal register use, it is argued that the lift points for older Primary school children requires further research. This is because previous studies (e.g., Guerrini, 2006; Rutkowski, 2015; Rutkowski & Chen-Hafteck, 2001; Rutkowski et al., 2007; Rutkowski & Miller, 2003a, 2003b, 2003c) mostly used the *SVDM* scale to test vocal register use for younger children.

(b) Using the *MAPM* System to Illustrate Children's Singing Behaviour from a Micro Perspective

As reviewed in Chapter 2, Section 2.4, previous studies frequently used software to 'objectively' measure vocal pitch-matching accuracy for pitch patterns. While most of this software used a measure of mean fundamental frequency to analyse vocal pitch-matching accuracy, this type of data is not musical, but acoustic (Welch, 1994). While some previous studies (e.g., Goetze, 1985) calculated the deviation between vocally matched pitches and standard pitches, the absolute value of deviation was usually applied to avoid being offset during data analyses. Consequently, these absolute values could not show the type of vocal pitch-matching errors (sharp or flat). Although this drawback was addressed in the study of Elmer and Elmer (2000), their approach could not reveal the size of the vocal pitch-matching errors.

However, these two weaknesses of approaches used in previous studies were remedied by the *MAPM* system that was newly created in the current study. Firstly, by using semitone errors to show vocal pitch-matching errors, the data became more musical. The semitone, as a regular musical term, seems to be easier to understand than cents and fundamental frequency for ordinary people. Secondly, the first numbering approach from '-8' to '8' of the *MAPM* system can show the type of semitone error of each of the vocally matched pitches. In addition, different colours can visually present the size and type of semitone errors of each vocally matched pitch. This approach is not only appropriate for vocal pitch-matching of patterns, but is also suitable for the analysis of a song.

The *MAPM* system also shares the advantages of approaches used in previous studies. For instance, an advantage of calculating deviations between a vocally matched pitch and a target pitch by cents in Goetze's (1985) study was that it could show pitches matched relatively accurately. In the current study, this advantage could be illustrated by observing the colour characteristic and the numbering approach of the *MAPM* system in a vocally matched song for a participant. When pitches were marked with the same colour and number by the *MAPM* system for a target song for a participant, these pitches used the same tonality. While Flowers and Dunne-Sousa (1990) analysed the number of pitches and patterns that matched absolutely and relatively accurately, using a standard within +/-50 cents, and divided songs into three categories based on tonalities within or out of three, their research questions could also be explored by using the *MAPM* system. The *MAPM* system could even reveal details of more than three tonalities used by a participant when singing a song.

Furthermore, using the *MAPM* system also allows analysis of other characteristics of vocal pitch-matching accuracy for songs, such as consistency of tonalities applied to sing a song, the relative accuracy of vocally matched musical intervals during song singing, the location of common pitch-matching errors (see Chapter 5, Sections 5.2 to 5.4), and also to test reliability of scores of vocal pitch-matching accuracies for a song as measured by a rating scale, such as the *VPMD* scale (see Chapter 5, Section 5.1). These functions have not been explored by the software used in previous studies (e.g., a software-based measurement used by Nichols et al., 2023).

However, it should be admitted that the *MAPM* system shares a similar characteristic in application with the software packages used in previous studies, which is that it is time-consuming, especially during the processes of noting each matched pitch by using Praat and Sing&See and marking colours for each vocally matched pitch. However, as it can reveal rich details of vocal pitch-matching accuracy for a song, the effort seems worthwhile.

(c) Comparing the *MAPM* System and the *VPMD* Scale

The *MAPM* system and the *VPMD* scale were both used to measure participants' vocal pitch-matching accuracy for songs from micro and macro perspectives. Differences and correlation across the two measurements are discussed below.

As Chapter 3, Figure 3.4 shows, ratings of the *VPMD* scale are generally defined by degrees of an accurate melody contour, changes of tonality, and vocal range of a sung song against a target melody. Although these diverse musical concepts are used, the measurement process is still subjective –

although relatively time efficient when dealing with large numbers of vocal products. In the current study, when measuring *N* = 1,539 children's singing performances using the *VPMD* scale, all singing performances were judged three times to overcome the problem of uncertain demarcation between a pair of ratings, although the uncertainty diminished after judged all singing performances had been judged twice. Only data measured for the third time were applied in the current study because of a more consistent definition of each of the ratings of the *VPMD* scale.

On the other hand, the colour characteristics of the *MAPM* system can be seen as more objective. It can visually show a degree of sharpness and flatness of semitone errors for each matched pitch in a song. Vocal pitch-matching accuracy of a whole target song can be shown by an overall accurate rate for a matched song using a total score of each of the matched pitches based on a '1-9' numbering system divided by the full score of a target song (see more details in Chapter 4, Section 4.2.3c). It can also be shown by distribution and frequency of diverse tonalities, and the length of each of the sung tonalities for a vocally matched target song (see Chapter 5, Figure 5.3). Overall, using the colour system can provide more details about vocal pitch-matching accuracy for a song for each participant than using the *VPMD* rating scale.

However, due to the time-consuming nature of the collection process, using the *MAPM* system to collect details of vocal pitch-matching accuracy for songs seems to be more suitable for a relatively small sample size. In contrast, the *VPMD* scale seems to be more suitable for a large sample size that can help to show a general tendency for vocal pitch-matching accuracy for songs for a group of people.

Additionally, results based on the *MAPM* system can support those measured by the *VPMD* scale. Chapter 5, Section 5.1 shows that scores measured by the two approaches (the *VPMD* scale and the *MAPM* system) were statistically significantly positively correlated. As Figures 5.2 and 5.3 illustrate, for participants measured with a higher rating of the *VPMD* scale, their singing performances measured by the *MAPM* system generally had fewer changes in tonality and each of the tonalities tended to last longer. Furthermore, the colour characteristics of the *MAPM* system suggest that each rating of the *VPMD* scale was valid when measuring vocal pitch-matching accuracy for a song for Chinese Primary school children. It also suggests that judgement using the *VPMD* scale in the current study was reliable. Consequently, the *MAPM* system, as a more detailed, note-by-note measurement which reveals many details of vocal pitch-matching accuracy for a song, helps to improve the perceived reliability of judgement based on the *VPMD* rating scales, especially with a small sample size. It is assumed that the *MAPM* system could also be used to test the reliability of another rating scale designed to measure vocal pitch-matching accuracy for songs.

The study of Leighton and Lamont (2006) and that of Hedden and Baker (2010) used both relatively subjective (e.g., perceptual analysis, or a rating scale) and 'objective' (e.g., software) measurements to analyse vocal pitch-matching accuracy for a song. The current study not only retained these two types of measurements, but also developed a novel approach, using colour characteristics to show the type and degree of semitone error of each vocally matched pitch of a song. The current study also had more participants and vocally matched pitches compared with those of the two above previous studies.

7.1.3 The Influence of Bio-Development on Children's Singing Behaviour

(a) Discussing Vocal Register Use for Songs by Age

The current study found that participant children generally increased the number of vocal registers they used to sing each of the three target songs between the ages of 6 and 11 (see Chapter 6, Figures 6.1 and 6.9). The above findings of the current doctoral study were generally in line with the studies of Mang (2006), Welch et al. (2009a), and Pereira and Rodrigues (2019). The three songs applied in the current study were similar with those of Mang (2006) and Welch et al. (2009a), but they were different from the singing task applied in the study of Pereira and Rodrigues (2019). This suggests that the development of vocal register use for songs with increasing age during the Primary school period does not vary significantly by different target songs with diverse complexities and cultures.

The current study found that the percentage of participants whose mean score of vocal register use was ≥ 3 on the *SVDM* scale improved from 69% at age 6 to 97% at age 11. This suggests that almost all children at age 11 would be at or above the criteria for a 'limited range singer' (Rutkowski, 1996, p. 365), defined as using an initial singing range to sing a song. This is due to the natural physical development of vocal folds in increasing size with age (*cf.* Titze, 1994).

The mean scores of vocal register use in the current Chinese study for participants across school

Grades 1 to 6 were between 3 and 4 on the *SVDM* scale. These means were closer to those reported by Welch et al. (2009a) (M = 3.26 to 3.60) than to means reported by Mang (2006) (M = 4.15 to 4.43) and Pereira and Rodrigues (2019) (M = 3.6 to 4.55). It seems that the different mean scores were not because of cultural difference *per se*, but were likely the product of diverse degrees of singing training. Children from the current study and children from the baseline of the national study in Welch et al. (2009a) had no extra singing training, while peers in Mang (2006) and Pereira and Rodrigues (2019) may have received more related training about vocal register use. Another explanation could be because of different understandings of the *SVDM* scale. For instance, the 'limited range singers' might sing a song with a major 6th and its highest pitch was F4 in the current study and the study of Welch et al. (2009a), while they might be identified mainly by a vocally matched range—e.g., a major 6th in the studies of Mang (2006) and Pereira and Rodrigues (2019). These assumptions require further study.

(b) Discussing Vocal Pitch-Matching Accuracy for Songs by Age

As with the results of vocal register use for the target songs, vocal pitch-matching accuracy for participant children in the current study generally improved with increasing age. This finding is generally in line with those reported by many previous studies (e.g., Davies & Roberts, 1975; Gould, 1969; Green, 1994; Mizener, 1993; Pereira & Rodrigues, 2019; Welch et al., 2009a). Since this characteristic of vocal pitch-matching accuracy for songs has been demonstrated across diverse Western (e.g., America, Australia, Portugal, and the UK) and Asian (e.g., China, including Hong Kong) cultures, this suggests that the impact of age – including natural physical and psychological development – on children's basic singing development is likely to be greater than that of culture (such as language differences). It can be conjectured that the increasing vocal pitch-matching accuracy for songs with age is because of the characteristics of a developing vocal mechanism with growing age, such as a more stable vocal mechanism, expanded length and more flexible vocal folds and stronger air support, during the Primary school period (see Chapter 2, Section 2.2.1).

It should be noted that the study by Mang (2006) did not find the significant difference in vocal pitch-matching for *Happy Birthday* between age 7 and age 9 children in Hong Kong, which might be because of the narrower age groups in the study and/or the single and more difficult song compared with those in the current study. It may also be because these ages mark a relative plateau in singing development competency unless there are specific pedagogical interventions (Papageorgi et al., 2022).

By age 11+, the current study found that most participants (76%) could generally achieve accurate vocal pitch-matching of a song. This was supported by findings from the earlier Chinese national study which reported that 65.9% of Grade 4th students generally matched a song accurately (MCCEMEC, 2020). These findings suggest that participants' singing behaviour in the current study may illustrate a mean level of singing behaviour for Primary school children across the country. This might be related to the fact that Hunan Province's GDP is at the medium level of the whole country (see Chapter 1, Figure 1.2), as a positive correlation between participants' singing behaviour and their socioeconomic status reported in the current study.

On the other hand, the current study found that participants across ages, including older peers, tended to match a more difficult song (e.g., *Happy Birthday*) less successfully than the other two simpler target songs. This partly supports the view of Mang (2006) who reported that a difficult target song (*Happy Birthday*) was the explanation for finding no statistically significant difference in vocal pitch-matching ability by age. However, the result in the current study is that the age difference was still statistically significant when matching the most difficult target song, *Happy Birthday*. This might be because of a wider age range of age in the current study (ages 6 to 11) than that in Mang's (2006) study (ages 7 and 9).

(c) Discussing Participants' General Singing Behaviour by Age

In the current study, the ratings for overall singing behaviour are the combined scores for participants' vocal register use and vocal pitch-matching accuracy for the three target songs. There was a significantly positive correlation between vocal register use and vocal pitch-matching accuracy for songs (see more details in Section 7.2.2).

Generally, the current doctoral thesis found that participant children's singing behaviour (including vocal register use and vocal pitch-matching accuracy) generally developed with increasing age during their Primary school period. This finding was also evidenced by results based on the longitudinal data (see Chapter 6, Section 6.4). Participants who were tested twice in successive years generally demonstrated an improved level of overall singing behaviour competency (including vocal register use and vocal pitch-matching accuracy for songs) between 2017-2018 to 2018-2019. This age-related improvement was not influenced by sex, and geographic location or parental income.

The current finding about Chinese participants' singing behaviour by age was also supported in the English national singing programme *Sing Up* data reported by Welch et al. (2009b) (see also the reanalysis in Papageorgi et al., 2022). Furthermore, Chinese participants' singing behaviour with developing age was particularly similar to that of English Primary school children without *Sing Up* training who acted as controls for the main programme intervention evaluation (see Welch et al., 2009b, Figure 13, p. 36). Children without extra singing training from both studies showed a rapid development from age 6 to age 9, but they presented a slower development from age 9 to 11 or 12 years old. Both studies involved similar-aged children who were asked to sing similar singing tasks (*Twinkle, Twinkle* and *Happy Birthday*), and these singing performances were measured by two identical rating scales (the *SVDM* and *VPMD* rating scales). These similar choices in research tools and tasks increased the possibility of comparing their results across these two studies. As both Chinese and English Primary school children's singing behaviour improved with age, this implies that age seems to have a greater influence than culture in the development of Primary school children's singing behaviour.

However, the development of singing behaviour with growing age reported in the current study was not evidenced by Mang (2006) who was the pioneer of using the *SVDM* and *VPMD* rating scales to define Primary school children's singing behaviour. Although the mean differences in singing behaviour between n = 64 age 7 and n = 56 age 9 did not reach a statistically significant level, there was a tendency that older children's singing behaviour to be better. Compared with Mang's (2006) study, the current study also included two other simpler songs, *Twinkle, Twinkle* and *Little Donkey* as well as *Happy Birthday* with a wider age group (ages 6 to 11) and with a greater number of participants. These differences may explain the different results by age between the current study and Mang (2006).

(d) The Interaction Effect between Age and Sex on General Singing Behaviour

The current study found that the age and sex interaction had a statistically significant effect on participants' singing behaviour. Girls' singing ability was generally more advanced than that of boys across different age groups. However, the sex difference was smaller for the youngest and oldest participants, while it was greater for the middle age group of Primary school participants. This was in line with findings by Welch et al. (2009b, see Figure 15 on p. 37). In the current study, although boys' singing behaviour tended to be poorer than that of girls, boys generally continued to develop their singing behaviour across the Primary school period, while girls tended to develop their singing behaviour at a slower speed after reaching a climax at age 9.

7.1.4 The Influence of School on Children's Singing Behaviour

The current study found that the variable 'school' had a statistically significant influence on participants' singing behaviour. Participants of School F generally reached an advanced singing behaviour earlier than their peers from the other five participating schools. However, for the two oldest age groups (ages 10+ to 11+), participants from all six participating schools reached similar levels of singing behaviour. This suggests that children from a school with rich singing activities and advanced teaching resources tended to reach a high level of singing behaviour earlier (see the school differences in the teaching resources for the music curriculum in Section 4.2.1a).

However, it seems that they did not continue to develop their singing behaviour, but sustained a high level of singing behaviour across the whole Primary school period, perhaps because of a ceiling effect in the assessment tools. This might also be because the aim of music education during the Primary school period is to develop a positive attitude towards life through music learning (see Chapter 1, Section 1.4.2c), instead of developing a high skill of singing.

Less advanced singing behaviour for younger participants from the other five schools might be because of less advanced teaching resources than those peers from School F (see Chapter 4, Section 4.2.1a). However, children from these five schools mostly had regular music classes, and together with their singing activities out of school and developing vocal mechanisms, they generally reached a similar level of singing behaviour with that of participants from School F.

Overall, this suggests that children's singing behaviour can develop more quickly in a rich (formal and informal) singing environment. When formal singing teaching is limited, children can still self-develop their singing behaviour to a certain level if some informal singing opportunity is provided.

7.2 The Influence of the Mesosystem on Children's Singing Behaviour

As reviewed in Chaper 3, Section 3.5.2, the mesosystem of Bronfenbrenner's Ecological Theory focuses on the interaction between two perspectives within the microsystem. The current mesosystem discusses the interactions between singing materials (including the task of singing musical intervals as well as a whole song) and aspects of bio-development (age and sex), as well as any correlation between vocal register use and vocal pitch-matching accuracy for songs.

7.2.1 The Interaction between Singing Materials and Bio-Development on Vocal Pitch-Matching Accuracy for Songs

In the current section, four interactions are discussed, including (a) the age difference in the consistency of a song's key centres, (b) the age difference in vocal pitch-matching accuracy for varied types of musical intervals, (c) the sex difference in vocal pitch-matching accuracy for varied types of musical intervals, and (d) the interaction between the complexity of songs and the singer's chronological age. A key difference from most previous studies (see a review in Chapter 2, Sections 2.5.3 and 2.5.4), is the analyses concerning the consistency of key centres and vocal pitch-matching accuracy for musical intervals when matching a target song.

(a) The Influence of Age on Consistency of Sung Key Centres

The current study found that older Primary school children tended to maintain a key centre better than younger children. This was in line with the result reported by Flowers and Dunne-Sousa (1990) who asked American children to sing a self-selected song. This might be because of the following reasons. Firstly, developing lung capacity with age can provide more air support during song singing, especially when matching a higher pitch (*cf.* Proctor, 1980). Secondly, developing vocal folds with growing age during the Primary school period, including thicker intermediate and deep layers of the lamina propria with an increasing number of elastic and collagenous fibres (Boseley et al., 2006; Hirano, 1988), enable the vocal folds to be more flexible when attempting to match different pitches, supported by other developing vocal mechanisms (see more details in Chapter 2, Section 2.2.1). Thirdly, children's singing experience tends to increase with growing age (Demorest & Pfordresher, 2015), which provides more opportunities to practise song singing either individually or in a group (Welch et al., 2009b).

