Article

National Association for Music Education

General Music Today
2017, Vol. 30(3) 13–17

DOI:10.1177/1048371317690864
journals.sagepub.com/home/gmt
SAGE

Constructing Singing Assessments for the Music Classroom

Bryan E. Nichols[1]

Abstract

As teachers are increasingly asked to respond to assessment and evaluation demands in their schools, the aim of this article is to provide suggestions and scales for singing accuracy. A description of each step in the assessment design process is provided, along with specific suggestions for a singing development scale and a singing accuracy scale.

Keywords

assessment, elementary general music, evaluation, music education, singing

Differentiated instruction and formative assessments are increasingly required elements of teaching in music classrooms. Music teachers are being asked to provide more and more evidence of student progress on an individualized level, tracking students' progress from year to year. To report data on student learning, music teachers may be asked to administer more individualized assessments in singing skills than ever before, and teachers depend on self-designed and researcher-designed tools for evaluating achievement. Several scales for singing accuracy have been put forth by researchers for studying specific characteristics in children (Salvador, 2010), yet teachers have few classroom tools for constructing tests that lead to sound pedagogical decisions and accurate academic reporting. The aim of this article is to bridge the gap between research and practice in singing assessment by providing examples and suggestions for measuring singing accuracy.

Singing instruction is an important component of the elementary general music curriculum as well as the central component of the secondary choral music curriculum. Teachers from these settings can attest to the phenomenon that some children match pitches easily, and others seem to have difficulty coordinating their voice to match what they hear (Rutkowski & Barnes, 2000). Unfortunately, students who experience singing problems may become timid singers or dislike singing altogether. Since self-perception and positive feedback have been linked to future musical participation (Clements, 2002), it is critical that teachers have ways to assess and encourage individual students' singing development.

The Assessment of Singing

Teachers can choose from song material and pitch sequences in the curriculum for assessment purposes, but it is important to point out that every interval and song varies in terms of difficulty level and other features. The term *assessment* is used in this article to refer to the general evaluation of student ability. For singing ability, certain intervals have been shown to be generally more difficult than others, and overall, interval-singing tasks may be less difficult to sing accurately than single pitch matching or pattern matching for singers at lower levels of development (Nichols, 2015). Similarly, some songs are more difficult than others, and yet another factor like singing alone or doubled by other voices may differentially affect performance of songs or other tasks.

"Best Practices" in Measurement

Evaluation and measurement in music should be founded in the sequential, step-by-step instruction that all teachers value. Measuring progress can be useful only if it is a frequent feature of instruction, not a haphazard "test" administered merely for school reporting purposes. In other words, music instruction should be frequent and consistent, and so should the evaluation of students' learning. Next, students should be evaluated on tasks related to those they are taught and practice in class, not tasks unrelated to the practical, applied work of making music. When possible, teachers can use tools already developed such as the researcher-developed scales, like those presented later in this article. Often these tools are already tested for reliability and other measurement features.

