

THE RELATIONSHIP BETWEEN PORTUGUESE CHILDREN'S USE OF SINGING VOICE AND SINGING ACCURACY WHEN SINGING WITH TEXT AND A NEUTRAL SYLLABLE

ANA ISABEL PEREIRA, & HELENA RODRIGUES
Universidade NOVA de Lisboa, Lisboa, Portugal

THE PURPOSE OF THIS STUDY WAS TO INVESTIGATE the relationship between Portuguese children's use of singing voice and their singing accuracy on the pitches belonging to the Singing Voice Development Measure (SVDM) criterion patterns (Rutkowski, 2015), as well as the influence on singing with a neutral syllable or text on both variables. Children aged 4 to 9 ($n = 137$) were administered the SVDM individually and three raters evaluated recordings of the children's singing, both for the use of singing voice (i.e., effective use of pitch range and register) and singing accuracy. Prior to data analysis, the validity and reliability of the measure was examined and assured. A significant relationship was found between both variables. Favoring the neutral syllable, significant differences were found in response mode for singing accuracy, but not for use of singing voice, suggesting that the use of neutral syllable in classroom singing activities might be beneficial to improve accuracy. Older children and girls obtained higher scores for the use of singing voice and accuracy. Within a common pitch range, children with higher SVDM scores sang accurately a higher number of pitches, suggesting that expanding children's use of singing voice might also improve singing accuracy.

Received: February 2, 2018, accepted April 28, 2019.

Key words: children's use of singing voice, singing accuracy, pitch-matching, neutral syllable, text

ALTHOUGH SINGING IS CONSIDERED A CORNERSTONE in the Portuguese National Curriculum of Music Education from 1st through 9th grades, and also in the Curricular Guidelines for Preschool Education, no studies have been conducted in Portugal regarding the assessment of children's ability to use the singing voice. Therefore, this study first explored the validity and reliability of the Singing Voice Development Measure (SVDM) (Rutkowski, 2015) with

Portuguese children. The SVDM is a measure of children's use of singing voice that encompasses five main categories, with nine scoring levels, defined on the basis of three voice registers. This measure has already been shown to be a valid measure of children's use of singing voice with participants from the USA (e.g., Guerrini, 2006; Kim, 2000; Levinowitz et al., 1998) and with participants from other countries (e.g., Mang, 2006; Rutkowski & Chen-Hafteck, 2001; Welch et al., 2011). On the other hand, the relationship between children's use of singing voice and their ability to sing accurately has received little study. Rutkowski (2015) explored this possible relationship and it is clear that further replication is needed. Thus, results of this study are also compared with Rutkowski's findings using the same methodologies.

Singing Voice Development Measure (SVDM)

Everyone has the potential to learn to sing. From early childhood through adolescence, the human voice has a distinctive anatomy and physiology that is capable of producing a diversity of singing behaviors (Welch, 2006). Assuming that singing is a developmental skill, it is important to understand its nature and acknowledge that children of the same chronological age can be in different phases of development. Therefore, singing performance assessments based on accuracy need to take children's psychological as well as physical development of the vocal apparatus into consideration rather than just age (Trollinger, 2003; 2007; Welch, 2006).

In order to establish a more consistent way of describing the various stages of children's singing-voice development, the Singing Voice Development Measure (SVDM) was developed, piloted and implemented by Rutkowski (1990, 1996, 1999). Its purpose is to measure the use of singing voice, which is related to vocal register rather than vocal accuracy. Although the label "vocal register" may be controversial in different domains of vocology as well within voice science and voice education communities (Finks, 1992; Thurman, Welch, Theimer, & Klitzke, 2004), it is important to acknowledge how Rutkowski defines and uses it in order to

contextualize its use throughout this article. The author defends that two singing behaviors have been used to assess the use of singing voice—vocal range and vocal register: “While vocal range is one indicator of singing development . . . this behavior alone does not provide a complete picture of use of singing voice” (Rutkowski, 2019, p. 630). The author then illustrates the difference between vocal range and vocal register with the following example: “. . . a 6th-grade girl may sing with a range of A3 to C5 but only be using the chest (lower) register. In that case, the range does not provide a complete profile of the child’s use of singing voice” (p. 630). Thus, vocal range is only related to the highest and lowest pitches someone can sing. Concerning vocal registers, Rutkowski considers the existence of the chest and head voice registers, though mentioning evidence that suggests that the child voice may have at least three registers (lower, middle, and upper as used in the SVDM). The frontiers between registers are “based on lift points aurally identified in the child voice” (Rutkowski, 2019, p. 631). In this context, the general definition of vocal registers by Titze (as cited in Thurman et al., 2004) may be considered: “. . . perceptually distinct regions of vocal quality that can be maintained over some ranges of pitch and loudness” (p. 2).

Even though the SVDM does not measure singing accuracy, Rutkowski (1990) found that children’s abilities to use their full voice in all vocal registers is a separate but important requisite behavior to achieve singing accuracy. In fact, Rutkowski (2015) found a significant positive relationship between the use of the singing voice and singing accuracy, revealing that a higher percentage of pitches were sung accurately when they belonged to the children’s accessible singing registers.

As mentioned before, Rutkowski (2015) identified at least three registers of the child voice and five main categories of use of singing voice. These are perceptual categories based on results of previous research and upon consultation with several elementary vocal music specialists. It would be difficult to analyze use of singing voice through existing technology because it is not based on pitch alone but vocal quality. Content validity was established so it can be inferred that the established categories do exist (see Rutkowski, 1990). Two types of singers only use the lower register: (a) the pre-singer, the child who does not sing but chants the song text; (b) the speaking range singer, the child who sustain tones with some sensitivity to pitch, usually A3 to C4. Two types of singers are able to use the middle register: (c) the limited range singer, or one who exhibits consistent use of limited range, usually D4 to F4; (d) the initial range singer, or one who uses his/her initial singing range consistently,

usually D4 to A4. One type of singer is able to use the upper register: (e) the singer, who has full singing range above and beyond B4. Further development of the measure included a total of nine possible scores by adding the possibility of inconsistent behaviors between each category (Rutkowski, 1996).