Furthermore, the current study found that the development of a consistency of sung tonalities by age during the Primary school period was not influenced by participant's sex, geographic location or inferred parental income. Overall, these findings suggest that the variables of a song's internal melodic complexity, participant sex, socioeconomic status (geographic location and income) and culture appear to have a limited impact on the tendency for greater sung key consistency with increasing age.

(b) The Influence of Age on Vocal Pitch-Matching Accuracy for Different Types of Musical Intervals

The current study found that older participants tended to have fewer semitone errors for vocal pitch-matching accuracy for different types of musical interval than younger participants, especially for larger musical intervals, such as a minor 6th and an octave (see Chapter 5, Figure 5.7). This finding is supported by studies by Sinor (1994), Moore et al. (1995/1996) and Feng et al. (2013). Sinor (1994) invited *N* = 96 children aged 3 to 5 in the United States to sing 4-note patterns in the United States. The study of Moore et al. (1995/1996) included *N* = 480 children aged 6 and 9 from Argentina, Poland, Spain and the United States who were asked to echo 16 musical intervals. Feng et al. (2013) included *N* = 130 children aged 3 to 6 from Jiangsu Province, China, who were asked to sing two composed songs. Older children's lesser difficulty in matching larger musical intervals might have related to their wider overall vocal range and their comfortable singing range (*cf*. Chapter 2, Section 2.5.1), and/or a more flexible vocal mechanism, with more air support, than younger children (see

Chapter 2, Section 2.2.1).

However, the current study – using target songs as singing material – also found that the age difference in vocal pitch-matching accuracy was less when matching a musical interval smaller than a perfect 5th, which is in line with the study of Feng et al. (2013). This might be because the participants across varied age groups shared a common comfortable singing range (*cf*. Chapter 2, Section 2.5.1) that allowed them to match musical intervals within this common pitch range relatively accurately.

(c) The Influence of Sex on Vocal Pitch-Matching Accuracy for Different Types of Musical Intervals

In terms of the sex difference in vocal pitch-matching accuracy for musical intervals involved in the three target songs, the current study found that girls tended to match most types of musical interval (except for the unison and a minor 6th) significantly more accurately than boys. The sex difference was even greater when matching a perfect 5th and an octave (see Chapter 5, Figure 5.8).

The general finding on the sex difference is supported by Moore et al. (1995/1996) who asked *N* = 480 children aged 6 and 9 from Argentina, Poland, Spain, and the United States to match 16 musical intervals. Their data are also in line with the finding reported in the current study that girls' singing behaviour was generally better than that of boys (see Chapter 6, Figure 6.5). The sex difference in vocal pitch-matching accuracy for different types of musical intervals, as measured by the *MAPM* system ,provided more details about the sex difference in the general song singing behaviour as measured by two rating scales in the current study.

However, the current sex difference in vocal pitch-matching accuracy for different types of musical intervals is not in line with the study of Sinor (1994). The different findings might be because the kindergarten children (aged 3-5) in the study by Sinor (1994) were much younger than the Primary school participants (aged 6-11) in the current study. As reported in the current study (see Chapter 6, Figure 6.5), the sex difference by means in singing behaviour was relatively smaller for ages 6 and 7 than for older participants.

Furthermore, the reduced sex difference in vocal pitch-matching accuracy for unison and a minor 6th in the current study might be because unison was the easiest musical interval for both boys and girls, while a minor 6th, nearly the most difficult interval, presented a similar vocal challenge for both sexes. However, the greater sex difference in vocal pitch-matching accuracy for a perfect 5th and an octave, biased towards girls, might be because Primary school girls generally sing more and, consequently, may have a more positive attitude towards singing than boys (Welch et al., 2009b). These biographical biases towards more personal singing practice might lead to a more advanced singing skill when asked to match a large musical interval in a song context.

(d) The Interaction between the Complexity of Songs and Age

As reported in Chapter 6, Figure 6.3 in the current study, vocal pitch-matching accuracies for *Twinkle, Twinkle* and *Little Donkey* were generally similar across different age groups, and their accuracy tended to be higher than that for *Happy Birthday*, particularly for older participants. This suggests that improvement in vocal pitch-matching accuracy for a more difficult song – in terms of its musical challenge (e.g., *Happy Birthday* with an upward octave leap) might be slower with increasing age during the Primary school period (*cf.* Mang, 2006). This situation might not be changed until children were given training in transferring between vocal registers.

7.2.2 Comparing Vocal Register Use and Vocal Pitch-Matching Accuracy for Songs

In the current study, vocal register use for the three target songs was measured by the *SVDM* scale (Rutkowski, 1996, 2015, 2018), while vocal pitch-matching accuracy for the songs was measured by the *VPMD* scale (Welch, 1998) and also a newly created system—the *MAPM* system.

According to Chapter 6, Sections 6.1 and 6.2, raw scores of vocal register use for the three target songs measured by the *SVDM* scale and those of vocal pitch-matching accuracy for the songs measured by the *VPMD* scale were statistically significantly positively correlated. This was in line with the related findings reported by Mang (2006) and Welch et al. (2009b). All of these findings suggest that the ability to match a target song more accurately might be related to increasing competency in using more vocal registers.

The above correlation might be because the ability to adjust vocal registers expanded their available vocal range when singing a song. When participants were not able to adjust their vocal registers, but needed to match a high pitch in the course of singing a song, the current study found that they generally sang the pitch flatter than the target. The cause-and-effect assumption about adjustment to vocal registers and vocal pitch-matching accuracy for songs was also supported by the findings on the most common vocal pitch-matching errors towards a lift point during singing. According to the findings reported in Chapter 5, Section 5.4.5, when matching a series of continuously upward pitches close to a lift point, many participants matched these pitches flat, i.e., under the target. Similarly, when matching a series of continuously downward pitches close to a lift point, many participants matched them sharp. It seems that these vocal pitch-matching errors related to Rutkowski's proposed register lift points might be because these participants did not adjust their vocal registers appropriately.

Furthermore, when comparing the same participants' vocal register use and the number of keys used to sing each of the three target songs, the current study found that there was a significantly negative correlation between these. When participants used more vocal registers to sing a target song, they tended to exhibit fewer changes in key centres. This might be because children's available vocal range was extended due to using more vocal registers – a likely biproduct of the physical development of the vocal mechanism with age. This wider vocal range may give pitches more opportunity to be matched accurately because the older participants' comfortable singing range is expanded compared to that of the younger participants. This was somewhat in line with the finding of Jones (1979) who reported that children aged 6 and 9 who could not sing in-tune changed key centres because of their limited vocal range.

Overall, these findings suggests that scores of vocal register use and vocal pitch-matching accuracy measured by the three above measures (the *SVDM* and *VPMD* rating scales and the *MAPM* system) can support each other.

However, there is an observable difference between scores of vocal register use and vocal pitch-matching accuracy for songs. The improvement in vocal register use was generally greater than that in vocal pitch-matching ability for songs (see Chapter 6, Section 6.4). This might be because vocal register use develops earlier than vocal pitch-matching for children (*cf.* Rutkowski, 2018).

In the current study, younger participants showed greater improvement, especially in vocal register use than vocal pitch-matching accuracy, for the three target songs compared to older participants (see Chapter 6, Section 6.4). This is partly supported by Welch et al. (2009b) and Pereira and Rodrigues (2019) who reported a statistically significant age difference in singing behaviour with age, with faster development in the younger Primary school age group. According to Titze (1994), this might be because the developing length of vocal folds with increasing age had less impact on fundamental frequency for older compared with younger children due to the interaction between length, tension and mass of vocal folds. Consequently, vocal range can increase at a faster pace for younger than for older children (*cf.* Titze, 1994).

7.3 The Influence of the Exosystem on Children's Singing Behaviour

As reviewed in Chapter 3, Section 3.5.2, proposed elements in the application of Bronfenbrenner's hypothesised exosystem did not appear to influence children's behaviour directly, but did so indirectly by influencing the environment of elements in the microsystem. In the current study, one element of the exosystem is socioeconomic status, as measured by geographic location and inferred family income. The current section starts by discussing the influence of socioeconomic status on children's general singing behaviour. Then, the interaction between this status and singing materials on vocal pitch-matching accuracy for different types of musical intervals, and the interaction of socioeconomic status and age on participants' general singing behaviour are discussed separately.

7.3.1 The Influence of Socioeconomic Status (Measured by Geographic Location and Family Income) on Children's General Singing Behaviour

Socioeconomic status in the current study is defined by geographic location (rural/urban areas) and inferred-family income (lower/higher income families), based on self-reporting by the children and informal interviews with local people during the fieldwork. Geographic location was used to define the socioeconomic status because disposable income per capita differs between rural and urban people across China (see Section 1.2.1). Inferred-family income was applied to explore further economic difference within a rural/urban area. Both geographic location and inferred family income could influence participants' singing behaviour through (a) the availability of specialist music teachers within the 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 (see Table 4.3)). Also, based on the current author's informal interviews with music teachers, urban music teachers tended to have more professional development and training opportunities than rural music teachers. Teachers reported that training opportunities tended to be located in urban areas, which implies that rural-based music teachers have to travel, often considerable distances by train or bus. 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 (the exception being a music teacher in School D, see Table 4.3). Furthermore, while all the music teachers in Schools E and F, located in urban areas, were music specialists, the parents of School F, who were generally working in a nearby hospital, were likely to be richer than the parents of School E, who were originally rural based and moved to the city to work. That the parents in School F were wealthier might explain why School F had an orchestra. Also, rural participants 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 to be slightly greater.

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 rural peers and children from lower-income families. This suggests that Chinese children with a higher socioeconomic status (e.g., urban area and/or higher-income families) tend to have better singing behaviour (e.g., singing more accurately in Western tonalities). The impact of socioeconomic status by geographic location and family income on participants' singing behaviour might be because schools located in higher socioeconomic areas (urban areas and/or higher-income families) were more likely to recruit better qualified music teachers, and the music curriculum in these schools was more likely to be undertaken within a school timetable (e.g., Lu, 2017; L. P. Wang, 2013; T. Wang, 2012; Zhang, 2017). This assumption was confirmed by the author for the six participating schools in the current study during the process of data collection.

Primary school children's singing behaviour is more likely to develop when teaching resources were guaranteed (*cf*. Pereira & Rodrigues, 2019). This was supported by Welch et al. (2009b), who reported that children who received *Sing Up* training (*n* = 5,832 singing assessments) showed a statistically significantly higher level of singing behaviour than peers who did not receive such training (*n* = 2,967 singing assessments).

In the current context of China, urban children generally receive better teaching resources than their rural peers, especially for courses that do not require a formal test, such as music teaching (e.g., L. P. Wang, 2013; T. Wang, 2012). The lack of specialist music teachers in rural areas in China has been a long-term problem with various causes, such as a relatively poor living environment, less attention to music from school leaders, a lack of teaching equipment, and fewer training opportunities (e.g., Lu, 2017; L. P. Wang, 2013). Consequently, many music lessons are taught by teachers whose main teaching responsibility is another area of the curriculum (e.g., Lu, 2017). This inequality of teaching resources for music lessons in Primary schools between rural and urban areas seems to be a common problem in China (see Chapter 1, Section 1.3.2). This also within a context of more rural children relocating to a city with their parents to find more working opportunities there (see Chapter 1, Section 1.3.1).

Income difference is related to economic variation within a region. It seems that across the world the difference in the economic development within a region tends to be smaller in a rich area but greater in a poorer area (see the report of the World Bank[77]). Income difference within a region in China might be caused by unequal local economic development[78]. This could be for various reasons, such as different trade barriers, labour supply across regions, the amount of international trade and foreign investment (Cai & Du, 2000; He & Liang, 2004). He and Liang (2004) argued that rapid overall economic growth increased the gap in economic development within and across Chinese regions. Consequently, the author assumes that the singing competence of Primary school children from other regions in China may also vary significantly according to local income variation

---

[77] LAC Equity Lab: Income Inequality - Urban/Rural Inequality (worldbank.org)

[78] 国务院印发《关于统筹推进县域内城乡义务教育一体化改革发展的若干意见》 _ 滚动新闻 _中国政府网 (www.gov.cn)

which indirectly impacts on the professionalism of music teachers, as unequal development within a region has been a common issue across China.

According to the current study, the interaction effect between geographic location and income had no statistically significant effect on participants' singing behaviour (see Chapter 6, Figures 6.9, 6.18 and 6.21). This suggests that the singing behaviour of participants from higher-income families within either rural or urban areas tended to be more advanced than that of their peers from lower-income families in the same areas. According to the mean differences and standard errors from Figures 6.9 and 6.18, urban children from higher-income families tended to have the most advanced singing behaviour, followed by urban children from lower-income families and rural children from higher-income families. Rural children from lower-income families were more likely to have the poorest singing behaviour, although it should be noted that the influence of age on the interaction effect between geographic location and income was not analysed in the current thesis. It was not sure whether the above tendency would be changed for different age groups.

Participants from the four above groups could be matched to four categories of children shown in Chapter 1, Section 1.3.1, as their schools were selected based on the guideline of the varied status of geographic location and income. Rural children from lower-income families tended to be rural children who were left behind and looked after by relatives. Rural children from higher-income families were more like to be children who were looked after by one or both of their birth parents. Urban children from lower-income families originally came from rural areas and came to a city to study with their parents who found a job in the city. Urban children from higher-income families were children who were likely to be born there.

Together with their mean singing behaviour reported above for these groups, it illustrates the great power of local socioeconomic status on Primary school children's singing behaviour in the context of Chinese society.

One aspect of the above findings is related to singing in the family home and this was not able to be researched in this study. Cohort-based evidence from Australia (Williams et al., 2015) suggests that where pre-school children experience a rich singing environment at home, they are more likely to be advanced developmentally on entering school. This cultural bias towards the use of music is not necessarily linked to income or poverty levels, but it can be conjectured that children with working parents may have less opportunity to experience singing at home and, consequently, enter school as less developed singers. Kirkpatrick (1962) found that children who entered Elementary school as 'singers' were more likely to have been sung to by their mothers at home.

7.3.2 The Interaction Effect between Socioeconomic Status and Singing Materials on Vocal Pitch-Matching Accuracy for Varied Types of Musical Intervals

The current study found that the interaction between geographic location and the type of sung musical interval had no significant impact on participants' vocal pitch-matching accuracy for different types of musical interval (see Chapter 5, Figure 5.9). However, the interaction between the type of sung musical interval and inferred family income had a statistically significant impact. Participants from higher-income families showed better vocal pitch-matching accuracy for two particular types of musical intervals: a perfect 5th and an octave, than peers from lower-income families (see Chapter 5, Figure 5.10). This suggests that the potential influence of family income was generally greater than that of geographic location on vocal pitch-matching accuracy for more difficult matched types of musical interval.

7.3.3 The Interaction Effect between Age and Socioeconomic Status on Children's General Singing Behaviour

The age and geographic location interaction, and the age and family income interaction had a statistically significant effect on participants' singing behaviour. In both rural/urban areas, and higher/lower-income families, older participants tended to have better singing behaviour than younger participants. However, the difference in singing behaviour by geographic location and income tended to be smaller for older participants (age ≥ 10) than younger participants (see Chapter 6, Figures 6.6, 6.7, 6.15, and 6.16).

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. For instance, according to the author's informal interviews with participants, listening to music on a phone is a common after school activity for children from School A where many parents had left the family home to work in another part of the country. These parents usually gave their children a phone to keep in contact. Consequently, the phone provided a convenient opportunity for children to listen to music when they remained in the family home and were looked after by their relatives. Some said that they listened to music when they missed their parents.

Overall, this may explain the small difference in singing behaviour for the oldest Primary school participants, regardless of diverse socioeconomic status. It also 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.

7.4 The Influence of the Macrosystem on Children's Singing Behaviour

As reviewed in Chapter 3, Section 3.5.2, the macrosystem implies a particular culture for the exosystem in Bronfenbrenner's ecological theory. In the current section, the singing behaviour of Chinese and English Primary school children's singing behaviour is compared from the two following perspectives: bio-development (age and sex) and socioeconomic status (geographic location and family income).

7.4.1 The Influence of Culture and the Interaction Effect between Culture and Bio-Development on Chinese and English Children's Singing Behaviour

As discussed in Section 7.1.3d, when comparing all Chinese participants' singing behaviour with that of English Primary school children who did not receive *Sing Up* training, children from both cultures improved their singing behaviour significantly with age (see Chapter 6, Section 6.5). Furthermore, the current study found that the means of singing behaviour across age groups between both cultural groups were similar. There was no statistically significant difference in singing behaviour by participants' age and culture interaction, nor the age, sex and culture interaction.

In terms of any sex differences, the current study is in line with the findings reported by Mang (2006), Welch et al. (2009b) and Pereira and Rodrigues (2019), who reported that girls generally show more developed singing behaviour than boys. The children in the three above studies were from Hong Kong, England and Portugal. Together with the children from Hunan Province in China, this suggests that the sex difference in singing behaviour is likely to be a shared characteristic for Primary school children in varied cultures where a regular music lesson has been provided in a Primary school system. Welch et al. (2009b) considered whether this sex difference in singing behaviour might be because boys thought that singing was a female activity, which led to them to be involved in fewer singing related activities.

Overall, these findings suggest that Chinese participants and English Primary school students without *Sing Up* training show similar development of singing behaviour by age and sex. This suggests that national culture may have a limited impact on any differences in children's singing competences. However, it should be noted that Primary school children from both countries generally received some form of regular music curriculum in Primary schools based on using equal temperament tuning. The current study included n = 1,539 singing performances from Chinese children, while that of Welch et al. (2009b) involved n = 3,498 singing performances provided by English children who did not receive *Sing Up* training. It might be helpful to remember that data from the two groups was based on two identical target songs (*Twinkle, Twinkle* and *Happy Birthday*) and was measured by the same two existing rating scales (the *SVDM* scale and the *VPMD* scale). The age range of the two groups was also close, which was from age 6+ to 11+.