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

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

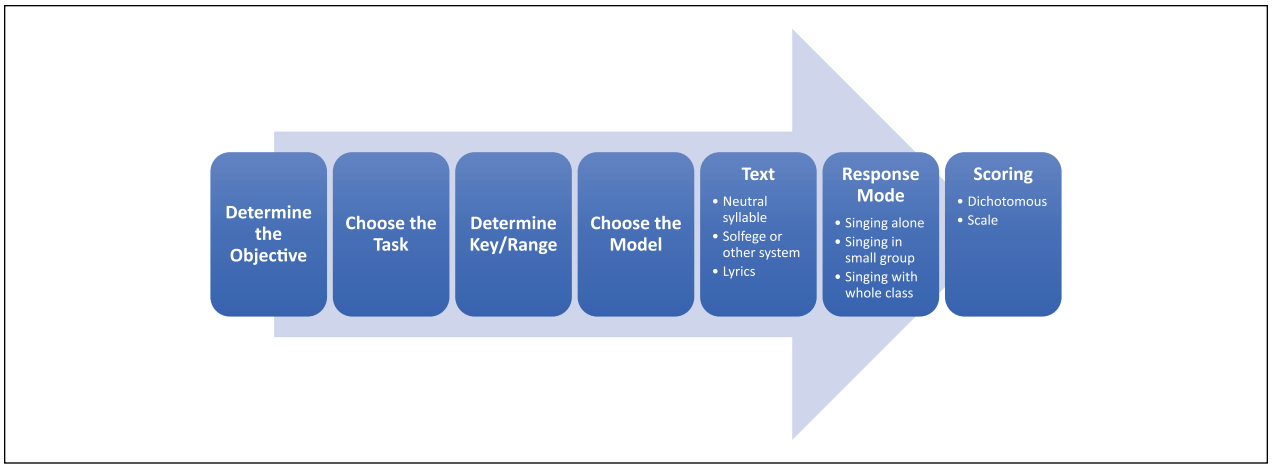

Figure 1. Flow of assessment choices.
Note. This figure demonstrates a teacher's a priori decision making for evaluation of student singing accuracy.

Establishing the Objective

A recommended pedagogy practice is to start with the desired final outcome, or the *objective*—what it is students will know or be able to do—and to plan based on this objective (backward design; Wiggins & McTighe, 2005). The next step is to determine how to measure success in each individual's ability to meet the objective. Thus, once the objective is determined, the tool for assessment can be chosen. Then—and only then—can a pedagogical sequence be developed. This author's opinion is that every curricular unit should be backward-designed in this way.

Singing Voice Development

Singing tests should be designed based on the purpose for assessment. If the purpose is to examine students' voice development, then tasks and scales appropriate for that construct should be used. The Singing Voice Development Measure (SVDM) is intended for the assessment of developmental characteristics such as range and use of singing versus speaking voice, which must be established before measuring for accuracy (Rutkowski, 1990). This measure has been shown to provide reliable score interpretations in Grades 1 to 5 (Levinowitz et al, 1998). If, alternatively, the purpose is to examine singing *accuracy*—the ability to sing specific pitches in tune—the appropriate tasks and scoring system should be constructed. In general, children should not be assessed for accuracy outside their singing range. If a child is an initial range singer, meaning he or she can sing up to A3 but not beyond, the child should not be assessed on a song with pitches above A3.

Constructing a Singing Accuracy Assessment

For teachers designing their own tests, there are many important decisions based on the purpose of the testing, including the type of task, key, range, text, singing alone or doubled, and scoring (see Figure 1). Normally these assessment decisions should be made during the lesson-planning process and can be aligned with the content standards relevant to the teacher's locale.

Selecting Singing Tasks

First, the specific pitch matching or song singing task type(s) must be chosen. Assessment tasks like single pitches, intervals, patterns, and song singing have been shown to be good discriminators of accuracy (Nichols, 2016; Roberts & Davies, 1975). If the purpose of student evaluation is formative assessment, and the teacher wishes to know which students can accurately sing a specific interval from concert repertoire to design the next day's lesson plan, the teacher may choose short phrases from the current repertoire. Or, the purpose may be to determine which patterns a child has mastered, and the teacher would present several recently practiced patterns for assessment. For pitch matching, choose at least three items for each task type used (i.e., single pitch, interval, pattern) for good reliability (Nichols, 2016).

For song singing, any song could be used to measure tuneful singing, but performance on any one song may not be representative of a student's overall ability. A teacher who is asked to present a summative assessment for grading may wish to use several tasks (Nichols, 2016). For instance, a teacher could use two songs, or several pitch matching items plus a song, since some songs are more difficult than others based on factors like range, specific

1	**Pre-singer** does not sing but chants the song text
1.5	**Inconsistent Speaking Range Singers** sometimes chants, sometimes sustains tones and exhibits some sensitivity to pitch but remains in the speaking voice range (usually A2 to C3)
2	**Speaking Range Singer** sustains tones and exhibits some sensitivity to pitch but remains in the speaking voice range (usually A2 to C3)
2.5	Inconsistent Limited Range Singer waivers between speaking and singing voices and uses a limited range when in singing voice (usually up to F3)
3	**Limited Range Singer** exhibits consistent use of limited singing range (usually D3 to F3)
3.5	**Inconsistent Initial Range Singer** sometimes only exhibits use of limited singing range, but other times exhibits use of initial singing range (usually D3 to A3)
4	**Initial Range Singer** exhibits consistent use of initial singing range (usually D3 to A3)
4.5	**Inconsistent Singer** sometimes only exhibits use of initial singing range, but other times exhibits use of extended singing range (sings beyond the register lift: B3-flat and above)
5	**Singer** exhibits use of extended singing range (sings beyond the register lift: B3-flat and above)

Figure 2. Singing Voice Development Measure.
Source. Rutkowski (1990).

intervals, tonality, and other features and thus represent singing ability differently (Wolf, 2005).

