Original Research Article

National Association for Music Education

Journal of Research in Music Education
1–13

DOI: 10.1177/0022429416666054
jrme.sagepub.com
SAGE

Task-Based Variability in Children's Singing Accuracy

Bryan E. Nichols[1]

Abstract

The purpose of this study was to explore the effect of task demands on children's singing accuracy. A 2 × 4 factorial design was used to examine the performance of fourth-grade children (N = 120) in solo and doubled response conditions. Each child sang four task types: single pitch, interval, pattern, and the song "Jingle Bells." The results indicated that children's singing accuracy varied by task type, with poorer performance on patterns and songs than on single pitches and intervals. Performance was significantly better for all tasks in the doubled condition than in the solo condition, and a significant interaction indicated task-based performance varied by response mode. Students who indicated some history of private lessons (n = 54) performed significantly better than those without. Internal reliability using five test items for each type of singing task was satisfactory. Application of the Spearman-Brown formula suggests that a minimum of three items can be included in each task in future research for a reliability coefficient of .75, and four items for a coefficient greater than .80. Performance on these singing tasks was significantly intercorrelated.

Keywords

singing accuracy, vocal pitch accuracy, assessment, doubling, unison

Singing proficiency and song acquisition are important elements of the elementary music curriculum, and musicianship skills, like music literacy and improvisation, often are emphasized in music classrooms. In addition, children are expected to develop in-tune singing to express a vocabulary of folk, game, and other songs. In-tune singing is required for successful musical participation at the elementary school level, and rubrics and scales have been developed for use in classrooms, though approaches

[1]University of Akron, Akron, OH, USA

Corresponding Author:
Bryan E. Nichols, PhD, University of Akron, 254 Guzzetta Hall, Akron, OH 44325-1002, USA.
Email: bnichols@uakron.edu

for summative assessment generally have not been reported (Salvador, 2010). Furthermore, the remediation of in-tune singing has received significant attention from music education researchers, with findings that some children seem to develop tuneful singing easily and naturally, while others require significant remediation (Rutkowski & Barnes, 2000). The assessment of in-tune singing (referred to here as singing accuracy) remains one of the biggest pedagogical challenges elementary music teachers face in the classroom.

In studies with children, singing accuracy has been shown to improve over time and with certain attempts at remediation, supporting the notion that in-tune singing should be taught and evaluated. In some cases, pitch-matching skills have been shown to improve across the school year, while another skill, song singing, has not (Demorest, Nichols, & Pfordresher, 2016; Welch, Sergeant, & White, 1997). Remediation efforts have been especially notable in study designs targeting component skills for singing, like breath support and range extension (Phillips, 1985). However, improvement in component skills does not always result in overall improvement in singing accuracy (pitch matching and melodic perception skills in Apfelstadt, 1984; breath control in Phillips & Aitchison, 1997). Problematically, it is difficult to compare results from studies that use different assessment tasks, such as single pitches, intervals, patterns, and song singing. Each of these may vary in terms of complexity because of the number of unique pitches they contain and may vary in terms of memory demands as a result of the total number of pitches each task type contains (Pfordresher & Brown, 2007).

Among the variables that affect singing accuracy, the type of task and the effect of doubling have received the most attention (Goetze, Cooper, & Brown, 1990). First, there are two distinct categories of task types: pitch matching and song material (Nichols, 2015). For pitch matching, students can be asked to sing individual pitches, intervals, and patterns. For song singing, students can be asked to sing phrases or entire songs. There is a moderate or strong positive relationship in singing accuracy between pitch matching and song singing, which has been found to be moderately correlated in a sample of kindergarteners (Demorest et al., 2016) and highly correlated in a sample of adults (Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). Each of these task types can be used successfully to discriminate between inaccurate and accurate elementary singers (Roberts & Davies, 1975).

