education sciences

MDPI

*Article*

# The Influence of Teaching Songs with Text and a Neutral Syllable on 4-to-9-Year-Old Portuguese Children's Vocal Performance

**Ana Isabel Pereira * and Helena Rodrigues**

Department of Musicology, School of Social Sciences and Humanities (FCSH), Centre for Music Studies (CESEM), NOVA University of Lisbon, 1069-061 Lisbon, Portugal; helenarodrigues@fcsh.unl.pt
* Correspondence: anapereira@fcsh.unl.pt

**Abstract**

Research on children's singing development is extensive. Different ages, approaches, and variables have been taken into consideration. However, research on singing with text or a neutral syllable is scarce, and findings are inconclusive. This study investigated the influence of singing with text and a neutral syllable on children's vocal performance. Children aged 4 to 9 ($n$ = 135) participated in two periods of instruction and assessment. In Period One, Song 1 was taught with text and Song 2 with a neutral syllable, and in Period Two, the text was added to Song 2. In each period, children were individually audio-recorded singing both songs. Three independent raters scored the songs' vocal performances using two researcher-designed rating scales, one for each song, which included the assessment of tonal and rhythm dimensions. Before data analysis, the validity and reliability of the rating scales used to assess vocal performance were examined and assured. The results revealed that 4-, 5-, and 7-year-olds sang Song 1 significantly better in Period One, and 4- and 5-year-olds sang Song 1 significantly better in Period Two. Thus, singing with text seems to favour younger children's vocal performance. Findings also revealed that girls scored significantly higher than boys for Song 1 in both periods, but not for Song 2 in Period One. The implications of incorporating songs with text and neutral syllables into music programs, as well as the instruments used to assess vocal performances, are discussed.

**Keywords:** children's vocal performance; rating scales; songs with text; songs with a neutral syllable; music education



Academic Editors: Alberto Cabedo-Mas, Siu Hang Eric Kong and Susan A. O'Neill

Received: 22 June 2025
Revised: 23 July 2025
Accepted: 29 July 2025
Published: 1 August 2025

**Citation:** Pereira, A. I., & Rodrigues, H. (2025). The Influence of Teaching Songs with Text and a Neutral Syllable on 4-to-9-Year-Old Portuguese Children's Vocal Performance. *Education Sciences*, *15*(8), 984. https://doi.org/10.3390/educsci15080984



## 1. Introduction

Research on children's singing development and teaching children to sing is extensive. Numerous variables that might influence children's singing ability have been identified: accompaniment, age, attitude, audiation, pitch discrimination, sex, home environment, vocal model, motor coordination, range and use of vocal registers, song acquisition, song literature, individual and group singing, and the use of text or a neutral syllable (see reviews in, e.g., Goetze et al., 1990; Nichols, 2015; Phillips, 2014).

Ongoing research continues to provide insights into children's singing development, including attitude toward singing (Papageorgi et al., 2022), the use of vocal or piano doubling (Nichols, 2021), and motor skills (Pfordresher, 2022). This study aims to contribute to discussions in the field, focusing on the use of text and neutral syllables when learning and singing songs. This topic has yielded inconclusive results, and, to our knowledge, it has not been explored in the literature in contemporary classrooms.

### 1.1. Singing with Text or with a Neutral Syllable

Concerning the use of text and neutral syllables, several studies have reported the dominance of text in song acquisition (Davidson, 1985; Davidson et al., 1981; Moog, 1976; Rutkowski, 1998). The importance of words in children's social lives might explain this dominance (Welch, 2012; Welch et al., 1995). Another explanation could be related to the fact that words (text) combine elements such as pitch, rhythm, and timbre, thus enriching the combination of vowels and consonants as if it were an additional musical timbre (Moog, 1976). Does text mask the properties of a melody, such as melodic contour or rhythm, thus affecting children's singing accuracy? Welch et al. (1998) suggested that the melody should be taught separately from the text to correct the text bias, at least for young developing singers. Teaching songs with a neutral syllable might favour children since it provides them with a greater focus on the intrinsically musical aspects of the melody, such as pitches, intervals, intensities, and it allows them to concentrate on audiating tonal and rhythm patterns, tonality, and meter (Gordon, 2003, 2012; Reifinger, 2012). However, empirical evidence is contradictory regarding the interaction between melody and text in song singing.

Pereira and Rodrigues (2019) summarized previous findings as follows: some studies have reported better performance with neutral syllables (Goetze, 1985); others, significantly higher scores with text for the younger group of 4-yearolds (Jacobi-Karna, 1996); others performed slightly better with neutral syllables (Levinowitz, 1989); others found no significant differences but slightly higher scores with text (Rutkowski, 1993); some reported no significant differences but performed better with neutral syllables (Smale, 1987); and some performed better with text (Welch et al., 1995). Methodological differences across these studies limit the comparability of their results. The use of different song materials (different tonalities, meters, and ranges) and instructional periods should also be acknowledged. The main procedural differences found are related to how the songs were presented and how children were asked to perform the task. For example, Goetze (1985) taught both songs with text and then repeated the same songs with a neutral syllable (loo, in this case), with data collection taking place immediately after the session, where children were asked to sing in both ways. Smale (1987) used one song and asked children to perform four singing tasks; these included singing the song's first phrase with text echoing the researcher, singing the song's first phrase with text in unison with the researcher, singing the song's first phrase with a neutral syllable echoing the researcher, and singing the song's first phrase with a neutral syllable in unison with the researcher. Jacobi-Karna's (1996) study is the only one that considers the influence of the teaching order of a song's melody and text on children's vocal performance. It presented two phases: during phase one (eight weeks, two 30 min classes per week), the same song was taught with text and a neutral syllable to two groups in a counterbalanced way, and in phase two (one week duration), for one group, the text was removed from the song, which was then presented with a neutral syllable; for the other group, the text was added to the melody. During the test phase, which took place the week after each phase, children were asked to sing the song by phrase and in its entirety.

### 1.2. Age and Sex Versus Singing

Tsang et al. (2011) suggest that there is evidence that maturation factors related to age contribute to better vocal performance, as improvements in memory capacity, motor skills, and control can also be observed. A meta-analysis by Svec (2018) demonstrated a trend in which accuracy tends to improve across the ages of 5, 6, 7, and 8, aligning with previous research (e.g., between preschoolers and fourth graders in the study by Geringer, 1983; or children aged 5 to 7 years old in the study by Welch et al., 1997). Demorest and Pfordresher (2015) found that kindergartners were more accurate than sixth graders and

adults when singing a familiar song from memory. However, the researchers pointed out that these differences might be due to the different songs used rather than skill differences. Interestingly, Svec (2015) reports that accuracy then decreases between the ages of 8 and 9, 10, and 11, suggesting that establishing a respectful and safe classroom environment before third grade is essential in mitigating or softening the impact of this decline. The crossover between age and singing with text or a neutral syllable has been investigated, although with contradictory results as mentioned before, with ages spanning between 3 and 8 years old (Goetze, 1985—4, 5, 6, and 8 years old; Jacobi-Karna, 1996—3, 4, and 5 years old; Smale, 1987, and Levinowitz, 1989—4 and 5 years old; Rutkowski, 1993—6 years old; Welch et al., 1995—4 to 8 years old).