Nevertheless, the summary of the current study does not seem to be in line with that of previous studies. As reviewed in Chapter 3, Section 3.2.3b, the studies of Mang (2006) and Welch et al. (2009b) reported that Primary school children's singing competency, including vocal register use and vocal pitch-matching accuracy for songs, varied statistically significantly by local cultural variables. However, the possible cultural influence was investigated in different ways across these two studies. Mang (2006) explored the influence of a tonal language on children's singing behaviour, while Welch et al. (2009b) compared the singing behaviour of varied ethnic groups.

To test the influence of a tonal language on children's singing behaviour, Mang (2006) compared the singing competence of $n = 60$ Cantonese monolingual speakers with that of $n = 60$ English bilingual speakers in Hong Kong. All the children sang *Happy Birthday*. She defined singing competency based on vocal register use and vocal pitch-matching accuracy for the target song, measured by the *SVDM* and *VPMD* rating scales respectively. Her study found that the singing competency for Cantonese monolingual children was statistically significantly higher than that for English bilingual children. Mang assumed that this might be because Cantonese is a tonal language, and there was a link between a tonal language and singing. However, it was not clear whether the advantage of singing for Cantonese children was because they used only one language compared rather than using two languages for the English bilingual group, as practicing singing in any of two languages might be more difficult than for one language.

Welch et al. (2009b) reported that children's singing competency statistically significantly varied by ethnicity, with Black and White children showing a statistically significantly better singing competency than Asian children whose families had originated mainly from India, Pakistan and Bangladesh. It is uncertain whether the singing behaviour of these Asian children, particularly as they were living in a different cultural environment, could represent the singing behaviour of Chinese children who were living on the Chinese mainland.

Furthermore, the results between Mang (2006) and Welch et al. (2009b) seem to conflict. While English bilingual children, who might be black or white, showed a lower level of singing behaviour than Cantonese monolinguals in the study of Mang (2006), the singing behaviour of black and white children generally showed higher singing behaviour in the study of Welch et al. (2009b). One difference is that the English bilingual children were the minority group in the community of Hong Kong in study of Mang (2006), while White children were in the majority group in England in the study of Welch et al. (2009b). This seems to suggest that the singing behaviour of children originally from the same ethnicity can vary in different contexts. When they are in the majority in their local community, they are more likely to show better singing behaviour.

However, when comparing singing competence between Chinese participants in the current study and English children (mainly White children) in the study of Welch et al. (2009b), both groups were the majority ethnic group in their local context. In more detail, all participants in the current study, located in the northeast and east of Hunan Province (see Figure 4.2), were Han Chinese. All participants were encouraged to speak Mandarin at their schools, although they might speak their local language at home. It was not clear whether their language background was different from their peers from other ethnic groups. However, it should be noted that over 90% (91.10% in 2024) of the population in China is Han[81].

Consequently, it can be conjectured that these children from the majority ethnic group were less likely to be affected by racial tensions (Waters et al., 2008). Furthermore, although music was part of the curricula in both China and England, it was not a main Primary school subject in either country. The standard amount of available time used to teach music in a state Primary school in both China[79] and the UK seems similar, being notionally nearly one to one and half hours per week in an academic term. The singing environment out of school for both groups might be also similar, as listening to and singing songs with a recording are popular in both countries. Lastly, Primary school children from different cultures seem to share similar physical development. For instance, the length of vocal folds, increasing with age (Titze, 1994), led to increasing vocal range (Welch, 1979b), ignoring the different ethnicities. It should be noted that Asian children in the English *Sing Up* study tended to have a higher speaking pitch compared to their Black and White peers, implying that their

[79] According to the latest Curriculum Program of Compulsory Education 【义务教育课程方案】published by the Ministry of Education of the People’s Republic of China in 2022, music curriculum is a part of art curriculum. The art curriculum included musical games and painting for Grades 1 and 2. It included music, painting, dancing, opera, and movie for Grades 3 to 7, mainly focused on music and painting. The standard rate of time for teaching art in the nine-years compulsory education was from 9% to 11%. Consequently, the maximum rate of time for teaching music might be around 4% to 5%.

vocal folds were physically smaller.

7.4.2 Comparing Primary School Children's Singing Behaviour in Hunan Province, China and England, UK by Socioeconomical Status (Measured by Geographic Location and Family Income)

As discussed in Section 7.3.1, Chinese Primary school children's singing behaviour varied significantly by socioeconomic status in the current study's analyses. However, it seems that the influence of socioeconomic status on singing behaviour may differ in China and England. The study of Welch et al. (2009b) is one of a limited number of previous studies to report the influence of deprivation on English children's singing competency. They reported a positive correlation between scores on the Index of Multiple Deprivation[80] for school localities and means of children's normalised singing scores. In other words, children from a school located in a higher deprivation area tended to show advanced singing development or at least equivalent to that of children in more affluent areas. Welch et al. explained that this might be because most English children from high-deprivation families were from inner urban areas of big cities (such as London and Manchester), and in which many schools involved plenty of music activities. Furthermore, according to the author's personal informal interview with the head of a Primary school where many music activities are undertaken in a high IMD rated area of Greater London, music was being used as a tool to develop children's confidence within both the school and the community, especially for children from minority groups.

Consequently, the varied socioeconomic contexts (e.g., ethnicities) of England and China suggest that we cannot compare their data directly using an identical standard. According to Rural-Urban Labour Statistics'[81] published by the International Labour Office in 2018, the mean difference in poverty rates in rural and urban areas was smaller in Europe than in East Asia, suggesting a smaller economic difference by geographic location (rural and urban areas) in Europe than China.

---

[80] See more details about the Index of Multiple Deprivation from the link: Index of Multiple Deprivation (IMD) 2007 - data.gov.uk

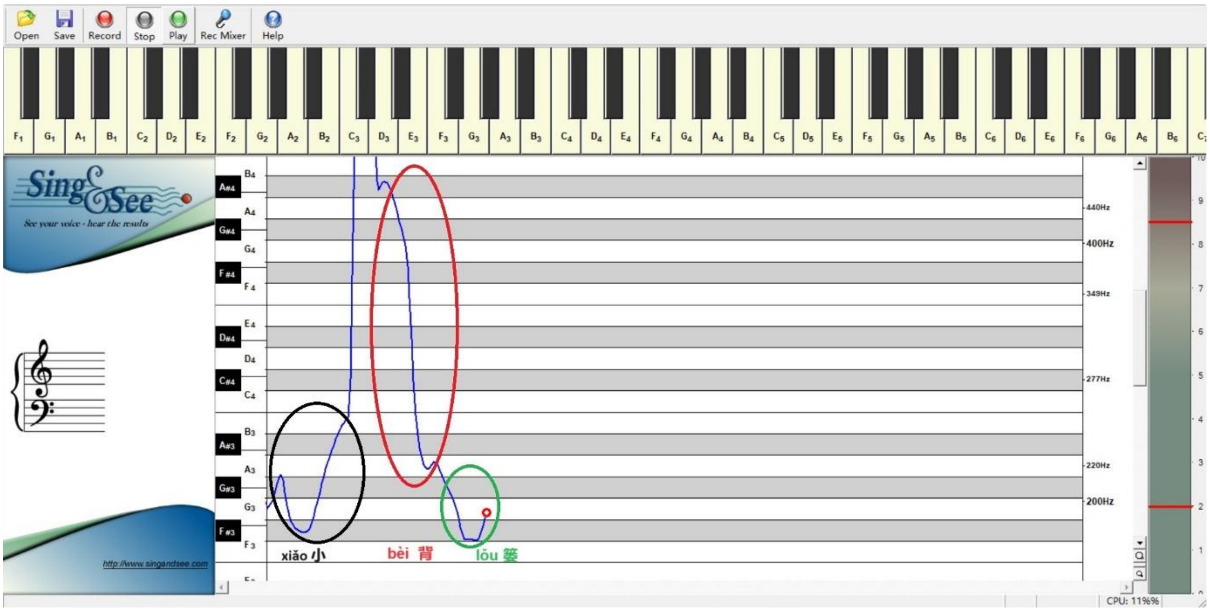
[81] See 'Rural-urban labour statistics' published by the International Labour Office (wcms_636038.pdf (ilo.org)).

Overall, the study of Welch et al. (2009b) suggested that a teaching resource (e.g., *Sing Up* experience) seems to have a more powerful impact on children's singing behaviour, when compared with the influence of the socioeconomic environment for English Primary school children. However, urban Chinese children and children from higher-income families in the current study usually attended a school with richer teaching resources, including for the music curriculum, and so the data from the current study is in line with the study of Welch et al. (2009b), in that the provision of appropriate teaching resources, especially staffing, can have a great impact on the development of children's singing behaviour. In a rich singing environment (such as urban Chinese children from higher-income families), children tend to develop their singing competency earlier.

7.4.3 The Influence of a Tonal Language on Vocal Pitch-Matching Accuracy for Song Singing

As all Chinese participants were Mandarin speakers, and Mandarin is a tonal language, it seems necessary to add a note about the possible influence of a native tonal language on participants' singing behaviour. To explore the possibility, characteristics of pitch height for three Chinese words, xiǎo bèi lǒu 小背篓 (a little back basket), as spoken by a Chinese girl who is eight years old, were explored in terms of spoken and singing behaviours on the screen of Sing&See. The tones of the three words are the third, fourth, and third, respectively. The melody for the three words is *mi-la,-re-do-la,*. The characteristic for each vocal behaviour is shown in Figure 7.1. As the figure illustrates, the degree of pitch movement for each of the three words tends to be greater for the spoken language compared to the sung version. Whilst the shape and degree of the movement of the pitches for a spoken language are mainly controlled by the tone and the speaker's spoken pitch centre, those for a singing voice are likely to be more consistent and dominated by the relative pitch height of the related melody and the speaker's comfortable singing range. Overall, the spoken tone seems to have limited influence on vocal pitch-matching accuracy for the singing behaviour.

Figure 7.1 Characteristics of the Pitch Height of Three Chinese Words between Spoken (top) and Sung (bottom) Behaviours, as displayed on Sing&See



7.5 Summary

Overall, the findings from the fieldwork reported above and their relationship to the literature review are summarised in the opening sections of Chapter 8 against the original research questions.

Chapter Eight

Summary and Conclusions

8.1 Summary

In the current section, three research questions raised in Chapter 1, Section 1.7 are answered initially based on the findings reported in Chapters 5 and 6. These findings are then summarised based on Bronfenbrenner’s ecological theory.

8.1.1 Research Questions and Related Answers

Research question 1:

- How can participants' singing behaviour of song singing be illustrated from a micro perspective?

The current study has created a new system, named the Melodic Analysis of Pitch-Matching (*MAPM*) system, to describe the detail of Chinese Primary school children's singing behaviour in terms of vocal pitch accuracy. By using the system, the type (sharp, flat, or no error) and degree of sharpness or flatness for each vocally matched pitch of a song for a participant can be shown visually by the colour characteristics of the system. By observing the colour characteristics illustrating vocal pitch-matching accuracy for a song for each participant, several following common pitch errors across age groups can be shown.

The most common pitch related errors are listed here: (1) Most of participants did not start with the most suitable key centre when they were allowed to self-select a key to sing a target song starting with a tonic. The most suitable key was generally one to two semitones flatter than the key they selected. (2) Vocal pitch-matching accuracy generally decreased when matching a larger size of musical intervals, although the tendency was less marked for older participants; girls' vocal pitch-matching accuracy for most types of musical intervals was generally similar to that of boys. However, girls tended to match larger musical intervals more accurately than boys. Geographic location seems to have no significant effect on participants' vocal pitch-matching accuracy for different types of musical intervals, but income does. Children from higher-income families tended to match a bigger musical interval more accurately. (3) The number of keys that an individual used to match pitch in a pair of target songs were significantly positively correlated. Older children tended to sing more consistently (changing fewer key centres) than younger children. This tendency was not influenced by the complexity of songs, sex or socioeconomic status (measured by geographic location and income). Participants matching a target song more consistently (using fewer key centres) tended to apply more vocal registers. (4) In addition, if any target song included two identical phrases, participants showed similar vocal pitch-matching errors for both these phrases. (5) For any target song that included a big upward musical interval, such as an octave, many participants matched the first pitch of the interval sharply. (6) For any target song including a series of small upward or downward musical intervals towards a lift point, many participants tended to match these pitches out-of-tune, if they did not change vocal register. In may be hypothesised that the above findings reported in the current study are likely to be common characteristics when Primary school children match a song.

Overall, these findings suggest that the second research aim – understanding Chinese Primary school children's singing behaviour from a more micro perspective – has been achieved.

Research question 2:

- What is the general singing behaviour in selected song singing for Primary school students in China?
- 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 these participants?
- What are the similarities and differences for data of the general singing behaviour of song singing between the current study from China and studies collected from the National Singing Programme *Sing Up* in England, UK?

This research question found the following answers: (1) Chinese participants' singing behaviour – including vocal register use and vocal pitch-matching accuracy – generally improved with increasing age. This tendency was not varied statistically significantly by songs with different complexities, especially for vocal register use. For vocal pitch-matching accuracy, the developmental tendency tended to be slower when matching the more difficult target song, *Happy Birthday*. (2) There was a significant positive correlation between scores of vocal register use and those of vocal pitch-matching for the three target songs. (3) When scores of vocal register use and vocal pitch-matching accuracy for the three target songs were normalised to create a picture of general singing behaviour for participants, the normalised singing scores (NSS) improved with increasing age. Girls' NSS generally were higher than those of boys. However, the sex difference was smaller for the youngest (age 6+) and the oldest (age 11+) participants. Similarly, the differences by geographic location, inferred family income and school in the NSS were smaller for the oldest participants.

The above findings suggest that the first research aim – understanding Chinese Primary school children's general singing behaviour – has been achieved.

When comparing the singing performances of the same two target songs sung by Chinese participants and English children without *Sing Up* experience from the national study reported by Welch et al. (2009a), children from both cultures generally improved their singing behaviour with increasing age. This tendency was not varied statistically by sex and culture. Means of the normalised singing scores (NSS) were similar for each age groups of Chinese participants and English children without *Sing Up* training. Overall, the current study found that the difference in the NSS between the children from the two cultures was limited. These suggests that the third research aim – comparing Chinese and English children's singing behaviour – has been reached.

Research question 3:

- What is relationship between the micro and macro perspectives in explaining participants' singing behaviour?

Overall, scores measured by the *MAPM* system provided details of singing behaviour which generally supported those measured by the two rating scales (the *SVDM* and *VPMD* scales) that described participants' singing behaviour from a macro (whole song) perspective. These additional *MAPM* details increased the reliability of findings reported in the current study. For instance, when participants' vocal pitch-matching accuracy for the three target songs was found to be improved with age measured by the *VPMD* scale, this improvement could be supported by more details revealed by the *MAPM* system. For instance, older children tended to match greater musical intervals more accurately. They also generally changed key centres fewer times when matching a target song than younger children.

8.1.2 A Summary of Children's Singing Behaviour Based on Bronfenbrenner's Ecological Theory

According to the findings reported in the current study, it seems that Chinese Primary school participants' singing behaviour was more affected by the first three systems of Bronfenbrenner's ecological theory: the microsystem, the mesosystem and the exosystem.

Two dominant elements of the microsystem are the research design (such as singing materials and measurement) and bio-development (age and sex). Many elements of a song can impact on children's singing behaviour, such as the selection of a starting pitch, the melodic contour, the size and number of each type of musical interval, whether having two identical phrases in a target song, and the combination of small musical intervals in a melodic phrase. Using the colour characteristics of the *MAPM* system to show vocal pitch-matching accuracy for these elements suggested that these elements are correlated and support each other. For instance, when a child did not know how to transfer between vocal registers, they might have a limited singing range, and have difficulty in matching a big musical interval (e.g., an octave) and/or matching a series of upward or downward musical intervals towards a lift point. In terms of the influence of bio-development, older children and girls generally showed better singing behaviour than young children and boys.

The mesosystem includes the interaction between research design and bio-development. For instance, older children generally used fewer key centres—singing more consistently and matched greater musical intervals more accurately. However, when matching a more difficult song, *Happy Birthday*, they generally showed a slower improvement than matching other two simpler target songs, *Twinkle, Twinkle* and *Little Donkey.*

The exosystem relates to participants' socioeconomic status, measured by geographic location and income. According to the current study, urban children and children from higher-income families tended to show more developed singing behaviour than their rural peers and children from lower-income families. However, when children are in the oldest age group (age 11+) in the Primary school period, it seems that the impact of socioeconomic status has more limited impact.

When comparing singing behaviour between Chinese participants and English children without *Sing Up* training in the national study conducted by Welch et al. (2009a), data of the NSS were similar across the two cultures. This suggests that culture – as a part of the macrosystem – may have a limited impact on Primary school children’s singing behaviour, when a music curriculum is provided using equal temperament (see Chapter 2, Section 2.1.1). Instead of the cultural difference, the amount of singing education seems to have greater power to influence the quality of children’s singing behaviour.

Due to the short longitudinal data collection period (two years) in the current study, the current study could not conclude an influence of the chronosystem (time) on children’s singing behaviour, although children tended to improve their singing with age, all other variables (such as sex, geographic location, inferred family income) being equal.

Overall, it seems that Bronfenbrenner’s ecological theory provides a useful perspective into most of the findings reported in the current study well.

8.2 Application of the Study Findings

The current study has two principal applications: research design for future studies and teaching singing in a Primary school in the real world for music teachers.