Formative Assessment—Informs teaching and for monitoring learning
Summative Assessment—Evaluates student learning at end of instructional period, often for reporting purposes

Range

Teachers may consider presenting melodies in a different key center than that presented in the source material to avoid pitches that are too low, such as middle C for some kindergarten children (Goetze, 1985; Wolf, 2005). If a song from the class repertoire is used for assessment, the key selected is likely to be the key in which the song has been taught. The key selected should align with how the range and tessitura of the song connects to singer development. The teacher will know whether the selected range is too high or low if the student can sing some of the pitches but not others, or if the student sings in a higher or lower octave. The astute teacher can differentiate between students who choose to do this and those who switch because they must (Hedden, 2012).

Not all students have developed full use of the voice range. As mentioned earlier, the SVDM is an evaluation tool for identifying the varying degrees of range and singing voice use. This construct was first described in five levels as (1) presinger (chants), (2) speaking range singer, (3) uncertain singer, (4) initial range singer, and (5) singer. It is the fifth level that represents the student who can use all parts of the singing voice range, below and above the register "lift" (Rutkowski, 1990). Some teachers may wish to use the expanded SVDM (see Figure 2) in conjunction with a singing accuracy measure to differentiate those students who are "poor-pitch singers" because they experience range limitations. Still, teachers should not assess singing on pitches that lie outside the range of the individual child.

Model

Once the range and register have been chosen, teachers must decide how test items will be presented to students. Students can be asked to echo pitch sequences by call-and-response vocal modeling. For elementary school–aged students, male teachers must decide whether to sing in the falsetto register or to sing in the male timbre, which may affect the results (Price, Yarbrough, Jones, & Moore, 1994). Teachers can also have students respond to pitch sequences from a piano or other instrument. Although children have been shown to respond best to another child's voice, and after that an adult female's voice (Green, 1990), it is difficult to administer classroom assessments for echo tasks using a child's voice. Using song tasks from memory eliminates this concern for vocal modeling since students simply sing back a song from memory, though the choice still must be made to establish a key signature (above) or to allow the student to self-select a starting pitch. However, a teacher could easily record a child who performs at a higher level—perhaps in an older grade—to use for the assessment.

Text

Next, students must sing using text or a neutral syllable, a decision often based on the role of text in the task used. If memorization is important to the purpose for testing, text may be included. For testing single pitches or intervals, the use of text may be unnecessary. If an echo task is conceptualized as a pitch matching task, neutral syllables

8	All melody is accurate and in tune, and key is maintained throughout.
7	Key is maintained throughout, and melody accurately represented, but some mistunings (though not enough to alter the pitch-class of the note)
6	Key is maintained throughout and melody mostly accurately represented, but some errors (notes mistuned sufficiently to be 'wrong').
5	Melody largely accurate, but singer's key drifts or wanders. This may be the result of a mistuned interval, from which the singer then continues with more accurate intervals but without returning to the original pitch.
4	Melody fairly accurate, or mostly accurate within individual phrases, but singer changes key abruptly, especially between phrases (e.g. adjusting higher-lying phrases down).
3	Singer accurately represents the contour of the melody but without consistent pitch accuracy or key stability.
2	Words are correct but there are contour errors. Pitches may sound almost random.
1	Singer sings with little variation in pitch, and may chant in speaking voice rather than singing.

Figure 3. Song singing scale from Wise & Sloboda (2008).

may be used. If an echo task is conceptualized as a phrase singing task, song text may be used. If the purpose for assessment does not prescribe the use of text, the teacher could rely on the research in this area to suggest the task that is chosen. Unfortunately, the effect of text on accuracy is mixed; some suggest the use of text sometimes may elicit better performance (Gault, 2002), but others suggest no difference (Levinowitz, 1987; Sims, Moore, & Kuhn, 1982; Smale, 1987) or that they are more accurate on neutral syllables (Goetze, 1985).