The pitch matching tasks—single pitch, interval, and pattern matching—have been used together as part of a summative evaluation of participants with varying results as to the effect of the type of task used (Nichols & Wang, 2016). For example, adult participants previously performed differently across the three pitch matching tasks. Those deemed by researchers to be "good" singers demonstrated more accurate performance on tasks with more pitches, whereas "poor" singers' accuracy decreased as the number of pitches per task increased (Pfordresher et al., 2010). Contrastingly, Wise and Sloboda (2008) found performance of self-described tone-deaf and non-tone-deaf groups to differ significantly, and unlike Pfordresher et al. (2010), performance of both groups decreased across the three tasks, respectively. The pitch-matching stimuli in Wise and Sloboda varied in the number of total pitches, whereas some other studies

presented each task type in four total pitches to control for potential memory effects (e.g., *do–do–do–do* for single pitch task, *do–do–re–re* for interval task; Demorest et al., 2016; Pfordresher et al., 2010). Kindergarten participants performed more accurately on intervals than on single pitches or patterns, suggesting a compromise between complexity (number of unique pitches) and tonal context. The additional pitches may assist the developing singer in replicating given notes perceived as pitch patterns (Demorest & Clements, 2007; Geringer, 1983).

Doubling—in which students are asked to sing in unison with others—is an important response variable for assessment because both doubling and solo response modes reflect how children are asked to respond in music classrooms. Children have been shown to have superior performance when doubled by another voice (Green, 1994), yet other researchers have come to the opposite conclusion: that children have been shown to have superior solo performance (Goetze, 1985; Goetze & Horii, 1989; Smale, 1987). Furthermore, additional research has demonstrated no significant difference between solo and doubled singing (Cooper, 1995; Smith, 1973). Thus, the previous research has been inconclusive regarding the effect of solo versus unison singing on accuracy assessments.

The conflicting reports on doubled singing are likely due to variation in the constituent studies' methodology. Singing along with peers may present different challenges than singing with the teacher or researcher, as found in some study designs, due to the effect of singing with an unfamiliar voice, discomfort while singing alone, or differences in pitch pattern difficulty and song selection. Previous studies have used a variety of conditions, including a doubling condition in which the participant was asked to sing along with the researcher (Smale, 1987), a prerecorded vocal model (Cooper, 1995), peers (Green, 1994; Smith, 1973), or peers plus the researcher (Goetze, 1985; Goetze & Horii, 1987). To further limit generalizability of previous research, each of these studies used a unique task, usually song singing (though different songs in each study). In addition, when songs or phrases were used as test material, some participants were asked to sing unfamiliar song material, whereas others were asked to sing a familiar song. When pitch matching was the singing task used in the test instrument, the pitch sequences varied in length and complexity, making it difficult to synthesize previous findings as to the effect of doubled singing in singing accuracy.

Students perform variably on different types of singing assessments, especially when doubled by other voices. An understanding of task-based performance in solo and doubled response modes would allow teachers to know whether students in the upper grades perform similarly on these specific tasks when doubled by another voice. Further, teachers can design better assessments if they know how many test items may be required for summative assessments of singing accuracy. The purpose of this study was to examine the performance of elementary music students on commonly used assessment tasks in two conditions: solo and doubled by one other voice. The comparative performance of students across a range of abilities, rather than "monotone" versus "normal" students in a remedial situation (Roberts & Davies, 1975), has not been studied in elementary students for solo or doubled singing. The research questions were as follows: (1) What is the relative task discrimination level and difficulty

level of commonly used solo and doubled singing assessment tasks? and (2) How many items within those tasks are required for reliable summative assessment of in-tune singing in fourth-grade students?

Method

I chose fourth-grade students as the target population to best represent the upper grade levels and to avoid the adolescent changing voice. The student participants came from public ($n = 2$) and private ($n = 4$) elementary schools in the U.S. Pacific Northwest. I recruited participants during their regular classroom time or during music class, and classroom teachers collected consent forms from parents following institutional review board–approved protocols. It was important to attempt to sample from a typical range of ability. Thus, all fourth-grade students in each school were invited to participate, and no child was prescreened for musical or other abilities. The consent form contained three background questions, including age and history of hearing diagnosis, to detect outliers. The third question asked for a history of private music instruction because previous research supports a link between private lessons and instrumental tuning accuracy (Yarbrough, Morrison, & Karrick, 1997).