8.2.1 Application of the Current Findings on Research

The current study includes the two following sections, (a) the implication of its research design for those of future studies and (b) the impact of its findings on the academic world.

(a) Application of the Current Study on the Research Design

The application of the research design focuses on singing materials – songs and the newly created measurement – the Melodic Analysis of Pitch-Matching (*MAPM*) system.

Firstly, the current study used three target songs with different complexities to explore Chinese Primary school children’s singing behaviour. It shows that children’s vocal pitch-matching accuracy varied according to the complexity of the songs. This suggests that future studies can use multiple songs with diverse levels of complexity to test children’s singing behaviour, especially for vocal pitch-matching accuracy for songs, and to offer insights into potential areas for focused singing development.

Secondly, the measurements of the current study involved established rating scales (the *SVDM* and *VPMD* scale) and a more detailed measurement (the *MAPM* system) to explore children's singing behaviour (vocal register use and vocal pitch-matching accuracy for songs). Using these two approaches simultaneously can not only describe a range of children's singing behaviours, but can also show singing behaviour (such as vocal pitch-matching accuracy) in more detail. As data measured by the *MAPM* system is more minutely measured, these results can be used to test the reliability of results measured by two macro-focused rating scales which are more summative perceptually. In addition, it is suggested that the *MAPM* system can be applied to test the reliability of other rating scales about vocal pitch-matching accuracy for songs, as it can show the details of vocal pitch-matching accuracy of each vocally matched pitch of a song. The songs should be using the system of equal temperament or pythagorean tuning, as their pitches can be shown in a visual piano of Sing&See.

Additionally, the newly created *MAPM* system provides an opportunity for researchers to explore singing behaviour in greater detail, especially for vocal pitch-matching accuracy for songs. This need not be limited to analysing children's singing behaviour, but should be also valid when analysing that of adults. More importantly, the details revealed by the *MAPM* system, which show common vocal pitch-matching errors for songs, should be able to help music teachers to improve children's singing behaviour more effectively (see Section 8.1.2). This shows the value of the current academic study for Primary school children's singing behaviour in a real world context.

(b) Academic Values of the Current Study's Findings

The current study reports Primary school children's singing behaviour based on a relatively large number of singing performances collected from *N* = 1,193 children, aged from 6+ to 11+ in Hunan Province, China. As limited fieldwork Chinese studies have been conducted previously, the current study provides an example of systematic data collection on children's singing which may have wider applicability. Furthermore, as limited previous studies conducted in the context of the Chinese mainland have been reported in English, it is hoped that the current study may help people from outside China understand Chinese Primary school children's singing behaviour, based on different complexities of target songs, age groups, sex and socioeconomic status (geographic location and inferred family income).

Using the same singing materials and measurements, the current thesis also compares results between China and England to explore children's singing behaviour under the influence of culture. Overall, it seems that the cultural influence on Primary school children's singing behaviour is relatively small, at least in this comparison. Music education in both China and England is based on Pythagorean tuning, and equal temperament. In contrast to culture, singing training opportunity, in the context of national policy and local economic support, seems to have a greater impact on children's singing behaviour and development. Consequently, it seems that findings on children's singing behaviour from China can be combined with related English-language studies (e.g., Henrich, 2006; Hollien, 1974; Kochis-Jennings et al., 2012, 2014) to provide a deeper understanding of children's singing behaviour as a whole. This also increases the possibility for advanced teaching strategies related to singing to be shared across cultures that use equal temperament to improve children's singing ability.

When using Bronfenbrenner's ecological theory to explain Chinese participants' singing behaviour, it seems that the theory can generally be matched to the data relatively well, especially for the first three systems (microsystem, mesosystem and exosystem). Consequently, the current study can stand as an example of using the ecological theory to describe Chinese children's singing behaviour.

Overall, the current study fills the gap left by the few studies in English which have paid attention to Chinese Primary school children's singing behaviour. The current study also explores cultural difference in children's singing behaviour, and uses Bronfenbrenner's ecological theory to explain Chinese Primary school children's singing behaviour and development.

8.2.2 Application of the Current Findings for Music Teachers' Singing Teaching

In the current section, the application of the current findings, based on the *MAPM* system and the two existing rating scales (the *SVDM* and *VPMD* scales), on singing teaching for music teachers working in a Primary school are described below.

(a) Application of the Findings based on the *MAPM* System for Music Teachers in Primary Schools

Common pitch errors found by using the colour characteristics of the *MAPM* system in the current study contributed to the following suggestions for music teachers in Primary schools who would like to improve children's vocal pitch-matching accuracy for songs.

Firstly, the current study found that around 70% of participants did not start with a key that they could sustain when matching *Twinkle, Twinkle* and *Little Donkey* whose starting pitch is a tonic. Half of them started with a key that was one or two semitones sharper than a key they could sustain subsequently. It is assumed that this phenomenon may be replicated for other songs starting with a tonic. If a song starts with a tonic, music teachers can pay attention to a key self-selected by Primary school children, as the key may be slightly sharper than another key that they can sustain better. In that case, music teachers may suggest a new key that is one to two semitones flatter than a key selected by children. Another inference is that being able to transpose the key of a song for children is a useful skill for the music teacher, as this would allow children to experience a greater sense of being in-tune, depending on the choice of key, as well as enhancing their growing sense of tonality and key structure.

Secondly, since participants expanded their vocal range when matching a series of upward intervals, it is assumed that music teachers can use a series of upward and downward musical intervals to expand children's vocal range. Furthermore, the significantly positive correlation between the number of keys used to sing any of two target songs reported in the current study suggests that the number of key centres applied to sing one target song by children can predict those used to sing another. As a result, it suggests that music teachers can start with a simple song, and use a small number of songs with different levels of complexity to train greater sung key consistency.

Thirdly, in terms of vocal pitch-matching accuracy for musical intervals, the current study found that big musical intervals were more likely to be matched poorly, even for older Primary school children. Furthermore, there was a tendency for Primary school participants to match a series of upward or downward intervals towards a lift point poorly. These two types of errors might be because the strategies the children used to transfer between vocal registers were limited. Consequently, it seems necessary for a music teacher to guide children to practise transferring between vocal registers before attempting to match melodies with big musical intervals and a series of upward or downward intervals towards a lift point.

In addition, the current study found that many participant children matched the first pitch of an ascending large musical interval sharp when singing a song. They also tended to repeat their vocal pitch-matching errors when vocally matching a song. Therefore, if observed in their own pupils, music teachers can focus on correcting these mistakes to improve Primary school children's vocal pitch-matching accuracy for songs.

Finally, two of the three target songs are popular across both China and the UK, and Primary school children across the two countries showed similar singing behaviour for these. Consequently, it seems that the above suggestions are not only suitable for music teachers working in Primary schools in China, but can also apply to music teachers from other countries which use the system of equal temperament.

(b) Application of General Singing Behaviour Reported in the Current Study for Music Teachers in China

While the mean of singing behaviour for the current study's participants is close to the related Chinese national report based on singing performances of Grade 4 students across the country (MCCEMEC, 2020), the current study's data also shows Chinese Primary school children's singing behaviour from a wider age range. This new data may, therefore, be more representative of the mean singing behaviour for Primary school children who come from other provinces in China whose GDP is similar to Hunan Province. If that is the case, the current data may provide information for these music teachers to understand Primary school children's singing behaviour (including vocal register use and vocal pitch-matching accuracy for songs) from different perspectives, including varied age groups, sexes and socioeconomic status (as measured by geographic location and inferred family income). As well as for an experienced music teacher, these findings may be equally useful for a future music teacher, or a music teacher who is just starting work, or one who usually teaches only one or two Primary school Grades, to see Primary school children's general singing behaviour from a developmental point of view in different contexts.

8.3 Limitations of the Current Study

In the current section, the limitations of the current study's selected geographic areas, measurements and other limitations are discussed below.

8.3.1 The Limitation of Selected Areas and Definition of Parents' Income

As the current data were collected from Hunan Province only, where the GDP is close to the median level across the country (see Chapter 1, Figure 1.2), data from the province may be not appropriate to show the singing behaviour of Primary school children from the top or bottom GDP areas in the country. Furthermore, while limited cultural difference in Primary school children's singing behaviour was found between Chinese and English Primary school children, it is unclear whether this would be the case in a third culture, especially one where a music curriculum is not provided in children's school curriculum and one which does not use equal temperament.

It should be noted that the inferred family income was defined by any information of parents' occupations in the area around the school, instead of the real family income. This was because the information on actual family income was unavailable. Furthermore, some data were missed for the information of parent's occupation, although it had been asked for in a form designed to collect participant children's information (see Appendix B). For instance, some participants from School F did not report their parents' occupations as they reported that they were not allowed to do so. Furthermore, it was discovered that many rural participants did not know their parents' occupation. The inferred-family income reported in the current study was based on the researcher's informal conversations with participants, other students in the schools, school teachers and other staff working in the schools. It is recognised that further research is needed to explore the possible impact of the family income variable in more detail as this is a limitation in the current study.

8.3.2 The Limitations of Measurements

The current section focuses on the limitations of the *SVDM* scale and the *MAPM* system.

Firstly, it is not clear whether the highest rating, '5' of the *SVDM* scale, means singing a song with a wide range (such as an octave), or using a 'head' register higher than Bb4. Although this uncertain definition of '5' in the scale seemed to have limited negative influence when measuring young participants' singing, as their vocal range, pitch height and lift points were more likely to match those suggested by Rutkowski (2015, 2018), the problem was more obvious when measuring singing performances of older participants whose voice had started to undergo voice change through puberty in the current study.

For instance, if a participant (usually in Grade 6) could match pitches of songs (especially in *Little Donkey*) from B3 to B4 with a similar timbre of the singing voice, it was difficult to decide which ratings, '4' or '5', should be given. It was more likely to be rated as '4' (suggested from C4 to A4 in Rutkowski, 1996, 2015, 2018) as this singing performance seems to use an 'initial singing range'. However, it could be measured as '5', as the pitch range (B3-B4) is slightly beyond the suggested lift point, Bb4. Consequently, these suggested pitches provided less help in deciding on older participants' ratings. This confusion has brought some uncertainty to data measured by the *SVDM* scale, because the scale involves two musical elements: vocal registers and vocal range. The conflict between these two concepts might be small for young children, but it would be more obvious for older children, especially boys in their puberty period.

In terms of the *MAPM* system, although it can reveal many characteristics of vocal pitch-matching accuracy for songs, it is a time-consuming process to get the final picture of vocal pitch-matching accuracy for each of the vocally matched pitches of a song, especially for a large group of participants. Furthermore, the *MAPM* system is designed for equal temperament that can be shown with the visual piano keyboard of Sing&See. It may be not suitable for analysing vocal pitch-matching accuracy of a song under the system of pure intonation (e.g., traditional Indian music), as pitches cannot then be shown in the visual piano of Sing&See. However, once the pitches' errors are marked based on pure intonation, it is assumed that the idea of the colour characteristics of the *MAPM* system can be applied to show the degree of pitch errors. In this case, the original semitone errors may need to be updated based on the pure intonation system. It may be that, in the future, the *MAPM* system can be automated through the use of IT which would speed up the transcription process.

8.3.3 Other Limitations

The current study also has other limitations. For instance, only English and Chinese previous studies were reviewed. Consequently, related studies published in other languages, such as Russian and Japanese, could not be accessed. Furthermore, the study's *N* = 1,539 singing performances were singing in unaccompanied solo. Participants' individual singing performances reported in the current study may be different from a group singing, or singing with an accompaniment.

8.4 Future Work

It is suggested that future studies can pay attention to the four following perspectives.

8.4.1 Future Studies Related to Analysing Vocally Matched Pitches of Songs by the MAPM System

The current study only uses the *MAPM* system to show some of the details of vocal pitch-matching accuracy for songs. Future studies can use the same approach to explore participants' vocal pitch-matching accuracy for target songs in more detail. The current section summarises five possible related research topics that can be analysed directly using the current data based on the *MAPM* system.

- Firstly, five common characteristics of vocal pitch-matching errors for the three target songs reported in Chapter 5, Section 5.4 can be explored further by independent variables, such as age, sex, and socioeconomic status, to understand better the development of these musical elements by bio-development and environment.

- Secondly, it seems that unison can help any following musical interval be better matched, as when observing the colour characteristics of the (in this instance) three target songs. Although this has not been summarised here, it is worth exploration.
- Thirdly, to respond to the study of Nichols et al. (2023), one future study can explore the influence of an octave, such as of *Happy Birthday,* on the vocal pitch-matching accuracy for following pitches and musical intervals to provide evidence based on Chinese children's singing.
- Fourthly, the influence of height on vocal pitch-matching accuracy for the same types of musical intervals when singing target songs can be explored in the future.
- Fifthly, future studies can explore the influence of melodic shape (upward and downward musical intervals) on children's vocal pitch-matching accuracy for songs.

Furthermore, as many pitch errors can be revealed by observing vocal pitch-matching errors for songs using the *MAPM* system, it may be then important to explore in future studies how to solve each of these problems in real singing lessons.

8.4.2 Future Studies Related to Overcoming Limitations of Measurement

As discussed in Section 8.3.2, the *SVDM* scale (Rutkowski, 1996, 2015, 2018) seems adequate to measure young children, but it may be difficult to use it to measure register use in older children due to their broader and lower pitch range. For instance, when an older boy whose voice was changing in puberty sang a target song accurately from D3 to D4, he might be given a low score '2.5' according to the current *SVDM* scale, because most of the matched pitches were below F4. However, the pitch range D3 to D4 may be also his initial comfortable singing range, which would be rated as '4' based on Rutkowski's scale. In order to improve the reliability of findings reported in the current study, *n* = 12 children whose voice changed and used a low pitch range to sing were not included in the current study. However, their vocal register use can be explored in the future.

Further studies can explore the details of pitch height, vocal range and lift points and their interactions for students whose voices are undergoing change during puberty in order to see how to adapt the current *SVDM* scale to match the emerging vocal register use of such students. When considering updating the *SVDM* scale, it should be considered whether specific pitches, such as D4 and A4, should be included in the description of the scale. These common pitches might help a judge make a quicker decision about a specific rating.

Furthermore, due to the confusion raised by the inclusion of two musical concepts (the application of vocal registers and vocal range) in the *SVDM* scale, a future study can explore whether the *SVDM* scale can apply just one of the concepts more explicitly to reduce any such confusion when measuring singing performance by the scale.

As mentioned above, in terms of the limitation of the *MAPM* system reported in Section 8.3.2, a future study can explore how to collect vocally matched pitches or calculate semitone errors of each vocally matched pitch of a song for a participant automatically. Although this may have to be based on the key set by the starting pitch, it can be transferred later based on a prevailing key. Future studies can also use the *MAPM* system to test the reliability of other rating scales designed to test people's vocal pitch-matching accuracy for songs.

8.4.3 Future Studies Related to a Wider Area and Age Groups

Future studies can explore the interaction among age, geographic location and income to see the influence of age on children's singing behaviour analysed by the interaction between geographic location and income.

The methodology of the current study can be transferred to studies undertaken in other provinces or areas in China to test the possible generalisability of its findings. GDP in these areas can be in the country's top, median or bottom levels. Such a transfer may help to test the validity of the findings reported in the current study which was based on *N* = 1,193 Primary school children in Hunan Province, China. Furthermore, a study using the same methodology can be applied to test the singing behaviour of Primary school children from a minority group in China to understand and compare their singing behaviour in a different cultural context.

Furthermore, the methodology of the current study can be applied to test the singing behaviour of younger (e.g., kindergarten) or older (Secondary or High school) children to explore a wider picture of the development of people's singing behaviour before they become adults. It can also be applied to test adults' singing behaviour to create a whole picture of human singing behaviour across life.

8.4.4 Exploring Other Unanalysed Data Collected in the Current Study

Owing to limited space, another three perspectives, the vocal range of the three target songs, speaking pitch centre and attitudes towards singing for N = 1,193 participants, are not reported in the current doctoral thesis. Analysing these already collected data in the future could help to understand the findings reported in the current study better.

Finally, other issues were raised during the singing test. Some second Graders in school D wanted to continue singing a favourite song after the test. Their extra self-selected singing performances were also audio recorded, and the other children usually kept quiet during the process, although in recess. Although these recordings were not formally analysed as they fell outside the parameters of the study, they provide a snapshot of these children's self-selected singing behaviour. The author found that most of the young participants who sang an extra song preferred to sing songs with a wider pitch range and more complicated musical intervals than the three target songs. The current author was told that they had learned these songs from their parents. However, these difficult songs tended to make their vocal pitch-matching accuracy poorer than the selected test items. Future studies can explore the relationship between Primary school children's singing behaviour for target songs and for self-selected songs.

References

Aherne, S. (2011). *An examination of the relationships among home musical environment, tonal music aptitude, and vocal performance achievement of kindergarten students* [Doctoral thesis, The University of Delaware]. https://doi.org/ 10.1017/S0265051715000364

Alberston, K. (1979). Teaching pronunciation with an oscilloscope. *Journal of Language Learning Technologies*, *13*. 25-30.

Anderson, J. D. (1991, Winter). Children's song acquisition: An examination of current research and theories. *The Quarterly, 2*(4). 42-49.

Apfelstadt, H. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of Kindergarten children. *Journal of Research in Music Education, 32*(1). 15-24. https://doi.org/10.2307/3345277

Arora, V., Jaiswal, S., Raina, A., & Kumar, S. (2023). Automatic detection and analysis of singing mistakes for music pedagogy. *Journal of Latex Class Files, 14*(8). 3-9. https://doi.org/10.36227/techrxiv.23269502.v1

Asztalos, A. (2021). Development of children's singing voice. *Studia Universitatis Babes-Bolyai Musica, 66*(1), 39-54.