Scoring

Finally, before testing students, the teacher must decide how to score the student responses. Teachers can count the number of accurate pitches to create a score, or use a rubric or scale. For song tasks, a scale like the one shown in Figure 3 can also be used (Wise & Sloboda, 2008). Teachers may need a cutoff score for the consideration of "in-tune" and will save time by scoring the pitch at the time of testing, rather than record audio for later scoring. Researchers sometimes record sung responses for later scoring by one or more judges; alternatively, researchers measure the actual cent deviation of the sung pitch from the given pitch as a way to specify the degree and direction of the pitch response. Teachers, however, operate under great time constraints and are likely to make scoring decisions during test administration. Teachers will do best to choose dichotomous scoring (right-or-wrong, in-tune or not) or choose a scale.

A Word on Doubling

Another area of conflicting research is whether testing students individually—rather than in groups when the singer's voice is doubled by others—would be the best way to evaluate student performance. Since many teachers see students only once a week or teach very large ensemble classes, one-on-one assessment may be impractical. For many, evaluating students while they perform as a part of a group is a logical solution to these time constraints. In this scenario, students sing with an added stimulus: the voices of their peers. If students are being tested based on their performance in an ensemble, doubled singing may be more ecologically valid since it represents the type of singing the test is intended to replicate. If determining individual student progress is the purpose, singing alone is a more a valid response mode. Unfortunately, there are mixed results on whether students sing more accurately individually or with others, but the chosen response mode should be dependent on the purpose for assessment.

Conclusion

Inaccurate singing among students presents a problem that music educators have a keen interest in identifying and improving. In order to do so, teachers need to properly evaluate students, whether using their own or others' tests and measures for students' singing. Importantly, these evaluations should incorporate a task or a set of tasks that best represent students' abilities based on the purpose for assessment.

Singing accuracy assessment is a challenging area of music assessment, given the many factors involved in testing this skill, but teachers can design useful classroom assessments that address content standards as they exist in local schools and states. Though various singing tasks have been shown to be useful in singing assessment, more than one task type may be necessary for summative assessment, which is sometimes "high-stakes" (grades, department results, etc.). Additionally, singing accuracy performance may be different when students sing alone versus along with another voice, so teachers should choose carefully whether to test students alone or in groups. With these suggestions, music teachers can more confidently implement assessments of singing for improving student growth and for purposes of reporting.

Declaration of Conflicting Interests

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

Funding

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

References

Clements, A. (2002). *The importance of selected variables in predicting student participation in junior high choir*. Available from ProQuest Digital Dissertations database. (AAT 3062930)

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.

Goetze, M. (1985). *Factors affecting accuracy in children's singing*. Retrieved from ProQuest Digital Dissertations database. (AAT 8528488)

Green, G. A. (1990). The Effect of vocal modeling on pitch matching accuracy of elementary schoolchildren. *Journal of Research in Music Education*, *38*, 225–231. doi:10.2307/3345186

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*, 52-62. doi:10.1177/8755123312438516

Levinowitz, L. M. (1987). *An experimental study of the comparative effects of singing songs with words and without words on children in kindergarten and first grade*. Retrieved from ProQuest Digital Dissertations database. (AAT 8716497)

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

Nichols, B. E. (2015). Critical variables in singing accuracy test construction: A review of literature. *Update: Applications of Research in Music Education*, *34*, 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*. Advance online publication. doi:10.1177/0022429416666054

Price, H. E, Yarbrough, C., Jones, M., & Moore, R. S. (1994). Effects of male timbre, falsetto, and sine-wave models on interval matching by inaccurate singers. *Journal of Research in Music Education*, *42*, 269-284. doi:10.2307/3345736

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. (1990). The measurement and evaluation of children's singing voice development. *Quarterly Journal of Music Teaching and Learning, 1*(1-2), 81–95.

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*, 40–47. doi:10.1177/8755123310378454

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

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

Wiggins, G., & McTighe, J. (2005). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

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.

Author Biography

Bryan E. Nichols is an assistant professor of music education at The University of Akron. His research interests include musical assessment and music literacy pedagogy.