Test Construction

An assessment was created employing four task types: single pitch, interval, pattern matching, and song singing. Individual test items within each task type were created to be moderately difficult based on previous research on item difficulty in elementary school samples (Sinor, 1984; Wolf, 2005). Each of the three pitch-matching tasks contained five items preceded by a practice item. The use of five items per task was chosen as a conservative estimate to assure internal reliability on the measure. The test stimuli ranged from D4 to A4 (Goetze, 1985; Wolf, 2005) and were presented by a prerecorded adult female vocal model. The model was instructed to sing with minimal vibrato (Yarbrough, Green, Benson, & Bowers, 1991) on the syllable [du] at 60 beats per minute. Prior to test administration, an analysis of the acoustic profile of the first pitch confirmed that the /u/ vowel was used; the mean of the middle 50% of the pitch duration of the first format was 447.02 Hz, and the mean of the second formant was 885.61 Hz.

The same recording was used for both the doubled and solo response conditions, and the participants sang along with the prerecorded stimuli after the initial presentation. Finally, each item was presented in four total notes to mediate memory and other presentation effects (Pfordresher & Brown, 2007). For example, an item in the single pitch task was given as D–D–D–D, and an item from the interval task was given as A–A–F#–F#. For the song task, participants were presented with the starting note from a recorded pitch pipe and instructed by the model to begin, “Ready, set, sing!” The song “Jingle Bells” was selected to represent song singing performance. Importantly, the decision was made to incorporate all items in two common response modes (solo and doubled). Every participant sang all items in each task and repeated them in the

doubled condition. To control for order effects, half the sample began in the solo response mode, and half began in the doubled response mode.

Scoring

All recordings were later edited so that the solo response mode preceded the doubled response mode to facilitate use of the scoring form and to remove the prerecorded instructions to reduce scoring time. The researcher and an independent judge scored 20% of the participants with high reliability ($r > .88$ for pitch matching and $r > .90$ for song singing). Next, the researcher scored half of the remaining participants, and the second judge scored the other remaining half. The judge training included instructions to consider pitches as accurate even if they were preceded by "scooping," or variability in the pitch before, during, or after it is sung (Pfordresher & Brown, 2007). Pitch matching was considered accurate if the participant sang closer to the given pitch than an adjacent pitch, which corresponds to a perceived plus-or-minus 50-cent threshold.

Each individual pitch in every set of four pitches was scored independently of the others and was coded 1 for accurate or 0 for inaccurate (for example, 1–1–0–0 indicated two correct pitches and two incorrect pitches, for a total item score of 2). The song-singing task was scored using an 8-point scale constructed for singing pitch accuracy (Wise & Sloboda, 2008). Scores from the pitch-matching and song-singing tasks were transformed to a scale from 0 to 1 to facilitate comparisons. Mean pitch-matching scores at the item level (x) were divided by four for this transformation (x') for a resulting scale in which the lowest possible score was 0 and the highest possible score was 1. The song-singing scale ranged from 1 to 8 (no possibility of a 0 score), so scores (w) were transformed with the expression, $w' = (w - 1)/7$, for a resulting scale in which the lowest possible score was 0 and the highest possible score was 1.

Participants' Background

Participants ($N = 120$) ranged in age from 9 years, 6 months, to 10 years, 10 months, at the time of testing, and the mean age was 10 years, 1 month ($SD = 3.73$ months). No one indicated a history of hearing difficulties. Fifty-five percent of participants indicated some history of private music lessons ($n = 66$). The range of lesson experience was 2 weeks to 54 months, and the mean duration was 21 months ($SD = 14.66$).

Results

One purpose of this study was to contrast performance between types of singing tasks; however, it was important to establish that these tasks are similar in that they can be used to assess the construct of in-tune singing. This unidimensionality of tasks was explored using the principal components method. All factors loaded at or above .72, which is considered high (Costello & Osborne, 2005). A one-factor model explained the variation in these results (see correlation matrix, Table 1). For order effect, there

Table 1. Correlation Matrix (Pearson's *r*) of Four Task Types in Solo and Doubled Response Modes.