Atterbury, B. W., & Silcox, L. (1993). The effect of piano accompaniment on Kindergartners' development singing ability. *Journal of Research in Music Education, 41*(1). 40-47. https://doi.org/10.2307/3345478

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia A group study of adults afflicted with a music-specific disorder. *Brain, 125*. 238-251.

Baken, R. J. (1998). An overview of laryngeal function for voice production. In R. T. Sataloff (Ed.), Vocal *health and pedagogy* (pp. 27-47). San Diego, Singular Publishing Group Inc.

Barrett, M. S. (2019). Singing and invented song-making in infants' and young children's early learning and development: From shared to independent song-making. In G. F. Welch, D. M. Howard & J. Nix (Eds.), *The Oxford handbook of Singing*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199660773.013.42

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

Bentley, A. (1968). *Monotone*. Music Education Research Papers No. 1. London: Novello.

Bentley, A. (1969). Measurement and development of musical abilities: Some research interests and findings. *Journal of Research in Music Education*, *17*(1). 41-46.

Bernardoni, N. H., Smith, J., & Wolfe, J. (2014). Vocal tract resonances in singing: Variation with laryngeal mechanism for male operatic singers in chest and falsetto registers. *The Journal of the Acoustical Society of America, 135*(1). 491-501. https://doi.org /10.1121/1.4836255

Bhachu, D. K. (2019). *Facilitating musical learning in Scottish Primary schools: An interview-based study of generalist Primary teachers', Primary music specialists' and community music practitioners' views and experiences* [Doctoral thesis, The University of Edinburgh]. https://hdl.handle.net/1842/36622

Boardman, E. L. (1964). *An investigation of the effect of pre-school training on the development of vocal accuracy in young children* [Doctoral thesis, The University of Illinois]. ProQuest Dissertations Publishing.

Boseley, M. E., & Hartnick, C. J. (2006). Development of the human true vocal fold: Depth of cell layers and quantifying cell types within the lamina propria. *Annals of Otology, Rhinology & Laryngology, 115*(10). 784-788.

Bradshaw, E., & McHenry, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice, 19*(3). 431-439. http://dx.doi.org /10.1016/j.jvoice.2004.07.010

British Educational Research Association [BERA]. (2018). Ethical guidelines for educational research (4th ed.). London.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *The American Psychologist, 32*(7). 513-531.

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *The American Psychologist. 34*(10). 844-850.

Brophy, T. S. (1997). Authentic assessment of vocal pitch accuracy in First through Third Grade children. *Contributions to Music Education, 24*(1). 57-70.

Buckton, R. (1977). A comparison of the effects of vocal and instrumental instruction on the development of melodic and vocal abilities in young children. *Psychology of Music, 5*(1). 36-47. https://doi.org /10.1177/030573567751006

Caro, D. H., McDonald, J. T., & Willms, J. D. (2009). Socio-economic status and academic achievement trajectories from childhood to adolescence. *Canadian Journal of Education, 32*(3). 558-590.

Cheadle, J. E. (2008). Educational investment, family context, and children's math and reading growth from Kindergarten through the Third Grade. *Sociology of Education. 81*(1). 1-31. https://www.jstor.org/stable/20452721

Chen-Hafteck, L. (1999a). Discussing text-melody relationship in children's song-learning and singing: A Cantonese-speaking perspective. *Psychology of Music. 27*(1). 55-70. https://doi.org /10.1177/03057356992710

Chen-Hafteck, L. (1999b). 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

Cobes, C. J. (1969). *The conditioning of a pitch response using uncertain singers* [Doctoral thesis, The Pennsylvania State University]. ProQuest Dissertations Publishing.

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

Cooper, N. A. (1992). *Selected factors related to children's singing accuracy* [Doctoral thesis, Indiana University]. ProQuest Dissertations Publishing.

Cooper, N. (1995). Children's singing accuracy as a function of grade level, gender, and individual versus unison singing. *Journal of Research in Music Education, 43*(3). 222-231. https://doi.org/10.2307/3345637

Cuadrado, A., & Rusinek, G. (2016). Singing and vocal instruction in Primary schools: An analysis from six case studies in Spain. *British Journal of Music Education, 33*(1). 101-115. https://doi.org/10.1017/S0265051715000273

Cuddy, L. L., Balkwill, L. L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: A study of "tone deafness" among university students. *Annals of the New York Academy of Sciences, 1060*(1). 311-324. https://doi.org /10.1196/annals.1360.026

Dalla Bella, S., & Berkowski, M. (2009). Singing proficiency in the majority: Normality and "phenotypes" of poor singing. *Annals of the New York Academy of Sciences, 1169*. 99-107. https://doi.org/10.1111/j.1749-6632.2009.04558.x

Daniélou, A. (1995). *Music and the power of sound: The influence of tuning and interval on consciousness.* Inner Traditions. (Rev. Ed. of *Introduction to the study of musical scales*. India Society, 1943).

Daniélou, A. (1999). Introduction to the study of musical scales. *Oriental Book Reprint Corporation*.

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.

Davies, A. D. M., & Roberts, E. (1975). Poor pitch singing: A survey of its incidence in school children. *Psychology of Music, 3*(2). 24-36. https://doi.org/10.1177/030573567532004

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.

Demorest, S. M., Kelley, J., & Pfordresher, P. Q. (2017). Singing ability, musical self-concept, and future music participant. *Journal of Research in Music Education, 64*(4). 1-16. https://www.jstor.org/stable/44631463

Demorest, S., Nichols, B., & Pfordresher, P. Q. (2018). The effect of focused instruction on young children's singing accuracy. *Psychology of Music, 46*(4). 488-499. https://doi.org/10.1177/0305735617713120

De Yarman, R. M. (1971). *An experimental analysis of the development of rhythmic and tonal capabilities of Kindergarten and First Grade children* [Doctoral thesis, The University of Iowa]. ProQuest Dissertations & Theses Global.

Deutsch, D. (2013). Absolute Pitch. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 141-182). Academic Press. https://doi.org /10.1016/ B978-0-12-381460-9.00005-5

Dittemore, E. E. (1968). *An investigation of some musical capabilities of elementary school students* [Doctoral thesis, The University of Iowa]. ProQuest Dissertations Publishing.

Drexler, E. N. (1938). A study of the development of the ability to carry a melody at the Preschool level. *Child Development, 9*(3). 319-33. https://www.jstor.org/stable/1125444

Elmer, S. S., & Elmer, F. J. (2000). A new method for analysing and representing singing. *Psychology of Music, 28*(1). 23-42. https://doi.org /10.1177/0305735600281003

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

Etopio, E. A. (2009). *Characteristics of early musical environments associated with preschool children's music skills* [Doctoral thesis, The State University of New York at Buffalo]. ProQuest Dissertations Publishing.

Ettekal, A., & Mahoney, J. L. (2017). Ecological systems theory. In K. Peppler (Ed.), *The SAGE encyclopedia of out-of-school learning* (pp. 239-241). SAGE Publications, Inc.

Flowers, P. J., & Dunne-Sousa, D. (1990). Pitch-pattern accuracy, tonality, and vocal range in Preschool children's singing. *Journal of Research in Music Education, 38*(2). 102-114. https://doi.org/10.2307/3344930

Frederickson, K. B. (1992). *The relationship of spatial ability and encoding ability to Kodaly hand signs and singing performance* [Doctoral thesis, Arizona State University]. ProQuest Dissertations Publishing.

Geringer, J. M. (1983). The relationship of pitch-matching and pitch-discrimination abilities of Preschool and Fourth-Grade students. *Journal of Research in Music Education, 31*(2). 93-99. https://doi.org/10.2307/3345213

Geringer, J. M., Nelson, J. K., & Kostka, M. C. (1980). Differential assessment of child and adult singing ranges. *Contributions to Music Education, 8*(8), 39-46. https://www.jstor.org/stable/24127432

Goetze, M. (1985). *Factors affecting accuracy in children's singing* [Doctoral thesis, University of Colarado at Boulder]. ProQuest Dissertations Publishing.

Goetze, M., & Horii, Y. (1989). A comparison of the pitch accuracy of group and individual singing in young children. *Bulletin of the Council for Research in Music Education, 99*. 57-73.

Gordon, E. (1971). *The psychology of music teaching.* Prentice Hall.

Gould, A. O. (1969). Developing specialized programs for singing in the Elementary school. *Bulletin of the Council for Research in Music Education, 17*(17). 9-22. https://www.jstor.org/stable/40317056

Gratton, M. (1989). The effect of three vocal models on uncertain singers' ability to match and discriminate pitches [Master's thesis, McGill University]. ProQuest Dissertations Publishing.

Green, G. A. (1994). Unison versus individual singing and elementary students' vocal pitch accuracy. *Journal of Research in Music Education, 42*(2). http://www.jstor.com/stable/3345495

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

Guilbault, D. M. (2004). The effect of harmonic accompaniment on the tonal achievement and tonal improvisations of children in kindergarten and first grade. *Journal of Research in Music Education, 52* (1). 64-76.

Fieldhouse, A. E. (1937). *A study of backwardness in singing among schoolchildren.* [Doctoral thesis, University of London].

Hallam, S. (2010). The power of music: Its impact on the intellectual, social and personal development of children and young people. *International Journal of Music Education, 28*(3). 269-289. https://doi.org/10.1177/0255761410370658

Harkey, B. L. (1978). *The identification of and the training of vocal range of three-year-old Preschool children* [Doctoral thesis, The Louisiana State University]. ProQuest Dissertations Publishing.

Hallam, S., & Himonides, E. (2022). *The power of music*. Open Book Publishers.

Hedden, D. G., & Baker, V. A. (2010). Perceptual and acoustical analyses of second graders' pitch-matching ability in singing a cappella or with piano accompaniment. *Bulletin of the Council for Research in Music Education, 184*. 35-48.

Henley, J. (2017). How musical are Primary generalist student teachers? *Music Education Research, 19*(4). 470-484. https://doi.org/10.1080/14613808.2016.1204278

Hennessy, S. (2000). Overcoming the red-feeling: The development of confidence to teach music in Primary school amongst student teachers. *British Journal of Music Education, 17*(2). 183-196. https://doi.org/10.1017/S0265051700000243

Hennessy, S. (2012). Improving Primary teaching: Minding the gap. In G. E. McPherson & G. F. Welch (Eds.), *The Oxford Handbook of Music Education, Volume 2* (pp. 625-628). Oxford University Press. http://doi.ord/10.1093/oxfordhb/9780199928019.013.0046_update_001

Henrich, N. (2006). Mirroring the voice from Garcia to the present day: Some insights into singing voice registers. *Logopedics Phoniatrics Vocology, 31*(1). 3-14. https://doi.org/10.1080/14015430500344844

Henrich, N., d'Alessandro, C., Doval, B., & Castellengo, M. (2005). Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *The Journal of the Acoustical Society of America, 117*(3). 1417-1430. https://doi.org/10.1121/1.1850031

Hirano, M. (1988). Vocal mechanisms in singing: Laryngological and phoniatric aspects. *Journal of Voice, 2*(1). 51-69.

Hirano, M. Kurita, S., & Nakashima, T. (1983). Growth, development and aging of human vocal folds. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology.* College-Hill Press.

Holden, H., & Button, S. (2006). The teaching of music in the Primary school by the non-music specialist. *British Journal of Music Education, 23*(1). 23-38. https://doi.org/ 10.1017/S0265051705006728

Hollien, H. (1974). On vocal registers. *Journal of Phonetics, 2*(2). 125-143. https://doi.org/10.1016/S0095-4470(19)31188-X

Honda, C., & Pfordresher, P. Q. (2023). An acoustic analysis and comparison between remotely collected and in-person laboratory collected data in vocal imitation of pitch. *Auditory Perception & Cognition*, *Volume 6* (pp. 1-18). https://doi.org/10.1080/25742442.2023.2210050

Hornbach, C. M., & Taggart, C. C. (2005). The relationship between developmental tonal aptitude and singing achievement among Kindergarten, First-, Second-, and Third-Grade students. *Journal of Research in Music Education, 53*(4). https://www.jstor.org/stable/3648430

Houlahan, M., & Tacka, P. (2015). *Kodaly today (Second edition)*. Oxford University Press.

Howard, D. M., & Welch, G. F. (1993). Visual displays for the assessment of vocal pitch matching development. *Applied Acoustics, 39*(4). 235-252. https://doi.org/10.1016/0003-682X(93)90008-T

Hutchins, S., & Peretz, I. (2012). Amusics can imitate what they cannot discriminate. *Brain and Language, 123*(3). 234-239. https://doi.org/10.1016/j.bandl.2012.09.011

Iacarino, T. K. (2018). Utilizing learning modalities to increase pitch-matching abilities of Fourth

Grade students [Master's thesis, Goucher College].

Illari, B. (2006, August 22-26). *Mapping musical development in Brazil: Children's musical practices in Maranhao and Para* [Paper presentation]. 9th International Conference on Music Perception and Cognition.

Jacobi-Karna, K. L. (1996). *The effects of the inclusion of text on the singing accuracy of Preschool children* [Doctoral thesis, The University of Arizona]. ProQuest Dissertations Publishing. https://www.jstor.org/stable/40318837

Jersild, A. T., & Bienstock, S. F. (1934). A study of the development of children's ability to sing. *The Journal of Educational Psychology, 25*(7). 481-503.

Jones, M. (1971). A pilot study in the use of a vertically-arranged keyboard instrument with the uncertain singer. *Journal of Research in Music Education, 19*(2). 183-194. https://doi.org/10.2307/3343822

Jones, M. (1974). *A study of the use of a vertically-arranged keyboard instrument with the uncertain singer*. [Doctoral thesis, The Florida State University]. ProQuest Dissertations Publishing.

Jones, M. (1979). Using a vertical-keyboard instrument with the uncertain singer. *Journal of Research in Music Education, 27*(3), 173-184.

Joyner, D. R. (1969). The monotone problem. *Journal of Research in Music Education, 17*(1), 115-124. https://doi.org/10.2307/3344198

Joyner, D. R. (1971). Pitch discrimination and tonal memory and their association with singing and the larynx [Master's thesis, University of Reading].

Kalmus, H., & Fry, D. B. (1980). On tune deafness (dysmelodia): frequency, development, genetics and musical background. *Annals of Human Genetics, 43*(4). 369-382. https://doi-org /10.1111/j.1469-1809.1980.tb01571.x

Kayes, G. (2015). Structure and function of the singing voice. In G. F. Welch, D. M. Howard & J. Nix (Eds.), *The oxford handbook of singing*. Oxford University Press.

Keidar, A., Hurtig, R. R., & Titze, I. R. (1987). The perceptual nature of vocal register change. *Journal of Voice*, *1*(3). 223-233. https://doi.org/10.1016/S0892-1997(87)80004-8

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

Kirkpatrick, W. C. (1962). *Relationships between the singing ability of Prekindergarten children and* *their home musical environment* [Doctoral thesis, University of Southern California]. ProQuest Dissertations Publishing.

Knight, S. D. (2010). *A study of adult 'non-singers' in Newfoundland* [Doctoral thesis, University of London].

Kochis-Jennings, K. A., Finnegan, E. M., Hoffman, H. T., & Jaiswal, S. (2012). Laryngeal muscle activity and vocal fold adduction during chest, chestmix, headmix, and head registers in females. *Journal of Voice, 26*(2). 182-193. https://doi.org/10.1016/j.jvoice.2010.11.002

Kochis-Jennings, K. A., Finnegan, E. M., Hoffman, H. T., Jaiswal, S., & Hull, D. (2014). Cricothyroid muscle and thyroarytenoid muscle dominance in vocal register control: Preliminary results. *Journal of Voice, 28*(5). 652. e21-652. e29. https://doi.org/10.1016/j.jvoice.2014.01.017

Krause, D. W. (1983). *The positive use of conducting stance and motion to affect vocal production and to assist musicality in the training of children's choirs* [Doctoral thesis, University of Southern California]. ProQuest Dissertations Publishing.

Large, J. W. (1969). Acoustical study of Isoparametric tones in the female chest and middle registers in singing. *The Journal of the Acoustical Society of America, 45*(1). 314-314. https://doi.org/10.1121/1.1971679

Lecanuet, J. P. (1996). Prenatal auditory experience. In I. Deliège & J. A. Sloboda (Eds.), *Musical beginnings: Origins and development of musical competence* (pp. 3-34). Oxford University Press.

Lecanuet, J. P., Granier-Deferre, C., Jacquet, A. Y., & Busnel, M. C. (1992). Decelerative cardiac responsiveness to acoustical stimulation in the near term fetus. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology, 44*(3-4), 279-303.

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. (1987). *An experimental study of the comparative effects of singing songs with words and without words on children in Kindergarten and First Grade* [Doctoral thesis, University Microfilms International]. ProQuest Dissertations Publishing.

Levinowitz, I. (1989). An investigation of preschool children's comparative capability to sing song with and without words. *Bulletin of the Council for Research in Music Education, 100*. 14-19.

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).

Liao, M. Y., & Davidson, J. W. (2007). The use of gesture techniques in children's singing. *International Journal of Music Education, 25*(1). 82-94. https://doi.org /10.1177/0255761407074894

Liao, M. Y., & Davidson, J. W. (2016). The effects of gesture and movement training on the intonation of children's singing in vocal warm-up sessions. *International Journal of Music Education, 34*(1). 4-18. https://doi.org /10.1177/0255761415614798

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology, 18*. R331-332.

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.