Research has yielded mixed results when investigating the accuracy differences in singing between boys and girls. As mentioned by Pereira and Rodrigues (2019), some studies reported no significant differences (Apfelstadt, 1984; Leighton & Lamont, 2006; R. S. Moore, 1994; Paney & Kay, 2015; Welch et al., 1995), and others revealed that girls sing more accurately than boys (Goetze & Horii, 1989; Mang, 2006; Trollinger, 2003; Yarbrough et al., 1991). Svec's (2018) meta-analysis revealed that singing instruction was approximately as effective for boys as it was for girls. However, instruction had a greater effect on girls, although the difference was not statistically significant. This researcher also supports the idea that conflicting results on sex may be attributed to other variables, such as the student's attitude towards the teacher or social pressures. Other researchers have suggested that differences between girls and boys, favouring girls, may be due to the predominance of female vocal models (Welch et al., 1997) or to differences in attitude and motivation (Petzold, 1969; Welch et al., 2012) A longitudinal study by Welch et al. (2012) revealed that, given an appropriate educational experience, boys can achieve results comparable to those of girls of the same age in terms of accuracy. Svec (2018) also suggests that "rather than demonstrating concern regarding differences between genders, boys and girls alike may benefit from being compared to themselves. Having knowledge of one's own singing development may encourage a sense of empowerment over one's unique ability" (p. 11). To our knowledge, the influence of sex on singing with text or a neutral syllable has not been approached as of yet.

### *1.3. Singing Assessment Instruments*

According to Rutkowski (2019), the measurement of children's singing development began in the 1960s. Since then, multiple investigations have been conducted focusing on children's ability to sing in tune, measuring different singing behaviours, such as the use of vocal registers, the accuracy of pitch matching in isolated pitches, intervals, patterns, and songs (see reviews in Goetze et al., 1990; D. Hedden, 2012; Mizener, 2008; Nichols, 2015; Philipps & Doneski, 2011; Phillips, 2014; Rutkowski, 2019).

Measurement tools can utilize technology (e.g., using software packages such as Multispeech reported by D. G. Hedden & Baker, 2010; the Continuous Response Digital Interface device by D. G. Hedden et al., 2008; the Sona-Speech Model 3600 by Miyamoto, 2005; the Multidimensional Voice Range Profile-Advanced in R. E. Moore et al., 2008; the TF32 by Pfordresher & Brown, 2007; CSpeech by Trollinger, 2003) or human raters through published or researcher-customized scales. The rating scales may be developed using a combination of levels (bi-level and multi-level) and criteria (single and multiple) (T. Brophy, 2000). According to Gordon (2002), the optimal rating scale should include two or more dimensions (e.g., tonal, rhythm, expression, improvisation, and technique, adjusted to vocal or instrumental assessments), with an optimal number of five criteria for each dimension—fewer than five criteria would reduce reliability, and more than five would compromise validity. For example, Gault (2002) and Levinowitz (1989) utilized the tonal and rhythm

dimensions to evaluate children's performance achievement on song singing. Moreover, rating scales can be continuous (three scoring levels: Davidova et al., 2019; four scoring levels: Runfola & Etopio, 2009; five scoring levels: Feierabend, 1984; Fixmer, 2022; Gault, 2002; Giga, 2005; Hale, 1977; Hornbach & Taggart, 2005; Jordan-DeCarbo, 1982; Lange, 1999; Levinowitz, 1989; seven scoring levels: Leighton & Lamont, 2006; eight scoring levels: Wise & Sloboda, 2006), or additive (Giga, 2005; Nichols, 2016; Sims et al., 1982), where the former is hierarchical and the latter resembles a checklist.

As Rutkowski (2019) mentions, a measurement instrument is more effective in both research and educational contexts when it assesses a single construct rather than multiple constructs. Yet, several scales measure two constructs at the same time: the use of the singing voice and vocal accuracy. When measuring singing behaviours, these should be considered as sequential constructs, as children often do not sing in tune because they do not use their voice appropriately (Rutkowski & Trollinger, 2005), i.e., they are not able to use the singing voice across all vocal registers, which is a necessary condition to become an independent singer. Nonetheless, an accuracy deficit could also be due to the vocal mechanism not being physiologically ready, for example, because of a lack of vocal strength (Phillips, 2014). The Singing Voice Development Measure (SVDM) by Rutkowski (2015) has been validated as a tool for assessing children's use of the voice for singing, with participants from different countries (e.g., Pereira & Rodrigues, 2019; Rutkowski & Chen-Hafteck, 2001; Welch et al., 2011). This study will not focus on the use of the singing voice, although the customized rating scales considered this and did not measure vocal accuracy per se, as explained further ahead.

### *1.4. Rationale of the Study*

The above-mentioned studies on the interaction between melody and text in song singing have yielded inconsistent results (Goetze, 1985; Jacobi-Karna, 1996; Levinowitz, 1989; Rutkowski, 1993; Smale, 1987; Welch et al., 1995). Notably, none have used a research design that is capable of determining whether children sing better with text when the melody is taught in isolation before the introduction of text, as opposed to teaching melody and text simultaneously. If children's vocal performance is improved by using one of these song-teaching strategies, then this practice could be used more consciously. The purpose of this study was to investigate the influence of two song-teaching strategies on children's vocal performance of two songs. Thus, two teaching periods were considered: in Period One, Song 1 was taught with text, and Song 2 was taught with a neutral syllable; in Period Two, Song 1 continued to be taught with text, and text was added to Song 2. The research questions are provided as follows:

RQ1—Do the vocal performances of Songs 1 and 2 in Period One differ by grade level or sex?
RQ2—Do the vocal performances of Songs 1 and 2 in Period Two differ by grade level or sex?

This article is part of a larger study that also investigates the use of singing voice (Pereira & Rodrigues, 2019) and the cognitive processes involved in song recognition, examining how text and melody influence perception.

## 2. Materials and Methods

### *2.1. Participants*

The initial sample consisted of 137 children (aged 4–9 years, 63 boys and 74 girls) attending a private school in Lisbon and belonging to families with medium to high income levels. To control for experimental mortality, participants were required to participate in all music classes or all minus one to be included in the study. Overall, 2 four-year-old girls did

not meet this criterion, and only 135 children were considered for data analysis, despite their participation in data collection. Table 1 presents the characterization of children by age and sex. The age groups will be referred to as K4, K5, 1st grade, 2nd grade, 3rd grade, and 4th grade, as the study intervention took place in intact classes organized by school level.

**Table 1.** Participants' characterization by age and sex.

| | K4 | K5 | 1st Grade | 2nd Grade | 3rd Grade | 4th Grade | Total |
|---|---|---|---|---|---|---|---|
| | N = 22<br>(16.3%)<br>M = 3.95<br>SD = 0.21 | N = 19<br>(14.1%)<br>M = 5.11<br>SD = 0.32 | N = 27<br>(20.0%)<br>M = 6.07<br>SD = 0.27 | N = 22<br>(16.3%)<br>M = 7.09<br>SD = 0.43 | N = 23<br>(17.0%)<br>M = 8.13<br>SD = 0.34 | N = 22<br>(16.3%)<br>M = 8.82<br>SD = 2.04 | N = 135<br>(100.00%)<br>M = 6.56<br>SD = 1.88 |
| Sex | N (%) | N (%) | N (%) | N (%) | N (%) | N (%) | N (%) |
| Girls | 13 (59.1) | 12 (63.2) | 10 (37.0) | 11 (50.00) | 12 (52.2) | 14 (63.6) | 72 (53.3) |
| Boys | 9 (40.9) | 7 (36.8) | 17 (63.0) | 11 (50.00) | 11 (47.8) | 8 (36.4) | 63 (46.7) |

### 2.2. Design and Procedure

This study employed a quantitative quasi-experimental design (Creswell & Creswell, 2018). A convenience sample was used, as the first author was the music teacher at the school. A counterbalanced design for song teaching was not considered (i.e., in Period One, Song 1 was presented with text and Song 2 with a neutral syllable for half of the groups, and Song 1 was presented with a neutral syllable and Song 2 with text for the other groups; adding the text in Period Two to the songs presented first with a neutral syllable). This is because among participants from different grade levels, there were several pairs of siblings; therefore, providing them with the same song-teaching strategy would be very difficult. In addition to this, over the years of teaching at that school, anecdotal observations of children's playtime revealed that they sometimes sang the songs taught in music classes, which would interfere with the pre-determined order of the teaching strategies. Both the intervention and data collection took place at the participants' school, a natural setting.