Comfortable Singing Pitches

Several researchers have investigated children’s singing ranges (Goetze, Cooper, & Brown, 1990; Welch, 1979). For example, results indicate that the children’s comfortable singing pitch tends to fall around C4, C#4, and D4 and that the vocal range tends to increase with age (Rutkowski & Miller, 2003a; Welch, 1979). On the other hand, a study by Levinowitz et al. (1998) did not find a significant difference in the use of children’s singing voices from 1st to 6th graders. Nevertheless, it is always important to provide an appropriate nurturing environment including singing tasks designed to match—and then to extend—current vocal behaviors (Welch, 2006).

As a few studies reveal, children’s preferred singing pitches might be related to the range of their speaking voices (Rutkowski, 2015; Welch, Rush, & Howard, 1991). Trollinger (2003) found for children aged 36 to 71 months that the higher the children’s overall speaking voice, the more accurate the ability to match pitches. The researcher also pointed out that “depending upon individual physical development, a child’s speech voice can also be his/her singing voice” (p. 89), which is physiologically related to under or improperly developed muscle memory, muscle strength, ligaments, and proper breathing that has not yet been acquired. She also hypothesized that social and cultural models, along with other factors identified in music education research, might contribute to vocal singing range development. For example, a social preference toward lower speaking voices may encourage children to vocalize in their speaking voice (the lower register as defined by Rutkowski) instead of exploring and using their singing voice.

Trollinger (2003) observed an interesting phenomenon related to children’s singing accuracy and their accessible voice registers. For example, she noted that a child who is just beginning to use the middle register (between D4 to A4) and/or the upper register (above B4) shows difficulty singing D4 and E4. Post hoc analysis in Rutkowski’s study (2015) supported this observation by showing that even when children sang in their accessible voice registers they did not sing accurately all the pitches within those registers. She has also noted when children begin to use a register new to them they often sing lower pitches sharp. For example, some

children who generally sing very well in tune often will still sing pitches near the bottom of their middle register (C4–D4) sharp if they are, in fact, appropriately using the middle register and not the chest register (Rutkowski, 2019).

SVDM Singing Task

Studies investigating children's ability to sing in tune differ in methodology. Namely, there are methodological differences with respect to the age of participants, evaluation tools, assessment measures, and item test difficulty. Moreover, the definition of accurate singing is not always consensual. Nichols (2015) pointed out the need to make a distinction between presentation and response variables. The first type is related to how a singing task is presented to students (e.g., model, use of text or neutral syllable). The latter encompasses the specific conditions under which students are asked to respond and have been considered satisfactory discriminators of accurate singing (Nichols, 2016). These involve pitch-matching tasks (single pitch, interval or patterns) and song singing tasks (phrase and entire song). In fact, research has shown that singing skill development might be considered along a hierarchy of vocal pitch-matching competence. For example, children's singing competency improves along the tasks of matching pitches/simple glides, patterns, all glides, songs, and complex glides (Welch, Sergeant, & White, 1995/1996). In a more recent study, it is suggested that singing skill development proceeds from matching single pitches, intervals, short patterns, towards singing familiar songs from memory (Demorest, Nichols, & Pfordresher, 2018).

The initial version of SVDM required children to sing a criterion song previously taught in music classes plus echoing five tonal patterns (three pitches each), which did not belong to the song but were related to its tonality (Rutkowski, 1990). The most current version of the test is based only on echoed tonal patterns (Rutkowski, 2015) given that children performed slightly better on patterns than songs (Rutkowski, 1990). These results are also corroborated by other studies (e.g., Demorest et al., 2018; Flowers & Dunne-Sousa, 1990; Guerrini, 2006; Nichols, 2016; Sims, Moore, & Kuhn, 1982; Welch et al., 1995/1996; Welch, Sergeant, & White, 1997). For example, in a study with fourth- and fifth-graders, Guerrini (2006) found, "many students learn to sing accurately by expanding their ranges first; after the range has been expanded, many students become more vocally accurate by singing short patterns (the researcher used the SVDM eight patterns of three pitches each), but may not be able to apply this skill to entire songs; and once

students are able to sing one song accurately, using notes above the lift, they appear to be able to sing other songs accurately" (p. 30). In another study, Flowers and Dunne-Sousa (1990) found that preschoolers consistently used larger vocal ranges in echoing patterns than when singing songs. With first graders, Van Zee (as cited in Nichols, 2016, p. 41) also concluded that the ability to match tonal patterns does not always transfer to singing the songs from which they were taken. In general, research has suggested that singing a song from memory (without echoing phrases or the whole song) involves a different set of cognitive skills (Demorest et al., 2018). Children might not use their singing voice when performing a song because this skill includes the memorization of text, rhythm and tonal patterns (Gault, 2002; Rutkowski, 1990).

Singing with Text or Neutral Syllable

Along with the type of response tasks, research has also identified other variables that need to be considered when measuring singing behaviors (see Hedden, 2012; Nichols, 2015). These included the type of vocal model; the type of accompaniment (melodic/harmonic); individual versus unison singing; breath-control training; the approach to teaching a song (immersion or phrase-by-phrase); use of visual, auditory and kinesthetic modes; individual, small group or large group instruction; and use of neutral syllable or text.