	Solo				Doubled			
Variable	Single	Interval	Pattern	Song	Single	Interval	Pattern	Song
Solo								
Single	—	.73	.72	.61	.70	.72	.72	.68
Interval		—	.74	.55	.77	.77	74	.66
Pattern			—	.65	.69	.68	.65	.66
Song				—	.52	.53	.57	.64
Doubled								
Single					—	.90	.89	.82
Interval						—	.92	.79
Pattern							—	.84
Song								—

was a significant difference in performance based on whether participants began in the solo condition or the doubled condition, as indicated by a response mode by form interaction, $F(1, 117) = 7.667$, $p = .007$, $\eta_p^2 = .012$ (see Figure 1).

Question 1: Discrimination and Difficulty of Assessment Tasks

A discrimination index was calculated to determine how well performance on each task differentiated high performance from poor performance. This index was expressed by the difference between the mean score of test items for the upper and lower scoring subgroups when participants were divided into three groups by their total score ranking. A total score was calculated using the sum of the task-level scores plus song-singing scores, and on the basis of these scores, participants were ordered from low to high. Subsequently, the participants were divided into three groups of 40 participants, and the mean performance of the lowest scoring third was subtracted from the mean performance of the highest-scoring third for each task (see Table 2). The task-level discrimination was interpreted to be satisfactory for each task to be included in the test (>.30 per Allen & Yen, 2001).

Participants demonstrated significantly more accurate performance in the doubled condition than in the solo condition, $F(1, 117) = 34.608$, $p < .0005$, $\eta_p^2 = .228$. The power to detect this moderately large effect (Cohen, 1988) was calculated to be 1.00, as expressed in the statistics software output and corroborated in a calculation using G*Power with alpha set at .05 (Faul, Erdfelder, Lang, & Buchner, 2007). The task difficulty was examined by an ANOVA with two within-subjects factors (response mode and task type) and one between-subjects factor (history of private music lessons), as seen in Table 3. The ANOVA with a Hyunh-Feldt correction determined that performance across task types was significantly different, $F(2.5, 292.447) = 17.942$, $p < .0005$, $\eta_p^2 = .133$. Post hoc tests using the Bonferroni correction revealed that the single pitch and interval task performances were