All children received music instruction from the same teacher once a week (30 min in kindergarten, and 60 min in first to fourth grades). Music classes were based on the music learning theory developed by Edwin Gordon (1927–2015) and also influenced by the approaches of Carl Orff (1895–1982) and Émile Jaques-Dalcroze (1865–1950). Thus, children's musical routines included singing, chanting, moving, dancing, playing, improvising, and creating, as well as reading and writing with the oldest children. Table 2 summarizes the study's conditions.

**Table 2.** Music instruction and data collection timeline.

| | Period One | Data Collection | Period Two | Data Collection |
|---|---|---|---|---|
| | 6 months total (including school breaks and holidays) | | | |
| Duration | 5 weeks | 5 weeks | 3 weeks | 5 weeks |
| Song 1 | with text | 135 audios | with text | 135 audios |
| Song 2 | with a neutral syllable | 135 audios | Text added | 135 audios |

For the study, two periods were defined for instruction and assessment spanning six months. Period One corresponds to a period of five music classes: Song 1 was taught with text, and Song 2 was taught with a neutral syllable; then, both songs' vocal performances were audio-recorded. Period Two corresponds to a period of three music classes, where Song 1 continued to be taught with text, and text was added to Song 2. The vocal perfor-

mances of the songs (Songs 1 and 2 with text) were collected following these classes. Songs were taught by rote and by immersion (also referred to as the whole-song method), which is considered a more appropriate procedure for short and simple songs (Phillips, 2014).

Data collection was individual; each child was escorted to a private room at the school, where her/his voice was recorded. The vocal performances were prompted by a preparatory sequence indicating the meter, tonality, key, tempo, and the first pitch of the song, which is a familiar procedure used in music classes. The average length of this data collection procedure was approximately 1.5 min. A total of 540 audio samples were collected—270 for each period. A Zoom H4N Pro Digital Multitrack Recorder (Zoom Corporation, Berlin, Germany) was used.

Informed consent was obtained from all participants' legal tutors, as well as children's assent. The anonymity of participants and the confidentiality of the data collected were ensured.

### 2.3. Songs

The songs were chosen based on the following criteria: non-familiar to eliminate previous contact with the material, to more accurately assess melody and text interactions, and to control how many times songs were presented during instruction; belonging to the children's cultural context, i.e., a western music-base; and both songs should be similar in tonality, meter, length, and vocal range. Several songbooks from different countries, including Portugal, were consulted during the selection process. Although the last criterion was difficult to accomplish, two songs were chosen: 'Let's play in the snow' by Doug Nichol (Valerio et al., 2000) and 'Stirring the soup' by Edwin Gordon (Gordon et al., 1993). A panel of five expert judges was tasked with comparing the two songs in terms of vocal performance difficulty, based on melodic contour, harmonic underpinnings, and rhythm patterns. They were also asked to verify both songs' texts concerning their prosody and their relationship to the musical phrases. The text for 'Let's play in the snow' had to be translated and adapted into Portuguese [Veste o casaco, vem brincar, vamos lá p'ra fora. Veste o casaco, vem brincar, já está a nevar]. A text was written for 'Stirring the soup', since it was published as a song without words. Two judges agreed that the harmonic underpinnings were sufficiently different to make the second song more difficult, two considered the songs to be of similar difficulty, and one answered that the first song was probably easier to sing. Facing these opinions, 'Let's play in the snow' was kept, and two versions of a new song were composed by the first author to replace 'Stirring the soup', although maintaining the same theme as the title suggested—'A sopa' (the soup). The songs were sent to a panel of 10 judges (including three from the first panel), who were asked to compare the two versions of 'A sopa' with 'Let's play in the snow'. All judges considered that both versions of 'A sopa' had the same difficulty level as 'Let's play in the snow'. Yet, two of them mentioned that one of the new song's versions presented a descending fifth interval at the penultimate bar that might be difficult to sing. However, this version was chosen to be part of the study, as the ending was sufficiently different from 'Let's play in the snow'. Additionally, in the previous panel, one judge suggested a minor change to the text, which was incorporated into the 'A sopa' final version. Furthermore, special consideration was given to both songs' texts, ensuring that all words were part of the children's vocabulary. Permission was asked from GIA Publications to use 'Let's play in the snow', which was granted. Both musical stimuli were in major tonality, with a triple meter, and in the key of D, ranging from C3 to A3. The song 'Let's play in the snow' is referred to as Song 1, and 'A sopa' as Song 2 (Figure 1).

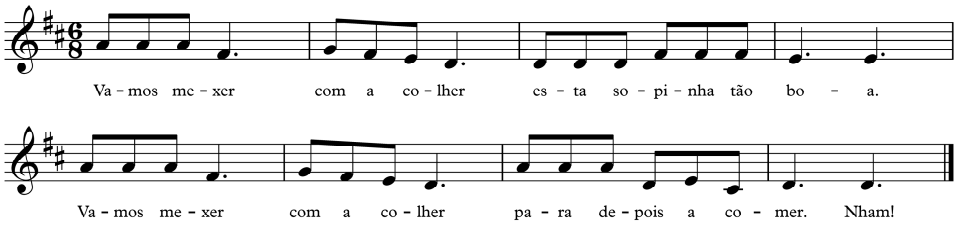



**Figure 1.** Song 2, 'A sopa'.

### *2.4. Instruments*

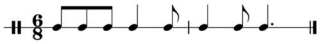
The rating scales used to assess children's vocal performance were specifically designed for this study and used in both periods (Figures 2 and 3). The rating scales were additive and consisted of two dimensions, each with five criteria (Gordon, 2002). Each dimension was rated on a scale of 0 to 5. The combined rating of both dimensions (ranging between 0 and 10) represented the composite score for the song's vocal performance. The validity and reliability of both scales are discussed in the results section.

**Tonal dimension**

a. First pitch is accurate (according to the auditory cue provided) (bar 1)
b. Tonic function pattern is accurate (bar 2)
c. Dominant function pattern to tonic resolution is accurate (bars 7–8)
d. Phrase a is accurate (bars 1–4)
e. Phrase b is accurate (bars 5–8)

**Rhythm dimension**

a. Consistency of tempo is maintained throughout phrase a (bars 1–4)
b. Consistency of tempo is maintained throughout phrase b (bars 5–8)
c. Rhythm pattern is accurate (bars 1–2)



d. Rhythm pattern is accurate (bars 3–4)

e. Rhythm pattern is accurate (bars 7–8)

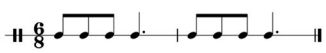
**Figure 2.** Rating scale for Song 1, 'Let's play in the snow'.