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

Marsh, K. (2012). Commentary: Music learning and teaching during childhood: Ages 5-12. In G. E. McPherson & G. F. Welch (Eds.), *The oxford handbook of music education*, Volume 1 (pp. 317-321). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199730810.013.0019

Martin, A. (2006). You sing like a girl? An exploration of 'boyness' through the treble voice. *Sex Education, 6*(2), 193-205.

Martin, B. A. (1991). Effects of hand signs, syllables, and letters on First Graders' acquisition of tonal skills. *Journal of Research in Music Education, 39*(2), 161-170.

McAllister, A., Sederholm, E., & Sundberg, J. (1993). Acoustic and perceptual analysis of vocal registers in children. *STL-QPSR*, *34*. 29-34.

McGlone, R. E., & Brown, W. S. (1969). Identification of the "shift" between vocal registers. *The Journal of the Acoustical Society of America*, *46*. 1033-1036.

McGraw, A. G. B. (1970). *An assessment of the effectiveness of vocalises in training Elementary school children to sing using head voice* [Doctoral thesis, The University of Georgia].

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

Michaud, B. G. (2014). *Male music teachers and singing fathers: Effects on and correlations with kindergarten children's singing abilities* [Doctoral thesis, Boston University]. ProQuest Dissertations Publishing.

Miller, K. (2020). *Song teaching and singing accuracy of Third Grade Elementary music students: An investigation using multilevel modeling* [Doctoral thesis, University of Kentucky]. ProQuest Dissertations Publishing.

Mizener, C. P. (1993). Attitudes of children toward singing and choir participation and assessed singing skill. *Journal of Research in Music Education, 41*(3). 233-245. https://doi.org/10.2307/3345327

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.

Moore, R. S. (1994). Effects of age, sex, and melodic/harmonic patterns on vocal pitch-matching skills of talented 8-11-year-olds. *Journal of Research in Music Education, 42*(1), 5-13.

Moore, R. S., Fyk, J., Frega, A. L., & Brotons, M. (1995/1996). Influences of culture, age, gender and two-tone melodies on interval matching skills of children from Argentina, Poland, Spain and the USA. *Bulletin of the Council for Research in Music Education, 127*, 127-135.

Moore, R. E., Keaton, C., & Watts, C. (2005). Role of pitch memory in pitch matching and pitch discrimination. *ASHA Leader, 10*(10), 4-4.

Moore, R. E., Estis, J., Gordon-Hickey, & S. Watts, C. (2008). Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *Journal of Voice, 22*(4), 399-407.

Murry, T. (1990). Pitch-matching accuracy in singers and nonsingers. *Journal of Voice, 4*(4), 317-321.

Nichols, B. E. (2013). *Task-based variability in children's singing accuracy* [Doctoral thesis, University of Washington]. ProQuest Dissertations Publishing.

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

Nichols, B., Hua, A., & Wang, Z. L. (2023). Defining "happy" in happy birthday: Singing accuracy a construct based on range and intervals. *Psychology of Music, 51*(4). 1414-1423.

Ockelford, A. (2000). Music in the education of children with severe or profound learning difficulties: Issues in current U.K. provision, a new conceptual framework, and proposals for research. *Psychology of Music, 28*(2). 197-217. https://doi-org/10.1177/0305735600282009

Pabon, P., Stallinga, R. Sodersten, M., & Ternstrom, S. (2014). Effects on vocal range and voice quality of singing voice training: The classically trained female voice. *Journal of Voice*, *28*(1). 36-51. https://doi.org/10.1016/j.jvoice.2013.06.005

Papageorgi, I., Saunders, J., Himonides, E., & Welch, G. (2022). Singing and social identity in young children. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.823229

Paulsen, E. (1895). Ueber die singstimme der kinder: Nach untersuchungen an den Kieler städtischen Schulen. *Pflugers Archiv, 61*(7-8). 407-426. https://doi.org/10.1007/BF01662068

Pedersen, D. M., & Pedersen, N. O. (1970). The relationship between pitch recognition and vocal pitch production in Sixth-Grade students. *Journal of Research in Music Education, 18*(3). 265-272. https://www.jstor.org/stable/3344466

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 syliable. *Music Perception, 36*(5). 468-479.

Peretz, I., Ayotter, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33*(2). 185-191. Https://doi.org/10.1016/S0896-6273(01)00580-3

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders the Montreal battery of evaluation of amusia. *Annals of the New York Academy of Science, 999*. 58-75. https://doi.org /10.1196/annals.1284.006

Persellin, D., & Bateman, L. (2009). A comparative study on the effectiveness of two song-teaching methods: Holistic vs. phrase-by-phrase. *Early Child Development and Care, 179*(6). 799-806. https://doi.org /10.1080/03004430902944841

Petzold, R. G. (1963). The development of auditory perception of musical sounds by children in the first six grades. *Journal of Research in Music Education, 11*(1), 21-43. http://www.jstor.org/stable/3344529

Pfordresher, P. Q. (2022). Singing accuracy across the lifespan. *Annals of the New York Academy of*

Sciences. 1-9. https://doi.org /10.1111/nyas.14815

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception, 25*(2). 95-115. https://doi.org/10.1525/mp.2007.25.2.95

Pfordresher, P., & Demorest, S. M. (2020). The prevalence and correlates of accurate singing. *Journal of Research in Music Education, 69*(1). 5-23. https://doi.org/10.1177/0022429420951630

Pfordresher, P. Q., & Larrouy-Maestri, P. (2015). On drawing a line through the spectrogram: How do we understand deficits of vocal pitch imitation? *Frontiers in Human Neuroscience, 9*. 271-271. https://doi.org/10.3389/fnhum.2015.00271

Phillips, K. H., & Aitchison, R. E. (1997). The relationship of inaccurate singing to pitch discrimination and tonal aptitude among third-grade students. *Contributions to Music Education, 24*(1). 7-22. https://www.jstor.org/stable/24126943

Pitt, J., & Welch, G. F. (2020). Music in early education and care settings for communication and language support. In M. S. Barrett & G. F. Welch (Eds.), *The Oxford handbook of early childhood music learning and development* (351–378). Oxford University Press.

Plumridge, J. M. (1972). *The range and pitch levels of children's voices in relation to published material for children's voices*. [Diploma Dissertation, University of Reading].

Porter, S. Y. (1977). The effect of multiple discrimination training on pitch-matching behaviours of uncertain singers. *Journal of Research in Music Education, 25*(1). 68-82. https://doi.org/10.2307/3344846

Proctor, D. F. (1980). *Breathing, speech, and song.* Springer-Verlag.

Ramsey, J. H. (1981). An investigation of the effects of age, singing ability, and experience with pitched instruments on the melodic perception of Preschool children [Doctoral thesis, The University of Iowa]. ProQuest Dissertations Publishing.

Ramsey, J. H. (1983). The effects of age, singing ability, and instrumental experiences on Preschool children's melodic perception. Journal of Research in Music Education, 31(2). 133-145. https://doi.org/10.2307/3345216

Reynolds, G. E. (1960). *Environmental sources of musical awakening in pre-school children* [Doctoral thesis, The University of Illinois].

Roberts, E., & Davies, D. M. (1975). Poor pitch singing: Response of monotone singers to a program of remedial training. *Journal of Research in Music Education, 23*(4). 227-239. https://doi.org/10.2307/3344852

Roberts, E., & Davies, A. M. (1976). A method of extending the vocal range of “monotone” schoolchildren. *Psychology of Music, 4*(1). 29-43. https://doi-org /10.1177/030573567641004

Roubeau, B., Henrich, N., & Castellengo, M. (2009). Laryngeal vibratory mechanisms: The notion of vocal register revisited. *Journal of Voice, 23*(4). 425-438. https://doi.org/10.1016/j.jvoice.2007.10.014

Runfola, M., Etopio, E., Hamlen, K., & Rozendal, M. (2012). Effects of music instruction on Preschoolers’ music achievement and emergent literacy achievement. *Bulletin of the Council for Research in Music Education,* (192). 7-27. https://www.jstor.org/stable/10.5406/bulcouresmusedu.192.0007

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). *The validity of the ‘Singing Voice Development Measure’ for measuring non-American children’s use of singing voice* [Paper presentation]. MENC: The National Association for Music Education, Minneapolis, MN.

Rutkowski, J. (1997). The nature of children’s singing voices: Characteristics and assessment. In B. A. Roberts (Ed.), *The Phenomenon of Singing* (pp. 201-209). Memorial University Press.

Rutkowski, J. (2014). *The comparative effectiveness of male and female singing models on kindergarten children’s use of singing voice achievement* [Paper presentation]. The NAfME Music Research and Teacher Education National Conference, St. Louis, MO.

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. (2016). *Continued investigation of the effect of a male singing model on kindergarten* *children's use of singing voice achievement* [Paper presentation]. The NAfMe Research and Music Teacher Education biennial conference, Atlanta GA.

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. (2019). Assessing singing voice development. In T. S. Brophy (Eds.), *The Oxford handbook of assessment policy and practice in music education* (pp. 629-652). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190248130.013.64

Rutkowski, H., & Chen-Hafteck, L. (2001). *The singing voice within every child: A cross-cultural comparison of First Graders' use of the singing voice* [Paper presentation]. The ISME Early Childhood Conference, Kingston, Canada.

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

Rutkowski, J., Chuang, M. J., & Gluschankof, C. (2007). *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 presentation]. The Society for Research in Music Education Symposium on Research in Music Education, Lawrence, KS

Rutkowski, J., & Miller, M. S. (2003a). The effect of teacher feedback and modelling on first graders' use of singing voice and developmental music aptitude. *Bulletin of the Council for Research in Music Education, 156*. 1-10.

Rutkowski, J., & Miller, M. S. (2003b). 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.

Rutkowski, J., & Miller, M. S. (2003c). A longitudinal study of elementary children's acquisition of their singing voices. *Update: Applications of Research in Music Education, 22*(1), 5-14.

Saarikallio, S. H., Randall, W. M., & Baltazar, M. (2020). Music listening for supporting adolescents' sense of agency in daily life. *Frontiers in Psychology, 10*. 2911-2911.

https://doi.org/10.3389/fpsyg.2019.02911

Saltari, R., & Welch, G. F. (2022). Exploring the culture of Greek children's musical games in the school playground: An ethnographic study. *Research Studies in Music Education.* https://doi.org/10.1177/1321103X211061978

Sato, K., Hirano, M., & Nakashima, T. (2001). Fine structure of the human newborn and infant vocal fold mucosae. *Annals of Otology, Rhinology & Laryngology, 110*(5). 417-424.

Saunders, J., Papageorgi, I., Himonides, E., Rinta, T., & Welch, G. (2011). *Researching the impact of the National Singing Programme 'Sing Up' in England: Diverse approaches to successful singing in Primary settings*. International Music Education Research Centre, iMerc, London.

Sergeant D.C. (1992). Towards a Specification of Poor Pitch Singing. In T. Murao & G. F. Welch (Eds.), *Onchi and singing development: A cross-cultural perspective (*pp. 63-73). David Fulton ISBN 1-85346-331-0

Sergeant, D. (2015). The developing voice. In G. F. Welch, D. M. Howard & J. Nix (Eds.), *The Oxford handbook of singing* (pp. 189-239). Oxford University Press.

Sethares, W. A. (2005). *Tuning, timbre, spectrum, scale* (2nd ed.). Springer.

Shelton, J. S. (1965). *The influence of home musical environment upon musical response of First-Grade children* [Doctoral thesis, George Peabody College for Teachers]. ProQuest Dissertations Publishing.

Shipp, T. (1975). Studies of larynx position in singing. *The Journal of the Acoustical Society of America, 58*(S1), S95-S95.

Simcox, E., & Allen, G. (1878). Note-deafness. *Mind, 3*(11). 401-404. https://www/jstor.org/stable/2246546

Sinor, E. (1984). *The singing of selected tonal patterns by Preschool children* [Doctoral thesis, Indiana University]. ProQuest Dissertations Publishing.

Sirin, S. R., (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3). https://www/jstor.org/stable/3515987

Smale, M. J. (1987). *An investigation of pitch accuracy of four- and five-year-old singers* [Doctoral thesis, University of Minnesota]. ProQuest Dissertations Publishing.

Smith, R. B. (1963). The effect of group vocal training on the singing ability of nursery school children.

Journal of Research in Music Education, 11(2). 137-141. https://doi.org/10.2307/3344153

Smith, R. S. (1973). *Factors related to children's in-tune singing abilities* [Doctoral thesis, West Virginia University]. ProQuest Dissertations Publishing.

Sundberg, J. (2018). The singing voice. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception*. Oxford University Press.

Svec, C. L. (2018). The effects of instruction on the singing ability of children ages 5 to 11: A meta-analysis. *Psychology of Music, 46*(3). 326-339. https://doi.org/10.1177/0305735617709920

Švec, J. G., Schutte, H. K., & Miller, D. G. (1999). On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges. *The Journal of the Acoustical Society of America, 106*(3). 1523-1531. https://doi.org/10.1121/1.427149

Tamburri, N., Sheets, D., Halliday, D., Smith, A., & MacDonald, S. (2021). Investigating the cortical correlates of singing: Potential neural benefits of choir for persons with dementia. *Innovation in Aging, 5*(Supplement_1), 129. https://doi.org/10.1093/geroni/igab046.498

Thurman, L., Welch, G., Theimer, A., & Klitzke, C. (2004, October 6-9). *Addressing vocal register discrepancies: An alternative, science-based theory of register phenomena* [Paper presentation]. Second International Conference the Physiology and Acoustic of Singing National Centre for Voice and Speech, Denver, Colorado, USA.

Tillotson, J. R. (1972). *A study of learning characteristics as identified in the music reading process* [Doctoral thesis, Northwestern University].

Titze, I. R. (1979). A physiological interpretation of vocal registers. *The Journal of the Acoustical Society of America, 66*(S1). S56-S56. https://doi.org/10.1121/1.2017833

Titze, I. R. (1988). A framework for the study of vocal registers. *Journal of Voice, 2*(3). 183-194. https://doi.org/10.1016/S0892-1997(88)80075-4

Titze, I. R. (1994). Mechanical stress in phonation. *Journal of Voice, 8*(2). 99-105. https://doi.org/10.1016/S0892-1997(05)80302-9

Titze, I. R., & Sundberg, J. (1992). Vocal intensity in speakers and singers. *The Journal of the Acoustical Society of America, 91*(5). 2936-2946. https://doi.org/10.1121/1.402929

Trehub, S. E. & Gudmundsdottir, H. R. (2019). Mothers as singing mentors for infants. In G. F. Welch, D. Howard & J. Nix (Eds), *The Oxford handbook of singing* (pp. 456-471). Oxford University Press.

Trollinger, V. L. (2003). Relationships between pitch-matching accuracy, speech fundamental frequency, speech range, age, and gender in American English-speaking Preschool children. *Journal of Research in Music Education, 51*(1). 78-94. https://www.jstor.org/stable/3345650

Turoy, A. K. W. (2017). Once a poor pitch singer, always a poor pitch singer? A bottom-up study of factors that may support singing development. *British Journal of Music Education, 35*(1). 91-103. https//doi.org/10.1017/S026505171700016X

Van den Berg, J. W., Vennard, W., Burger, D., & Shervanian, C. C. (1960). *Voice production: The Vibration larynx*. Instructional Film.

Vaudreuil, R. Bronson, H., & Bradt, J. (2019). Bridging the clinic to community: Music performance as social transformation for military service members. *Frontiers in Psychology, 10*. 119-119. https://doi.org/10.3389/fpsyg.2019.00119

Vennard, W., & Hirano, M. (1970). Physiological basis for vocal registers. *The Journal of the Acoustical Society of America, 47*(1A). 120-120. https://doi.org/10.1121/1.1973979

Wassum, S. (1979). Elementary school children's vocal range. *Journal of Research in Music Education, 27*(4). 214-226. https://www/jstor.org/stable/3344709

Waters, E., Ashbolt, R., Gibbs, L., Booth, M., Magarey, A., Gold, L., Kai Lo, S., Gibbons, K., Green, J., O'Connor, T., Garrard, J., & Swinburn, B. (2008). Double disadvantage: The influence of ethnicity over socioeconomic positing on childhood overweight and obesity: Findings from an inner urban population of primary school children. *International Journal of Pediatric Obesity*, *3*(4). 196-204. https://doi.org/10.1080/17477160802141846

Watts, C., Moore, R., & McCaghren, K. (2005). The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *Journal of Voice, 19*(4). 534-543.

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. (1983). *Improvability of poor pitch singing: Experiments in feedback* [Doctoral thesis, University of London]. ProQuest Dissertations Publishing.

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. (1986). A developmental view of children's singing. *British Journal of Music Education, 3*(3), 295-303.

Welch, G. F. (1994). The assessment of singing. *Psychology of Music, 22*(1), 3-19. https://doi.org/10.1177/03057356942210

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. (2006). Singing and vocal development. In G. E. McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 441-461). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198744443.003.0024

Welch, G. F. (2012). The arts and humanities, technology and the 'English Baccalaureate': STEAM not STEM. *Journal of Music, Technology & Education, 4*(2-3), 245-250.

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. (2017). The identities of singers and their educational environments. In R. MacDonald, D. Miell, & D. Hargreaves (Eds.), *Oxford handbook of musical identities* (pp. 543-565). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679485.003.0030

Welch, G. F. (2022). Solo voice. In G. E. McPherson (Ed.). *The Oxford Handbook of Music Performance, Volume 2* (pp. 378-398). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190058869.001.0001

Welch, G. F., Biasutti, M., MacRitchie, J., McPherson, G. E., & Himonides, E. (2020). Editorial: The impact of music on human development and well-being. *Frontiers in Psychology, 11*, 1246.