Regarding this last variable, during a revision process of the measure, Rutkowski (1999) investigated if there were differences between children's use of the singing voice with neutral syllable and text. No differences between these response modes were found for the participants (first graders). Nevertheless, she noted individual differences in favor of neutral syllable, suggesting that "some children sing better when they are not asked to sing words, perhaps since singing words is too closely related to their speech patterns, while others do not find singing words a problem" (p. 208). So, the current version of the SVDM includes both types of performances in the administration method. In fact, a recent reanalysis of data from Rutkowski and Miller (2003b) with first graders, and Rutkowski (2014) with kindergartners by Rutkowski (2015) revealed significant main effects for response mode (text or neutral syllable) for the use of singing voice and singing accuracy.

However, results from previous studies have been inconsistent regarding whether children sing better with a neutral syllable and/or text. For example, studies have shown: better with neutral syllable (Goetze, 1985); significantly higher scores with text for the youngest group

(Jacobi-Karna, 1996); slightly better with neutral syllable (Levinowitz, 1989); no significant differences but slightly higher scores with text (Rutkowski, 1993); no significant differences but better with neutral syllable (Smale, 1987); and better with text (Welch et al., 1995/1996). Nevertheless, different procedures have been used to investigate the effect of this variable, so the comparison across studies is limited.

Age and Gender Versus Singing Tasks

Research on singing development has shown that accuracy improves with age; for example with preschoolers and fourth-grade (Geringer, 1983); first- to sixth-grade (Petzold, 1963); 5 to 7-years-old (Welch et al., 1997); and K3, seventh- and eight-grades (Yarbrough et al., 1991). A few studies have used a different or revised battery of singing tasks to assess and compare children's singing accuracy (Demorest & Pfordresher, 2015; Leighton & Lamont, 2006; Welch et al., 1997). On the other hand, some researchers suggest that singing accuracy may be related to experience rather than general development (Demorest & Pfordresher, 2015). In a study comparing kindergartners, sixth-graders, and adults, Demorest and Pfordresher (2015) found a considerable improvement from kindergarten to sixth-grade children in the single pitch, interval pitch, and pattern pitch tasks. However, sixth-graders performed better than adults on interval and pattern tasks but did not differ on the single-pitch task. When singing a familiar song from memory, kindergartners were the most accurate. Yet, as the researchers point, these results might be due to differences in song material and not to skill differences.

Studies investigating differences in singing accuracy between girls and boys have shown mixed results. Some researchers have reported no significant differences; for example, with kindergartners (Apfelstadt, 1984; Welch et al., 1995/1996); with 4- to 8-years-old (Leighton & Lamont, 2006); with 8 years-old (Paney & Kay, 2015); and with 8- to 11-years-old (Moore, 1994). Others have found that girls sing more accurately than boys; for example, with first-, second-, third-, and fifth-grades (Green, 1994); with kindergarten, first- and third-grades (Goetze & Horii, 1989); with 7- to 9-years-old (Mang, 2006); with 36 to 71 months (Trollinger, 2003); with K3, seventh- and eight-grades (Yarbrough, Green, Benson, & Bowers, 1991). In addition, results from a longitudinal study by Welch, Saunders, Papageorgi, and Himonides (2012) suggest that boys can be as successful as girls from the same age if appropriate educational experience is provided.

As with the Welch et al. (1997) study, children's vocal pitch accuracy might be task-specific. For example, results from a 3-year longitudinal study found that "boys and girls enter school at age 5 with broadly similar vocal pitch matching abilities and this similarity is maintained across the first 3 years of schooling, except in relation to song singing" (Welch et al., 1997, p. 158). Using a revised test battery from Welch et al. (1997), similar results were found in a different longitudinal study among 4- to 8-years-old (Leighton & Lamont, 2006). Although certain tasks such as single pitch matching might favor girls, pitch accuracy seems not to be affected by gender differences. Some researchers have suggested that differences between girls and boys might be due to attitude and motivation (Petzold, 1969) or, at least for song singing, a cultural origin rather than biological one is in play; that is, that this particular task might be associated with the gender of the predominantly female vocal role model (Welch et al., 1997). On the other hand, the way a singing task is presented might also impact children's singing accuracy. Results from a study where third-graders played a computer game that scored the accuracy of their song singing revealed that boys saw a larger improvement during trials than girls (Paney & Kay, 2015). The authors pointed out that boys are usually more motivated to play video games than girls and that a score is a better motivator for them, thus "perhaps using this approach could help improve negative attitudes towards singing" (Paney & Kay, 2015, p. 48). Of course, if sampling procedures, singing tasks, and scoring systems were standardized, comparisons across age groups, gender, and different cultural backgrounds would be easier.

Aims

The purpose of this study was to investigate the relationship between children's use of singing voice and their singing accuracy on the pitches belonging to the SVDM criterion patterns, as well as the influence on singing with neutral syllable or text on both variables. Furthermore, prior to data analysis the validity and reliability of the SVDM with Portuguese children was explored.

The research questions were:

1. Is the SVDM a valid and reliable measure with Portuguese children?
2. Is there a relationship between children's use of singing voice and singing accuracy?
3. Do children's scores on the SVDM and singing accuracy differ by response mode, grade level and gender?



FIGURE 1. Patterns used in the experiment. The melody was written by Rutkowski (1996), with text adapted for Portuguese with permission.

4. Do children sing accurately all the pitches that fall within their accessible voice registers?

Method

The design of this study was descriptive, not experimental, so cause and effect was not investigated and should not be assumed. One hundred and thirty-seven children (4- to 9-years-old, 63 boys and 74 girls) from kindergarten to fourth grade attending a private school in Lisbon and belonging to families with medium/high income levels participated in this study. All children received music instruction from the same general music teacher once a week for 30 minutes in kindergarten and 60 minutes in first- to fourth-grade.