**Tonal dimension**

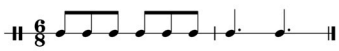
a. First pitch is accurate (according to the auditory cue provided) (bar 1)
b. Dominant-tonic interval is accurate (bar 7)
c. Leading tone – tonic interval is accurate (anacruse to bar 8)
d. Phrase a is accurate (bars 1–4)
e. Phrase b is accurate (bars 5–8)

**Rhythm dimension**

a. Consistency of tempo is maintained throughout phrase a (bars 1–4)
b. Consistency of tempo is maintained throughout phrase b (bars 5–8)
c. Rhythm pattern is accurate (bars 1–2)



f. Rhythm pattern is accurate (bars 3–4)



g. Rhythm pattern is accurate (bars 7–8)

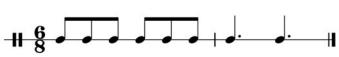


**Figure 3.** Rating scale for Song 2, 'A sopa'.

### 2.5. Data Analyses

Two experts in the field and the curriculum used—Music Learning Theory-based (Gordon, 2003, 2012)—assessed the content and construct validity of the rating scales (Boyle & Radocy, 1987; T. Brophy, 2000; Fautley, 2010). Three raters—music teachers who regularly work with children aged 4 to 9 years old—evaluated the randomized recordings of the children's vocal performances on both songs. Each rater was recommended to listen and rate every child's vocal performance on the tonal dimension before passing to the other dimension (Gordon, 2002). Intraclass correlation coefficients (ICC) were used to assess inter-rater reliability. Descriptive statistics were calculated for the song scores by grade level and sex in both periods. A mixed ANOVA was run with both songs' composite scores as the within-participants factor and grade level as the between-participants factor. The best option to analyze both grade level and sex would be to use sex as a second between-subjects factor. However, the normality assumption for the subgroups was not met, so sex was excluded from these mixed ANOVAs and was analyzed using a *t*-test for independent samples. Since the groups are dissimilar in size, the effect size considered was Hedges' *g* (Ellis, 2010; Marôco, 2014). Calculations revealed that effect sizes for Song 1 are high ($|0.5 \leq g < 1|$) and those Song 2 are medium ($|0.2 < g < 0.5|$), considering Periods One and Two. Data analysis was performed using the Statistical Package for the Social Sciences (SPSS, v.20).

## 3. Results

### 3.1. Reliability of the Rating Scales

Inter-rater reliabilities were computed for Song 1 and Song 2 on each dimension (tonal and rhythm) and for the composite scores of each rating scale, in both periods. The high coefficient values found indicated that the three raters consistently used each measure (Table 3).

**Table 3.** Intraclass correlation coefficient [ICC(2,k)] for Songs 1 and 2 by dimension and period.

| | **Period One** | | **Period Two** | |
|---|---|---|---|---|
| Dimension | Song 1 (n = 135) | Song 2 (n = 135) | Song 1 (n = 135) | Song 2 (n = 135) |
| Tonal | 0.88 | 0.98 | 0.96 | 0.95 |
| Rhythm | 0.82 | 0.86 | 0.87 | 0.80 |
| Composite | 0.88 | 0.95 | 0.94 | 0.93 |

For Song 1, reliability coefficient values ranged from ICC(2,k) = 0.82 to 0.96, and for Song 2 from ICC(2,k) = 0.80 to 0.98. Although still high, the ICCs are lower for Song 1 in Period One.

### 3.2. Effects of Singing with Text and a Neutral Syllable on Song Vocal Performance

Considering the high reliability coefficient values found, six new variables were computed based on the mean of the three scores set, i.e., the score set of each rater for both periods and songs: Rhythm Song 1; Tonal Song 1; Composite Song 1; Rhythm Song 2; Tonal Song 2; and Composite Song 2. These variables were computed for both periods.

Period One

Means and standard deviations for Period One are presented by grade level (Table 4) and sex (Table 5).

**Table 4.** Means (standard deviations) of Songs 1 and 2 scores by grade level in Period One.

| | Song 1 | | | Song 2 | | |
|---|---|---|---|---|---|---|
| **Grade Level** | **Tonal** | **Rhythm** | **Composite** | **Tonal** | **Rhythm** | **Composite** |
| K4 ($n$ = 22) | 1.80 (1.50) | 2.03 (1.52) | 3.83 (2.37) | 0.70 (1.71) | 1.35 (1.22) | 2.05 (1.43) |
| K5 ($n$ = 19) | 2.63 (1.67) | 3.44 (1.02) | 6.07 (2.22) | 1.67 (1.78) | 1.96 (1.28) | 3.63 (2.26) |
| 1st ($n$ = 27) | 2.21 (1.81) | 2.96 (1.24) | 5.17 (2.54) | 1.80 (1.76) | 2.96 (1.40) | 4.77 (2.72) |
| 2nd ($n$ = 22) | 3.80 (1.26) | 3.98 (0.88) | 7.79 (1.66) | 3.53 (1.33) | 3.58 (0.99) | 7.11 (1.85) |
| 3rd ($n$ = 23) | 3.29 (1.95) | 3.33 (1.48) | 6.62 (3.25) | 2.86 (2.12) | 3.86 (1.42) | 6.71 (3.19) |
| 4th ($n$ = 22) | 3.38 (1.76) | 3.76 (1.24) | 7.14 (2.56) | 3.23 (1.89) | 3.74 (1.29) | 6.97 (2.90) |
| Total ($n$ = 135) | 2.84 (1.79) | 3.24 (1.39) | 6.07 (2.77) | 2.30 (1.93) | 2.94 (1.56) | 5.23 (3.08) |

Note: The highest score for Tonal and Rhythm dimensions is 5 points; the highest score for the Composite is 10 points.

**Table 5.** Means (standard deviations) of Songs 1 and 2 scores by sex in Period One.

| | Song 1 | | | Song 2 | | |
|---|---|---|---|---|---|---|
| **Sex** | **Tonal** | **Rhythm** | **Composite** | **Tonal** | **Rhythm** | **Composite** |
| Girls ($n$ = 72) | 3.35 (1.63) | 3.61 (1.20) | 6.96 (2.47) | 2.58 (1.96) | 3.04 (1.48) | 5.62 (3.13) |
| Boys ($n$ = 63) | 2.25 (1.81) | 2.80 (1.46) | 5.05 (2.77) | 1.97 (1.85) | 2.82 (1.66) | 4.79 (2.99) |
| Total ($n$ = 135) | 2.84 (1.79) | 3.24 (1.39) | 6.07 (2.77) | 2.30 (1.93) | 2.94 (1.56) | 5.23 (3.08) |

Note: The highest score for tonal and rhythm dimensions is 5 points; the highest score for the composite is 10 points.

In Period One, there was a significant within-participants main effect for composite scores $F(1, 129) = 31.9$, $p < 0.001$, $h^2_p = 0.20$. Children scored significantly higher on Song 1, with a mean difference = 0.90, $SE = 0.16$. There was a significant between-participants effect for grade level $F(5, 129) = 12.1$, $p < 0.001$, $h^2_p = 0.32$. Games-Howell post hoc comparisons indicated a significant difference between K4 and all the other grades ($p < 0.05$); between K5 and second ($p = 0.001$) and fourth ($p = 0.035$); and between the first and fourth grades ($p = 0.001$). In these instances, the oldest children of each pair were more accurate than the youngest. There was also a significant interaction, $F(5, 129) = 6.13$, $p < 0.001$, $h^2_p = 19$. Pairwise comparisons were run for all grade levels and significant differences in favour of Song 1 were found for K4, $F(1, 21) = 15.4$, $p = 0.001$, $h^2_p = 0.42$, mean = 1.79, $SE = 0.46$, K5, $F(1, 18) = 22.3$, $p < 0.001$, $h^2_p = 0.55$, mean = 2.43, $SE = 0.52$, and second grade, $F(1, 21) = 4.82$, $p = 0.039$, $h^2_p = 0.19$, mean = 0.68, $SE = 0.31$.

Regarding sex, a $T$-test for independent samples was conducted to determine if the composite scores for Song 1 and Song 2 differed between girls and boys. Significant differences were found for Song 1, $t(133) = 4.22$, $p < 0.001$, $g = 0.73$, where girls scored higher than boys (Girls: mean = 6.96, $SE = 0.29$; Boys: mean = 5.06, $SE = 0.35$), but not for Song 2, $t(133) = 1.56$, $p = 0.12$, $g = 0.27$ (Girls: mean = 5.62, $SE = 0.37$; Boys: mean = 4.79, $SE = 0.38$).

Period Two

Means and standard deviations for Period Two are presented by grade level (Table 6) and sex (Table 7).