Welch, G. F., & Henley, J. (2014). Addressing the challenges of teaching music by generalist Primary school teachers. *Revista da ABEM, 22*(32).

Welch, G. F., Himonides, E., Saunders, J., Papageorgi, I., Rinta, T., Preti, C., Stewart, C., Lani, J., Vraka, M., & Hill, J. (2008). *Researching the first year of the National Singing Programme in England: An initial impact evaluation of children's singing behaviours and singer identity*. London: Institute of Education, University of London.

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., & Papageorigi, I. (2010). *Researching the impact of the National Singing Programme 'Sing Up' in England*. International Music Education Research Centre.

Welch, G. F., Himonides, E., Saunders, J., Papageorgi, I., & Sarazin, M. (2014). Singing and social inclusion. *Frontiers in Psychology, 5*, Article 803.

Welch, G. F., Howard, D. M., & Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music, 17*(2), 146-157.

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., Rush, C., & Howard, D. M. (1991). A developmental continuum of singing ability: Evidence from a study of five-year-old developing singers. *Early Child Development and Care, 69*(1), 107-119.

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.

Welch, G. F., Sergeant, D. C., & White, P. J. (1997). Age, sex, and vocal task as factors is singing ''in tune during the first years of schooling. *Bulletin of the Council for Research in Music Education, 133*. 153-160. https://www.jstor.org/stable/40318855

Welch, G. F., Sergeant, D. C., & White, P. (1998). The role of linguistic dominance in the acquisition of song. *Research Studies in Music Education, 10*(1), 67-74. https://doi.org/10.1177/1321103X9801000106

Welch, G. F. & Sundberg, J. (2002). Solo voice. In R. Parncutt & G. McPherson (Eds.), *The science and psychology of music performance: Creative strategies for teaching and learning* (pp. 377-398).

Oxford University Press.

Western, B. A. G. (2002). *Fundamental frequency and pitch-matching accuracy characteristics of First Grade general music students* [Doctoral thesis, The University of Iowa]. ProQuest Dissertations Publishing.

Whidden, C. (2009). *The adult non-singer: Connection, context and culture* [Doctoral thesis, University of Calgary].

Williams, K. E., Barrett, M. S., Welch, G. F., Abad, V., & Broughton, M. (2015). Associations between early shared music activities in the home and later child outcomes: Findings from the longitudinal study of Australian children. *Early Childhood Research Quarterly, 31*. 113-124.

Wilson, M. S. L. (1970). *A study of the child voice from six to twelve* [Doctoral thesis, University of Oregon]. ProQuest Dissertations Publishing.

Wilson, P. H. (2006). *Does real-time visual feedback improve pitch accuracy in singing?* [Master thesis, University of Sydney].

Wise, K. J., & Sloboda, J. A. (2008). Establishing an empirical profile of self-defined ''tone deafness'': Perception, singing performance and self-assessment. *Musicae Scientiae, 12*(1), 3-26. https://doi.org/10.1177/1029864908012001

Wolf, D. L. (2005). A hierarchy of tonal performance patterns for children ages five to eight years in Kindergarten and Primary Grades. *Bulletin of the Council for Research in Music Education, 163.* 61-68. http://www.jstor.com/stable/40311596

Woodward, S. C. (2019). Fetal, neonatal, and early infant experiences of maternal singing. In G. F. Welch, D. M. Howard & J. Nix (Eds.), *The oxford handbook of singing* (pp. 431-454). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199660773.013.41

Wu, S. M. (2005). *A survey of Taiwanese parents' attitudes toward early childhood music education and their participation in music activities at home* [Doctoral thesis, University of Southern California].

Wu, Y. T. (2018). *Musical development of young children of the Chinese diaspora in London* [Doctoral thesis, University College London]. ProQuest Dissertations Publishing.

Wu, Y. T., & Welch, G. F. (2022). Early childhood and musics of the diaspora. In M. S. Barrett & G. F. Welch (Eds.), *The Oxford handbook of early childhood music learning and development* (pp. 414-

438). Oxford University Press.

Wurgler, P. S. (1990). *A perceptual study of vocal registers in the singing voices of children*. [Doctoral dissertation, the Ohio State University]. ProQuest Dissertations Publishing.

Yank Porter, S. (1977). The effect of multiple discrimination training on pitch-matching behaviours. *Journal of Research in Music Education, 25*(1), 68-82.

Yang, Y. & Welch, G. (2016). Pedagogical challenges in folk music teaching in higher education: A case study of Hua'er music in China. *British Journal of Music Education, 33*(1), 61-79.

Young, W. T. (1971). An investigation of the singing abilities of kindergarten and first grade children in east Texas. *ERIC* EDO 69431.

Zhang, Z. Y. (2017). Effect of vocal fold stiffness on voice production in a three-dimensional body-cover phonation model. *The Journal of the Acoustical Society of America, 142*(4), 2311-2321.

Chinese Literature

Cai, F. & Du, Y. [蔡昉 & 都阳]. (2000). Convergence and divergence of regional economic growth in China – Enlightenment on the western development strategy [中国地区经济增长的趋同与差异一对西部开发战略的启示]. *Economic Research Journal* [*经济研究*], *10*. 30-37.

Cao, Y. G. [曹勇冠]. (2017). The general situation and reflection of Primary school music education in Tacheng [新疆多民族聚居地区塔城市小学音乐教育现状与反思] [Master's thesis, Shanghai Conservatory of Music].

Chen, Z. L. [陈长玲]. (2019). The problems of vocal music teaching in the training of Kindergarten teachers [幼儿教师培养过程中的声乐教学问题]. *China Music Education* [*中国音乐教育*], *9*. 22-25.

Chen, Y. H. & Fang, C. C. [陈友华 & 方长春]. (2007). Social stratification and educational diversion – An empirical study on the fairness of degree of arrangement for a system of 'Reginal and Nearby Enrollment' for compulsory education [社会分层与教育分流一一项对义务教育阶段"划区就近入学"等制度安排公平性的实证研究]. *Jiangsu Social Sciences* [*江苏社会科学*], *1*.

Cui, Q. L. [崔巧玲]. (2018). Problems of vocal pitch-matching accuracy during song singing and solutions for young Primary school students [低年级小学生唱歌音准问题的具体表现及对策]. *Art Evaluation* [*艺术评鉴*].

Du, J. & Yang, Y. [杜涓 & 杨阳]. (2019). Research on the current situation and countermeasures of music education in Diqing rural Primary schools – Taking Xialongga Primary school in Weixi County as an example [迪庆州乡村小学音乐教育现状调查及对策研究一 以维西县下笼噶小学为例]. *Song of the Yellow River* [*黄河之声*], *21*. 104-105.

Fang, C. C. & Feng, X. T. [方长春 & 风笑天]. (2008). Family background and academic achievement – a study of class differences in compulsory education [家庭背景与学业成就一义务教育中的阶层差异研究]. *Zhejiang Social Sciences* [浙江社会科学], *8*.

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*.

Guan, X. [关心]. (2014). *The concert and changes of social life in the period of the Republic of China: 1912-1945 – With school concert activities as the focus* [*民国音乐会与社会生活变迁：1912—1945 – 以学校音乐会活动为中心*] [Doctoral thesis, Nankai University].

Guo, L. [郭琳]. (2013). Exploring the creative characteristics of Tibetan popular music [探究藏族流行音乐的创作特点]. T*heoretical Preface* [*理论前言*], *5*.

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.

He, C. F. & Liang, J. S. [贺灿飞 & 梁进社]. (2004). The time and space change in the difference between China's reginal economies: Commercialization, globalization and urbanization [中国区域经济差异的时空变化：市场化、全球化与城市化]. *Management World* [*管理世界*], *8*. 8-17.

Hei, L. [黑力]. (2015). Discussing vocal pitch-matching issue during song singing [论歌唱的音准问题]. *Modern Music* [*当代音乐*], *17*. 9-12.

Hong, Y. X. [洪银兴]. (2022). Regional common prosperity and inclusive development [区域共同富裕和包容性发展]. *Economic Perspective* [*经济学动态*], *6*. 3-10.

Lei, G. Y. [雷光耀]. (2006). Discussing vocal pitch-matching accuracy during sightseeing and training [谈视唱的音准与训练]. *Chinese Music* [*中国音乐*], *2*. 146-149.

Li, W. [李伟]. (2010). *An overview on music education equity from the differences between urban and* *rural areas – On the bases of practical investigation and analysis to the middle schools in Harbin and Longjiang County* [*从我国城乡差异看音乐教育公平—基于对哈尔滨市和龙江县地区中学实地调查与分析*] [Maser's thesis, Harbin Normal University].

Li, X. P. [李响萍]. (2008). A study on the equity of school choice behaviour and education opportunity distribution in compulsory education stage – An empirical analysis based on the educa2tion choice expenditure of 18 urban households in China [义务教育阶段择校行为与教育机会分布公平性研究—基于中国 18 个城市居民家庭教育选择支出的实证分析]. *Educational Research* [*教育研究*], 3.

Li, Y. [李煜]. (2006). Institutional change and the mechanism of education inequality [制度变迁与教育不平等的产生机制]. *Social Sciences in China Press* [*中国社会科学杂志社*], *4.*

Li, Z. L. & Qiu Z. Q. [李忠路&邱泽奇]. (2016). How does family background affect children's academic achievements—Analysis on the difference of family socio economic status in compulsory education [家庭背景如何影响儿童学业成就？-- 义务教育阶段家庭社会经济地位影响差异分析]. *Sociological Studies* [*社会学研究*], *31*(4). 121-144.

Liu, D. W. [刘大巍]. (2001). A research on vocal pitch-matching accuracy during song singing [歌唱音准问题研究]. *JiaoXiang-Journal of Xi'an Conservatory of Music* [*交响—西安音乐学院学报（季刊）*], *20*(4). 41-48.

Liu, J. M. [刘精明]. (2008). Inequal opportunity and its changes in the compulsory education field in China [中国基础教育领域中的机会不平等及其变化]. *Social Sciences in China Press* [*中国社会科学杂志社*], *5.*

Liu, Y. X. [刘亚雄]. (2018). Don't forget to cultivate people's original intention, and use art to warm the heart – Interview with Professor Guo Shengjian, director of the aesthetic education development and research center of Hunan Normal University [勿忘育人初心，用艺术温润心灵一访湖南师大美育发展与研究中心主任郭声健教授]. *Hunan Education (Version D), 4*. 34-37.

Lu, T. T. [卢婷婷]. (2017). Reflections on the ways to train music education teachers in Primary and Secondary schools under the background of the New Curriculum Reform [新课改背景下中小学音乐教育师资培养途径的思考]. *Journal of Teaching and Management* [*教学与管理*], *15*, 68-70.

Ma, D. [马达]. (2002). The staging and development characteristics of the history of Chinese school music education in the 20th Century. *Art Research* [*艺术研究*], *3*.

Ma, Y. M. [马宇民]. (2010). “Only study is honourable”? – Parents’ dual expectations of education [“唯有读书高”? – 家长对教育的二元期待] [Maser’s thesis, Peiking University].

Mei, L. H. [魅丽华]. (2017). Discovering elements influencing children’s vocal pitch-matching accuracy and solutions [儿童歌唱中发音不准因素及纠正策略探析]. *Education Teaching Forum* [*教育教学论坛*], *20*. 81-83.

Ministry of Education of the People’s Republic of China [中华人民共和国教育部]. (2011). *Music Curriculum Standards for compulsory education (version 2011)* [*义务教育音乐课程标准（2011 年版）*]. Peking University Press [北京大学出版社].

Ministry of Education of the People’s Republic of China [中华人民共和国教育部]. (2022). *Curriculum Standards for art in compulsory education (version 2022)* [*义务教育艺术课程标准（2022 年版）*]. Peking University Press [北京大学出版社].

Pu, L. H. [浦利华]. (2016). New interpretation of “students dislike music class”. [对“学生不喜欢音乐课”的新角度阐释]. *China Music Education* [*中国音乐教育*], *12*, 11-15.

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*]. Ministry of Education of the People’s Republic of China.

Sun, A. N. [孙爱娜]. (2009). How to solve the problem of vocal pitch-matching accuracy for songs [如何解决唱歌中音准偏差的问题]. *Science & Technology Information* [*科技信息*], 574 & 577.

Wang, B. [王冰]. (2017). A practical study on Orff music therapy for children with autism [奥尔夫音乐治疗方法对孤独症儿童的实践研究]. *Medicine and Philosophy* [*医学与哲学*], *38*(1), 74-76.

Wang, L. P. [汪丽萍]. (2013). Characterization and attribution: A study on the imbalance of music teachers in counties – Taking Chongyang County, Hubei Province as an example [表征与归因：县域音乐师资失衡问题研究--以湖北省崇阳县为例]. *Journal of Wuhan Conservatory of Music, China* [*黄钟（中国·武汉音乐学院学报）*], *3*. 164-169 & 192.

Wang, T. [王婷]. (2012). Analysis of the current situation of music education in Primary and Secondary schools [中小学音乐教育现状评析]. *Journal of the Chinese Society of Education* [*中*

国教育学刊], *6*, 97-98.

Wang, X. P. [王秀萍]. (2004). A study on the value judgement of music curriculum [音乐课程价值取向论证]. *Musicology in China* [*中国音乐学（季刊）*], *1*, 116-125.

Wang, Y. G., Xu, J. Q. & Ding, J. H. [王永固，许家奇 & 丁继红]. (2020). Education 4.0 Global Framework: School education in the future and mode transformation – World Economic Forum "School of the future: Defining a new education model for the fourth industrial revolution interpretation of the report" [教育 4.0 全球框架：未来学校教育与模式转变—世界经济论坛《未来学校：为第四次工业革命定义新的教育模式》]. *Journal of Distance Education* [*远程教育杂志*], *38*(3). https://doi.org/10.15881/j.cnki.cn33-1304/g4.2020.03.001

Wie, W. J. [魏文娟]. (2018). On the significance of WeChat public platform for music teaching in primary schools – Interpretation of the report [浅谈微信公众平台对小学音乐教学的意义]. *Journal of the Northern Music* [*北方音乐*], *38*(18). 211.

Wen, D. M. [文东茅]. (2006). School selection in the stage of compulsory education in cities and its influence on the disadvantaged groups [我国城市义务教育阶段的择校及其对弱势群体的影响]. *Peking University Education Review* [*北京大学教育评论*], *2*.

Wu, Y. X. [吴愈晓]. (2013). Inequality of educational opportunities between urban and rural residents in China and its evolution (1978 – 2008) [中国城乡居民的教育机会不平等及其演变 (1978 – 2008)]. *Social Sciences in China Press* [*中国社会科学杂志社*], *3.*

Xiang, Y. [项阳]. (2010). A debate and analysis among music for rites, elegant music, and percussion music [礼乐·雅乐·鼓吹乐之辨析]. *Journal of the Central Conservatory of Music* [*中央音乐学院学报*], *1*.

Xie, J. X. [谢嘉幸]. (2005). A comprehensive study of contemporary music pedagogy in China [我国当代音乐教育学研究综述]. *Journal of Nanjing Arts Institute (Music & Performance)* [*南京艺术学院学报（音乐与表演）*], *3*. 7-24.

Xu, K. H. [徐考辉]. (2018). Exploration of practical vocal class: From lip-synch to light singing and unaccompanied singing [实效歌唱教学探究：从"假唱"到"轻唱""清唱"]. *China Music Education* [*中国音乐教育*], *12*. 20-24.

Xu, S. S. [胥珊珊]. (2019). Research on Chinese music education from the perspective of mapping knowledge domain analysis [知识图谱分析视域中的中国音乐教育研究（2009-2018）年

（上）]. *China Music Education* [*中国音乐教育*], *9.* 28-33.

Xu, W. [徐娃]. (2021). A study on the strategies to solve the problem of vocal pitch-matching issues in singing teaching [歌唱教学中解决音准问题的策略研究]. *New Curriculum Guidance* [*新课程导学*], *13*. 34-35.

Yan, B. L. [阎宝林]. (2020). Tone training – change in chorus [音色训练]. *China Music Education* [*中国音乐教育*], *4*. 44-50.

Yang, D. P. [杨东平]. (2006). The ideal and reality of educational equity in China [中国教育公平的理想与现实]. *Peking University Press* [*北京大学出版社*].

Yang, F. [杨帆]. (2019). The combination of music, dance, and poetry in music and dance in Tang dynasty [论唐代乐舞中音乐、舞蹈、诗歌的结合]. *Home Drama* [*戏剧之家*], *1*. 92.

Yang, H. N. [杨鸿年]. (1982a). Discussing vocal pitch-matching problem in a chorus (part 1) [论合唱音准问题（一）]. *Journal of the Central Conservatory of Music* [*中央音乐学院学报*], *2*. 22-32. https://doi.org/10.16504/j.cnki.cn11-1183/j.1982.02.006

Yang, H. N. [杨鸿年]. (1982b). Discussing vocal pitch-matching problem in a chorus (part 2) [论合唱音准问题（二）]. *Journal of the Central Conservatory of Music* [*中央音乐学院学报*], *3*. 46-55. https://doi.org/10.16504/j.cnki.cn11-1183/j.1982.03.016

Yang, L. M. [杨立梅]. (2015). Exploration on students' singing ability improvement in class [探索课堂教学中提高学生的准确歌唱能力]. *China Music Education* [*中国音乐教育*], *3*. 14-16.

Yao, X. G. & Zhang, H. F. [姚先国 & 张海峰]. (2008). Education, human capital and regional economic differentials [教育、人力资本与地区经济差异]. *Economic Research Journal* [*经济研究*], *5*. 47-57.