The SVDM, an instrument designed to measure children's use of singing voice (not vocal accuracy), was administered by the music teacher (the first author) (see the SVDM in the Appendix). The text used for the original patterns, in English, was adapted to Portuguese, respecting rhythmic accents and prosody, as indicated in Figure 1 (translation: John's balloon is blue as the sea. It twirls in his hand without fleeing away).

In preparation for administration, the children practiced the patterns in the classroom setting as a large group as recommended by Rutkowski (1990). Each pattern was one bar of notated music (see Figure 1) for a total of eight patterns of three pitches each. Two different response modes were used: The children echoed the patterns on text and also on a neutral syllable. For individual testing, each child was escorted to a private room at the school, where her/his voice was audio recorded. The music teacher first established the key by playing a "D" using the tuning fork, sang the first pattern, and then cued the child to echo-sing the pattern. The same procedure was continued for all eight patterns and was employed twice for each child since two response modes were considered as recommended by Rutkowski (1990, 2015). In order to control the order effect, half of the children first echoed all of the patterns on text and then

all of the patterns on the neutral syllable "Bah"; the other half of the children echoed the patterns on neutral syllable first. The average length of the test administration was one minute and a half.

Aiming to assess the use of singing voice according to the SVDM, three raters—music teachers who work regularly with children from 4- to 9-years-old—evaluated the randomized recording of children's pattern-singing with text and the neutral syllable "Bah." The raters first evaluated all the recordings for the use of singing voice with neutral syllable and then all the recordings for text. The highest score possible for use of singing voice was 5 according to the SVDM scoring system (see Appendix).

Additionally, the same singing recordings were also rated for pitch singing accuracy within all the patterns of the measure and in the same order as the use of singing voice ratings (first neutral syllable and then text). Singing accuracy was determined by scoring 1 if the child matched the pitch and 0 if not. The highest score possible for the entire measure singing accuracy was 24 (eight patterns of three pitches each).

Results

VALIDITY OF SVDM WITH PORTUGUESE CHILDREN

Prior to analysis of the data, validity and reliability on both SVDM and singing accuracy on the pitches belonging to the SVDM criterion patterns was established. Judges used all scoring levels of the SVDM for both response modes (text and neutral syllable), thus supporting the existence of all the singing behaviors among this sample of Portuguese children. Although a lower incidence of pre-singers (1), inconsistent speaking range singers (1.5), and speaking range singers (2) was found, the results support the validity of the measure.

Considering both response modes, interjudge reliabilities were calculated for the SVDM [neutral syllable and text: ICC(2, k) = .95] and singing accuracy [neutral syllable: ICC(2, k) = .95; text: ICC(2, k) = .98]. Coefficient values were high indicating that the three judges

TABLE 1. Means (Standard Deviations) for SVDM and Singing Accuracy Scores by Grade Level and Response Mode

GRADE	<i>n</i>	SVDM		Singing Accuracy	
		Neutral Syllable	Text	Neutral Syllable	Text
K	43	3.71 (0.86)	3.67 (0.88)	10.58 (6.73)	9.48 (6.01)
1 st	27	3.88 (0.98)	3.76 (0.96)	12.94 (7.17)	10.27 (6.65)
2 nd	22	4.48 (0.67)	4.32 (0.57)	17.59 (5.21)	15.07 (4.75)
3 rd	23	4.21 (0.98)	4.22 (0.98)	14.07 (8.19)	13.88 (8.46)
4 th	22	4.45 (0.79)	4.55 (0.69)	18.18 (6.00)	17.91 (5.80)
Total	137	4.10 (0.92)	4.03 (0.90)	13.98 (7.30)	12.63 (7.06)

Note. Highest SVDM score would be 5; highest Accuracy score would be 24 points

used SVDM and rated the singing accuracy in a consistent manner.

An additional question was whether there would be differences in both SVDM and singing accuracy scores between children relating to the order of response mode (neutral syllable first or text first). This potential order effect was analyzed with a repeated-measures ANOVA with children's scores as the within-subject variable and presentation order as the between-subjects variable. There was no significant between-subjects effect for either SVDM, $F(1, 135) = 0.93, p = .34, h^2_p = .007$, or singing accuracy, $F(1, 135) = 1.47, p = .23, h^2_p = .01$. No interaction effect was found. A significant within-participants main effect for response mode was found for singing accuracy favouring the neutral syllable, $F(1, 135) = 20.24, p < .001, h^2_p = .13$, mean difference = 1.36, $SE = 0.30$, but not for the SVDM, which will be discussed further ahead.

RELATIONSHIP BETWEEN CHILDREN'S USE OF SINGING VOICE AND SINGING ACCURACY

Pearson correlations were run on all children's SVDM and singing accuracy scores. Strong and significant correlations were found between both measures in both response modes: neutral syllable, $r(137) = .88, p < .001$; text, $r(137) = .82, p < .001$.

IMPACT OF RESPONSE MODE, GRADE LEVEL, AND GENDER ON SVDM SCORES AND SINGING ACCURACY

Scores for both SVDM and singing accuracy by grade level are shown in Table 1. Differences in scores were analyzed using a mixed ANOVA with response mode (neutral syllable/text) as the within-participants factor and grade level (kindergarten, 1st, 2nd, 3rd, and 4th grades) as the between-participants factor for both use of singing voice and singing accuracy. For the use of singing voice, there was no significant within-participants main effect for response mode, $F(1, 132) = 0.59, p = .44, h^2_p = .004$, but a significant between-participants

effect was found for grade level, $F(4, 132) = 5.65, p < .001, h^2_p = .15$. Scheffé post hoc comparisons on grade level revealed significant differences between kindergartners and 2nd grade ($p = .029$) and 4th grade ($p = .007$), where the oldest exhibited a higher classification concerning the use of singing voice than the kindergartners. No interaction was found.