**Table 6.** Means (standard deviations) of Songs 1 and 2 scores by grade level in Period Two.

| | Song 1 | | | Song 2 | | |
|---|---|---|---|---|---|---|
| **Grade Level** | **Tonal** | **Rhythm** | **Composite** | **Tonal** | **Rhythm** | **Composite** |
| K4 ($n$ = 22) | 2.08 (1.71) | 2.77 (1.53) | 4.85 (2.78) | 1.55 (1.88) | 2.11 (1.50) | 3.65 (2.96) |
| K5 ($n$ = 19) | 2.88 (1.55) | 3.81 (1.27) | 6.68 (2.33) | 2.33 (2.07) | 2.84 (1.01) | 5.18 (2.60) |
| 1st ($n$ = 27) | 2.49 (1.96) | 3.59 (1.28) | 6.09 (2.68) | 2.80 (1.64) | 3.49 (1.32) | 6.30 (2.44) |
| 2nd ($n$ = 22) | 4.18 (1.16) | 4.03 (1.16) | 8.21 (1.74) | 4.24 (1.26) | 4.05 (1.05) | 8.29 (1.79) |
| 3rd ($n$ = 23) | 3.41 (1.98) | 4.03 (1.15) | 7.43 (2.86) | 3.51 (1.76) | 3.72 (1.33) | 7.23 (2.58) |
| 4th ($n$ = 22) | 3.76 (1.77) | 4.18 (1.00) | 7.94 (2.29) | 3.88 (1.76) | 3.97 (0.91) | 7.85 (2.21) |
| Total ($n$ = 135) | 3.12 (1.85) | 3.73 (1.30) | 6.85 (2.71) | 3.06 (1.93) | 3.38 (1.37) | 6.44 (2.88) |

Note: The highest score for tonal and rhythm dimensions is 5 points; the highest score for the composite is 10 points.

**Table 7.** Means (standard deviations) of Songs 1 and 2 scores by sex in Period Two.

| | Song 1 | | | Song 2 | | |
|---|---|---|---|---|---|---|
| **Sex** | **Tonal** | **Rhythm** | **Composite** | **Tonal** | **Rhythm** | **Composite** |
| Girls ($n$ = 72) | 3.59 (1.62) | 3.93 (1.09) | 7.52 (2.35) | 3.36 (1.86) | 3.54 (1.25) | 6.91 (2.70) |
| Boys ($n$ = 63) | 2.57 (1.95) | 3.50 (1.49) | 6.07 (2.89) | 2.71 (1.97) | 3.19 (1.49) | 5.91 (3.00) |
| Total ($n$ = 135) | 3.12 (1.85) | 3.73 (1.30) | 6.85 (2.71) | 3.06 (1.93) | 3.38 (1.37) | 6.44 (2.88) |

Note: The highest score for tonal and rhythm dimensions is 5 points; the highest score for the composite is 10 points.

In Period Two, there was a significant within-participants main effect for composite scores, $F$ (1, 129) = 9.07, $p$ = 0.003, $h^2_p$ = 0.066. Children scored significantly higher on Song 1, with a mean difference = 0.45, and $SE$ = 0.15. There was a significant between-participants effect for grade level, at $F$ (5, 129) = 9.06, $p < 0.001$, $h^2_p$ = 0.26. Scheffè post hoc comparisons indicated a significant difference between K4 and second ($p < 0.001$), third ($p$ = 0.002), and fourth grades ($p < 0.001$), where the oldest children of each pair were more accurate than the youngest. There was also a significant interaction, $F$ (5, 129) = 3.68, $p$ = 0.004, $h^2_p$ = 12. Pairwise comparisons were run for all grade levels and significant differences in favour of Song 1 were found for K4, $F$ (1, 21) = 13.9, $p$ = 0.001, $h^2_p$ = 0.40, mean = 1.20, $SE$ = 0.32, and K5, $F$ (1, 18) = 10.5, $p$ = 0.005, $h^2_p$ = 0.37, mean = 1.51, $SE$ = 0.46.

As for sex, a $T$-test for independent samples was run to check if both Song 1 and Song 2 composite scores were different between girls and boys. Significant differences were found for Song 1, $t$(119.7) = 3.17, $p$ = 0.002, $g$ = 0.55 (Girls: mean = 7.52, $SE$ = 0.28; Boys: mean = 6.07, $SE$ = 0.36), and also for Song 2, $t$(133) = 2.04, $p$ = 0.045, $g$ = 0.35 (Girls: mean = 6.91, $SE$ = 0.32; Boys: mean = 5.91, $SE$ = 0.38). Girls scored significantly higher than boys.

## 4. Discussion

The purpose of this study was to examine whether children's vocal performances of two songs differed by grade level or sex, across two instructional periods. In Period One, one song (Song 1) was taught with text and another (Song 2) with a neutral syllable. In Period Two, both songs were taught with text. The findings contribute to ongoing research on the role of instructional strategies, age, and sex in the development of children's singing.

### *4.1. Vocal Performance Differences in Period One (RQ1)*

Results from Period One indicated that younger children (K4, K5, and second grade) performed significantly better on Song 1, which was taught with text, than on Song 2, which used a neutral syllable. This suggests that presenting a song with text may facilitate more accurate vocal performance among younger children. These findings align with those

of Jacobi-Karna (1996) and Pereira et al. (2016), who also reported better performance outcomes when songs were presented with text, particularly among kindergarten-aged children. In contrast, older children (third and fourth graders) did not show statistically significant differences in performance between the two songs.

Sex differences in Period One showed that girls outperformed boys overall, though no significant sex effect was found for Song 2 (neutral syllable). The previous literature has yielded mixed results regarding sex in vocal performance; some studies have found no significant differences (Goetze et al., 1990), while others have identified superior accuracy in girls (Goetze, 1985; Jordan-DeCarbo, 1982; Welch et al., 1997). The findings in this study may suggest that singing with a neutral syllable presents similar challenges to both sexes, supporting the notion that sex-related differences in singing may be task-specific (Leighton & Lamont, 2006).

Although Levinowitz (1989) found tonal advantages for the song taught with a neutral syllable, her study did not analyze children by grade level or sex, and it used continuous rather than additive rating scales, limiting comparison. Other studies investigating children's vocal performance with text or a neutral syllable did not analyze sex (e.g., Goetze, 1985; Jacobi-Karna, 1996; Lange, 1999; Smale, 1987), which also limits comparison.

### *4.2. Vocal Performance Differences in Period Two (RQ2)*

In Period Two, both Song 1 and Song 2 were taught with text, and second-grade students again demonstrated higher composite performance scores than their third- and fourth-grade peers. These findings corroborate those of Hornbach and Taggart (2005), who reported that second graders performed better (albeit not significantly) than third graders on a song taught with text. While those authors attributed the result to differences in social rather than musical development, such an explanation may not fully apply here, given that anecdotal observations of these participants reveal frequent and positive singing experiences across grade levels.

K4 and K5 students continued to perform significantly better on Song 1 in Period Two, though the performance gap between grade levels was less pronounced than in Period One. These findings suggest that while younger students benefit consistently from singing with text, older students' performance may reach its maximum level or be influenced by other factors, such as motivation or task familiarity. Nonetheless, the findings corroborate the results of previous studies, suggesting that age influences vocal performance, where the oldest children perform better than the youngest (e.g., Geringer, 1983; Goetze, 1985). Nonetheless, singing instruction that includes appropriate vocal development guidance plays a crucial role in vocal performance, regardless of age (Svec, 2018; Welch et al., 2012).