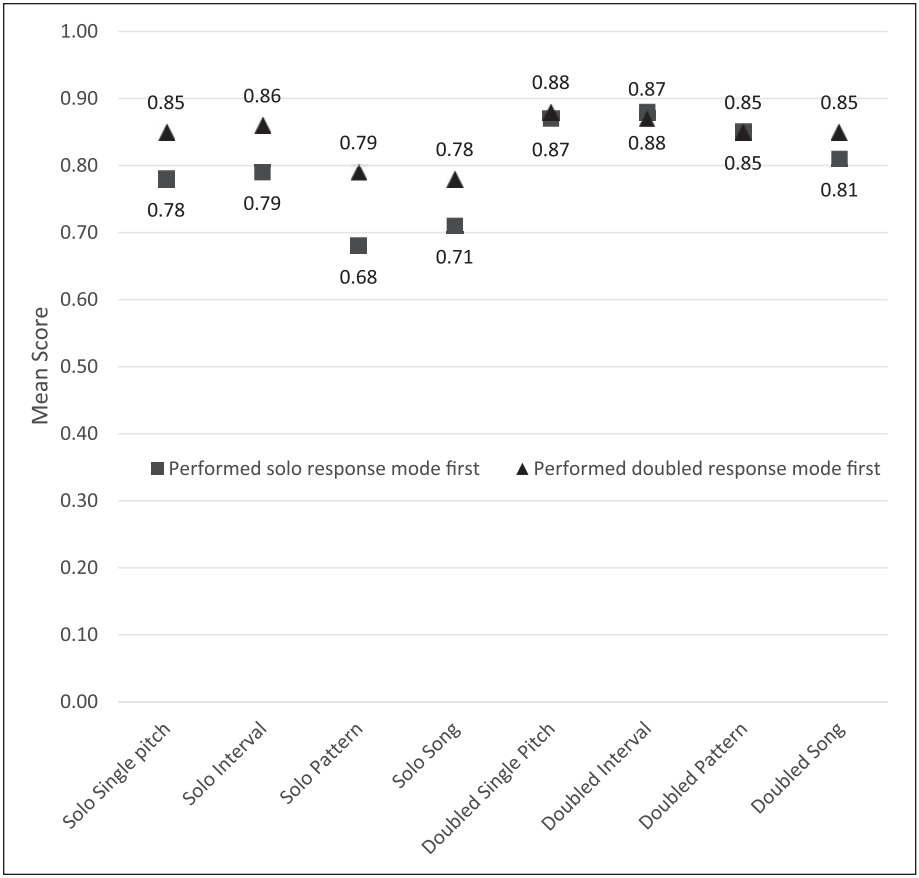

Figure 1. Task performance by presentation order.
Note. This figure illustrates that solo task performance was higher when it was presented to subjects after the doubled condition.

Table 2. Discrimination Index by Task.

Variable	Single Pitch	Interval	Pattern	Song
Solo				
Low mean	.50	.62	.48	.54
High mean	1.00	.96	.92	.94
D index	.49	.34	.44	.40
Doubled				
Low mean	.64	.66	.61	.61
High mean	1.00	.98	.98	.98
D index	.36	.32	.37	.37

Note: n = 40 for low group and n = 40 for high group.

Table 3. Mean Transformed Scores and Standard Deviations by Response Condition and History of Lessons.

Variable	Single Pitch	Interval	Pattern	Song
Solo				
All (*N* = 120)	.82 (.29)	.83 (.23)	.74 (.24)	.75 (.23)
History of lessons (*n* = 66)	.88 (.24)	.87 (.17)	.77 (.22)	.81 (.19)
No history (*n* = 54)	.74 (.33)	.78 (.28)	.68 (.26)	.67 (.25)
Doubled				
All (*N* = 120)	.88 (.21)	.88 (.21)	.85 (.24)	.82 (.22)
History of lessons (*n* = 66)	.94 (.16)	.92 (.15)	.91 (.18)	.89 (.15)
No history (*n* = 54)	.79 (.32)	.81 (.25)	.77 (.28)	.74 (.26)

Note: *n* = 120 for pitch-matching tasks and *n* = 119 for song-singing tasks.

significantly more accurate than pattern and song task performances ($p < .0005$). There was also a significant Task × Response Mode interaction, $F(3, 351) = 3.990$, $p = .008$, $\eta_p^2 = .033$. Post hoc tests using the Bonferroni correction indicated a significant cubic relationship ($p = .009$), meaning performance increased then decreased across tasks, respectively (rather than being linear). Participants with a history of private music lessons performed significantly more accurately overall than others, $F(1, 117) = 11.287$, $p = .001$, $\eta_p^2 = .088$. There was not a significant interaction, meaning that task variability was not differentiated by history of lessons (see Table 3).

Question 2: Items Required for Valid Assessment

To determine the appropriate number of items to include for the assessment of each task type, I calculated Cronbach's alpha for each task as a measure of internal consistency. Since each task type was tested with satisfactory internal consistency using five items (alpha > .80), I applied the Spearman-Brown formula to determine the reliability coefficient if fewer than five items were used to assess each task type. A minimum coefficient of .75 was deemed acceptable (Allen & Yen, 2001), and all pitch-matching tasks met this criterion if a minimum of three items were used for testing each task type.

Discussion

The purpose of this study was to examine task-based performance on fourth graders' singing accuracy while singing alone or when doubling a recorded voice. The results indicated a hierarchy of task difficulty that teachers may find useful for developing or remediating singing skills in their students. For this fourth-grade sample, the order of difficulty from easiest to hardest was (a) doubled intervals, (b) doubled single pitches, (c) doubled patterns, (d) solo intervals, (e) solo single pitches and doubled song ("Jingle Bells"), (f) solo song ("Jingle Bells"), and (g) solo pattern.