For singing accuracy, there was a significant within-participants main effect for response mode, $F(1, 127) = 19.60, p < .001, h^2_p = .13$. Children were significantly more accurate using the neutral syllable (mean difference = 1.35, $SE = 0.30$). There was a significant between-participants effect for grade level, $F(4, 132) = 7.69, p < .001, h^2_p = .19$. Games-Howell post hoc comparisons indicated a significant difference between kindergartners and 2nd ($p < .001$) and 4th grades ($p < .001$); and between 1st and 2nd ($p = .039$) and 4th grades ($p = .006$), where the oldest children of each pair were more accurate than the youngest. There was also a significant response mode by grade level interaction, $F(4, 132) = 2.85, p < .026, h^2_p = .08$. Pairwise comparisons were run for all grade levels and significant differences in accuracy in favour of the neutral syllable were found for kindergartners, $F(1, 42) = 4.44, p = .041, h^2_p = .096$, 1st grade, $F(1, 26) = 13.02, p = .001, h^2_p = .33$, and 2nd grade, $F(1, 21) = 8.52, p = .008, h^2_p = .30$.

Both neutral syllable and text scores are shown as a function of gender, the SVDM and singing accuracy in Table 2.

As for grade level, differences in scores were analyzed using a mixed ANOVA with response mode (neutral syllable/text) as the within-participants factor and gender as the between-participants factor for both use of singing voice and singing accuracy. For the use of singing voice, there was no significant within-participants main effect for response mode, $F(1, 135) = 0.75, p = .39, h^2_p = .006$, as presented before, and no significant between-participants effect for gender,

TABLE 2. Means (Standard Deviations) for SVDM and Singing Accuracy Scores by Gender and Response Mode

GENDER	n	SVDM		Singing Accuracy	
		Neutral Syllable	Text	Neutral Syllable	Text
Girls	74	4.19 (0.83)	4.17 (0.83)	14.94 (6.94)	14.07 (7.03)
Boys	63	3.92 (1.00)	3.86 (0.96)	12.84 (7.60)	10.93 (6.76)
Total	137	4.10 (0.92)	4.03 (0.90)	13.98 (7.30)	12.63 (7.06)

Note. Highest SVDM score would be 5; highest Accuracy score would be 24 points

TABLE 3. SVDM Groups by Grade Level and Gender

GRADE	SVDM groups								Total
	5		4		3		2		
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	
K	7	4	9	6	9	4	2	2	43
1 st	4	3	4	7	2	5	0	2	27
2 nd	6	7	3	4	2	0	0	0	22
3 rd	10	4	1	3	1	2	0	2	23
4 th	12	3	1	5	0	0	1	0	22
Total	39	21	18	25	14	11	3	6	137

$F(1, 135) = 3.89, p = .051, h^2_p = .028$. No interaction was found.

For singing accuracy, there was a significant within-participants main effect for response mode, $F(1, 135) = 21.29, p < .001, h^2_p = .14$, in favor of the neutral syllable, as seen before. There was a significant between-participants effect for gender, $F(1, 135) = 4.97, p = .027, h^2_p = .036$. Girls were significantly more accurate than boys (mean difference = 2.62, $SE = 1.18$). No interaction was found.

CHILDREN'S SINGING ACCURACY WITHIN VOICE REGISTER

To further explore the relationship between the use of singing voice and singing accuracy, children were divided into groups based on their SVDM score. Since no significant differences were found between response modes, the mean between the neutral syllable and text scores was used to classify children into the groups. Rutkowski's (2015) criteria were used to classify the groups, but the group designation was changed in order to match as closely as possible the SVDM rounded number: Group 5 was classified as singers ($SVDM \geq 4.5$) and could sing all 8 patterns; Group 4, the initial range and inconsistent singers ($3.5 \leq SVDM < 4.5$) could sing 6 of the 8 patterns (see measures 1, 3, 4, 5, 6, and 8 in Figure 1); Group 3, limited range and inconsistent initial range singers ($2.5 \leq SVDM < 3.5$) could only sing 2 patterns (see measures 4 and 5 in Figure 1);

Group 2, speaking range singer, inconsistent speaking range singer and pre-singer was not able to sing any of the patterns ($SVDM \leq 2$). The criterion for grouping considered the patterns with all the pitches within the children's vocal register.

As can be seen in Table 3, each group included children of different ages and gender. A higher number of children belong to groups 5 and 4 (60 and 43, respectively), whereas Group 2 is the smallest (9). In addition, among upper grades (2nd, 3rd, and 4th grades) there are fewer children belonging to the groups with less use of singing voice. A higher number of girls belong to group 5 (39), and a higher number of boys belong to group 4 (25).

It should be noted that children in groups 4 and 3 could access a few pitches belonging to other patterns. For example, all the D4's belonging to the patterns 1, 2, 3, 6, 7, 8 were also available for group 3 children to sing and were rated as accurate if they matched the pitch. Singing accuracy was then assessed considering all the pitches accessible for each group by response mode (see Table 4).