In terms of sex effects, girls again demonstrated superior performance across both songs. These results echo earlier findings (Goetze, 1985; Welch, 2012). The literature has highlighted that instructional methods and contextual factors, such as teacher–student dynamics, motivation, and task design, may influence boys' vocal engagement and accuracy (Paney & Kay, 2015; Petzold, 1963; Svec, 2015; Welch et al., 1997). For example, Welch et al. (1997) stated that "boys' increasingly inaccurate vocal matching in songs is a negative by-product of the extent to which these males identified with music as a subject taught by female teachers" (p. 156). Yet, as Svec (2018) suggests, the focus should not be on sex differences, but on how to provide appropriate singing instruction—a notion also supported by Welch et al. (2012). Moreover, although research suggests a maturity gap favouring girls at young ages in terms of brain development (Lim et al., 2015), social skills (Hajovsky et al., 2022), and emotional expression (Fomby & Cherlin, 2011), maturation timing can vary between individuals of any sex. Thus, regardless of sex, pedagogical

interventions should acknowledge that a wide range of developmental pathways exists and is influenced by contextual and biological variables.

### 4.3. Implications

The findings emphasize the importance of tailoring vocal instruction strategies to the developmental stage and individual learning profile of each child. Among these strategies, existing or newly created singing games should be considered, as they could help improve pitch accuracy, among other skills, while also attributing ownership of the activity to the children (Byrne et al., 2024). The consistent advantages seen with songs taught using text, particularly among younger children, suggest that text content may support pitch and tonal stability, possibly through increased engagement or contextual memory cues. Yet, for others, the use of a neutral syllable might be less demanding in terms of text memory load.

Furthermore, this study highlights the value of using multi-dimensional rating scales to assess vocal performance. By evaluating, at least, tonal and rhythmic aspects, in addition to pitch accuracy, educators and researchers can gain a richer understanding of children's musical strengths and challenges. This aligns with prior recommendations for more holistic assessment practices (Gault, 2002; Gordon, 2002).

### 4.4. Limitations and Future Directions

This study has a few limitations. First, the findings can only be generalized to settings and contexts that are like those examined in the present research. Secondly, there is the absence of a control group or a counterbalanced design. Future research could also adapt the design, wherein Song 1 and Song 2 are both presented with a neutral syllable during Period One. Text would be added to Song 1 during Period Two, while Song 2 continued with a neutral syllable. Thirdly, in the design of this study, it should be acknowledged that Song 1 had more instructional time. Additionally, Song 2's first pitch was set in the dominant key, whereas Song 1's start was in the tonic key. This may have contributed to a slightly more challenging performance level, especially if the voice register was not fully accessed or comfortable for children.

Although a limited number of longitudinal studies are available in the literature (e.g., Leighton & Lamont, 2006; Rutkowski & Miller, 2002; Welch et al., 2011), most of the existing research relies on cross-sectional or "snapshot" designs, including the current study. As noted by several scholars, longitudinal approaches are essential for gaining a comprehensive understanding of the development of musical behaviors over time. Future research on singing with text and a neutral syllable may consider a longitudinal investigation to track children's behaviors related to vocal performance, including the use of the singing voice. Limited access to certain vocal registers may hinder a child's ability to sing in tune. Rutkowski (2015), in a pattern-singing task, observed that even when children access all vocal registers, pitch accuracy across those registers is not guaranteed.

Future research may also include assessing children's music aptitude, as conceptualized by Gordon (1983, 2001a, 2001b). The literature reports mixed findings regarding the relationship between musical aptitude and singing achievement, which might be due to methodological differences. Some studies have identified moderate to strong relationships (Guerrini, 2004; Jaffurs, 2000; Jones, 1993; Phillips & Aitchison, 1997, 1999); others have suggested a predictive relationship (Rodrigues, 1997), while some have found no significant relationship (Atterbury & Silcox, 1993; Hornbach & Taggart, 2005; Mota, 1997; Phillips et al., 2002; Rutkowski, 2015). It may be valuable to explore whether tonal and rhythm aptitude scores are related to tonal and rhythm rating scores in vocal tasks. Additionally, the interaction between vocal performance and linguistic development deserves further exploration, particularly given Levinowitz's (1989) suggestion that language skills may

be linked to singing ability. Moreover, the relationship between phonological awareness and music aptitude has been investigated by Culp (2017), and future research could explore these variables in conjunction with the vocal performance of songs with text and a neutral syllable.

Moreover, future research should investigate the influence of teacher sex, instructional models, and student self-concept in vocal music education, especially as they pertain to sex engagement patterns, as well as age engagement patterns (e.g., second-grade students demonstrated higher composite performance scores than older children for both songs) observed in this and other studies. Future studies could also investigate gender and age sensitivities when comparing songs in different meters and tonalities (the songs used in this study were in triple meter and major tonality). A sociocultural perspective may also shed light on how factors such as ethnicity, socioeconomic status, oral traditions, and peer influence—as well as broader cultural norms, identity development, and community values—shape children's engagement and performance in vocal music contexts. Another avenue for research is to complement children's vocal performance assessment by incorporating a dataset of pedagogical singing games. In the context of this study, it could be advantageous to use game songs that include text and a combination of neutral syllables or onomatopoeia, for example. While helpful in assessing vocal performance, these games also enable children to engage in child-directed music-making experiences that foster meaningful participation, learning, self-determination, and musical development (Byrne et al., 2024; Roberts, 2018; Zosh et al., 2018). T. S. Brophy (1997) had already shown that singing games were valid and reliable for use in the authentic assessment of vocal pitch accuracy. Thus, accommodating vocal assessment through a child-centred playful pedagogy should also be of concern to teachers. Lastly, while the additive rating scales used here proved helpful, converting them into continuous measures may enhance sensitivity and enable more nuanced analysis of performance data.

## 5. Conclusions

In summary, this study finds that younger children, especially K4, K5, and second graders, tend to perform better on songs presented with text, and that girls generally outperform boys in terms of vocal performance, but not when singing with a neutral syllable. However, interaction effects suggest that both song-teaching strategies—text and neutral syllable—might have pedagogical value (Jacobi-Karna, 1996; Levinowitz, 1989; Welch, 2012) due to children's individual learning preferences. Incorporating both approaches in early music education may support broader participation and musical development, potentially influencing children's long-term engagement (Demorest et al., 2016). To this end, it is also essential to maintain a playful learning environment from an early age, allowing both musical games and guided musical play, which, by nature, prioritize body and movement.

**Author Contributions:** Conceptualization, A.I.P. and H.R.; methodology, A.I.P. and H.R.; formal analysis, A.I.P.; investigation, A.I.P.; data curation, A.I.P.; writing—original draft preparation, A.I.P.; writing—review and editing, A.I.P. and H.R.; supervision, A.I.P. and H.R.; project administration, A.I.P.; funding acquisition, A.I.P. and H.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was developed within the framework of CESEM (Centre for Music Studies, UID/EAT-00693/2013). Financial support was provided by FCT (Portuguese Foundation for Research and Technology) through a PhD grant (PD/BD/114489/2016), co-financed under the European Social Fund (ESF) and the Ministry of Science, Technology and Higher Education (MCTES). The APC was partly funded by NOVA FCSH.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by a Review Board at NOVA FCSH (7 March 2016, no code), as the Institutional Ethics Committee was established after the study had begun, and by FCT (Portuguese Foundation for Research and Technology), with a contract signed on 1 April 2016.

**Informed Consent Statement:** Child assent and informed consent were obtained from the legal guardian of the participants involved in the study.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** We would like to acknowledge the parents and children who participated in this study. We would also like to extend our gratitude to the School Board and all the staff, the experts involved in song selection, and the judges who participated in the evaluation tasks.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

Apfelstadt, H. (1984). Effects of melodic instruction on discrimination accuracy of children. *Journal of Research in Music Education*, *32*(1), 15–24. [CrossRef]

Atterbury, B. W., & Silcox, L. (1993). A comparison of home musical environment and musical aptitude in kindergarten students. *Update: Applications of Research in Music Education*, *11*(2), 18–22. [CrossRef]

Boyle, J. D., & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences*. Schirmer Books.