Ye, H. & Wu X. G. [叶华 & 吴晓刚]. (2011). Fertility decline and the trend in educational gender inequality in China [生育率下降与中国男女教育的平等化趋势]. *Sociological Studies* [*社会学研究*], *5*.

Ye, L. H. & Xing, L. F., [叶丽慧&邢丽芬]. (2010a). Controlling vocal pitch and singing lively – vocal pitch-matching issues during singing teaching and solutions (part 1) [把握音准 生动歌唱—歌唱教学中存在的"音准"问题及对策研究（上）]. *China Music Education* [*中国音乐教育*], *8*. 16-18.

Ye, L. H. & Xing, L. F., [叶丽慧&邢丽芬]. (2010b). Controlling vocal pitch and singing lively – vocal pitch-matching issues during singing teaching and solutions (part 2) [把握音准 生动歌唱—歌唱教学中存在的"音准"问题及对策研究（下）]. *China Music Education* [*中国音乐教育*], *9*. 18-21.

Yin, S. Y. [殷思瑶]. (2020). Exploration of targeted pitch teaching in lower Primary school [精准"把握"精益"配方"— 小学低学段学生音准问题探析]. *China Music Education* [*中国音乐教育*], *4*. 24- 27.

Zhang, B. B. [张蓓蓓]. (2017). Analysis and research on the current situation of music education practice in rural Primary and Secondary schools in Guanzhong area. [关中地区农村中小学音乐教育实践现状分析与研究]. *Journal of Teaching and Management* [*教学与管理*], *15*. 28-30.

Zhang, D. L. [张丹丽]. (2016). The music education gap between city and country from the aspect of stage and popularization [从城乡音乐教育的差距看音乐教育阶段性和地域普及性—以安徽省六安市城乡学校为例]. *China Music Education* [*中国音乐教育*], *8*. 33-37.

Zhang, H. L. [张红丽]. (2019). The current situation and improvement strategies of music education in rural Primary schools [农村小学音乐教育发展现状及提高策略]. *Modern Economic Information* [*现代经济信息*], *20*. 436.

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.

Zhao, Y. D. & Hong Y. B. [赵延东，洪岩璧]. (2012). Social capital and education acquisition – from the perspective of network resources and social closure [社会资本与教育获得—网络资源与社会闭合的视角]. *Sociological Studies* [*社会学研究*], *5*.

Zi, L. P. & Zhu Y. B. [资利萍&朱咏北]. (2006). The philosophical basis of the three dimensional goal of music curriculum – Responding to Wu Bin's "Focus on Music" [音乐课程三维目标的哲学分析—回应吴斌的《关注音乐》]. *China Music Education* [*中国音乐教育*] , *6*. 17-19.

Appendixes

Appendix A: The English version of Information letters and ethics form for music teachers and participant children's parents

University College London

Consent Form Version 2; 21 March 2017

Institute of Education, 20 Bedford way, London, UK WC1H 0AL

Tel: 004420 7612 600

MUSIC TEACHERS CONSENT FORM:

Project: An investigation of Chinese Primary school children's singing behaviour, development and engagement

Name of Principal Investigator: Can Lu, PhD student

Please tick Yes or No

	Yes	No
1. I confirm that I have read and understood the information sheet for the above study and have had the opportunity to ask questions and any questions have been answered satisfactorily.	☐	☐
2. I understand that my students' participation is voluntary and that they are free to withdraw at any time, with or without giving any reason, without affecting students' health care or legal rights.	☐	☐
3. I understand the research will involve students having (i) an individual singing assessment, which includes (a) counting backward from either 10 or 20, (b) vocalising pitch glissandi across their vocal pitch range, and (c) singing three criterion songs (*Twinkle, Twinkle; Happy Birthday; Little Donkey*); and (ii) completing a questionnaire about children's attitudes to singing;	☐	☐

(iii) In addition, it would help if I could observe music lessons to understand more about their music learning in school.

4. I understand that the following data will be collected: (i) spoken pitch centre, comfortable singing range, vocal register, and vocal pitch matching accuracy; (ii) children's singing attitudes in school, home, informal settings, and their self-identity (emotional engagement with singing and self as a singer), and social inclusion; (iii) students' activities and teachers' activities in music lessons.

Yes ☐ **No** ☐

5. I understand that I may need to help the investigator to make a timetable and organise students to take part in the study.

Yes ☐ **No** ☐

6. I understand that the singing assessment will be audio recorded; the pupil questionnaire will be done on paper; and my music lesson will be directly observed either with or without video recording.

	Yes	**No**
With video	☐	☐
Without video	☐	☐

7. I understand that I can ask the investigator the results for my students, and that these will be made available as soon as possible after the data collection.

Yes ☐ **No** ☐

8. I understand that the data and notes will be stored by responsible individuals from UCL-IoE in a secure, password protected location. The data will be shared with my two supervisors only. I give permission for these individuals to have access to my students' records of singing assessment, data of questionnaire and records of my music lesson.

Yes ☐ **No** ☐

9. I understand the data will be managed by the principal investigator safely and securely such that no individual teacher, nor pupil, nor school can be identified in any subsequent reporting.

Yes ☐ **No** ☐

10. I am aware that any video material with permission will be viewed and analysed only by those directly involved with the research.

Yes ☐ **No** ☐

11. I understand that personal information and school details will be protected by the principle investigator and will not be published for any reason.

Yes ☐ **No** ☐

12. I allow the investigator to publish anonymized findings under an academic context.

Yes ☐ **No** ☐

_________________________ _______________

School Grade level

_________________________ ____________________

Name of teacher (Signature) Date

____

Signature of researcher

Thank you! *Best wishes.*

University College London

Consent Form Version 2; 21 March 2017

Institute of Education, 20 Bedford way, London, UK WC1H 0AL

Tel: 004420 7612 600

PARENTAL CONSENT FORM:

Project: An investigation of Chinese Primary school children's singing behaviour, development and engagement

Name of Principal Investigator: Can Lu, PhD student

Please tick Yes or No

	Yes	No
1. I confirm that I have read and understood the information sheet for the above study and have had the opportunity to ask questions and any questions have been answered satisfactorily.	☐	☐
2. I understand that my child's participation is voluntary and that we are free to withdraw at any time, with or without giving any reason, without affecting my child's health care or legal rights.	☐	☐
3. I understand the research will involve my child having (i) an individual singing assessment, which includes (a) counting backward from either 10 or 20, (b) vocalising pitch glissandi across their vocal pitch range, and (c) singing three criterion songs (*Twinkle, Twinkle; Happy Birthday; Little Donkey*); and (ii) completing a questionnaire about children's attitudes to singing; (iii) in addition it would help if I could observe of music lessons to understand more about their music learning in school.	☐	☐
4. I understand that the following data will be collected: (i) spoken pitch centre, comfortable singing range, vocal register, and vocal pitch matching accuracy; (ii) children's singing attitudes in school, home, informal settings, and their self-	☐	☐

identity (emotional engagement with singing and self as a singer), and social inclusion; (iii) students' activities and teachers' activities in music lesson.

Statement	Yes	No
5. I understand that the singing assessment will be audio recorded; questionnaire will be done on paper; and music lessons may be directly observed with video recording if permission is granted.	☐	☐
6. I understand that the data and notes will be stored at by responsible individuals from UCL-IoE in a secure, password protected location. The data will be shared with my two supervisors only. I give permission for these individuals to have access to my child's singing records of singing assessment, data of questionnaire and music activity.	☐	☐
7. I understand the data will be managed by the principal investigator safely and securely such that no individual teacher, nor pupil, nor school can be identified in any subsequent reporting.	☐	☐
8. I am aware that any video material with permission will be viewed and analysed only by those directly involved with the research.	☐	☐
9. I understand that personal information and school details will be protected by the principle investigator and will not be published for any reason.	☐	☐
10. I allow the investigator to publish anonymised findings under an academic context.	☐	☐

	Yes	No
11. I agree that my child can take part in the above study.	☐	☐

Name of Child (Signature) School Grade - Class

Name of Parent (Signature) Date

Signature of researcher

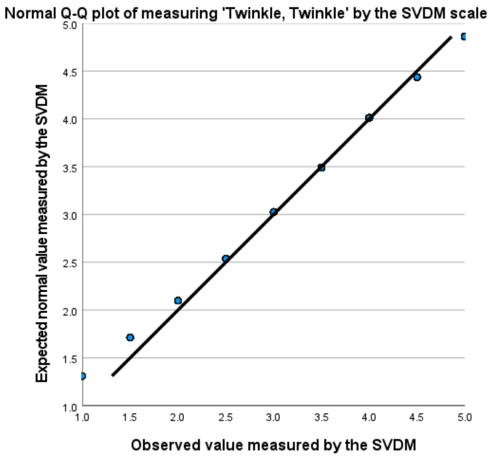
Thank you! Best wishes.

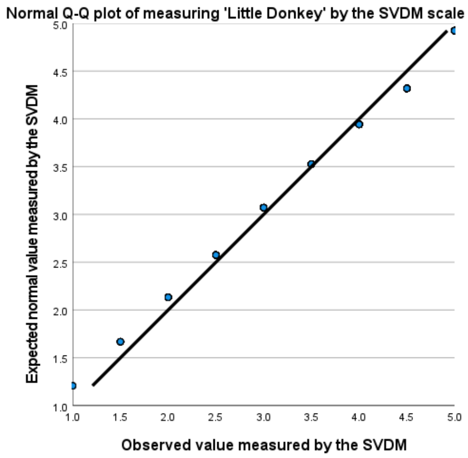
Appendix B: Participant children's information sheet

姓名	性别	出生年月（不知道可不填）	学校	年级	父母的工作
	男（或） 女				
身高 （不知道可不	体重 （不知道可不	是否参加课外音乐培训?	如有参加课外音乐培训，是否是歌唱培训?		
		是 否 请在选项上画圆圈。如果选否，之后的一题不填哦！！	是 否 请在选项上画圆圈。如果选否，所参加的课外音乐培训项目是: ________		

Appendix C: Distribution of raw scores measured by two rating scales for Chinese participants

Figure C1 Q-Q Plots Testing Normal Distribution of Raw Data on Vocal Registers for Each of the Three Criterion Songs, Using the Rutkowski Scale (1996)





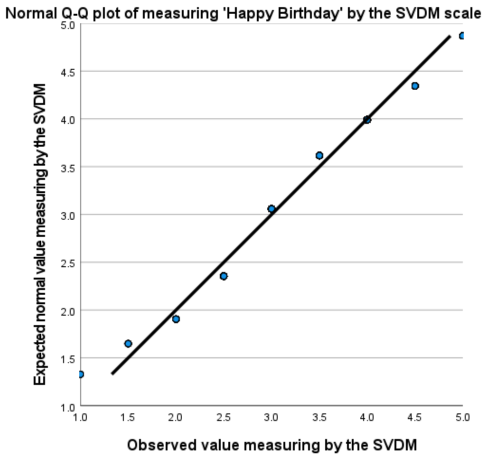
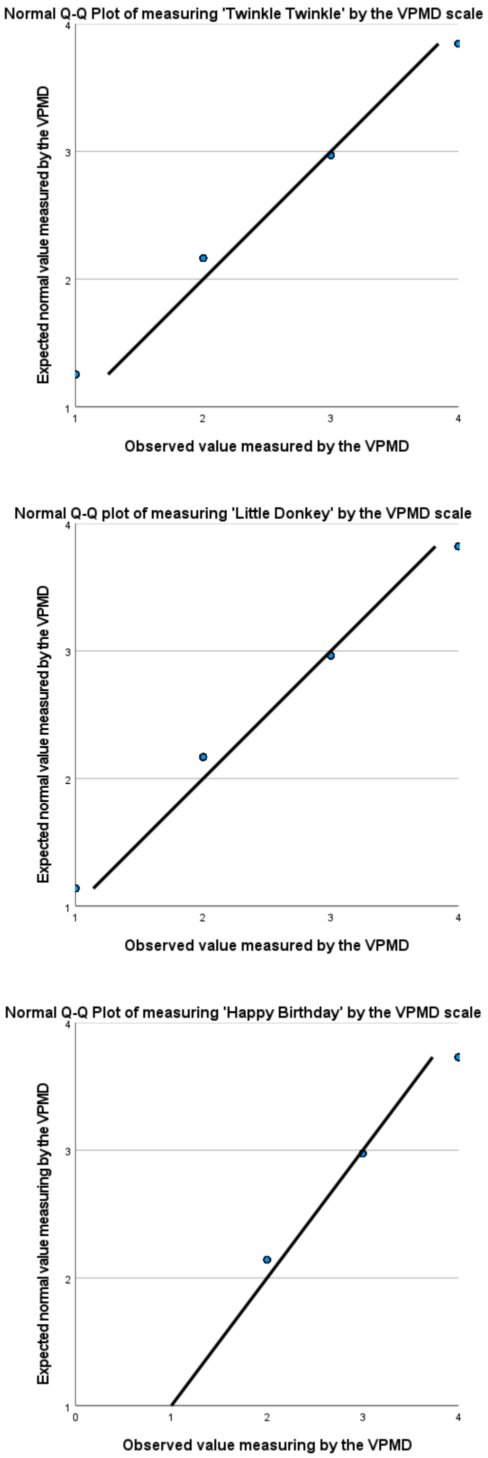
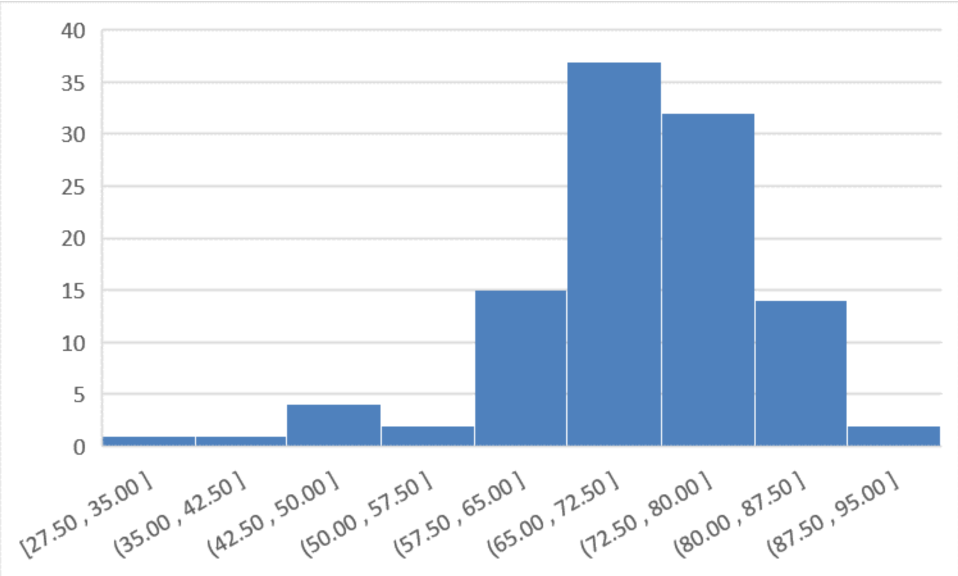


Figure C2 Q-Q Plots Testing Normal Distribution of Raw Vocal Pitching-matching Development (VPMD) Scores for Each Criterion Song



Appendix D: Distribution of normalised singing scores based on by two rating scales for English participants without *Sing Up* experience

Figure D1 Distribution of *Normalised Singing Scores Based on Two Existing Rating Scales for Twinkle, Twinkle and Happy Birthday for English Participants without Sing Up Experience*



Note. The above figure was drawn based on data provided by the team from University College London, Institute of Education who conducted the research for the National Singing Programme *Sing Up*, led by Professor Graham F. Welch

Appendix E: Analyses of a music textbook

Table E1 A Summary of N = 18 Songs Required to Learn in a Chinese Music Textbook for Grade 4 Students

No.	Chinese name	English translation	The choice of key	Vocal range	Register(s) required for the material	Chinese/Western/ others	No. of tonalities used in the song
1	茉莉花	Jasmine flower	Eb	a perfect 12th	chest to head	Chinese	1
2	知了	Ciada	Bb	an octave	middle to head	Chinese	1
3	小小足球赛	A little football match	C	an octave	chest to head	Chinese	1
4	我的家乡日喀则	My hometown Rikaze	Bb	a major 11th	chest to head	Chinese	1
5	卓玛	Zhuoma	Eb	a minor 11th	chest to head	Chinese	1
6	踩雨	Trampling on the rain	B	an octave	middle to head	Chinese	1
7	大雨和小雨	Heavy rain and light rain	D	a perfect 5th	chest to middle	Chinese	1
8	可爱的家	A lovely home	D	an octave	chest to head	Western melodies (America) + Chinese lyrics	1
9	吉祥三宝	Three lucky people	F	an octave	chest to middle	Chinese	1
10	小小鼓号	Little drummer	E	a major	chest to	Chinese	1

	手			9th	middle		
11	狐假虎威	A fox masquerading as a tiger	Eb	a major 11th	chest to head	Chinese	1
12	铁路修到苗家寨	A railway repairs to Miaojia stockade village	Eb	a perfect 12th	chest to head	Chinese	1
13	田野的呼唤	The call of the field	D	an octave	chest to head	Western melodies (Italy) + Chinese lyrics	1
14	我心爱的小马车	My beloved little horse carriage	D	an octave	chest to head	Chinese	1
15	远方钟声	Distant bell sound	D	an octave	chest to head	Japanese melodies + Chinese lyrics	1
16	西风的话	Words of the west wind	F	a major 9th	chest to middle	Chinese	1
17	如果幸福你就拍拍手	If you are happy, clap your hands	F	a minor 7th	chest to middle	melody - unknown author, Japanese lyrics translated into Chinese	1
18	童心是小鸟	Childlike innocence is like a bird	Eb	a major 11th	chest to head	Chinese	1