Considering all the pattern pitches, the highest possible score for group 5 (which could sing all the pitches) was 24; for group 4 (which exhibits a vocal range usually up to A4) was 22 (of a total of 24); and for group 3 (which exhibits a vocal range usually up to F4) was 16 (of a total of 24 pitches). When constraining the

TABLE 4. Means (Standard Deviations) for Singing Accuracy by SVDM Groups and Accessible Patterns

SVDM GROUPS	All patterns		Patterns 1, 3, 4, 5, 6, 8		Patterns 4, 5	
	Neutral syllable	Text	Neutral syllable	Text	Neutral syllable	Text
5 ($n = 60$)	19.45 (3.25)	18.07 (4.56)	14.85 (2.46)	13.82 (3.47)	4.95 (1.25)	4.79 (1.25)
4 ($n = 43$)	13.91 (5.12)	11.98 (4.26)	11.63 (4.25)	10.17 (3.87)	3.96 (1.72)	3.85 (1.51)
3 ($n = 25$)	5.63 (4.31)	4.96 (3.17)	4.95 (3.93)	4.28 (2.74)	2.55 (1.85)	2.64 (1.54)
2 ($n = 9$)	–	–	–	–	–	–

Note: Group 5, singers ($SVDM \geq 4.5$); Group 4, initial range and inconsistent singers ($3.5 \leq SVDM < 4.5$); Group 3, limited range and inconsistent initial range singers ($2.5 \leq SVDM < 3.5$); Group 2, speaking range singers, inconsistent speaking range singers and pre-singers ($SVDM \leq 2$).

number of patterns to those available to group 4 (i.e., patterns 1, 3, 4, 5, 6, and 8), the highest possible score for groups 5 and 4 was 18 (3 pitches, 6 patterns); and for group 3, 14. As for patterns 4 and 5, the highest possible score for groups 5, 4, and 3 was 6 (3 pitches, 2 patterns). Results revealed that children did not sing accurately all the pitches within their usable voice registers. Even for patterns 4 and 5, with all the pitches accessible to all the groups, accuracy decreased with decreased overall vocal registers.

In order to investigate if significant differences existed among SVDM groups for singing accuracy concerning patterns 4 and 5 (with accessible pitches to all SVDM groups), a mixed ANOVA with response mode (neutral syllable/text) as the within-participants factor and SVDM group (5, 4, 3) as the between-participants factor was run. There was no significant within-participants main effect for response mode, $F(1, 125) = 0.18$, $p = .66$, $h^2_p = .001$, but a significant between-participants effect was found for the SVDM groups, $F(2, 125) = 28.86$, $p < .001$, $h^2_p = .32$. Games-Howell post hoc comparisons indicated significant differences between all groups, with higher mean differences between group 5 and 3 ($p < .001$), followed by group 4 and 3 ($p = .002$), and then group 5 and 4 ($p = .001$). No interaction effect was found. Thus, considering the pitches accessible to all children, those who use more vocal registers are more likely to sing a higher number of pitches in tune than the children who use fewer vocal registers.

Discussion

Validity and reliability of the SVDM with Portuguese children was established, supporting the results of previous studies using the measure outside the USA (e.g., Rutkowski & Chen-Hafteck, 2001; Welch et al., 2011) and increasing the populations for which the measure is suitable. Validity was supported since the behaviors described on the SVDM scoring levels were also found among Portuguese children. As in Rutkowski (2015),

a lower frequency of responses for pre-singers, inconsistent speaking range singers, and speaking range singers was also found. Reliabilities of raters using this measure have ranged from .85 to 1.00 (as cited in Rutkowski, 2015, p. 287), and high reliabilities of .95 for both neutral syllable and text were reported in this study for the SVDM.

No significant differences in scores by response mode (neutral syllable versus text) for the use of singing voice were found, which did not corroborate Rutkowski's findings (2015). This fact may be explained by these Portuguese children's familiarity with hearing and singing patterns (and songs) both with the neutral syllable and text as a part of their regular music classroom activities. However, significant differences were found for singing accuracy in favor of the neutral syllable as in Rutkowski's study. In fact, Rutkowski (1990) pointed out that "while the text for the pattern performances seems interesting, and easily remembered by the children, the vowel and consonant combinations were not easily sung. The use of a neutral syllable, rather than words, would alleviate this problem" (p. 92). This finding might also be related to the idea that substantive music learning occurs when children are exposed to songs and chants presented with a neutral syllable as claimed by Gordon (2012).

As in Rutkowski (2015), a significant relationship between use of the singing voice and singing accuracy was found in this study. Indeed, children who access more vocal registers tend to be more accurate (i.e., children who belong to Group 5 and exhibit a vocal range up to B4 and beyond show higher accuracy scores). The results presented here reveal that differences between SVDM groups' accuracy get smaller when considering common pitch ranges. Moreover, overall accuracy scores show that even if the pitches belong to their accessible voice registers, children do not sing all of them in tune. These findings are important to consider in music education and singing assessments because low achievement might be a result of not yet accessing the

voice register needed for the task or of not having comfort managing all the pitches within their accessible voice register. Trollinger (2003) posited that children's voices go through a negotiation process when more pitches become accessible. For example, when some children begin to access higher pitches they have difficulty singing lower pitches. So, the ability to sing in tune is, of course, important, but it should not be an issue of concern until the children have access to a wider vocal range and can be considered as singers according to the SVDM classification. Therefore, this study reinforces a previous recommendation by Rutkowski (2015) about measuring children's use of the singing voice and vocal accuracy as separate but related behaviors.

It is interesting to observe that younger children can also be classified as singers (Group 5), which reinforces the fact that the SVDM is not a sequential scale because children do not need to go through all the listed stages. On the other hand, it is also possible that those younger children classified as singers already have progressed through all the stages, but at an accelerated rate. A longitudinal study would be needed to provide evidence for either of these possibilities. Although previous research has shown that children tend to expand their voice ranges and singing accuracy within school age (e.g., Demorest & Pfordresher, 2015; Welch et al., 1997; Yarbrough et al., 1991), vocal behaviors might not correspond to chronological age. As Levinowitz et al. (1998) noted, "... we cannot rely solely on maturation for students' voices to reach their peak. Moreover, it seems that we should give attention to the underdeveloped singer in particular if we are to increase the number of students using their singing voice in a facile manner" (p. 44).