Brophy, T. (2000). *Assessing the developing musician: A guide for general music teachers*. GIA Publications.

Brophy, T. S. (1997). Authentic assessment of vocal pitch accuracy in first through third grade children. *Contributions to Music Education*, *24*(1), 57–70.

Byrne, R., Murphy, R., Ward, F., & McCabe, U. (2024). Towards a model of playful music learning for primary classrooms: Recommendations based on a review of literature. *British Journal of Music Education*, *41*(3), 242–253. [CrossRef]

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Culp, M. E. (2017). The relationship between phonological awareness and music aptitude. *Journal of Research in Music Education*, *65*(3), 328–346. [CrossRef]

Davidova, J., Zavadska, G., & Rauduvaitė, A. (2019). Levels and indicators for assessment of singing voice of 6–8 year-old children. *Pedagogika*, *133*(1), 149–160. [CrossRef]

Davidson, L. (1985). Tonal structures of children's early songs. *Music Perception*, *2*(3), 361–373. [CrossRef]

Davidson, L., McKernon, P., & Gardner, H. (1981). The acquisition of song: A developmental approach. In *Documentary report of the Ann Arbor symposium: Applications of psychology to the teaching and learning of music*. Music Educators. National Conference.

Demorest, S. M., Kelley, J., & Pfordresher, P. Q. (2016). Singing ability, musical self-concept, and future music participation. *Journal of Research in Music Education*, *64*(4), 405–420. [CrossRef]

Demorest, S. M., & Pfordresher, P. Q. (2015). Singing accuracy development from K-adult: A comparative study. *Music Perception*, *32*(3), 293–302. [CrossRef]

Ellis, P. D. (2010). *The Essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press.

Fautley, M. (2010). *Assessment in music education*. Oxford University Press.

Feierabend, J. M. (1984). *The effects of specific tonal pattern training on singing and aural discrimination abilities of first grade children* [Doctoral dissertation, Temple University]. University Microfilms No. DA8410243.

Fixmer, D. (2022). The effect of self-evaluation, self-recording along with self-listening, and modeling on 2nd graders' melodic singing accuracy. *Research Perspectives in Music Education*, *22*(1), 31–45. [CrossRef]

Fomby, P., & Cherlin, A. J. (2011). Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin*, *72*(2), 181–204. [CrossRef]

Gault, B. (2002). Effects of Pedagogical approach, presence/absence of text, and developmental music aptitude on the song performance accuracy of kindergarten and first-grade students. *Bulletin of the Council for Research in Music Education*, *152*, 54–63.

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. [CrossRef]

Giga, I. (2005). *Efeitos da pedagogia musical ward no desenvolvimento musical e desempenho vocal de crianças do 1º ciclo do ensino básico* [Doctoral dissertation, Universidade de Évora].

Goetze, M. (1985). *Factors affecting accuracy in children's singing* [Unpublished doctoral dissertation, University of Colorado, Boulder]. Dissertation Abstracts International 46: 2955A.

Goetze, M., Cooper, N., & Brown, C. J. (1990). Recent research on singing in the general music classroom. *Bulletin of the Council for Research in Music Education*, *104*(104), 16–37.

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. E. (1983). *Audie*. GIA Publications.

Gordon, E. E. (2001a). *Medidas intermédias de audiação musical*. CEGOC.

Gordon, E. E. (2001b). *Music aptitude and related tests: An introduction*. GIA Publications.

Gordon, E. E. (2002). *Rating scales and their uses for measuring and evaluating achievement in music performance*. GIA Publications.

Gordon, E. E. (2003). *A music learning theory for newborn and young children* (3rd ed.). GIA Publications.

Gordon, E. E. (2012). *Learning sequences in music: A contemporary music learning theory*. GIA Publications.

Gordon, E. E., Bolton, B., Hicks, W., & Taggart, C. (1993). *The early childhood music curriculum experimental: Songs and chants without words, book one*. GIA Publications.

Guerrini, S. C. (2004). The relationship of vocal accuracy, gender, and music aptitude among elementary students. *Visions of Research in Music Education*, *4*, 4.

Hajovsky, D. B., Caemmerer, J. M., & Mason, B. A. (2022). Gender differences in children's social skills growth trajectories. *Applied Developmental Science*, *26*(3), 488–503. [CrossRef]

Hale, M. R. (1977). An experimental study of the comparative effectiveness of harmonic and melodic accompaniment in singing as it relates to the development of a sense of tonality. *Bulletin of the Council for Research in Music Education*, *53*, 23–30.

Hedden, D. (2012). An overview of existing research about children's singing and the implications for teaching children to sing. *Update: Applications of Research in Music Education*, *30*(2), 52–62. [CrossRef]

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*(184), 35–48. [CrossRef]

Hedden, D. G., Johnson, C., & Johnson, C. M. (2008). The effect of teaching experience on time and accuracy of assessing young singers' pitch accuracy. *Bulletin of the Council for Research in Music Education*, *178*, 63–72. [CrossRef]

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), 322–331. [CrossRef]

Jacobi-Karna, K. L. (1996). *The effects of the inclusion of text on the singing accuracy of preschool children* [Doctoral dissertation, The University of Arizona].

Jaffurs, S. E. (2000). *The relationship between singing achievement and tonal music aptitude* [Master's thesis, Michigan State University]. Dissertations & Theses@CIC Institutions.

Jones, M. (1993). An assessment of audiation skills of accurate and inaccurate singers in grades 1, 2, and 3. *Update: Applications of Research in Music Education*, *11*(2), 14–17. [CrossRef]

Jordan-DeCarbo, J. (1982). Same/different discrimination techniques, readiness training, pattern treatment, and sex on aural discrimination and singing of tonal patterns by kindergartner. *Journal of Research in Music Education*, *30*(4), 237–246. [CrossRef]

Lange, D. M. (1999). *The effect of the use of text in music instruction on the tonal aptitude, tonal accuracy, and tonal understanding of kindergarten students* [Unpublished doctoral dissertation, Michigan State University].

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. [CrossRef]

Levinowitz, L. M. (1989). An investigation of preschool children's comparative capability to sing songs with and without words. *Bulletin of the Council for Research in Music Education*, *100*, 14–19.

Lim, S., Han, C. E., Uhlhaas, P. J., & Kaiser, M. (2015). Preferential detachment during human brain development: Age- and sex-specific structural connectivity in diffusion tensor imaging (DTI) data. *Cerebral Cortex*, *25*(6), 1477–1489. [CrossRef]

Mang, E. (2006). The effects of age, gender and language on children's singing competency. *British Journal of Music Education*, *23*(02), 161. [CrossRef]

Marôco, J. (2014). *Análise estatística com o SPSS Statistics* (6th ed.). ReportNumber.

Miyamoto, K. A. (2005). The effects of a remedial singing method on the vocal pitch accuracy of inaccurate elementary singers. *Research and Issues in Music Education*, *3*(1), article 2.