Performance on one type of singing task is predictive of performance on other tasks for more accurate singers but not so for inaccurate singers. Thus if the teacher improves a student's ability on one task type, the student may or may not improve in accuracy on other types of tasks. When ability is not developed on certain tasks, teachers can use tasks in which students excel for transferring skill to those areas. To build confidence in some children, teachers can begin by practicing doubled singing, then progressing to solo singing exercises. Also, teachers can advance from single pitches and intervals to patterns, which were shown to be more difficult overall. The results of this study are consistent with Van Zee (1984), who grouped students by specific symptoms of poor-pitch singing so that each could be specifically remediated; it was possible and useful to differentiate instruction for students in an effort to remediate singing skills.

The results of this study indicated that fourth-grade students' performance was better on single pitch and interval tasks than on patterns or singing a song from memory. The fundamental aspects of pitch matching must be in place before one can successfully sing a song. The results were similar to previous studies that compared multiple tasks in children (Demorest et al., 2016; Roberts & Davies, 1975; Welch et al., 1997; Welch, Sergeant, & White, 1995) and in adults (Pfordresher et al., 2010). Several authors have suggested that tonal memory may play a role in a person's ability to sing accurately (Jones, 1971; Joyner, 1969; Petzold, 1963), and the results support previous findings. The test items in the present study were presented in four notes, including the single pitch and interval matching tasks, which also has implications for tonal memory. This result would suggest that complexity, and to a certain extent, memory, may play a role in singing accuracy development.

Doubled singing could have helped some students sing more accurately compared to solo singing because they used their perceptual skills to adapt their singing to the accurately sung stimuli. The stimuli in this study were provided by a prerecorded track, which was used to ensure identical and accurate stimuli presentations for every child. For testing in the classroom, students may be more likely to perform well when singing along with accurate singers, like the model used in this design, than with a poor-singing peer.

The participants in this study came from various schools and may therefore represent abilities developed by various music teachers. Previous studies sampling students from one school may report results biased by one teacher's instructional style or method. The results from the fourth graders in this study were similar to those found by Green (1994) with students in Grades 1, 2, 3, and 5. In that study, students were asked to sing the song "Bow Wow Wow" alone and again in groups of eight, and they sang more accurately in the group singing condition. The results in the current study differed from reports of no significant difference between solo and doubled singing, perhaps because the doubling was done by the researcher (Cooper, 1995) or in groups of peers (Smith, 1973). The current results also differed from studies reporting more accurate solo singing than doubled singing (Goetze, 1985; Smale, 1987). One difference between those designs and the current study is that both of those studies used song phrases in the design rather than pitch matching or whole songs. In the case of Goetze (1985), students were taught two new melodic

phrases and were recorded as they imitated the melody in response to the researcher's voice. The current study emphasized the students' general ability to sing whole songs, and thus students were asked to recall a song from memory, not in call-and-response. Additionally, the decision was made for song singing to be presented in solo and doubled conditions but not in echo. Students were not primed by vocal modeling prior to song performance.

Doubled performance was more stable across the four tasks than in the solo response mode, generally. In the solo response mode, these fourth graders were like younger children in previous studies who performed single pitches better than some other tasks (Demorest et al., 2016; Welch et al., 1995, 1997). Overall, the most accurately performed solo task was not sung as accurately as the least accurate doubled task. When participants in the low-scoring grouping were examined, solo singing was more accurately sung on the interval task. Some upper elementary students appear to be better performers on simpler tasks when asked to sing alone but may appear to be better on tasks of moderate complexity when asked to sing doubled by other voices. Furthermore, teachers may need to practice intervals and patterns rather than single pitches for beginning singers. There was a significant cubic relationship in the interaction between task condition and response mode. Thus, task variability may be dependent on the specific response mode used for assessment.