Even though children outside of school mostly hear and sing music with text and they are used to speaking and communicating with text, overall results revealed higher singing accuracy scores when singing with a neutral syllable. As pointed out by Rutkowski (1990), singing with text may be too closely related to speech and so children may use their lower register when singing with text. Singing with a neutral syllable does not have that same association resulting in higher SVDM scores. Moreover, reproducing a different set of words for each pattern might add more complexity to the task than always echoing the same neutral syllable. It seems that the use of a neutral syllable enables children to focus on pitch without being distracted by the text words (Levinowitz, 1989). On the other hand, perhaps the fact that these children were familiar with hearing and singing patterns both with the neutral syllable and text in music class might also have contributed to these findings, i.e.,

because singing on a neutral syllable is part of their regular practice.

The results of this study corroborate previous findings regarding gains in pitch accuracy with age (Demorest & Pfordresher, 2015; Geringer, 1983; Petzold, 1963; Welch et al., 1997; Yarbrough et al., 1991). Differences in gender favoring girls are also aligned with previous findings (Green, 1994; Goetze & Horii, 1989; Mang, 2006; Trollinger, 2003; Yarbrough et al., 1991). Concerning the use of singing voice, this study's findings also suggest it improves with age. Rutkowski (2015) did not find significant differences between kindergarten and 1st graders, the current study points to significant differences only between kindergartners and 2nd and 4th graders, thus suggesting that the use of more vocal registers is more likely to happen beyond the 2nd grade.

The SVDM has been used together with other measures, proving to be a helpful tool while investigating singing voice behaviors (Mang, 2006; Welch et al., 2011). Proposals such as the Seattle Singing Accuracy Protocol (SSAP) (Demorest et al., 2015), which comprises a series of tasks to measure singing accuracy, could be used in conjunction with the SVDM in order to establish a more complete profile of children's singing voice achievement. Future work should map children's singing development using a longitudinal approach as already suggested by many researchers (Demorest & Pfordresher, 2015; Leighton & Lamont, 2006; Welch et al., 2011).

It is important to acknowledge individual differences in children's vocal behaviors to provide them with suitable developmentally appropriate singing activities. This can only be accomplished if music teachers implement appropriate measures in their classrooms that have proved to be valid and reliable, such as the SDVM. As results from this study show, teachers also need to be aware of the importance of focusing instruction first on the use children have of their singing voices (vocal registers as used by the author of SVDM) rather than on how accurately they are singing.

Author Note

This research was developed in the framework of CESEM (Sociology and Musical Aesthetics Research Center – UID/EAT-00693/2013) and financial support was provided by FCT (Fundação para a Ciência e Tecnologia) 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). I would also like to acknowledge the

assistance of Joanne Rutkowski in advising aspects of this project and reviewing the manuscript.

Correspondence concerning this article should be addressed to Ana Isabel Pereira, Centro de Estudos de

Sociologia e Estética Musical (LAMCI-CESEM), Faculdade de Ciências Sociais e Humanas – NOVA FCSH, Av. de Berna, 26C, 1069-061 Lisboa, Portugal. E-mail: anaisabelpereira@campus.fcsh.unl.pt

References

- APFELSTADT, H. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. *Journal of Research in Music Education*, 32, 15–24. DOI: 10.2307/3345277
- DEMOREST, S. M., NICHOLS, B., & PFORDRESHER, P. Q. (2018). The effects of focused instruction on young children's singing accuracy. *Psychology of Music*, 46, 488–499. DOI: 10.1177/0305735617713120
- DEMOREST, S. M., & PFORDRESHER, P. Q. (2015). Singing accuracy development from K-adult: A comparative study. *Music Perception*, 32, 293–302. DOI: 10.1525/MP.2015.32.3.293
- DEMOREST, S. M., PFORDRESHER, P. Q., DALLA BELLA, S., HUTCHINS, S., LOUL, P., RUTKOWSKI, J., & WELCH, G. (2015). Methodological perspectives on singing accuracy: An introduction to the special issue on singing accuracy (part 2). *Music Perception*, 32, 266–271. DOI: 10.1525/MP.2015.32.3.266
- FINKS, P. L. (1992). *Vocal registration: History, analysis, and modern pedagogical applications* (Master thesis). Eastern Illinois University. Retrieved from <http://thekeep.eiu.edu/theses/2194>
- 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, 102–104. DOI: 10.2307/3344930
- 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 pre-school and fourth-grade students. *Journal of Research in Music Education*, 31, 93–99. DOI: 10.2307/3345213
- GOETZE, M. (1985). Factors affecting accuracy in children's singing. *Dissertation Abstracts International*, 46(10), 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, 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*, 52, 23–30.
- GORDON, E. E. (2012). *Learning sequences in music: A contemporary music learning theory*. Chicago, IL: GIA Publications, Inc.
- GREEN, G. A. (1994). Unison versus individual singing and elementary students' pitch accuracy. *Journal of Research in Music Education*, 42, 105–114. DOI: 10.2307/3345186
- 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.
- 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. DOI: 10.1177/8755123312438516
- JACOBI-KARNA, K. L. (1996). The effects of the inclusion of text on the singing accuracy of preschool children. *Dissertation Abstracts International*, 57(11), 4682A.
- KIM, J. (2000). Children's pitch matching, vocal range, and developmentally appropriate practice. *Journal of Research in Childhood Education*, 14, 152–160. DOI: 10.1080/02568540009594760
- LEIGHTON, G., & LAMONT, A. (2006). Exploring children's singing development: Do experiences in early schooling help or hinder? *Music Education Research*, 8, 311–330. DOI: 10.1080/14613800600957461
- 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.
- 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, 35–47. DOI: 10.2307/3345758
- MANG, E. (2006). The effects of age, gender and language on children's singing competency. *British Journal of Music Education*, 23, 161–174. DOI: 10.1017/S0265051706006905
- MOORE, R. S. (1994). Effects of age, sex, and melodic/harmonic patterns on vocal pitch-matching skills of talented 8-11-year-olds. *Journal for Research in Music Education*, 42, 5–13. DOI: 10.2307/3345332
- 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. DOI: 10.1177/8755123315576764
- NICHOLS, B. E. (2016). Task-based variability in children's singing accuracy. *Journal of Research in Music Education*, 64, 309–321. DOI: 10.1177/0022429416666054