Mizener, C. P. (2008). Our singing children. *General Music Today*, *21*(3), 18–24. [CrossRef]

Moog, H. (1976). The development of musical experience in children of pre-school age. *Psychology of Music*, *4*(2), 38–45. [CrossRef]

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. [CrossRef] [PubMed]

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. [CrossRef]

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

Nichols, B. E. (2015). Critical variables in singing accuracy test construction: A review of literature. *Update: Applications of Research in Music Education*, *35*(1), 39–46. [CrossRef]

Nichols, B. E. (2016). Task-based variability in children's singing accuracy. *Journal of Research in Music Education*, *64*(3), 309–321. [CrossRef]

Nichols, B. E. (2021). Effect of vocal versus piano doubling on children's singing accuracy. *Psychology of Music*, *49*(5), 1415–1423. [CrossRef]

Paney, A. S., & Kay, A. C. (2015). Developing singing in third-grade music classrooms: The effect of a concurrent-Feedback computer Game on pitch-matching skills. *Update: Applications of Research in Music Education*, *34*(1), 42–49. [CrossRef]

Papageorgi, I., Saunders, J., Himonides, E., & Welch, G. F. (2022). Singing and social identity in young children. *Frontiers in Psychology*, *13*, 1–13. [CrossRef]

Pereira, A. I., & Rodrigues, H. (2019). The relationship between Portuguese children's use of singing voice and singing accuracy when singing with text and a neutral syllable. *Music Perception*, *36*(5), 468–479. [CrossRef]

Pereira, A. I., Rodrigues, H., & Ávila, P. (2016). The relationship between kindergarten children's vocal performance and recognition of songs with and without words. In T. Zanto (Ed.), *Proceedings of the 14th international conference on music perception and cognition* (pp. 549–550). Causal Productions. Available online: https://icmpc.org/icmpc14/files/ICMPC14_Proceedings.pdf (accessed on 20 December 2024).

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. [CrossRef]

Petzold, R. G. (1969). Auditory perception by children. *Journal of Research in Music Education*, *17*(1), 82–87. [CrossRef]

Pfordresher, P. Q. (2022). Singing accuracy across the lifespan. *Annals of the New York Academy of Sciences*, *1515*(1), 120–128. [CrossRef]

Pfordresher, P. Q., & Brown, S. (2007). Poor pitch singing in the absence of tone-deafness. *Music Perception*, *25*(2), 95–116. [CrossRef]

Philipps, K. H., & Doneski, S. M. (2011). Research on elementary and secondary school singing. In R. Colwell, & P. R. Webster (Eds.), *MENC handbook of research on music learning* (volume 2: Applications, pp. 176–232). Oxford University Press.

Phillips, K. H. (2014). *Teaching kids to sing* (2nd ed.). Schirmer.

Phillips, K. H., & Aitchison, R. E. (1997). The relationship of singing accuracy to pitch discrimination and tonal aptitude among third-grade students. *Contributions to Music Education*, *24*(1), 7–22.

Phillips, K. H., & Aitchison, R. E. (1999). Second-year results of a longitudinal study of the relationship of singing instruction, pitch accuracy, and gender to aural acuity, vocal achievement, musical knowledge, and attitude towards singing among general music students. *Contributions to Music Education*, *26*(1), 67–85.

Phillips, K. H., Aitchison, R. E., & Nompula, Y. P. (2002). The relationship of music aptitude to singing achievement among fifth-grade students. *Contributions to Music Education*, *29*(1), 47–58.

Reifinger, J. L. (2012). Strategies to develop students' conceptual understanding and singing accuracy of pitch. *General Music Today*, *26*(2), 14–19. [CrossRef]

Roberts, J. C. (2018). Self-determination theory and children's singing games in and out of the classroom: A literature review. *Update: Applications of Research in Music Education*, *36*(3), 12–19. [CrossRef]

Rodrigues, H. (1997). *Avaliação da aptidão musical em crianças do 1º ciclo de escolaridade* [Doctoral dissertation, Universidade de Coimbra].

Runfola, M., & Etopio, E. A. (2009). The nature of performance-based criterion measures in early childhood music education research, and related issues. In T. S. Brophy (Ed.), *Proceedings of the 2009 symposium on assessment in music education: Frameworks, models, and designs* (pp. 395–411). GIA Publications.

Rutkowski, J. (1993, March). *The use of words versus a neutral syllable for evaluating children's use of singing voice*. Eastern Regional Meeting of the Music Educators' National Conference, Springfield, MA, USA.

Rutkowski, J. (1998). The nature of children's singing voices: Characteristics and assessment. In B. A. Roberts (Ed.), *Sharing the voices: The phenomenon of singing* (pp. 201–209). Memorial University Press.

Rutkowski, J. (2015). The relationship between children's use of singing voice and singing accuracy. *Music Perception*, *32*(3), 283–292. [CrossRef]

Rutkowski, J. (2019). Assessing singing voice development. In T. Brophy (Ed.), *Handbook of assessment practice and policy in music education* (Vol. 2, pp. 629–651). Oxford University Press. Available online: http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780190248130.001.0001/oxfordhb-9780190248130-e-64 (accessed on 20 December 2024).

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

Rutkowski, J., & Miller, M. S. (2002). A longitudinal study of elementary children's acquisition of their singing voices. *Update: Applications of Research in Music Education*, *22*(1), 5–14. [CrossRef]

Rutkowski, J., & Trollinger, V. (2005). Experiences: Singing. In *The musical lives of young children* (pp. 78–97). Pearson Prentice Hall.

Sims, W. L., Moore, R. S., & Kuhn, T. L. (1982). Effects of female and male vocal stimuli, tonal pattern length, and age on vocal pitch-matching abilities of young children from England and the United States. *Psychology of Music, Special Issue: Proceedings of the Ninth International Seminar on Research in Music Education*, 104–108.

Smale, M. J. (1987). *An investigation of pitch accuracy of four- and five-year-old singers* [Doctoral dissertation, University of Minnesota].

Svec, C. L. (2015). *The effects of instruction on the singing ability of children ages 5 to 11: A meta-analysis* [Doctoral dissertation, University of North Texas]. Available online: https://www.proquest.com/openview/497b3e96a7adbc1c1c708a28db49cd62/1?pq-origsite=gscholar&cbl=18750 (accessed on 20 December 2024).

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. [CrossRef]

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. [CrossRef]

Tsang, C. D., Friendly, R. H., & Trainor, L. J. (2011). Singing development as a sensorimotor interaction problem. *Psychomusicology: Music, Mind and Brain*, *21*(1–2), 31–44. [CrossRef]

Valerio, W., Reynolds, A., Bolton, B., Taggart, C., & Gordon, E. E. (2000). *Jump right in: The music curriculum. Teacher's edition, book 1*. GIA Publications.

Welch, G. F. (2012). Singing and vocal development. In G. E. McPherson (Ed.), *The child as musician: A handbook of musical development* (2nd ed., pp. 441–461). Oxford University Press. [CrossRef]

Welch, G. F., Himonides, E., Saunders, J., Papageorgi, I., Rinta, T., & Pretti, C. (2011). Researching the impact of the national singing programme 'sing up' in England. *Psychomusicology: Music, Mind and Brain*, *21*(1/2), 83–97. [CrossRef]

Welch, G. F., Saunders, J., Papageorgi, I., & Himonides, E. (2012). Sex, gender and singing development: Making a positive difference to boys' singing through a national programme in England. In S. Harrison, G. F. Welch, & A. Adler (Eds.), *Perspectives on males and singing* (pp. 37–54). Springer. [CrossRef]

Welch, G. F., Sergeant, D. C., & White, P. J. (1995). The singing competencies of five-year-old developing singers. *Bulletin of the Council for Research in Music Education*, *127*, 155–162.

Welch, G. F., Sergeant, D. C., & White, P. J. (1997). Age, sex and vocal task as factors in singing 'in-tune' during the first years of schooling. *Bulletin of the Council for Research in Music Education*, *133*, 153–160.

Welch, G. F., Sergeant, D. C., & White, P. J. (1998). The role of linguistic dominance in the acquisition of song. *Research Studies in Music Education*, *10*(1), 67–74. [CrossRef]

Wise, K. J., & Sloboda, J. A. (2006). Establishing an empirical profile of self-defined "tone deafness". *Musicae Scientiae*, *XII*(1), 3–23. [CrossRef]

Yarbrough, C., Green, G., Benson, W., & Bowers, J. (1991). Inaccurate singers: An exploratory study of variables affecting pitch-matching. *Bulletin of the Council for Research in Music Education*, *107*, 23–34.

Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Solis, S. L., & Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology*, *9*, 1–12. [CrossRef]