The superior solo interval matching performance in the lower scoring fourth graders is similar to previous results found in kindergarten students (Demorest et al., 2016), which may suggest a pitch sequencing compromise in developing singers between complexity (not too long) and context (the two pitches in the interval task may have provided helpful contextual cues). In the solo condition, those scoring above 90% overall indicated similar scores for the single pitch task, followed by decreasing subsequent performance on the other tasks, respectively. Wise and Sloboda (2008) found self-described tone-deaf and non-tone-deaf singers decreased in accuracy across the single, interval, and pattern matching tasks, which was similar to the high-performing fourth graders in the present study. The present data suggested the development level of higher performing fourth graders may be more similar to adults than to their lower performing peers. The high group singers have developed proficiency at executing single pitches, whereas lower performing peers were dependent on the context provided in the interval condition.

Participants with a history of private instruction demonstrated more accurate performance. Previous research supports a link between musical experience and tuning accuracy (Yarbrough et al., 1997). Students who take private lessons have an individualized musical experience different from the group skill building that can be achieved in an ensemble setting. Musical experience, such as private lesson experience, may contribute to musical skills, such as vocal pitch accuracy. Or, it is also possible that students who exhibit greater musical potential are more likely to take private lessons in the first place, and their higher scores could be due to inherent differences in these students rather than to a history of private lessons.

The assessment tasks used in this study may indicate ability differently in students with more or less in-class experience singing certain task types and response modes.

The proportion of time spent singing intervals and patterns in schools compared to time spent song singing may vary across the school year and is difficult to measure because of the song singing that may occur in the general classroom, private lessons, the home, and other domains (e.g., student-initiated informal singing). Within the music classroom, this proportion likely depends on the individual music teacher and on the methodologies and instructional techniques used in the music classroom.

Task discrimination was defined as the degree to which singing tasks can be employed in a test to separate participants by ability. The data indicated task discrimination at acceptable levels for a sample of children assumed to represent typical fourth graders in schools. These results replicated those of a previous study in which the discrimination power of the same tasks was compared in the lower elementary grades (Roberts & Davies, 1975). The previous study purposefully sampled from two distinct ability levels (an accurate sample and an inaccurate sample); the data in the present study demonstrated that the single pitch, interval, pattern matching, and song singing discriminate performance satisfactorily in a sample of typical students that was not prescreened for singing ability. Furthermore, the current study replicated the previous discrimination findings to upper elementary students. Teachers must construct assessments incorporating tasks used during the teaching process. Importantly, for summative assessment, at least three items per task type must be used for reliable assessment.

Successful solo singing is an important outcome for the music classroom. To address this, teachers can prime improved solo performance by "scaffolding" activities, such as doubled singing combined with individual rather than group feedback. These results suggested that participants were primed for improved solo performance by singing along with another voice first. Students who are still developing their basic voice skills may improve their ability to follow contours and stay in the given key by first practicing along with an accurate model before realizing improvement in solo singing skills. Teachers must emphasize solo and small-group singing experiences in the classroom to develop these skills in all students.

Declaration of Conflicting Interests

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

Funding

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

References

Allen, M. J. & Yen, W. M. (2001). *Introduction to Measurement Theory*. Long Grove: Waveland Press.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Cooper, N. A. (1995). Children's singing accuracy as a function of grade level, gender, and individual versus unison singing. *Journal of Research in Music Education*, *43*, 222–231. doi:10.2307/3345637

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*. Retrieved from http://pareonline.net/getvn.asp?v=10&n=7

Demorest, S. M., & Clements, A. (2007). Factors influencing the pitch matching of junior high boys. *Journal of Research in Music Education*, *55*, 190–203. doi:10.1177/002242940705500302

Demorest, S. M., Nichols, B. E., & Pfordresher, P. Q. (2016). *The effect of focused instruction on young children's singing accuracy*. Manuscript in preparation.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. doi:10.3758/bf03193146

Geringer, J. (1983). The relationship of pitch matching and pitch-discrimination abilities of preschool 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*. Retrieved from ProQuest Digital Dissertations. (AAT 8528488)

Goetze, M., Cooper, N., & Brown, C. J. (1990). Recent research on singing in the elementary music classroom. *Bulletin of the Council for Research in Music Education*, *104*, 16–37. Retrieved from http://www.jstor.org/stable/40318355