- 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. DOI: 10.1177/8755123314548047
- 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, 21–43. DOI: 10.2307/3344529
- PETZOLD, R. G. (1969). Auditory perception by children. *Journal of Research in Music Education*, 17, 82–87. DOI: 10.2307/3344191
- RUTKOWSKI, J. (1990). The measurement and evaluation of children's singing voice development. *The Quarterly*, 1(1-2), 81–95.
- RUTKOWSKI, J. (1993). *The use of words versus a neutral syllable for evaluating children's use of singing voice*. Paper presented at the Eastern Regional meeting of the Music Educators' National Conference, Springfield, MA.
- RUTKOWSKI, J. (1996). The effectiveness of individual/smallgroup singing activities on kindergartners' use of singing voice and developmental aptitude. *Journal of Research in Music Education*, 44, 353–368. DOI: 10.2307/3345447
- RUTKOWSKI, J. (1999). The nature of children's singing voices: Characteristics and assesment. *Canadian Music Educator*, 40(3), 43–47.
- RUTKOWSKI, J. (2014). *The comparative effectiveness of male and female singing models on kindergarten children's use of singing voice achievement*. Paper presented at 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*, 32, 283–292. DOI: 10.1525/mp.2015.32.3.283
- RUTKOWSKI, J. (2019). Assessment of singing voice development. In T. S. Brophy (Ed.), *Handbook of assessment practice and policy in music education* (Vol. 2, pp. 629–651). New York: Oxford University Press.
- 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. (2003a). A longitudinal study of elementary children's acquisition of their singing voices. *Update: Applications of Research in Music Education*, 22(1), 5–14. DOI: 10.1177/8755123302022001040
- RUTKOWSKI, J., & MILLER, M. S. (2003b). The effect of teacher feedback and modeling on first graders' use of singing voice and developmental music aptitude. *Bulletin of the Council for Research in Music Education*, 156, 1–10.
- 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*, 104–108.
- SMALE, M. J. (1987). An investigation of pitch accuracy of four- and five-year-old singers. *Dissertation Abstracts International*, 48(08), 2013A.
- THURMAN, L., WELCH, G., THEIMER, A., & KLITZKE, C. (2004, October). *Adressing vocal register discrepancies: An alternative, science-based theory of register phenomena*. Paper presented at the Second International Conference The Physiology and Acoustics of Singing. National Center for Voice and Speech, Denver, Colorado. Retrieved from <http://www.ncvs.org/pas/2004/pres/thurman/ThurmanPaper.htm>
- TROLLINGER, V. L. (2003). Relationship 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, 78–94. DOI: 10.2307/3345650
- TROLLINGER, V. L. (2007). Pediatric vocal development and voice science: Implications for teaching singing. *General Music Today*, 20(3), 19–25. DOI: 10.1177/10483713070200030105
- WELCH, G. J. (1979). Vocal range and poor pitch singing. *Psychology of Music*, 7(2), 13–31. DOI: 10.1177/030573567972002
- WELCH, G. J. (2006). Singing and vocal development. In G. McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 311–329). Oxford, UK: Oxford University Press.
- WELCH, G. F., HIMONIDES, E., SAUNDERS, J., PAPAGEORGI, I., RINTA, T., PRETTI, C., et al (2011). Researching the first year of the National Singing Programme Sing Up in England: An initial impact evaluation. *Psychomusicology*, 21(1/2), 83–97. DOI: 10.1037/h0094006
- WELCH, G. J., 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, 107–119. DOI: 10.1080/0300443910690111
- 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). London, UK: Springer. https://doi.org/10.1007/978-94-007-2660-4_3
- WELCH, G. F., SERGEANT, D. C., & WHITE, P. J. (1995/1996). The singing competences 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). Ages, sex, and vocal tasks as factors in singing “in tune” during the first years of schooling. *Bulletin of the Council for Research in Music Education*, 133, 153–160.
- 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.

Appendix Singing Voice Development Measure

The below categorizations for the Singing Voice Development Measure were developed by Rutkowski (1996):

1 = “Pre-singer” does not sing but chants the song text.

1.5 = “Inconsistent Speaking Range Singer” sometimes chant, sometimes sustains tones and exhibits some sensitivity to pitch but remains in the speaking voice range, lower register (usually A3 to C4).

2 = “Speaking Range Singer” sustains tones and exhibits some sensitivity to pitch but remains in the speaking voice range, lower register (usually A3 to C4).

2.5 = “Inconsistent Limited Range Singer” wavers between lower and middle registers and uses a limited range when in middle register (usually up to F4).

3 = “Limited Range Singer” exhibits consistent use of limited range (usually D4 to F4).

3.5 = “Inconsistent Initial Range Singer” sometimes only exhibits use of limited range, but other times exhibits use of initial range (usually D4 to A4).

4 = “Initial Range Singer” exhibits consistent use of initial range (usually D4 to A4).

4.5 = “Inconsistent Singer” sometimes only exhibits use of initial range, but other times exhibits use of extended singing range, upper register (sings beyond the lift to upper register: B4^b and above).

5 = “Singer” exhibits use of consistent extended range, upper register (sings beyond the lift to upper register: B4^b and above).