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. Retrieved from http://www.jstor.org/stable/40318325

Green, G. A. (1994). Unison versus individual singing and elementary students' vocal pitch accuracy. *Journal of Research in Music Education*, *42*, 105–114. doi:10.2307/3345186

Jones, M. (1971). A pilot study in the use of a vertically-arranged keyboard instrument with the uncertain singer. *Journal of Research in Music Education*, *19*, 183–194. doi:10.2307/3343822

Joyner, D. R. (1969). The monotone problem. *Journal of Research in Music Education*, *17*, 115–124. doi:10.2307/3344805

Nichols, B. E. (2015). Critical variables in singing accuracy construction. *Update: Applications of Research in Music Education*. Advance online publication. doi:10.1177/8755123315576764

Nichols, B. E., & Wang, S. (2016). The effect of repeated attempts and test-retest reliability in children's singing accuracy. *Musicae Scientiae*. Advance online publication. doi:10.1177/1029864916638914

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

Phillips, K. (1985). The effects of breath-control training on the singing ability of elementary students. *Journal of Research in Music Education*, *33*, 179–191. doi:10.2307/3345579

Phillips, K., & Aitchison, R. (1997). Effects of psychomotor instruction on elementary general music students' singing performance. *Journal of Research in Music Education*, *45*, 185–196. doi:10.2307/3345579

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness." *Music Perception*, *25*, 95–115. doi:10.1525/mp.2007.25.2.95

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America*, *128*, 2182–2190. doi:10.1121/1.3478782

Roberts, E., & Davies, A. (1975). Poor pitch singing: response of monotone singers to a program of remedial training. *Journal of Research in Music Education*, *23*, 227–239. doi:10.2307/3344852

Rutkowski, J., & Barnes, P. J. (2000, March). *Validation of the Singer Accuracy Measure: Versions 2.1 and 2.2.* Paper presented at the Music Educators National Conference, Washington, DC.

Salvador, K. (2010). How can elementary teachers measure singing voice achievement? A critical review of assessments, 1994–2009. *Update: Applications of Research in Music Education*, *29*(1), 40–47. doi:10.1177/8755123310378454

Sinor, E. (1984). *The singing of selected tonal patterns by preschool children*. Retrieved from ProQuest Digital Dissertations. (AAT 8501456)

Smale, M. J. (1987). *An investigation of pitch accuracy of four-and five-year-old singers*. Retrieved from ProQuest Digital Dissertations. (AAT 8723851)

Smith, R. (1973). *Factors related to children's in-tune singing abilities*. Retrieved from ProQuest Digital Dissertations. (AAT 7411404)

Van Zee, N. (1984). An investigation study of young children's vocal problems and remedial needs. *Missouri Journal of Research in Music Education*, *5*, 55–71.

Welch, G., Sergeant, D., & 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. Retrieved from http://www.jstor.org/stable/40318780

Welch, G., 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. Retrieved from http://www.jstor.org/stable/40318855

Wise, K. J., & Sloboda, J. A. (2008). Establishing an empirical profile of self-defined "tone deafness": Perception, singing performance and self-assessment. *Musicae Scientiae*, *12*, 3–26. doi:10.1177/102986490801200102

Wolf, D. (2005). A hierarchy of tonal performance patterns for children ages five to eight years in kindergarten and primary grades. *Bulletin of the Council for Research in Music Education*, *163*, 61–68. Retrieved from http://www.jstor.org/stable/40311596

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. Retrieved from http://www.jstor.org/stable/40318418

Yarbrough, C., Morrison, S. J., & Karrick, B. (1997). The effect of experience, private instruction, and knowledge of directional mistunings on the tuning performance and perception of high school wind players. *Bulletin of the Council for Research in Music Education*, *134*, 31–42. Retrieved from http://www.jstor.org/stable/40318878

Author Biography

Bryan E. Nichols is an assistant professor of music at the University of Akron. His research interests are related to vocal production and perception skills in children and adults.

Submitted November 20, 2014; accepted January 22, 2016.